Structural and Functional Studies of *Escherichia coli* Kinases and Phosphatases

By

Jimin Zheng

A thesis submitted to the Department of Biochemistry

in conformity with the requirements for

the degree of Doctor of Philosophy

Queen’s University

Kingston, Ontario, Canada

K7L 3N6

May 2009

Copyright © Jimin Zheng, 2009
Abstract

Phosphorylation/dephosphorylation is likely the most crucial chemical reaction taking place in all living organisms. It is the basis for the regulatory control of many diverse biological events triggered by extracellular effectors. Moreover, it is a ubiquitous element of intracellular signal transduction pathways that regulates a wide range of processes. While protein phosphorylation has been extensively characterized in eukaryotes, far less is known about its emerging counterparts in prokaryotes. This study involved determination of the crystal structures and functional characterization of two protein kinases, YihE and AceK (also a protein phosphatase), and two nucleotide pyrophosphatases, YjjX and YhdE. X-ray crystallographic structure determination combined with bioinformatics analyses, mutageneses and biochemical experiments, both in vitro and in vivo, were utilized for the functional characterization of each protein. YihE was found to be a previously unknown kinase component of a new type of bacterial phospho-relay mechanism, thus adding kinase activity as another response to the Cpx sensing system that functions to maintain cellular homeostasis. AceK, which possesses both kinase and phosphatase activities, modifies isocitrate dehydrogenase (ICDH) to regulate the flux of isocitrate into the glyoxylate cycle. The structures of AceK alone and in complex with its substrate, ICDH, provided us with information to explain the mechanisms underlying its bifunctionality and its molecular switch. Through structural comparison and, particularly, functional characterization, we revealed that YjjX is a novel ITPase/XTPase responsible for the removal of non-canonical nucleotides from the cell during oxidative stress in Escherichia coli. YhdE, identified as a novel dTTPase, was observed to retard cell growth and form a filamentous phenotype when overexpressed in the cell, suggesting that YhdE is involved in the control of cell growth and division by regulating the cell nucleotide pool for DNA synthesis. In summary, this research has made a substantial
contribution to the investigation of bacterial phosphorylation and dephosphorylation systems that respond to various environmental conditions.
Acknowledgements

I would like to express my deepest gratitude to the many people with whom I have worked at Queen's University during the years. Without their support, it would have been impossible to accomplish what I have achieved to date.

First and foremost, I would like to thank my supervisor, Dr. Zongchao Jia, who has provided me with a tremendous amount of support, both academically and personally. Dr. Jia has always been available to give me the advice that I needed from time to time, and has always been willing to share his opinions with me. His intelligent, educational approaches and easy-going personality as an outstanding professor of crystallography are precious, and have inspired me greatly. I am grateful to him for giving me the freedom and independence to pursue my academic goals. His wisdom and generous spirits have touched me more than he can possibly know.

I would like to thank all my fellow laboratory members, both present and past, for their time and assistance with my project. Special thanks go to Dr. Melanie Adams and Dr. Michael Suits. As members of the Montreal-Kingston Bacterial Structure Genomics Initiative, they have given me great support sharing their knowledge on numerous aspects, from protein expression and purification to data collection, and have contributed valuable resources and suggestions towards my project. In particular, I wish to thank Dr. Daniel Lee who has been a great friend and an important source of spiritual support for the past several years, as well as providing many useful discussions. His friendship and encouragement have seen me through the most dismal moments of my research with a gentle, open, and willing nature that enabled me to be myself in the
moments I least wanted to be. I appreciate Dr. Qilu Ye for her generosity in sharing laboratory
knowledge; the communication and discussions I have had with her about crystallization have
provoked great ideas for my research. I thank Brent Wathen and Dr. Mona Rahman, who
generously edited my manuscript and other important documents. Brent would put his own work
aside and help me with his heart in the most timely manner. They are likely the kindest people I
have met in Canada. I would also like to thank my colleague, Dr. Vinay Singh, for his concern
and assistance. His passion for research and spark for ideas inspired me throughout my research.

I am deeply grateful to Dr. Nancy Martin, and Dr. Graham Côté for their valuable guidance and
useful discussions. I appreciate their generosity and assistance in using laboratory equipment and
accessories. I would also like to thank Dr. Robert Campbell for his “all-around” computational
and problem-solving abilities, from which I benefited greatly. I am grateful to Hans Metz who
has provided a great deal of technical support in maintaining our equipment and keeping our X-
ray machine running all the time.

I would like to acknowledge the technical staff at the beam line at the National Synchrotron
Light Source of Brookhaven National Laboratory, as well as the technical support members at
the Cornell High Energy Synchrotron Source for their assistance and follow-up support. They
provided me with valuable knowledge and advice on the practices of X-ray crystallography.

I deeply thank my parents and my parents-in-law, who accepted my long stay away from home
and provided me with deep love and encouragement. In particular, I am grateful to my parents-
in-law, who have sacrificed most of their time, patiently taking care of my dear son even when in
poor health. I also acknowledge my brother, my sister-in-law and my aunt, for their enthusiastic encouragement and support.

To Yang Yang, my loving wife, for her intensity of care and understanding, dedication to the truth and encouragement, both in my project and in the life we share together, without which neither my project, nor our marriage would have been possible. She has given me more than she has received, in ways only a husband can truly know. I thank her wholeheartedly.

Thank you all.
# Table of Contents

Abstract .......................................................................................................................................... ii  
Acknowledgements ..................................................................................................................... iii  
Table of Contents ......................................................................................................................... vi  
List of Abbreviations ................................................................................................................... xi  
List of Tables .............................................................................................................................. xiii  
List of Figures ............................................................................................................................. xiv

## Chapter 1: General Introduction ............................................................................................... 1  
  1.1 Protein phosphorylation and its roles in prokaryotes .................................................. 2  
  1.2 The *E. coli* Cpx system ............................................................................................... 5  
  1.3 Diversion of the citric acid cycle into the glyoxylate bypass ............................................. 6  
  1.4 Nucleotide binding protein ............................................................................................. 9  
  1.5 Housekeeping enzymes and Ham1/Maf superfamily ................................................ 13  
  1.6 Structure determination of *E. coli* protein kinases and pyrophosphatases by X-ray crystallography and functional characterization ........................................................ 16

## Chapter 2: Crystal structure of a novel prokaryotic Ser/Thr kinase and its implication in the Cpx stress response pathway ................................................................................................................................. 21  
  2.1 Abstract ............................................................................................................................... 22  
  2.2 Introduction ......................................................................................................................... 23  
  2.3 Results and discussion ........................................................................................................ 25  
    2.3.1 Structure determination and analysis ........................................................................ 25  
    2.3.2 Overall structure ..................................................................................................... 26  
    2.3.3 Structure comparison ............................................................................................. 29  
    2.3.4 Kinase activity of YihE/RdoA ................................................................................. 30  
    2.3.5 Proposed ATP binding site ...................................................................................... 35  
    2.3.6 Substrate binding site .............................................................................................. 37  
    2.3.7 *In vivo* characterization of YihE/RdoA ................................................................... 38  
  2.4 Experimental procedures .................................................................................................... 43  
    2.4.1 Cloning, expression and purification ................................................................. 43
Chapter 3: Structure of Escherichia coli isocitrate dehydrogenase kinase/phosphatase and its complex with isocitrate dehydrogenase

3.1 Abstract ................................................................. 48
3.2 Introduction ............................................................ 49
3.3 Results and discussion ............................................... 54
  3.3.1 Characterization and structure determination .................. 54
  3.3.2 Overall structure of AceK ........................................ 58
  3.3.3 Structure comparison ............................................. 61
  3.3.4 ATP binding site .................................................. 62
  3.3.5 AMP binding site and interaction between KD and RD .......... 65
  3.3.6 Previous mutational study on AceK activities .................... 72
  3.3.7 AceK/ICDH complex structure and interaction between the two proteins ......................................................... 73
  3.3.8 Comparison between the apo and complex ICDH structures .... 77
  3.3.9 Comparison of the apo and complex AceK structures ........... 79
  3.3.10 Phosphorylation site ............................................. 81
  3.3.11 Proposed regulatory mechanism for the AceK activity switch .... 82
3.4 Conclusion ........................................................................ 85
3.5 Materials and methods ................................................... 86
  3.5.1 Cloning, expression and purification ................................ 86
  3.5.2 Preparation of the AceK/ICDH complex ........................ 87
  3.5.3 AceK crystallization ................................................ 88
  3.5.4 Complex crystallization .......................................... 88
Chapter 4: Identification of an ITPase/XTPase in *Escherichia coli* by Structural and Biochemical Analysis

4.1 Abstract .......................................................... 93
4.2 Introduction ...................................................... 94
4.3 Results ............................................................. 95
  4.3.1 Overall Structural Characterization ......................... 95
  4.3.2 Structure Comparison ......................................... 98
  4.3.3 Active Site ..................................................... 102
  4.3.4 Identification of YjjX as an ITPase/XTPase ................. 102
  4.3.5 *In Vitro* Pull-Down Assay and Mass Spectrometry .... 104
4.4 Discussion .......................................................... 108
4.5 Experimental Procedures .......................................... 113
  4.5.1 Materials .......................................................... 113
  4.5.2 Cloning, Expression, and Purification ......................... 113
  4.5.3 Crystallization and Diffraction Data Collection ............ 115
  4.5.4 Structure Determination and Refinement ................. 116
  4.5.5 Phosphatase Assays ............................................. 116
  4.5.6 Kinetic Analysis ................................................... 117
  4.5.7 *In Vitro* Pull-Down Assay ................................... 117
  4.5.8 Peptide Fingerprinting by Mass Spectrometry ............. 118

Chapter 5: Structural and biochemical characterization of YhdE, a novel dTTP/UTP pyrophosphatase from *Escherichia coli* ............................................. 120

5.1 Abstract .......................................................... 121
5.2 Introduction ...................................................... 122
5.3 Results ............................................................. 124
6.5 Conclusions................................................................................................................ 165

References.................................................................................................................................. 168
List of Abbreviations

APS  advanced photon source
BNL  Brookhaven National Laboratory
CCD  charge coupled device
CHESS Cornell High Energy Synchrotron Source
COG  Clusters of Orthologous Groups
HEPES 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
ICDH  Isocitrate dehydrogenase
IPTG  isopropyl-β-thiogalactoside
ITP  inosine triphosphate
IUBMB International Union of Biochemistry and Molecular Biology
LB  Luria Burtani Broth
MAD multiwavelength anomalous dispersion (diffraction)
MES  2-morpholinoethanesulfonic acid
MIR(AS) multiple isomorphous replacement (with anomalous scattering)
MS/MS tandem mass spectrometry
NADH reduced β-nicotinamide adenine dinucleotide
NADPH reduced β-nicotinamide adenine dinucleotide phosphate
NBD nucleotide binding domain
NCS  non-crystallographic symmetry
Ni-NTA nickel nitrilotriacetate agarose
NSLS National Synchrotron Light Source
NTP nucleoside triphosphate
PAGE  polyacrylamide gel electrophoresis
PDB Protein Data Bank
PEG  polyethylene glycol
pNPP p-nitrophenyl phosphate
r.m.s.(d.) root mean square (deviation)
SAD single-wavelength anomalous diffraction
SCOP structural classification of protein
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SIR(AS)</td>
<td>single isomorphous replacement (with anomalous scattering)</td>
</tr>
<tr>
<td>TB</td>
<td>terrific broth</td>
</tr>
<tr>
<td>TCS</td>
<td>two-component system</td>
</tr>
<tr>
<td>TPP</td>
<td>thiamine pyrophosphate</td>
</tr>
<tr>
<td>XTP</td>
<td>xanthine triphosphate</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1. Crystallographic data and refinement statistics of YihE structure ....................... 27
Table 3.1 Activators and inhibitors of AceK’s kinase/phosphatase activity ................................. 57
Table 3.2 Statistics from X-ray diffraction data of AceK crystals ........................................... 57
Table 4.1 Crystallographic Data and Refinement Statistics of YjjX structure ................................. 97
Table 4.2 Kinetic Data of YjjX with Various Substrates .......................................................... 105
Table 4.3 Metal Cofactor of YjjX ...................................................................................... 106
Table 5.1 Kinetic Data of YhdE with dTTP and UTP ............................................................ 132
Table 5.2 Statistics of X-ray diffraction data of YhdE E33A ...................................................... 135
List of Figures

Figure 1.1 the Cpx stress response ................................................................. 8
Figure 1.2 The citric acid cycle and the glyoxylate bypass mediated by AceK. .......... 10
Figure 1.3 The transcription of the aceBAK and icIR operons. .............................. 11
Figure 1.4 Protein kinase fold ............................................................................ 14
Figure 2.1 The overall structure of YihE .............................................................. 28
Figure 2.2 Structural alignment of YihE with choline kinase and APH(3’)-IIIa .......... 31
Figure 2.3 Autophosphorylation and kinase activity of RdoA/YihE ....................... 32
Figure 2.4 Protein sequence alignment of YihE, RdoA, and homologues .................. 34
Figure 2.5 Quantification of kinase activity of YihE and PKC ............................... 36
Figure 2.6 Electrostatic surface potentials of YihE .............................................. 39
Figure 2.7 RdoA expression and rdoA, cpxR null mutant analysis ........................ 42
Figure 3.1 Size-exclusion purification of AceK and SDS PAGE analysis ............... 55
Figure 3.2 Three forms of AceK crystals and AceK/ICDH complex crystals .......... 56
Figure 3.3 The overall structure of AceK .............................................................. 60
Figure 3.4 Structural comparison between FGFR2 kinase and the AceK kinase domain .... 63
Figure 3.5 The analysis of ATP binding site in AceK ............................................ 66
Figure 3.6 The analysis of AMP binding site in AceK ............................................ 69
Figure 3.7 The ATPase activity of AceK is harboured in its ATP binding site ........... 71
Figure 3.8 Overall complex structure of AceK and ICDH ...................................... 74
Figure 3.9 The interface between AceK SRL and the ICDH active site cleft .............. 76
Figure 3.10 Structural comparison between ICDH from the AceK/ICDH complex structure and the closed form of apo ICDH ................................................................. 78
Figure 3.11 Structural comparison of AceK from the AceK/ICDH complex structure and the apo structure ................................................................. 80
Figure 3.12 Interface between the ICDH phosphorylation loop and the AceK active site cleft ............................................................................................................ 83
Figure 4.1 A stereodiagram of the overall Structure of the YjjX monomer ............... 99
Figure 4.2 Dimer comparison of YjjX, Mj0226, and Maf ...................................... 100
Figure 4.3 The stereodiagrams showing structural comparison of the YjjX monomer with the Mj0226 and the Maf monomers ................................................................. 101
Figure 4.4 Electrostatic surface and the active site geometry comparison .......................... 103
Figure 4.5 Inhibition of ITPase activity of YjjX............................................................... 106
Figure 4.6 Tryptic digestion and peptide fingerprinting of in vitro pulled-down proteins .............................................................................................................. 107
Figure 4.7 Computational docking of ITP in the active site of YjjX................................. 111
Figure 5.1 Purification of YhdE................................................................................. 126
Figure 5.2 Optimization of reaction condition for YhdE .............................................. 128
Figure 5.3 Mass spectrometry analysis of YdhE reaction products.................................. 129
Figure 5.4 dTTP/UTPase pyrophosphatase activity of YhdE......................................... 131
Figure 5.5 Crystal structure of the Maf-like protein from Trypanosoma brucei .......... 132
Figure 5.6 YhdE_E33A crystals and its typical diffraction images ............................... 134
Figure 5.7 Dimer comparison of YhdE E33A with the Maf protein ............................... 137
Figure 5.8 Structural comparison and dTTP binding site of YhdE E33A ...................... 138
Figure 5.9 EM of cells overexpressing YhdE ............................................................... 141
Figure 5.10 Growth curve .......................................................................................... 142
Figure 6.1 Sequence alignment of Maf/Ham1 proteins in the nucleotide binding loop ..... 159
Figure 6.2 Structural comparison of Ham1/Maf proteins ............................................. 162
Chapter 1

General Introduction
1.1 Protein phosphorylation and its roles in prokaryote

Cells sense extracellular signals and invoke an appropriate intracellular response by triggering cascades of events that alter the activities of various signaling molecules. Inter- and intracellular signaling is often achieved by the interconversion between functionally distinct phosphorylated and dephosphorylated forms of relevant proteins. The newly acquired phosphate group introduces a negative charge, which is capable of forming multiple hydrogen bonds and has a tendency to associate with positively-charged amino acid side chains such as arginine and lysine. Typically, phosphorylation occurs near the protein surface, and is often associated with enzymatic reversibility, a change in enzyme activity, alteration of protein cellular location, or changes in protein-protein interactions. Protein (de)phosphorylation is therefore a critically important regulatory mechanism in all organisms, regulating various cellular functions such as gene expression, cell growth and differentiation, apoptosis, membrane transport, metabolic pathways, and cell movement. In eukaryotic organisms, phospho-proteins interact to form an integrated information processing network, which is able to coordinate cellular processes in response to a wide range of internal and external signals and has been found to be a remarkably versatile and sophisticated mechanism for exerting regulatory control. Protein kinases that convert their substrates to a functional phosphorylated state comprise the largest gene family in an eukaryotic organism's genome and are becoming important drug targets in therapeutics research. Based on the substrate specificity, protein Ser/Thr kinases and protein Tyr kinases are the two major subdivisions of the eukaryotic kinase superfamily. Since the first discovery of a protein kinase in 1956, extensive studies on known eukaryotic protein kinases have revealed that Ser/Thr and Tyr kinases are all related by their 250-300 amino-acid kinase domain. Indeed, X-ray crystal structures have shown that eukaryotic Ser/Thr and Tyr kinases actually share
similar catalytic domains, which indicates that a similar phosphor-transferring mechanism is shared among eukaryotic protein kinases.

Although there is a rich body of knowledge about the kinases and phosphatases that control the interconversion of phospho- and dephospho-proteins in eukaryotic organisms, an understanding of the regulation of cellular processes by protein (de)phosphorylation in prokaryotic organisms is still in its infancy. Due to lack of research in understanding their physiological functions, substrate specificities, and downstream targets, only a small fraction of prokaryotic kinases and phosphatases have been characterized thus far \(^8\). In fact, until quite recently it was assumed that Ser/Thr and Tyr phosphorylation only existed in eukaryotes \(^9\) as a relatively late evolutionary event to regulate more complicated cell activity \(^1\). However, with the progress in genomic sequencing and homology comparison over the past decade, the study of prokaryotic protein phosphorylation has finally taken off \(^10\), revealing that prokaryotic phosphorylation is also of major importance to bacteria cells \(^11\). A recent, large-scale microbial genomic sequencing project identified 20 known and novel eukaryotic protein kinase-like families in the microbial genome \(^12\). The surprising pervasiveness of eukaryotic-like kinase and phosphatase genes across the prokaryotic phylogeny suggests that the evolution of protein phosphorylation took place much earlier than was previously thought \(^1\). As a result, this revelation has sparked a flurry of research into phosphorylating systems in prokaryotes \(^10\).

To date, protein phosphorylation events have been implicated in the regulation of a number of processes in prokaryotic organisms, including the expression of stress response genes, catabolite repression, coordination of nitrogen and carbon metabolism, synthesis of secondary metabolites, and infection mechanisms of pathogens \(^13\). Complete genome sequencing of many prokaryotes organisms has provided the opportunity for structural and functional studies of
protein kinases and phosphatases that regulate their critical processes. Thus far, four general systems of protein phosphorylation in prokaryotes have been studied. The first confirmed prokaryotic phosphorylation mechanism, the “two-component system” (TCS), involves histidine and aspartic acid, and is an important and efficient signal transduction mechanism. In response to an input stimulus, a sensor histidine kinase autophosphorylates a histidine residue using ATP, and this same phosphoryl group is then transferred to an aspartic residue of a separate receiver protein, triggering downstream cellular responses. The majority of the sensor histidine kinases are homodimeric proteins, consisting of a widely diverse extracellular or periplasmic sensory domain and a conserved cytoplasmic domain. The Cpx system, for example, is a typical bacterial TCS that will be discussed below. The second bacterial phosphorylation system is the phosphoenol pyruvate:carbohydrate phosphotransferase system (PTS), where five protein domains transfer a phosphoryl group from a phosphoenol pyruvate molecule through a chain of intermediate proteins (via histidine and, in one case, serine residues) to eventually phosphorylate a sugar. Initially, histidine phosphorylation involved in the TCS and PTS was thought to be unique to prokaryotes. However, homologues of the TCS have since been uncovered in yeast, plants, and rat mitochondria; and the PTS has been found in lower eukaryotic organisms. The third bacterial phosphorylation system is highly similar to the eukaryotic ATP/GTP-dependent system, where kinases specifically phosphorylate Ser, Thr, and Tyr residues. In addition, two bacterial Ser/Thr kinase/phosphatase bi-functional enzymes that are involved in cell metabolism have been found in bacteria (AceK and HPr kinase/phosphatase) that do not have eukaryotic counterparts. In fact, AceK, a major component of the present research, was the first Ser/Thr protein kinase identified in prokaryotes.
In contrast to protein kinase classification, protein phosphatase classification is currently only possible for eukaryotes, as very little is known about their prokaryotic counterparts. At present, there are three classes of protein phosphatases in eukaryotes 19-21: 1) Ser/Thr phosphatases that undergo metal-assisted catalysis such as PP1 22 and PP2A 23; 2) the PTP superfamily of proteins that have the signature motif CX5R 24, an example of which is PTP1B 25; and 3) proteins such as FCP1 26 and SCP 27 with signature motif DXDXT/V that utilize the newly discovered Asp-based catalysis.

1.2 The *E. coli* Cpx system

When unable to flee from a suboptimal environment, bacteria have multiple regulatory pathways to protect themselves from various types of environmental stressors such as changes to oxygen levels, temperature, pH, osmolarity, and the intensity and wavelength of light 28. In many cases, bacteria sense the status of the external environment via two-component histidine kinase signal transduction pathways. One such pathway, the Cpx stress response system, responds to potentially disruptive conditions within the cell envelope, such as the presence of misfolded proteins in the periplasm 29. These misfolded proteins often arise as a result of changes in the surrounding pH that force proteins off their normal folding pathways. The Cpx pathway is also involved in the attachment of bacterial cells to surfaces, and in several additional virulence mechanisms. Cpx-pathway genes encode proteins that contribute to protein degradation, protein modification, protein folding, cell wall biosynthesis, and bacterial pathogenesis. In the Cpx TCS, the sensor histidine kinase, CpxA, is activated by the presence of misfolded periplasmic proteins, or, alternatively, undergoes NlpE-dependent activation; following this, the signal is transmitted to the response regulator CpxR through histidine/aspartate phosphotransfer reactions.
Phosphorylated CpxR acts as a transcriptional activator of genes encoding factors such as periplasmic proteases (Figure 1.1)\textsuperscript{30}. This is a classical response system whose activity is determined by its phosphorylation status: upon envelope stress, CpxA functions as a kinase to activate CpxR; in the absence of stress, CpxA functions as a phosphatase to deactivate CpxR\textsuperscript{31}. Therefore, CpxA functions as both kinase and phosphatase to regulate the CpxR’s function. Although the phenotypes of the Cpx stress are well-studied, molecular mechanisms governing the entire Cpx-mediated signaling network require further elucidation in order to fully understand the system\textsuperscript{28,29,32}.

1.3 Diversion of the citric acid cycle into the glyoxylate bypass

The survival of most organisms depends on their ability to successfully adapt to constant, often rapid changes in their environment. These changes can translate into growing conditions that range from “feast” to “famine”. As a result, organisms have typically evolved a series of mechanisms to monitor their surroundings and adjust their physiology accordingly. Many bacteria utilize an important adaptation in their fundamental metabolism to switch between energy production and growth. This switch diverts the citric acid cycle (the Krebs’ cycle) – the final common pathway for the complete oxidation of fuel molecules in all organisms – to a bypass called the glyoxylate cycle (Figure 1.2A).

When \textit{E. coli} grow on carbon sources containing more than two carbon atoms per molecule, such as glucose, the citric acid cycle is exploited to completion\textsuperscript{33,34}. However, when substrates are limited to two-carbon sources such as acetate, the completion of all oxidative decarboxylation steps in the citric acid cycle would result in the net loss of all carbon atoms as
CO₂. In these limiting conditions, the glyoxylate bypass is activated to produce the four-carbon metabolic intermediate succinate for biosynthesis at the cost of energy production (Figure 1.2B).

In order to grow using only acetate as their carbon source, *E. coli* need to activate their acetate switch. This results in a change in enzyme expression that ultimately shifts their acetate posture from excretion to utilization 35. The acetate switch is made up of two components: the first is competition for isocitrate by isocitrate dehydrogenase (ICDH) and isocitrate lyase (ICL), while the second is competition for coenzyme A (HS-CoA) by α-ketoglutarate dehydrogenase (α-KGDH) and phosphotransacetylase (PTA) 35. HS-CoA diverted through α-KGDH results in energy generation, while HS-CoA diverted through PTA leads to biosynthesis 35. For efficient growth, the ratio of HS-CoA and isocitrate flux through energy generation and biosynthetic pathways must be synchronized. While the mechanism by which the HS-CoA flux is regulated remain unclear, the regulation of the isocitrate flux at the isocitrate branch point has been well characterized 33,34.

The first step of the glyoxylate bypass is catalyzed by ICL, which competes with IDCH for isocitrate. Regulation of this important branch point depends on a bifunctional enzyme, AceK (isocitrate dehydrogenase kinase/phosphatase), which is expressed along with two other essential enzymes in the glyoxylate bypass in many bacteria 36-38 (Figure 1.3). AceK regulates the activity of ICDH by (de)phosphorylating S113 on ICDH 39,40. Under normal circumstances, ICDH is activated via dephosphorylation by AceK. As an active enzyme, ICDH has a much higher affinity for isocitrate than ICL does. Therefore, much of the isocitrate goes through the full citric acid cycle 33. However, once the preferred carbon source is consumed, the acetate switch is activated, and acetate is utilized rather than excreted 33. AceK starts to phosphorylate ICDH to inactivate this enzyme. Therefore, isocitrate enters into glyoxylate bypass. AceK is known to be
Figure 1.1 The Cpx stress response. The Cpx system can be induced in an NlpE-dependent or NlpE-independent manner. Whereas adhesion to hydrophobic surfaces requires NlpE for activation of CpxA, misfolded PapE does not. Upon activation of CpxA by either signal, CpxA phosphorylates CpxR, which then acts as a transcriptional activator of genes encoding factors involved in envelope biogenesis. CpxA targets periplasmic proteases (such as degP). (adapted from Ruiz’s paper\textsuperscript{31}.)
essential for this change: experiments have shown that AceK knockouts do not utilize acetate, and growth stops once other carbon sources are depleted. The role of AceK is to deactivate ICDH and promote carbon flux through ICL; this is supported by the sustained growth of *E. coli* AceK knockouts on acetate when ICL is over-expressed. Given the importance of the citric acid cycle and its bypass, there is an intricate network of regulators that act directly upon AceK as well as on the substrate to provide another level of regulation. For example, the presence of AMP and ADP, which represents energy depletion, stimulates the phosphatase activity and inhibits the kinase activity of AceK. Subsequently, the resulting dephosphorylated ICDH acts to drive energy production. Other regulators, such as NADP+, do not act directly on AceK, but on ICDH to regulate the AceK kinase and phosphatase activities. So far, many features of AceK have not been clarified. For example, what triggers AceK’s kinase activity when cell grows on acetate media, and what is the mechanism that switches AceK from phosphatase activity to kinase activity? However, due to the lack of AceK structure information, the research on AceK has been blocked for almost ten years. In this study, we report the AceK’s individual structure and the complex structure with ICDH. They provide a great deal of structural information for understanding the AceK’s kinase and phosphatase mechanisms.

1.4 Nucleotide binding proteins

The nucleotide binding domain (NBD) is an essential feature in any kinase or nucleotide phosphatase that aids in phosphorylation or dephosphorylation. Nucleotide binding proteins are not limited to phosphatases or kinases, but instead constitute one of the largest protein families that are involved in almost all cellular processes. Studying nucleotide binding domains can
Figure 1.2 The citric acid cycle and the glyoxylate bypass mediated by AceK. A) In the complete citric acid cycle, isocitrate is converted to α-ketoglutarate and carbon dioxide by the enzyme isocitrate dehydrogenase (IDH). When acetate is the sole carbon source, AceK phosphorylates isocitrate dehydrogenase (IDH) and inactivates the enzyme. The inactivation of IDH increases the conversion of isocitrate to glyoxylate and succinate by isocitrate lyase. Glyoxylate is then combined with another molecule of Acetyl-CoA to form malate, a downstream product of the citric acid cycle, thus completing the glyoxylate bypass. B) Compared to the complete citric acid cycle, the glyoxylate cycle largely reduces energy output but allows the net production of succinate for subsequent gluconeogenesis and growth. Both steps that produce carbon dioxide are skipped in the glyoxylate cycle, thus preserving the carbon source for cell growth.

**Net reaction of the citric acid cycle:**

\[
\text{Acetyl-CoA} + 3 \text{NAD}^+ + \text{FAD} + \text{GDP} + \text{P}_i + 2\text{H}_2\text{O} \rightarrow \text{CoA} + 3 \text{NADH} + \text{FADH}_2 + \text{GTP} + 2\text{CO}_2 + 3\text{H}^+
\]

**Net reaction of the glyoxylate cycle:**

\[
2 \text{Acetyl-CoA} + \text{NAD}^+ + 2\text{H}_2\text{O} \rightarrow \text{succinate} + 2\text{CoA} + \text{NADH} + \text{H}^+
\]
Figure 1.3 The transcription of the *aceBAK* and *iclR* operons. The aceBAK operon encodes three genes, the malate synthase AceB, the isocitrate lyase AceA, and the isocitrate dehydrogenase kinase/phosphatase AceK. These three enzymes are involved in the glyoxylate cycle. (Adapted from 42.)
shed light on the manner by which they specifically recognize particular nucleotides, and the mechanism of (de)phospho-hydrolysis and phospho-transfer.

In the case of protein kinase, since the discovery of protein kinase (rabbit skeletal muscle glycogen phosphorylase) in 1956, there has been an exponentially increasing number of protein kinases and their substrates characterized in a wide variety of eukaryotes ranging from fungi to mammals. Protein phosphorylation in eukaryotes mainly involves ATP-dependent Ser/Thr and Tyr protein kinases (ePKs). All members of this well-defined superfamily exhibit extensive sequence similarity. Indeed, X-ray crystal structures have revealed that eukaryotic kinases actually share similar catalytic domains. All eukaryotic protein kinases, Ser/Thr and Tyr kinases alike, are related by their 250-300-residue kinase domains, where 12 conserved sub-domains fold into a common catalytic core architecture. The overall structure consists of two domains, with an N-terminal lobe providing some residues for ATP binding, and a C-terminal lobe containing a catalytic loop and the protein substrate binding site; the catalysis occurs at the cleft between the two lobes. Several residues highly conserved among these ePKs are scattered throughout the catalytic core: in the N-terminal lobe, these are essentially the glycine-rich motif (GXGXXG), involved both in nucleotide binding and in catalysis, and a lysine residue found invariably and known to interact with the α- and β-phosphates of bound MgATP. In the C-terminal lobe, there is a Brenner's motif [HNDX_{4}N] found in most of ePK. The most conserved residues in this motif are an aspartate residue, presumed to act as a catalytic base, and two other residues, one asparagine and one aspartate, each involved in chelating one of the two ATP-liganded Mg^{2+} ions in PKA. This cluster of three residues forms a catalytic triad which is invariably found in all members of the ePK family. Recently, genome sequencing programs have allowed identification of several putative bacterial members belonging to this
family, since they display most of the conserved residues mentioned above \(^{18}\), which provide a great way to identify the protein kinases in bacteria.

Not just limited to protein kinases, previous sequence comparison studies and crystallographic analysis of nucleotide binding proteins have revealed that an appreciable proportion of proteins that bind ATP or GTP share a number of conserved sequence motifs. Typically, these include the Walker A sequence motif and the glycine-rich 'P-loop' that typically forms a flexible loop between a \(\beta\)-strand and an \(\alpha\)-helix. The structure of an ATP-binding cassette reveals that the Walker A amino acid backbone forms hydrogen bonds with the ATP phosphate. Further, there is a Walker B motif in which an aspartate at the C-terminus of NBDs binds to an \(\text{Mg}^{2+}\) ion required for catalysis \(^{51}\). Although the Walker A motif is well studied, not all ATP- or GTP-binding proteins or other nucleotide binding proteins contain this motif. There are numerous ATP- or GTP-binding proteins that contain the P-loop, such as the ATP synthase \(\alpha\) and \(\beta\) subunits, myosin heavy chains, ABC transporters, and the Ras family of GTP-binding proteins \(^{52-56}\). A number of NBD proteins escape detection, however, because the structures of their nucleotide-binding sites are completely different from that of the P-loop. Examples of such proteins include the E1-E2 ATPases and the glycolytic kinases.

1.5 Housekeeping enzymes and Ham1/Maf superfamily

Housekeeping enzymes in the nucleotide binding protein family play the important role of clearing dangerous non-canonical nucleotides like ITP or XTP which are by-products of normal cellular metabolism. Incorporation of these nucleotides during transcription causes mismatches and a decrease in the stability of the RNA-DNA complex resulting in a lower elongation rate and
All protein kinases have a similar protein fold that comprises two lobes: one lobe consists of mainly β-sheet structure (blue) and the other lobe consists of α-helices (green, orange and yellow). This lobe structure forms an ATP-binding cleft that constitutes the active site. The crystal structure of cyclin-dependent kinase-2 (CDK2) (PDB code: 1qmz) shows this representative fold. ATP is bound in the cleft (adapted from 57.).
early termination. A variety of non-canonical NTPs are produced by oxidation, deamination or other modifications of nucleotides. The nucleotide pool cleaning enzymes are able to detect non-canonical nucleotides present in the background of over 1000-fold excess of canonical NTPs. \(^{58}\)

There are at least four structural superfamilies of housekeeping NTP pyrophosphatases targeting non-canonical NTPs: MutT-related (Nudix) hydrolases, dUTPase, ITPase (Maf/HAM1) and all-α NTP pyrophosphatases (MazG). Although these enzymes have their own specific affinities for various natural substrates (8-oxo-dGTP, dUTP, dITP, 2-oxo-dATP), they have a common ability to recognize and bind nucleotides. The Walker A nucleotide binding motif is not found in these enzymes, which indicates that they may have adapted a distinct mechanism for nucleotide substrate binding and hydrolysis. \(^{58}\) A comparison of the nucleotide binding features in these enzymes is anticipated to provide insights into the mechanism(s) that allows them to select their substrates with such high specificity.

One superfamily of housekeeping enzymes is the Maf/Ham1 nucleotide binding superfamily, which originally includes two protein subfamilies: the Ham1 protein family and the Maf protein family. These two protein families have similar structures despite low sequence homology. The Ham1 protein is an ITP pyrophosphatase, which was first discovered in human erythrocytes. In 1999, the Ham1 protein structure of Mj0226 from Methanococcus jannaschii was determined and shown to hydrolyze ITP and XTP. \(^{59}\) The Maf protein, a septum-formation related protein from Bacillus subtilis, was found to have an ITPase fold by structural comparison, but the precise function of these proteins has not been determined. \(^{60}\) The crystal structure of Maf has revealed that it binds to the β- and γ-phosphates of dUTP. Crystal structures have also shown that both proteins contain a mixed α/β fold, as well as two lobes separated by a large cleft that is thought to be a nucleotide binding site. The nucleotide binding
ability of Maf as well as its structural homology to the members of the Ham1 protein family led to the hypothesis that Maf is also a specific nucleotide phosphatase. Thus far, there are seven members with determined crystal structures in the Maf/Ham1 superfamily, and this provides a strong foundation for systematically studying the function(s) of the Maf proteins and the nucleotide binding mechanism of Maf/Ham1 superfamily proteins. YhdE (a Maf protein homologue) from *E. coli* was selected for functional analysis in this research. Detecting conservation through sequence and structure comparisons is anticipated to shed light on the structural motifs that are important for nucleotide binding.

1.6 Structure determination of *E. coli* protein kinases and pyrophosphatases by X-ray crystallography and functional characterization

For my PhD project, two protein kinase (YihE, AceK) and two nucleotide pyrophosphatases (YjjX, YhdE) from *E. coli* were selected for structure-function studies.

**YihE:** The *yihE* gene from *E. coli* is located immediately upstream of *dsbA*, a disulfide oxidoreductase that plays an important role in protein folding in many Gram-negative bacteria. Previous studies in *S. enteria* and *E. coli* demonstrated that the transcription of *yihE* is regulated by the Cpx stress response pathway and is involved in the extra-cytoplasmic stress response. It has also been postulated that the inactivation of YihE in *Shigella* may lead to inefficient production of UDP-glucose. UDP-glucose is required for LPS synthesis, as it regulates the expression of the galETK operon. Sequence analysis of YihE predicts that it is a putative homoserine kinase, though this has not been confirmed experimentally. Understanding
the function of YihE will help explain how it coordinates with DsbA in response to the activation of the Cpx stress response pathway.

**AceK: Isocitrate dehydrogenase kinase/phosphatase.** The study of bacterial phosphorylation systems was advanced by the discovery of a phosphorylating activity in *E. coli* that regulates isocitrate dehydrogenase (ICDH) \(^67\). This was the first prokaryote phosphorylation system to be identified in bacteria, and is the only known Ser/Thr phosphorylation system/pathway in *E. coli*. AceK is a 66.5 kDa protein that possesses both kinase and phosphatase activities \(^68,69\). This unusual phosphorylation-dephosphorylation system modifies the S113 residue on ICDH; it is this modification that regulates the flux of isocitrate through the glyoxylate bypass. ICDH competes with ICL to either direct isocitrate through the Krebs’ cycle or into the glyoxylate bypass. When the organism is grown on acetate, ICDH is in its inactive phosphorylated form, which inhibits the Krebs’ cycle \(^70\). Alternatively, a change of carbon source to glucose or pyruvate results in the dephosphorylation of ICDH and the initiation of the Krebs’ cycle. AceK also demonstrates ATPase activity independent of ICDH \(^71\). Structurally, AceK contains a eukaryotic-kinase-like ATP-binding motif, which is required for both kinase and phosphatases function \(^72\). Interestingly, however, sequence comparison shows no similarities between AceK and the eukaryote Ser/Thr protein kinases. There is a possibility, then, that AceK, as a highly unusual bifunctional protein, may possess a novel kinase/phosphatase structural fold and (de)phosphorylation mechanism. The function of AceK and its involvement in the regulation of the Krebs’ cycle and the glyoxylate bypass is well-characterized, but its structural and mechanistic qualities have remained relatively unknown. Structural studies should facilitate analysis of AceK at both a macro- and micro-scale. For example, the determination of the crystal structure is expected to confirm whether AceK
contains individual kinase, phosphatase and ATPase domains. A crystal structure should also provide insights into the coordination of AceK's kinase, phosphatase and ATPase activities. Of particular interest, structure determination is expected to reveal what sort of changes, if any, the active site undergoes as it switches between kinase and phosphatase activity. Determination of the structure “caught” in both kinase and phosphatase modes would provide information on the manner by which this bifunctionality is achieved. In addition, structures of AceK complexed as either a kinase or a phosphatase with its substrates would be invaluable for clarifying its phosphorylation and dephosphorylation mechanisms.

**YjjX:** The yjjX gene from *E. coli* was predicted to encode a 20 kDa protein with 184 residues. It is positioned in between the *trpR* and *gpmB* genes, the former of which encodes the Trp repressor protein while the latter encodes a phosphoglycerate mutase. The yjjX gene, however, appears to be under the control of a promoter independent of these flanking genes and thus no functional information could be directly garnered from the genomic location of this ORF. Amino acid sequence analysis placed the YjjX protein in the COG1986 and DUF84 Pfam protein families. Close homologues could be identified in 12 archaeal, 33 bacterial, and several eukaryotic species. Inspection of the sequence homologues permitted their classification into one of two distinct categories: (i) hypothetical proteins of unknown functions and (ii) nucleotidyl transferases. Extensive literature analysis identified a single report that suggested a putative function for the YjjX protein. Genetic screening for genes related to thiamin biosynthesis and degradation suggested the involvement of YjjX in resistance to bacimethrin or 4-amino-2-trifluoromethyl-5 hydroxymethylpyrimidine (HMP). Both compounds are highly toxic to the cell and inhibit the growth of *E. coli* when low micromolar concentrations are added to the culture.
medium. Despite this result, the mechanism of YjjX in this detoxification process remained unclear. The structure of YjjX was determined in my M.Sc. project (PDB accession code 1U5W), but its biochemical analysis was not carried out. The tertiary structure of YjjX is highly similar to the nucleotide binding proteins Mj0226 (PDB accession code 1B78), a nucleotide pyrophosphatase from *Methanococcus jannaschii*, and Maf (PDB accession code 1EX2), a dUTP-binding protein from *Bacillus subtilis*, despite the fact that the sequence alignments of these proteins indicated only 19% identity. YjjX was classified into the Ham1/Maf superfamily in the SCOP database (http://scop.mrc-lmb.cam.ac.uk/scop-1.69/data/scop.b.d.gj.e.html). In my Ph.D. program, I investigated YjjX function, based on the clue from a close inspection of the structural and chemical features of aforementioned structural homologous.

**YhdE:** The *yhdE* gene from *E. coli*, encoding a 21.5 kDa protein, is located in the *mre* (murein region) operon, which consists of five genes, *mreB*, *mreC*, *mreD*, *yhdE* and *rng*. The genes of *mreB*, *mreC* and *mreD* in the operon were found to be essential for maintaining the natural rod-like shape of bacteria such as *B. subtilis* and *E. coli*. Specifically, MreB is a prokaryotic actin-like homolog found beneath the cell surface in a helical array that regulates cell width and maintains cell shape. MreB also plays a role in chromosomal DNA segregation and cell division. Downstream from the *mre* and *yhdE* genes is *rng*, which encodes RNaseG, a non-essential ribonuclease specific for adenine and uracil rich regions. RNaseG is homologous to the amino-terminal part of RNaseE, a protein involved in the regulation of the ratio of FtsZ/FtsA that is important for septum formation and cell division. Of the five *mre* genes, only the function of *yhdE* remains unknown at present. Overexpression of Maf, the YhdE homolog in *B. subtilis*, results in cells with a filamentous phenotype that lack septum formation.
suggesting that this homolog is involved in cell division. The three-dimensional X-ray crystallographic structure of Maf has been determined both on its own and bound to dUTP, displaying some structural similarity to Maf and the *Methanococcus jannaschii* Mj0226 dNTP pyrophosphatase and to *E. coli* YjjX. Notably, when tested, Maf did not show any ability to cleave phosphate bonds in either dUTP or dATP. Given its ability to bind nucleotides and its similarity to Mj0026, we suspect that YhdE might be able to cleave phosphate bonds of ribonucleotides, or other deoxyribonucleotides thus far untested. Structural and biochemical analyses of YhdE are expected to resolve the function of this protein in cell division.

In summary, analyses of these kinases and phosphatases structures are anticipated to enhance our understanding of kinases and phosphatases in terms of evolution and classification, protein substrate recognition, signal transduction pathways, protein engineering, and catalytic domains and overall folds. These four targets, YihE, AceK, YjjX, YhdE, were expressed, purified and then crystallized under various conditions. The structures of YihE, AceK, YjjX and YhdE, as well as the complex structure of AceK with its substrate ICDH, have been determined and the biochemical functions of YihE, YjjX, and YhdE have been explored.
Chapter 2

Crystal structure of a novel prokaryotic Ser/Thr kinase and its implication in the Cpx stress response pathway

Preface: This chapter was published in Molecular Microbiology

Jimin Zheng, Chunhua He, Vinay Kumar Singh, Nancy L. Martin and Zongchao Jia


Jimin Zheng was responsible for YihE and its mutants’ expression, purification, crystallization, data collection, structure solution and refinement, and some kinase activity experiments in the investigation of YhdE and its mutants’ function. The yhdE gene and its mutants were cloned by John Wagner from Biotechnology Research Institute, Canada. Dr. Nancy L. Martin’s Lab, our collaborator from Department of Microbiology and Immunology, Queen’s University, carried out the enzyme kinetics analysis and the phenotype analysis of YihE and its homologue, RdoA from Salmonella enteric. The manuscript was written by Jimin Zheng and Dr. Nancy L. Martin with editorial input from Dr. Zongchao Jia.
2.1 Abstract

The Cpx signalling system of *Escherichia coli* and *Salmonella enterica* senses extracytoplasmic stress and controls expression of factors that allow the bacterium to adapt to these stressors and thereby enhance survival. Many of the Cpx-responsive genes products are of unknown function. We determined the crystal structure of one of these gene products, called YihE in *E. coli*, which exhibits a eukaryotic kinase fold. Functional assays established that both YihE and the *S. enterica* YihE homologue, RdoA, undergo autophosphorylation and phosphorylate protein substrates at Ser/Thr residues *in vitro*, demonstrating that YihE/RdoA is a novel Ser/Thr protein kinase in prokaryotic cells. Phenotypic analysis of *yihE/rdoA* null strains indicates that this kinase is most abundant in stationary phase, and is important for long-term cell survival and for expression of surface appendages in both a Cpx-independent and -dependent manner. YihE/RdoA is therefore a previously unknown kinase component of a new type of bacterial phosphorelay mechanism, adding kinase activity as another response to the Cpx sensing system that functions to maintain cellular homeostasis.
2.2 Introduction

Bacteria experience rapid, sometimes extreme, changes in their environment while establishing infection in a new host or moving from one host to the next. Sensory systems capable of detecting changes in chemical concentrations, light, viscosity, osmolarity, temperature, and the presence of a host organism allow bacteria to respond by either moving to a more favourable location or adapting to the immediate surroundings. The primary bacterial sensing mechanism is a two-component system (TCS) consisting of a sensor protein (histidine kinase) and response regulator (phosphorylated DNA binding protein) \(^{91,92}\). Examples of bacterial multistep phosphorelays also exist that control sporulation in *Bacillus* sp. \(^{93,94}\) and the Rcs phosphorelay in *Enterobacteriaceae*, which influences pathogenicity and biofilm formation \(^{95}\). Ser/Thr or Tyr phosphorylation is recognized as a signalling mechanism employed by prokaryotes and archaea, where both unique and eukaryotic-like kinases have now been documented \(^{96-98}\). More than 600 Ser/Thr/Tyr bacterial kinases have been identified through genome sequencing projects and homology-based comparisons with eukaryotic kinases and phosphatases \(^{98,99}\), and less frequently through direct experimental evidence. The first bacterial serine protein kinase structure solved was that of HPr kinase/phosphorylase (HPrK/P) from *Lactobacillus casei* and *Bacillus subtilis* \(^{100,101}\), which acts in catabolite repression as the sensor enzyme. Not found in *Escherichia coli* or other Gram negative bacteria, the structure of HPrK/P is also unlike typical eukaryotic Ser/Thr protein kinases \(^{102}\). Generally, in putative Ser/Thr kinases, only the nucleotide binding region and the core catalytic domain, known as the Brenner's motif (H-X-D-X4-N;\(^{48}\)), are highly conserved between eukaryotic and prokaryotic kinases \(^{103}\).
Understanding the impact of activating a TCS requires knowing which cellular components are up- or downregulated, as well as an analysis of how those cellular components contribute to increased cell fitness. *E. coli* and *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) use a modified TCS system to sense extracytoplasmic stress called the Cpx pathway, containing a periplasmic protein (CpxP) that normally acts to inactivate the Cpx pathway in addition to the canonical TCS sensor kinase and response regulator. This stress response system senses perturbations in the cell envelope, such as misfolded protein in the periplasm, and upregulates the expression of genes encoding proteins to degrade or modify the damaged protein. The Cpx system plays important physiological roles in protein folding, cell envelope integrity, and pathogenesis. One of the target genes activated by the Cpx pathway, *yihE* in *E. coli* and its *S. typhimurium* homologue, *rdoA* (96% sequence similarity), has been annotated as a putative homoserine kinase. In these bacteria and close relatives, the *yihE/rdoA* gene is located immediately upstream of *dsbA*, a disulphide oxidoreductase that plays an important role in protein folding in many Gram negative bacteria. *YihE/RdoA* is co-transcribed with the disulphide oxidoreductase DsbA in *E. coli* and affects DsbA levels post-transcriptionally in *S. typhimurium*, therefore even though DsbA is commonly referred to as a Cpx regulon member, Cpx-dependent *dsbA* expression is associated with YihE/RdoA. In *S. typhimurium*, RdoA is also implicated in the control of flagellar phase variation. Microarray analysis of a *yihE* null strain of *Shigella flexneri* showed changes in transcript levels of many genes, where one of the most affected transcripts originated from the *galETK* operon and led to inefficient production of UDP-glucose, thereby affecting lipopolysaccharide (LPS) biosynthesis. Ultimately, the wide range of genes affected by loss of *yihE* in *S. flexneri* and its Cpx-dependent regulation point to an important role for YihE/RdoA in normal cell functioning with a need for
additional YihE/RdoA activity when the cell is under stress. We have found highly conserved yihE/rdoA homologues in the genomes of over 120 different bacteria representing 46 genera.

In the present work, we have determined that the crystal structure of the YihE protein shows a kinase-like fold similar to choline kinase and aminoglycoside phosphotransferase. In vitro functional assays demonstrate that YihE/RdoA is a novel Ser/Thr protein kinase. Phenotypic analysis of yihE/rdoA strains indicated YihE/RdoA is involved in the regulation of bacterial adhesion in association with the Cpx stress response. YihE/RdoA therefore participates in a new type of bacterial phosphorylation mechanism, combining His/Asp phosphorylation in sensing stress (TCS) and the transcriptional upregulation of a Ser/Thr protein kinase to maintain cellular function. To our knowledge, YihE is the first E. coli Ser/Thr protein kinase structure to be elucidated and it is the first demonstration of a Ser/Thr kinase involved in the Cpx stress response.

2.3 Results and discussion

2.3.1 Structure determination and analysis

The final structure of YihE was determined at 2.8 Å resolution by the method of single-wavelength anomalous dispersion (SAD). There is one YihE molecule in the asymmetric unit, in agreement with gel filtration and dynamic light scattering data which show that YihE is a monomer in solution (data not shown). The final $R_{\text{factor}}$ is 21.5% and $R_{\text{free}}$ is 27.9% respectively. Only the N-terminal three residues are disordered. Detailed data and refinement statistics are provided in Table 2.1. The Ramachandran analysis shows 98% of residues in the most favoured regions and the remaining 2% of residues in allowed area as defined by PROCHECK.
2.3.2 Overall structure

The overall structural architecture of YihE is that of a typical bilobal protein kinase with a smaller, predominantly β-sheet domain in the N-terminal region, and a larger, predominantly α-helical domain in the C-terminal region (Figure 2.1A). The two domains are linked by a short (∼10 residues) hinge, creating an open cleft (cleft 1, Figure 2.1A) that corresponds to the phosphotransfer region in all kinases. The N-terminal domain begins with an eight-residue loop, followed by helix A that lies above a well-conserved kinase structural motif composed of a five-stranded twisted antiparallel β-sheet (Figure 2.1A). Helix B encloses cleft 1 to form a pocket, believed to be the ATP binding site. This N-terminal domain structure is typical of eukaryotic kinases, constituting part of the 'essential' kinase fold present in all protein kinases.\textsuperscript{117}

The rest of the essential kinase structure is within the C-terminal domain. It is mainly composed of α-helices that can be subdivided into two lobes surrounding a second larger cleft (cleft 2, Figure 2.1A). The left lobe consists of four α-helices and a long hairpin-shaped loop that incorporates four short stretches of antiparallel β-strands (Figure 2.1). The hairpin loop contains many of the conserved, functionally important residues of typical protein kinases. For example, strands 9 and 10 house a conserved phosphotransferase domain, the Brenner's motif [HNDX\textsubscript{4}N]. A five-helix bundle is found in the right lobe. Together the two lobes create an approximately 28-Å-long, 23-Å-wide and 18-Å-deep cleft (cleft 2, Figure 2.1A), which is a putative substrate binding site.

The linker region connecting the two domains reflects a need for mechanical flexibility in protein kinases to facilitate the entering of ATP molecules. This hinge often contains less restricted amino acids such as glycine, exemplified by residues 108–115 in YihE (Figure 2.1A) akin to cAPK.\textsuperscript{118}
### Table 2.1. Crystallographic data and refinement statistics YihE structure

<table>
<thead>
<tr>
<th>Crystal</th>
<th>λ (Å)</th>
<th>D&lt;sub&gt;min&lt;/sub&gt; (Å)</th>
<th>Observed reflections</th>
<th>Unique reflections</th>
<th>% complete</th>
<th>I/σ (all/last shell)</th>
<th>% R&lt;sub&gt;merge&lt;/sub&gt; (all/last shell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se-Met crystal at inflation</td>
<td>0.979</td>
<td>2.8</td>
<td>97 262</td>
<td>20 263</td>
<td>100</td>
<td>4.8</td>
<td>8.8/39.8</td>
</tr>
</tbody>
</table>

**Refinement statistics**

- **Native P43212**
  - Cell dimensions: a = b = 90.97 Å, c = 110.42 Å (65–2.8 Å)
  - Unique/free reflections: 20 263/1921
  - R<sub>factor</sub>/R<sub>free</sub> (%): 21.5/27.9
  - Protein/water atoms: 2672/146
  - r.m.s.d. bond lengths (Å): 0.007
  - r.m.s.d. bond angles (°): 1.35

The summation is over all measurements.

\[
R_{\text{merge}} = \frac{\sum |I(k) - \langle I \rangle|}{\sum I(k)}, \quad R_{\text{factor}} = \frac{\sum ||F_o|| - ||F_c||}{\sum ||F_o||}, \quad R_{\text{free}} = \frac{\sum ||F_o^{\text{free}}|| - ||F_c^{\text{free}}||}{\sum ||F_o^{\text{free}}||}.
\]

\( F_o \) is the observed structure factor, \( F_c \) is the calculated structure factor based on the model. SAD data were collected in inverse beam mode. No σ cut-off was applied to the data and 10% of reflections were excluded from refinement for calculation of \( R_{\text{free}} \). Last shell is from 2.80 Å to 2.89 Å.
Figure 2.1 A) The overall structure of YihE. The N-terminal domain begins with an eight-residue loop, followed by helix A that lies above a well-conserved kinase structural motif (yellow) composed of a five-stranded twisted antiparallel β-sheet. There are two lobes in the C-terminal domain (red). B) Ribbon representation of the interaction between two symmetry-related YihE molecules in the crystal. The green-coloured C-terminal tail of one YihE is inserted into cleft 2 of the neighbouring YihE molecule, forming a natural protein–peptide complex structure.
2.3.3 Structure comparison

A structural homology search of YihE was performed using Dali and YihE was found to have structural similarity with protein kinases. Most interesting was the significant similarity with choline kinase (PDB 1NW1) from Caenorhabditis elegans, and with aminoglycoside 3'-phosphotransferase [APH(3')-IIIa, PDB 1J7I] from Enterococcus faecalis, with Z-scores of 16.9 and 14.1 respectively. Sequence alignments between YihE with either of these proteins exhibit only 17% identity suggesting a distant evolutionary relationship. Choline kinase participates in the phosphatidylcholine biosynthetic pathway, an important constituent of eukaryotic cell membranes that can be cleaved to produce a variety of second messengers. APH(3')-IIIa phosphorylates several aminoglycoside antibiotics at the 3' and/or 5' hydroxyl, inactivating them. Superimposing the YihE structure with choline kinase and APH(3')-IIIa showed that the root mean square deviations (r.m.s.d.) are 3.2 Å and 2.9 Å for Cα atoms respectively (Figure 2.2A), demonstrating that YihE belongs in the category of 'atypical' kinases with choline kinase and APH(3')-IIIa. Choline kinase and APH(3')-IIIa are both small molecule kinases with clear similarities to the ATP binding domain of eukaryotic protein kinases (ePKs) such as cAPK. However, APH(3')-IIIa has a small C-terminal domain with a small cleft compared with ePKs. Choline kinase has a very large and complex C-terminal domain with no obvious substrate binding cleft and possesses other features distinct from those of the typical protein kinases. In contrast, the C-terminal domain of YihE has a more open and larger cleft that would be able to bind larger substrates such as peptides. The conserved residue distribution in YihE is also different from those in choline kinase and APH(3')-IIIa, indicating that YihE would have different substrate specificity than either of these kinases. Specifically, the putative substrate binding cleft (cleft 2, Figure 2.1A) formed between two lobes within the C-terminal domain of
YihE is filled by a hairpin loop in choline kinase\textsuperscript{114}. Also the loop between β-strands 1 and 2 in choline kinase bends more into the cleft (cleft 1) compared with YihE. In general YihE has more open, wider and deeper clefts than choline kinase though the overall structural fold is similar. The larger opening in YihE is likely to be necessary for protein–protein interaction. Compared with YihE, APH(3')-IIIa lacks an elongated α-helix (helix K) and has a shorter helix I (Figure 2.1A), which makes a smaller lobe on the right side of the C-domain. Therefore the cleft (cleft 2) is much shorter than YihE. Although these differences seem small in comparison with the overall structure, they are most likely responsible for functional variations.

2.3.4 Kinase activity of YihE/RdoA

The structure of YihE clearly revealed that it has a typical kinase fold. To first test the ability of YihE/RdoA to autophosphorylate, RdoA and YihE were incubated with reaction mixtures containing [γ\textsuperscript{33}P]-labelled ATP in the absence of substrate. Both full-length MBP-RdoA and RdoA were labelled with [γ\textsuperscript{33}P] (81 kDa and 38.5 kDa bands, Figure 2.3A: lane 6), showing they were autophosphorylated. Two additional phosphorylated protein bands between 81 and 38.5 kDa are consistent with partly degraded MBP-RdoA bands. MBP-paramyosin was incubated with reaction mixtures containing [γ\textsuperscript{33}P]-ATP to exclude the effect of MBP and Factor Xa on the autophosphorylation assay. Neither cleaved MBP-paramyosin nor full-length MBP-paramyosin was phosphorylated (Figure 2.3A: lanes 5 and 7). Purified YihE was also shown to be autophosphorylated (Figure 2.3B) and both kinases were autophosphorylated on serine and threonine residues (Figure 2.3C), but not tyrosine residues (data not shown). Kinases are often activated via autophosphorylation, although YihE does not contain the canonical activation loop region often found in eukaryotic kinases. Several additional mechanisms where
Figure 2.2 A) Structural alignment of YihE (green) with choline kinase (magenta, right) and APH(3')-IIIa (cyan, left). B) Structural overlap of YihE (cyan) and APH(3')-IIIa (green; 1J7U) at the putative ATP binding site (see also Figure 2.1A). The side-chains of six highly conserved and catalytically important residues are highlighted as stick models. One-letter amino acid codes followed by a number is used for APH(3')-IIIa and amino acid codes in brackets correspond to YihE. AMPPNP molecules are shown as ball and stick models and the Mg$^{2+}$ ions as magenta spheres.
**Figure 2.3** Autophosphorylation and kinase activity of RdoA/YihE. A. Lanes 1–3: silver-stained SDS-PAGE of MBP–RdoA fusion protein before (lane 1, 20 µg protein) and after (lane 2, 2 µg protein) purification, and after cleavage with Factor Xa (lane 3, 10 µg protein). Lane 4: immunoblot of sample in lane 3 probed with anti-RdoA antibodies. Lanes 5–10: autorad of samples labelled with $[^{33}\text{P}]$-ATP. B. Lane 1: Purified YihE and myelin basic protein. Lanes 2 and 3: autorad of samples labelled with $[^{33}\text{P}]$-ATP. C. Immunoblots of samples from kinase assays probed with anti-phosphothreonine (left) or anti-phosphoserine (right) antibodies. a = MBP-RdoA; b = RdoA or YihE; c = myelin basic protein.
autophosphorylation activates kinase activity are known, such as displacement of an inhibitory protein or enhanced ability to interact with substrates\textsuperscript{120,121,121}. These mechanisms commonly occur through conformational changes induced by autophosphorylation of the kinase. The functional significance of autophosphorylation of YihE/RdoA awaits the outcome of mutagenesis studies to locate and inhibit the autophosphorylation of specific amino acid residues. When the kinase assay was carried out in the presence of myelin basic protein, a kinase substrate that has been shown to be phosphorylated by many eukaryotic and prokaryotic protein kinases, both RdoA and YihE were able to specifically transfer $[\gamma^{33}\text{P}]$ to the myelin basic protein (Figure 2.3A and B). Western immunoblotting using anti-phosphoserine or anti-phosphothreonine antibodies demonstrated that YihE/RdoA phosphorylates serine and threonine residues on myelin basic protein (Figure 2.3C). Although YihE clearly has Ser/Thr protein kinase activity, sequence analysis shows little similarity between YihE and ePKs, except for seven key conserved and three homologous residues that are dispersed over the sequence (data not shown). Specifically, the signature sequence H/Y-R-D-L/I-K-P-X-N for Ser/Thr kinases or H-R-D-L-R/A-A-A/R-N for tyrosine kinases, which is located between $\beta7$ and $\beta8$, are not conserved in YihE. Moreover, the G-X-G-X-X-G sequence (Walker A motif) between $\beta1$ and $\beta2$ that is often presented in ePKs, is absent in YihE, further obscuring the homology between YihE and protein kinases. There are, however, highly conserved regions that are predicted to participate in phosphotransfer. Figure 2.4 shows an alignment of the protein sequences of \textit{S. typhimurium} RdoA, \textit{E. coli} YihE, and \textit{S. flexneri} RdoA along with 12 predicted YihE homologues from a wide range of bacteria. The largest and most well conserved region corresponds to residues 196–235 of RdoA and includes a putative ATP binding domain, a Brenner's motif that is found in both aminoglycoside phosphotransferases and protein kinases\textsuperscript{48}. 
Figure 2.4 Protein sequence alignment of YihE, RdoA, and homologues. The protein sequences of 15 YihE homologues were aligned using clustalw. Only those residues different from those in three of the *S. typhimurium*, *E. coli*, and *S. flexneri* sequences are shown. Blanks indicate the same amino acid as the upper three sequences. **S.ty**, *S. typhimurium*; **E.co**, *E. coli*; **S.fl**, *S. flexneri*; **E.ca**, *Erwinia carotovora*; **Y.me**, *Yersinia medievalis*; **V.pa**, *Vibrio parahaemolyticus*; **P.ha**, *Pseudoalteromonas haloplanktis*; **S.on**, *Shewanella oneidensis*; **P.ae**, *Pseudomonas aeruginosa*; **G.me**, *Geobacter metallireducens*; **T.de**, *Thiobacillus denitrificans*; **B.fu**, *Burkholderia fungorum*; **R.ge**, *Rubrivivax gelatinosus*; **B.ba**, *Bdellovibrio bacteriovorus*; **L.co**, *Leptospira interrogans sevar Copenhageni*. 
2.3.5 Proposed ATP binding site

The present YihE structure does not contain bound substrate; however, the high degree of structural similarity between YihE and APH(3')-IIIa makes it possible to postulate where the ATP binding site is located and to identify key residues that play a role in interaction with ATP for phosphotransfer. Based on the structural alignment between YihE and APH(3')-IIIa in complex with AMPPNP (PDB 1J7U)\textsuperscript{115}, these molecules have very similar ATP binding pockets.

The superimposition of the proposed ATP binding segments in YihE and APH(3')-IIIa gives rise to an r.m.s.d. for main chain atoms of 1.5 Å (Figure 2.2B). This same region has been shown to bind ATP in all known structures of ePKs. The conserved phosphotransferase Brenner's motif [197-RLHGDCAG-206] is located on the connecting loop between strands 7 and 8 of YihE (Figure 2.1A), with the corresponding sequence motif also found in APH(3')-IIIa [186-FSHGDLGDSN-195]. In APH(3')-IIIa, residue H188 (YihE:His199) from this motif plays a structural role in forming the ATP binding site. The histidine residue hydrogen bonds with the main chain amide of D190 (YihE:Asp201) and the main chain carbonyl of I207 (YihE:Val216). These two sets of interactions help orient the side-chain of the catalytically important residues D190 (YihE:Asp201) and D208 (YihE:Asp217). D208 (YihE:Asp217) directly co-ordinates two magnesium ions for catalysis. Mutation of YihE Asp217 to alanine caused a complete loss of kinase activity (Figure 2.5) as would be expected for a residue involved in positioning cations that are required for ATP activity. The kinase activity of YihE was also compared with that of protein kinase C (PKC), a kinase known to phosphorylate myelin basic protein\textsuperscript{122}. Under the conditions tested, PKC was only three times more active than YihE lending further evidence to YihE's ability to act as a protein kinase. Additional residues can be predicted to be involved in phosphotransfer although conformation of their roles in catalytic activity will be provided in
Figure 2.5 Quantification of kinase activity of YihE and PKC. ADP generation was measured and is indicated as a change in relative fluorescence per min per ng of kinase. D217A is a mutant of YihE where residue 217 is changed to alanine.
future mutational analyses carried out when the native substrate for YihE has been identified. The conserved Asp residue (D190) in APH(3')-IIIa (YihE:Asp201) acts as a catalytic base for deprotonation of the substrate hydroxyl group, allowing for efficient attack of the γ-phosphate of ATP. N195 (YihE:Asn206) of the Brenner's motif also co-ordinates one of the magnesium ions. Finally, there is an invariant Ser residue (S27) in APH(3')-IIIa (YihE:Ser36) that directly interacts with the β-phosphate of ATP through the Ser hydroxymethyl group during formation of the metaphosphate-like transition state (Figure 2.2B) \(^{123}\). Previous work demonstrated that APH(3')-IIIa operates by a Theorell-Chance kinetic mechanism \(^{124,125}\). This sequential mechanism consists first of ATP binding followed by binding and direct phosphorylation of the substrate, release of the phosphorylated substrate, and rate-limiting dissociation of ADP. Further experimentation will be necessary to determine if the similarity of spatial arrangement of these residues in YihE to those in APH(3')-IIIa (Figure 2.2B) allows YihE to bind ATP and catalyse phosphotransfer reactions utilizing a similar mechanism.

2.3.6 Substrate binding site

Electrostatic surface potential analysis of YihE has shed some light on potential substrate interaction. Of particular interest was the large open cleft (cleft 2) positioned between two structural lobes in the C-terminal domain (Figure 2.1A). This cleft is connected to the front end of the nucleotide binding pocket, close to the predicted position of the γ-phosphate of ATP which is negatively charged, as a result of the conserved Asp clustering (Asp201, Asp217, Asp219, Asp220) (Figure 2.6A). The cleft extends far inside the molecule to form a deep pocket beside the ATP binding pocket in which the conserved residue Arg222 sits at the bottom of the pocket. Sequence alignment analysis of over 100 YihE homologues from Archaea and other prokaryotic organisms (data not shown) showed that a number of amino acids are highly conserved in this
cleft region of the structure, including Arg270, Arg273, Tyr277, Trp280 from helix I, Gln228 and Met232 from helix G; Ser 36, Tyr37 and Glu38 for the loop between β1 and β2. In comparison with the corresponding area of APH(3')-IIIa, the cleft in YihE is much less negatively charged, again indicating a very different type of substrate interaction.

Due to crystal packing, the C-terminal six-residue tail (321-LQLTPMY-328) of one molecule makes a sharp bend (~90°) from the preceding segment (Figure 2.1B) and inserts into cleft 2 of the symmetry-related molecule, forming a natural protein–peptide complex (Figure 2.6B). The peptide is surrounded by conserved residues in the cleft, including Ser36, His203, Arg273, Tyr277 and Trp280. The side-chain of peptide residue Met327 is inserted deep into the pocket, positioning the C-terminal Tyr328 near the γ-phosphate of ATP. The features of the pocket in cleft 2 and the insertion of the Met327 side-chain suggest that this deep pocket may play a role in recognizing long side-chains, such as lysine, and help locate the substrate hydroxyl in the proper position for phosphorylation. This natural protein–peptide complex structure provides strong evidence that YihE can recognize and bind to peptides and that the large cleft in the C-domain is the potential substrate binding site.

2.3.7 In vivo characterization of YihE/RdoA

RdoA is expressed at all growth phases, but is most highly expressed in stationary phase (Figure 2.7A), indicating the RdoA plays a greater role when cells are less actively dividing. RdoA and YihE are localized to the cytoplasm (data not shown), a location consistent with a need for ATP as a phosphate donor. Both YihE and RdoA are more highly expressed upon activation of the Cpx stress response pathway as a result of binding of phosphorylated CpxR to a conserved DNA binding motif \(^{126}\) in the \(yihE/rdoA\) promoter region \(^{63,107}\) and loss of
Figure 2.6 A) Electrostatic surface potentials of YihE reveal the substrate binding interfaces. Blue and red colours indicate positive and negative electrostatic potentials calculated using GRASP (Nicholls et al., 1991) respectively. AMPPNP from APH(3')-IIIa complex structure (1J7U) was placed in cleft 1 by superimposing the structures, which indicates ATP binding pocket in YihE. B) The C-terminal tail and another YihE molecule form a protein–peptide complex. The green mesh covers the six-amino-acid peptide. The side-chain of Met327 in the peptide is anchored in the deep pocket in cleft 2.
RdoA function causes slight activation of the Cpx pathway (data not shown). Together these results indicate that YihE/RdoA are necessary for 'normal' cell functioning, but are also important components of the Cpx stress response. Many of the well-studied regulon members that are positively controlled by phosphorylated CpxR, such as DsbA, HtrA, are also always present in the cell and act as post-translational modifiers, as does YihE. Long-term cell survival studies showed that the Cpx pathway is not essential, as previously shown\textsuperscript{127}, but loss of RdoA function substantially lessens survival by 3 log orders (Figure 2.7B). Basal levels of RdoA, expressed independently of CpxR, appear to normally allow growth under starvation conditions; however, the \textit{rdoA cpxR} strain survives even less well than \textit{rdoA}, suggesting that RdoA can act upon or with a component expressed under tight control of the Cpx pathway to enhance survival under starvation conditions. Alternatively, RdoA may affect a regulatory protein which is normally involved in surviving starvation, but the lack of the Cpx response and this regulatory protein makes survival more difficult.

Curli and other cell surfaces appendages such as pili play a role in biofilm formation and consequently in the ability of bacteria to establish infection\textsuperscript{128}. It was previously demonstrated that curli expression is regulated in part via the Cpx pathway where phosphorylated CpxR acts as a repressor at the \textit{csgD} promoter\textsuperscript{129,130}. Here it is shown that RdoA also affects curli expression. Both the \textit{rdoA} and \textit{cpxR} mutants allow expression of similar levels of curli under normally non-inducing growth conditions while curli levels in these mutants are not significantly different from wild type under curli-inducing conditions (Figure 2.7C). 'Cpx ON' refers to a strain containing the NlpE expression plasmid pND18\textsuperscript{63,107}. Overexpression of NlpE artificially turns on the Cpx pathway and here this strain is used as a control to show curli levels when the Cpx
pathway is stimulated. Combining the \textit{rdoA} and \textit{cpxR} null alleles in the curli assay shows that slightly higher levels of curli are made, suggesting that the mechanism by which CpxR normally represses curlin synthesis is distinct from an RdoA-mediated mechanism. This added level of control may allow curlin expression to be more rapidly or fully turned off as environmental conditions dictate.

In addition to those effects described here, YihE/RdoA affects LPS biosynthesis in \textit{S. flexneri} \textsuperscript{131}, and flagellar phase variation \textsuperscript{107}. The effect on LPS synthesis occurs through the \textit{galETK} operon\textsuperscript{65}, which is downregulated in the absence of YihE in \textit{S. flexneri}. Many additional genes were shown to be up- or downregulated in a YihE null strain\textsuperscript{65}, but only the effect on the \textit{galETK} operon has been followed up to date. Such a broad repertoire of phenotypes suggests that either RdoA interacts with several different targets, which seems somewhat unlikely, or with a protein with broad regulatory functions that then controls several different cellular processes. Future experimental approaches will address the nature of RdoA function by focusing on determining the natural substrate of this kinase. Although YihE is structurally similar to APH(3')-IIIa, neither \textit{E. coli} nor \textit{S. typhimurium} are naturally resistant to aminoglycosides and overexpression of YihE/RdoA does not confer resistance (data not shown), indicating that inactivation of aminoglycosides is not the target of YihE/RdoA activity. The large differences in the substrate interaction domain between APH(3')-IIIa, choline kinase, and YihE, and the YihE peptide interaction characteristics strongly support protein rather than small molecule kinase activity. YihE/RdoA is therefore a novel signal transduction regulator that has adopted a eukaryotic-like signal transduction mechanism: by phosphorylating Ser/Thr residues on its substrates, YihE/RdoA is able to affect gene expression and/or protein function.
Figure 2.7 RdoA expression and rdoA, cpxR null mutant analysis. A. SDS-PAGE of whole cell lysates corresponding to indicated OD_{600} (upper panel) and immunoblot (lower panel) probed with anti-RdoA antibodies (30 µg protein per lane). B. Long-term survival assay showing percent viable bacteria compared with day 1 over 20 days. C. Analysis of curlin production per OD unit as measured by change in absorbance due to cell-associated binding of Congo Red under inducing (28°C) and non-inducing (37°C) growth conditions. Cpx ON is strain SL1344 producing NlpE in order to activate the Cpx pathway.
Ser/Thr protein kinases have not been extensively characterized in prokaryotic cells. The study of phosphorylation-mediated signal transduction in prokaryotic cells has focused mainly on TCS systems as these mechanisms provide a way to move signals across the bacterial cell membrane. As the details of signalling pathways are uncovered, it is obvious that even 'simple' bacteria use signalling networks rather than linear paths of signal transduction. The combination of a TCS system with a Ser/Thr protein kinase has been recently found in Myxococcus xanthus in the regulation of a transcription factor that is essential for differentiation. YihE/RdoA functions represent a branch of the Cpx signal transduction response that extends the Cpx signal beyond traditional TCS-mediated gene regulation.

2.4 Experimental procedures

2.4.1 Cloning, expression and purification

N-terminal cleavable (TEV protease) His<sub>6</sub> tagged YihE was cloned by standard procedures. Recombinant YihE was expressed in BL21 E. coli and purified to near-homogeneity using a two-step protocol, Ni<sup>2+</sup>-affinity and size exclusion. Selenomethionine-substituted YihE was expressed in DL41 E. coli in LE Master medium supplemented with 50 mg l<sup>−1</sup> of D-L-selenomethionine (Fisher-Acros, USA) and purified as for the native protein. The final yield was ~20 mg of native protein and ~10 mg of selenomethionine labelled protein per litre E. coli culture. RdoA was cloned and expressed in E. coli K-12 TB1 using the cytoplasmically expressed pMAL expression system and purified as described by the manufacturer (New England Biolabs, USA). PCR-based site-directed mutagenesis was used to create the D217A His<sub>6</sub>-tagged YihE mutant. This mutant was expressed and purified as for wild type and its
stability was checked using differential scanning calorimetry where it was shown to be as stably folded as the wild-type YihE.

2.4.2 Crystallization and diffraction data collection

The N-terminal His$_6$ tag of YihE was cleaved using AcTEV protease and the resultant protein was further purified to remove the His$_6$ tag peptide and AcTEV protease prior to crystallization according to the manufacturer (Invitrogen, Canada). Preliminary crystallization conditions were screened by the sparse matrix method$^{134}$. Optimal crystallization conditions for the both native and selenomethionine proteins were 0.1 M Tris-HCl buffer at pH 8.5 containing 0.8–1.0 M lithium sulphate as precipitating agent at room temperature. Protein concentration was 10 mg ml$^{-1}$ and hanging drop vapour diffusion was used. Diffraction data were collected at 100 K. Diffraction data were collected at the beamline X29 at Brookhaven National Laboratory using an ADSC QUANTUM 4 CCD detector. The space group is tetragonal P4$_3$2$_1$2 with cell dimensions of $a = b = 90.97$ Å, $c = 110.42$ Å.

2.4.3 Structure determination and refinement

Single-wavelength anomalous dispersion data collected at the inflection wavelength permitted the heavy-atom positions to be located and initial phases determined by SOLVE$^{135}$. The initial YihE structure was partially traced by RESOLVE$^{136}$. Additional manual tracing and model building were carried out using Xfit in XtalView$^{137}$. Crystallography & NMR System (CNS;$^{138}$ was used for refinement.

2.4.4 Kinase activity assays
YihE, MBP-RdoA and MBP-paramyosin (New England Biolabs, USA) were used. The His$_6$ tag was cleaved from YihE while MBP-RdoA or MBP-paramyosin was partially cleaved with Factor Xa. Phosphorylation of myelin basic protein was used to assay kinase activity as described previously \cite{139} using 0.1–5 µg of kinase, 5–15 µg of myelin basic protein and 5 µCi [$\gamma^{33}P$]-ATP per reaction. Reactions were incubated at 20°C for 30 min and terminated by adding 6× SDS-PAGE loading buffer. Autophosphorylation activity assays were set up identically, excluding myelin basic protein. Reactions were electrophoresed on 12% polyacrylamide gels and stained using colloidal Coomassie \cite{140} or silver \cite{141}, or air dried and signals digitally captured using a Personal Molecular Imager FX (Bio-Rad, USA). Quantification of kinase activity for YihE and PKC (Sigma P-1609) was carried out using the ADP Quest General Purpose Kinase Assay (DiscoverX, CA). Reactions were set up as recommended by the manufacturer with myelin basic protein as the phosphate-accepting substrate (in excess). The assays were incubated at 37°C. The rate of ADP generation was calculated as the change in relative fluorescence units per min per ng kinase.

**2.4.5 Western immunoblotting**

Anti-RdoA antibodies were raised in chickens and used to detect RdoA from whole cell lysates. Detection of phosphoserine and phosphothreonine residues was carried out using a 1:100 dilution of polyclonal anti-phosphoserine or anti-phosphothreonine antibodies (Stressgen Bioreagents, Canada) after performing kinase reactions as described above using cold ATP. Proteins were electrophoresed, transferred to nitrocellulose and detected using the SuperSignal West Pico chemiluminescent substrate (Pierce, USA).
2.4.6 Long-term cell survival

Overnight cultures of SL1344 (wild-type), SL1344rdoA, SL1344cpxR and SL1344rdoA
\textit{cpxR} were diluted to 1:100–1:120 (based on \textit{OD}_{600} readings) to equalize starting cell
concentrations in 50 ml fresh Luria–Bertani (LB). Cultures were then incubated at 37\(^{\circ}\)C with
aeration for 20 days. Periodically, 1 ml of culture was removed and serial dilutions plated on LB
plates to determine the \textit{cfu ml}^{-1}. Viable counts were compared with day 1 to calculate the per
cent survival.

2.4.7 Curlin assay

Quantitative curlin assays were carried out as described previously\textsuperscript{142} using cells grown at
22\(^{\circ}\)C (inducing conditions) for 48 h or 36\(^{\circ}\)C (non-inducing conditions) for 24 h. SL1344 (wild-
type), SL1344\textit{rdoA}, SL1344\textit{cpxR}, SL1344\textit{rdoA cpxR}. SL1344 pND18 (\textit{NlpE} expressing plasmid)
\textsuperscript{107} was used as the negative control for curlin expression.

Acknowledgements

The authors would like to thank A. Manu-Boateng for assistance with kinase assays. Support
from Drs. M. Cygler, A. Matte, and J. Wagner is greatly appreciated. Canadian Institutes of
Health Research and Natural Sciences and Engineering Research Council of Canada provided
funding. V.K. Singh is the recipient of a Canadian Institutes of Health Research fellowship in
Transdisciplinary Cancer Research. Z.J. is a Canada Research Chair in Structural Biology.
Chapter 3

Structure of *Escherichia coli* isocitrate dehydrogenase kinase/phosphatase and its complex with isocitrate dehydrogenase

*This chapter will be submitted for publication*

Jimin Zheng and Zongchao Jia*

Department of Biochemistry, Queen’s University, Kingston, ON, K7L 3N6, Canada

*Correspondence e-mail: jia@queensu.ca

Jimin Zheng was responsible for protein expression, purification, crystallization, data collection, structure solution and refinement, and all biochemical experiments in the investigation of AceK function. The *aceK* and *icdH* genes were cloned by John Wagner. The manuscript was written by Jimin Zheng with editorial input from Dr. Zongchao Jia and with additional help from Brent Wathen.
3.1 Abstract

The *Escherichia coli* (*E. coli*) isocitrate dehydrogenase kinase/phosphatase (AceK) (EC 2.7.11.5) is a unique bifunctional protein that phosphorylates and dephosphorylates isocitrate dehydrogenase (ICDH), resulting in the inactivation and activation of ICDH, respectively. Responding to changes in the environment, this reversible phosphorylation directs isocitrate, an intermediate of the citric acid cycle, to either go through the full Krebs’ cycle or enter the glyoxylate bypass, thereby allowing *E. coli* to survive in both nutrient-rich and nutrient-poor conditions. Because of this prominent role in energy acquisition, understanding how AceK mediates *E. coli* metabolism in glucose-limited conditions is critical for developing a novel strategy to avoid water- and food-borne *E. coli* contamination. To clarify the regulatory mechanism of AceK in the central metabolism of *E. coli*, we determined both the structure of AceK as well as the complex structure of AceK with ICDH. The AceK structure reveals an eukaryotic protein kinase fold domain coupled with a novel regulatory domain. We identified two loops, one for substrate recognition and another for regulation, which may play important roles in substrate specificity and switching the enzymatic activity between phosphorylation and dephosphorylation. From our analysis, we propose a regulatory model that governs AceK activity switch that is compatible with previously reported biochemical data.
3.2 Introduction

*E. coli* is commonly found in the large intestine of warm-blooded animals; it is also able to survive in both soil and water, making it an ideal indicator organism to test environmental samples for fecal contamination of (drinking) water. In contrast to nutrient-rich environment of the large intestine, growth conditions in aqueous environments are much harsher, and concentrations of utilizable carbon sources such as glucose are very low. Hence, *E. coli* growth outside of a host is generally carbon/energy-limited. To successfully adapt to constant or rapid changes in its environment, most notably the change from “feast” to “famine” growing conditions, *E. coli* has evolved a series of mechanisms to monitor its surroundings and adjusts its physiology accordingly. By way of these mechanisms, *E. coli* is able to switch between a physiological program that permits rapid growth in the presence of abundant nutrients to one that enables survival in the absence of such nutrients.

One such switch, the “acetate switch,” was defined as the point at which the organism switches its enzymatic machinery from acetate excretion to acetate utilization. This switch is triggered based on growth conditions. For example, when *E. coli* grow on carbon sources containing more than two carbon atoms per molecule, such as glucose, the citric acid cycle (the Krebs’ cycle) is exploited to completion, resulting in acetate excretion. However, when the primary carbon source is fully consumed and only two-carbon molecules such as acetate remain as the sole source of carbon, the acetate switch is turned on, resulting in direct acetate uptake and use in the formation of acetyl-CoA. This latter mechanism is catalyzed by acetate kinase (AK), phosphotransacetylase (PTA), and/or acetyl-CoA synthetase (ACS). The subsequent oxidation of acetyl-CoA through the Krebs’ cycle requires a bypass to avoid complete oxidative decarboxylation that would result in the net loss of all carbon atoms as CO₂. This bypass, termed
the glyoxylate cycle, avoids two steps in the Krebs’ cycle that are responsible for this carbon to CO₂ transformation (see Figure 1.2 in Chapter 1). Under nutritionally challenging conditions, the glyoxylate bypass is activated to produce the four-carbon metabolic intermediates succinate and malate for subsequent biosynthesis at the cost of energy production.\textsuperscript{33,34,147}

The acetate switch is made up of two components. The first is the competition for isocitrate between isocitrate dehydrogenase (ICDH) and isocitrate lyase (ICL). The second component is the competition for coenzyme A (HS-CoA) between α-ketoglutarate dehydrogenase (α-KGDH) and PTA.\textsuperscript{35} HS-CoA diverted through α-KGDH results in energy generation, while HS-CoA diverted through PTA leads to biosynthesis.\textsuperscript{35} For efficient growth, the routing of HS-CoA and isocitrate into energy generating and biosynthetic pathways must be tightly synchronized. Though the regulation of HS-CoA utilization is not understood, the regulation of isocitrate at the isocitrate branch point has been well characterized, with one branch leading to the Krebs’ cycle and the other leading to the glyoxylate bypass. The bypass, therefore, is engaged when ICL out-competes ICDH for isocitrate.\textsuperscript{33,34,147} However, as an active enzyme, ICDH has a much higher affinity for isocitrate than ICL does. Therefore, in the absence of any regulatory steps, much of the isocitrate will go through the full citric acid cycle. The central feature in the activation of the acetate switch is the inactivation of a large fraction of ICDH, which results in increased levels of isocitrate that then promotes the use of the glyoxylate bypass.\textsuperscript{33} As the first discovered (in)activation regulation system in prokaryotes that uses reversible phosphorylation,\textsuperscript{147} ICDH is modified by a unique bifunctional regulatory enzyme, ICDH kinase/phosphatase (AceK), which is expressed in conjunction with two other essential enzymes (ICL and malate synthase (MS)) that are part of the glyoxylate bypass in many bacteria (see Figure 1.3 in Chapter 1).\textsuperscript{36-38}
AceK, a 66.5kDa protein that exists as a homodimer inside the cell \(^{148,149}\), is significantly larger than typical eukaryotic protein Ser/Thr kinases, which contain only 250-300 residues. Moreover, AceK does not share significant sequence homology with any eukaryotic protein kinases or phosphatases. However, it contains a eukaryotic-kinase-like ATP-binding motif, including the highly conserved Lys336 residue \(^{40,150}\). While the importance of AceK is highlighted in all introductory biochemistry textbooks, which usually devote one or two full pages to its role in the glyoxylate bypass, currently little else is known about AceK, largely due to the lack of structural information.

One of the most interesting features of AceK is that it possesses both kinase and phosphatase activity. This highly unusual bifunctionality feature has been observed for only one other protein (HPr kinase/phosphatase \(^{151}\)) thus far. While the structure of HPr kinase/phosphatase is available \(^{152}\), the regulatory structural mechanism between its kinase and phosphatase activity remains unresolved. In addition to possessing both kinase and phosphatase activities, AceK also exhibits ATPase activity \(^{149}\). Whereas most kinases typically exhibit weak ATPase activity, the ATPase activity of AceK is 2.5 and 7 times greater than its kinase and phosphatase activities, respectively. Given this unusually strong ATPase activity, it has been postulated that AceK may ultimately be involved in the mechanistic control of its own kinase/phosphatase bifunctionality \(^{40}\). Indeed, AceK’s phosphatase activity is dependent on ATP/ADP \(^{38}\). Previous evidence has shown that all three functions occur at the same active site \(^{153}\). It is currently unknown whether the active site undergoes a conformational change as it switches between kinase and phosphatase activities. However, one previous model postulates that the ICDH phosphatase employs the kinase back reaction as part of its phosphatase reaction. Catalysis of both reactions at the same site would allow many of the same amino acid to be used
for the recognition of two similar substrates\textsuperscript{153}.

Due to the importance of the citric acid cycle and its bypass, there is an intricate network of regulators that act not only upon AceK (Table 3.1), but also on the substrate (ICDH) to provide another level of regulation\textsuperscript{154}. For example, both AMP and ADP, two molecules which represent energy depletion, act directly on AceK to stimulate its phosphatase activity and inhibit its kinase activity. Subsequently, the resulting dephosphorylated ICDH acts to drive energy production. This feature is similar to the enzyme AMP-activated protein kinase (AMPK), which regulates cellular metabolism in response to the availability of energy in the human body. The AMP/ATP ratio regulates how AMPK acts on it downstream targets\textsuperscript{155,156}. Compared to AMPK, AceK appears to have a more complicated regulatory mechanism for sensing changes in cell metabolism. The ICDH cofactor NADP\textsuperscript{+}, for example, can induce a conformational change in ICDH that drastically reduces its susceptibility to AceK phosphorylation/dephosphorylation\textsuperscript{157,158}.

Unlike most other regulatory phosphorylations that typically induce conformational movements at allosteric sites\textsuperscript{159}, the phosphorylation of ICDH by AceK occurs near the active site of ICDH at S113\textsuperscript{157,159}. The crystal structures of the dephosphorylated\textsuperscript{160} and the phosphorylated forms of ICDH were both determined and showed that phosphorylation does not induce any significant conformational changes in the ICDH structure\textsuperscript{161}. The phosphorylation of S113 inactivates ICDH by blocking isocitrate binding, mainly by electrostatic repulsion but also secondarily by steric effects\textsuperscript{159,162}. One puzzle, however, that came from the ICDH structure was the fact that the phosphorylation site is actually largely buried within the protein, leaving the question of how AceK could reach this site and catalyze the transfer of phosphate from ATP to S113 or remove phosphate from phosphor-S113 unclear\textsuperscript{159}. It was hypothesized that in order to
bind to and transfer a phosphate to the S113 residue, IDCH may first have to undergo a conformation change around S113, either spontaneously on its own or induced by AceK, that would render S113 accessible for (de)phosphorylation. Another interesting feature of AceK is that in contrast to many cAMP-dependent protein kinases, AceK is not able to phosphorylate either proteolytic fragments derived from ICDH or a synthetic peptide corresponding to the sequence around the phosphorylation site; rather, it is only able to recognize the intact protein substrate, suggesting that substantial part of ICDH and AceK may be important for enzyme-substrate recognition and specificity.

Thus, while the importance of AceK for E. coli metabolism has been well established, many of the functional and structural details about this protein have not been clarified. Most importantly, there are two issues that need to be resolved: the structural mechanism by which AceK switches between kinase and phosphatase activity, and the enzyme-substrate recognition between AceK and ICDH. The resolution to these issues ultimately requires the structures of AceK on its own and in complex with ICDH. Thus far, despite thirty years of effort, AceK has not been successfully crystallized. Here, we report the novel structures of AceK in different modes and also the complex structure of AceK with its substrate, ICDH.

Our AceK structure reveals an eukaryotic kinase-like fold domain (KD) with an ATP molecule bound and buried, and a regulatory domain (RD) with a novel structural fold. AMP, an allosteric regulator, was observed in an allosteric regulatory pocket between these domains. From the complex structure, AceK and ICDH are found to interact directly through the active sites of the two proteins. A significant conformational change was observed in the phosphorylation loop of ICDH, moving it from the interior to the surface of ICDH. This conformation change renders the phosphorylation site more accessible to the active site of AceK.
However, the phosphorylation site still remains more than 13 Å away from the bound ATP molecule, suggesting that an even larger conformational change is required before a full understanding of the phosphor-transferring mechanism can be achieved.

3.3 Result and Discussion

3.3.1 Characterization and structure determination

The N-terminal hexahistidine-fusion constructs of full length AceK and ICDH were expressed in *E. coli*, and they were purified using Ni$^{2+}$-affinity and gel-filtration chromatography (Figure 3.1). AceK, in the presence of ATP, was crystallized into three different crystal forms, all in similar conditions (Figure 3.2). Three datasets were collected at CHESS and BNL (Table 3.2), with the best of the three diffracting to 2.9 Å. Based on SAD data, we were able to build and refine the AceK structure in a tetragonal space group (Figure 3.2). $R_{\text{cryst}}$ and $R_{\text{free}}$ were 21.3% and 26.8%, respectively. The asymmetric unit contained two AceK molecules. 566 out of 578 residues of each molecule were included in the structure, along with a total of 68 solvent molecules. An external loop (residues 497-503) could not be included because of weak or altogether absent electron density (Figure 3.3A). 98% of the residues are found within the most favorable regions of the Ramachandran plot, with the remaining 2% found in allowed areas, as defined by PROCHECK.$^{116}$

Previous studies have predicted that the interaction between AceK and ICDH involves extensive molecular contacts, and that ICDH might undergo a conformational change in the vicinity of its active site that fosters enzyme-substrate recognition.$^{165}$ This suggests that AceK and ICDH may form a stable complex *in vitro*, improving the likelihood of complex structure
Figure 3.1 Curve overlap diagram of size-exclusion purification of AceK, ICDH and the AceK+ICDH complex using the AKTA FPLC system (Amersham Biosciences) with a size exclusion Hiload Superdex S200 16/60 column in a buffer containing 20 mM Hepes pH 7.0, 2 mM DTT, 100 mM NaCl, and 10% glycerol. AceK$_2$ was pooled and concentrated to 5 mg/ml for crystallization trials. SDS-PAGE gel analysis shows purified AceK$_2$ protein after size-exclusion chromatography and fractions of AceK+ICDH peak.
Figure 3.2 A) Three forms ($P4_{1}2_{1}2$, $P3_{2}2_{1}$, $P2_{1}2_{1}2_{1}$) of AceK crystals and typical diffraction images for each. B) A crystal of AceK-ICDH complex and typical diffraction image. The components from the dissolved crystals were identified by SDS-PAGE gel analysis, which confirmed that the crystal contained both AceK and ICDH.
Table 3.1 Activators and inhibitors of AceK’s kinase/phosphatase activity.40,150,157

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Effects on AceK activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>isocitrate</td>
<td>↓ phosphatase</td>
</tr>
<tr>
<td>NAD⁺</td>
<td>↓ kinase</td>
</tr>
<tr>
<td>3-phosphoglycerate</td>
<td>↑ phosphatase</td>
</tr>
<tr>
<td>α-ketoglutarate</td>
<td>↓ kinase</td>
</tr>
<tr>
<td>AMP</td>
<td></td>
</tr>
<tr>
<td>ADP</td>
<td></td>
</tr>
<tr>
<td>citrate</td>
<td>↓ kinase</td>
</tr>
<tr>
<td>fructose 6-phosphate</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2 Statistics from X-ray diffraction data.

<table>
<thead>
<tr>
<th>Crystal form</th>
<th>Form A</th>
<th>Form B</th>
<th>Form C</th>
<th>AceK/ICDH complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.9789</td>
<td>0.9789</td>
<td>0.9179</td>
<td>0.9789</td>
</tr>
<tr>
<td>Space group</td>
<td>P4₁2₁₂</td>
<td>P3₂₁</td>
<td>P2₁2₁₂</td>
<td>P6₁</td>
</tr>
<tr>
<td>Unit-cell parameters (Å)</td>
<td>a = b = 124.61, c = 267.49</td>
<td>a = b = 152.71, c = 45.19</td>
<td>a = 64.20, b = 134.18, c = 187.23</td>
<td>a = b = 196.80, c = 156.46</td>
</tr>
<tr>
<td>Total number of reflections</td>
<td>580446</td>
<td>209544</td>
<td>67225</td>
<td>44356</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>38980 (3863)</td>
<td>49741 (5024)</td>
<td>21542 (2093)</td>
<td>12508 (1208)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30-3.2</td>
<td>20-2.9</td>
<td>50-3.0</td>
<td>30-2.7</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.3 (99.4)</td>
<td>94.4 (96.6)</td>
<td>93.7 (90.9)</td>
<td>93.2 (87.6)</td>
</tr>
<tr>
<td>&lt;I/σ&gt;</td>
<td>21.8 (2.4)</td>
<td>18.6 (2.3)</td>
<td>27.2 (2.6)</td>
<td>19.2 (2.6)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>14.9 (14.0)</td>
<td>4.2 (4.1)</td>
<td>3.1 (2.8)</td>
<td>3.5 (3.4)</td>
</tr>
<tr>
<td>Molecules per AU</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>10.4 (67.2)</td>
<td>9.7 (62.1)</td>
<td>7.6 (57.1)</td>
<td>10.2 (94.7)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>21.3/26.8</td>
<td>33.1/38.2</td>
<td>24.8/30.1</td>
<td></td>
</tr>
<tr>
<td>Solvent content</td>
<td>0.68</td>
<td>0.68</td>
<td>0.49</td>
<td>0.59</td>
</tr>
</tbody>
</table>

*Rmerge = \sum_{hkl} I_i(hkl)-\langle I(hkl)\rangle/\sqrt{\sum_{hkl} I_i(hkl)}* where I_i(hkl) is the intensity of the i^{th} observation of reflection hkl and \langle I(hkl)\rangle is the weighted average intensity of i observations of reflection hkl. Two datasets for crystal form A were collected on the A1 and F1 beamlines at CHESS with the wavelengths of 0.9789 Å and 0.9179 Å respectively. One dataset for crystal form B was collected on the A1 beamline and one dataset for crystal form C was collected on the F1 beamline. The AceK/ICDH complex dataset was collected at the A1 beamline. Values in parentheses are for the highest resolution shell.
determination. Attempts have been made in the past to co-crystallize AceK and ICDH \textsuperscript{163}, but none have been successful. In our size-exclusion experiment, both AceK and ICDH elute in the same fraction (Figure 3.1), with estimated molecular mass of approximately 110 kDa. SDS gel analysis of this fraction indicates that AceK and ICDH form a stable complex in solution. For co-crystallization trials, we used a mixture of full length purified AceK and ICDH with a 1:1 molar ratio in the presence of 1 mM ATP. Hexagonal crystals were obtained in PEG 300 at pH 6.5. The crystal structure of the complex was solved by molecular replacement using the individual structures of AceK (PDB code: 3eps) and ICDH (PDB code: 1sjs) as search models. The complex structure has been refined to a resolution of 3.0 Å with $R_{\text{cryst}}$ and $R_{\text{free}}$ values of 24.8% and 30.2% respectively. Data collection and refinement statistics are given in Table 3.2. The crystallographic asymmetric unit contains two AceK and two ICDH molecules, arranged as a symmetric dimer of two AceK-ICDH complexes, having a total molecular weight of approximately 220 kDa.

### 3.3.2 Overall structure of AceK

The crystal structure of AceK is composed of two distinct domains (Figure 3.3). One domain resembles eukaryotic protein kinases and contains an ATP molecule; therefore, it is likely the kinase domain (KD). The second domain of AceK does not resemble any known protein fold. Interestingly, AMP, a known inhibitor of AceK kinase activity and an activator of AceK phosphatase activity \textsuperscript{40,150}, is found in a pocket between these two domains. Previous studies have indicated that the ATP binding site is likely the active site where the phospho-transfer occurs \textsuperscript{166}, whereas the AMP binding site we see in our structure is most likely an
allosteric site. Therefore, we have tentatively named the second domain of AceK the “regulatory domain” (RD) to indicate that it may form a scaffold for regulator binding.

The AceK RD domain occurs in the N-terminal half of AceK (residues 2-290), and it is mainly composed of α-helices (Figure 3.3A). It begins with two parallel elongated α-helices (αI, αJ), which are 35 and 26 residues in length respectively. They are arranged in the form of a large hairpin, and are followed by two short parallel α-helices (αK, αL) that form a second, smaller hairpin. The smaller hairpin overlays the larger one and is connected to a four-stranded antiparallel β sheet by a 28-residue loop region. This loop region, in turn, is sandwiched between several isolated α-helices, eventually becoming a 27-residue long α-helix (αH) that links the two domains (Figure 3.3A).

The structure of the AceK KD domain, located in the C-terminal half of the sequence (residues 307-574), adopts the classic kinase bi-lobe fold, with the ATP-binding cleft clamped by the two lobes (Figure 3.3A). The N-terminal lobe consists mainly of a twisted, five-stranded antiparallel β sheet (β1- β5) and two α-helices (αA and αB). The ATP binding site is covered by the five-stranded β sheet, located and deeply buried in two extended random loops: Loop-β1β2 (residues 315-320) and Loop-β2αB (residues 340-350). The larger C-terminal lobe is predominantly α-helical with some stretches of antiparallel β-strands. Under the ATP molecule, two antiparallel β strands (β7-β8) form a hairpin structure. Interestingly, a long flexible loop stretches out of the C-terminal lobe, partial residues in which (residue 497-503) could not be traced out due to the lack of the electronic density, indicating a flexible region. This loop together with the loops in the N-terminal (Loop-β1β2 and Loop-β2αB) forms a large cleft, which was proved to be the substrate binding site from our complex structure (see below). We were able to build the external loop residues 497-503 missed in AceK structure in the complex
Figure 3.3 The overall structure of AceK. The active site of AceK includes a buried ATP molecule. The substrate binding cleft is shown with a black arrow. The structure also includes a bound AMP at the interface of the two domains. AceK contains two domains. The kinase domain (KD – yellow) resembles eukaryotic protein kinases. The other domain of unknown function (RD – magenta) does not have any structural homologues.
structure which is located at AceK/ICDH complex interface. We have name this loop (residues 487-514) the substrate recognition loop (SRL).

### 3.3.3 Structure comparison

Dali\(^{119}\) was used to search for AceK structural homologs. Our initial search for full length AceK homologs did not produce any matches. We next searched for homologs to the AceK KD and RD separately. For the RD, no clear structural homolog was found, suggesting that the regulatory domain may possess a novel structural fold. The AceK KD was found to be structurally similar to eukaryotic Ser/Thr protein kinases and to the Tyr protein kinase domain of signal transduction receptors. Dali provided more than 100 hits within these protein families, such as the interleukin-1 receptor\(^{167}\), Fibroblast growth factor receptor 2 (FGFR2)\(^{168}\) and the Proto-oncogene tyrosine protein kinase receptor\(^{169}\). Because most protein kinases in this family share a similar kinase fold, and because FGFR2 contains the ATP analog AMPPCP in its ATP binding site, we selected this kinase for further structural comparison with AceK (Figure 3.4). We find that the FGFR2 structure (PDB accession code 3cly) is structurally very similar to AceK KD, having a Z-score of 9.5 and an r.m.s.d. of 3.6 Å between them (as reported by Dali). Sequence alignments between AceK RD and FGFR2 exhibit only 14.9% identity, suggesting that there may be a distant evolutionary relationship between these proteins.

A structural comparison of AceK KD and FGFR2 shows that the conserved structural elements, particularly in the ATP binding region, are in a good alignment (Figure 3.4). However, several regions in the N-terminal lobe display significant variations, particularly in helices \(\alpha A\), \(\alpha B\), \(\alpha H\), and \(\beta 4\), \(\beta 5\) and \(\beta 6\). The most striking difference between these two proteins is the AceK Loop-\(\beta 2\alpha B\) section (Figure 3.4, red colour), which covers the ATP binding site; this loop is
entirely absent in FGFR2. This loop renders the ATP binding site of AceK inaccessible, whereas it is accessible in FGFR2. In the C-terminal lobes, the α-helices of these proteins do not align well, except for the region near the ATP binding site. Interestingly, the activation loop region of FGFR2 (Figure 3.4, blue colour) is completely absent in AceK KD. Instead, the SRL loop (Figure 3.4, red colour) occupies this region. This loop travels a significant distance away from the active site, with approximately a 20 Å shift. This conformation has not been observed previously in other protein kinases.

3.3.4 ATP binding site

The present AceK structure contains a bound ATP in its KD with well-defined electron density (Figure 3.5A). Although the structural comparison between AceK KD and FGFR2 showed some variance amongst their N- and C-lobes, the high degree of structural similarity at the ATP binding site of these two proteins makes it possible to identify key residues that are involved in ATP interactions and to postulate an ATP phospho-transferring mechanism. The superposition of the corresponding ATP binding site segments gives rise to an r.m.s.d. value of 1.49 Å between the main chain atoms of AceK and FGFR2. Notably, this region has been shown to bind ATP in all known structures of eukaryotic protein kinases (ePKs). Regarding the specific AceK residues involved in ATP binding, an analysis of the residues in AceK and those in other ePKs has revealed a restricted sequence homology between the region 315-340 of AceK and the corresponding ATP-binding residues in ePKs. This region contains: (a) a glycine-rich motif (the P-loop), described as a multifunctional element involved in both nucleotide binding and catalysis and (b) a conserved lysine residue, known to interact with and stabilize α- and β- phosphates of bound ATP. Our structure is consistent with these details, having (a) a loop labelled
Figure 3.4 Overall structural comparison between FGFR2 kinase (PDB accession code: 3cly – light green) and the AceK kinase domain (KD, PDB accession code: 3eps – grey). The labelling for AceK is the same as in Figure 3.3.
Loop-β1β2 (residues 315-APGIRG-320) that aligns well with the P-loop of FGFR2; and (b) residue K336, corresponding to K517 in FGFR2, locates in the ATP binding site that interacts with the ATP α-phosphate (Figure 3.5B).

In FGFR2, a conserved phosphotransferase motif, termed the Brenner's motif [622-ClHrDLAARN-631] is located on a hairpin loop under the ATP binding site (Figure 3.5C). Residue H624 in this motif plays a structural role in forming the ATP binding site. H624 hydrogen bonds with the main chain amide of D626 (AceK: D457) and the main chain carbonyl of A643. These two sets of interactions help orient the sidechain of the catalytically important residues D626 and D644. Residues N631, D644 and E534 directly co-ordinate two magnesium ions for the catalysis (Figure 3.5C). This cluster of residues forms a catalytic triad which is invariably found in all ePKs. Researchers have previously proposed a location for an AceK “catalytic triad” based on sequence analysis and mutagenesis studies. However, from the actual AceK structure, it is clear that this proposed “catalytic triad” is unrelated to the AceK ATP binding site. Instead, the corresponding sequence motif and “catalytic triad” was found at [453-IFPGDMLFKN-462] (Figure 3.5B), which is an incomplete Brenner's motif. D457 and N462 in this motif correspond to D626 and N631 in FGFR2, while D475 on the inner loop of the AceK ATP binding site corresponds to D644 in FGFR2 (Figure 3.5D). However, the histidine residue in Brenner's motif is absent in AceK; P455 is in its place instead. In other ePKs, this histidine residue is responsible for orienting two catalytic Asp residues to coordinate the ATP molecule and magnesium ion for catalysis. Because AceK lacks this histidine residue, and because it has a flexible Gly residue next to its catalytic D457, we speculate that this D457 has additional freedom of movement that allows it to shift in and out of the ATP pocket, perhaps
allowing it to participate in a nucleophilic attack for AceK ATPase activity. This may provide a clue to AceK’s bifunctionality. In addition, FGFR2 has two magnesium ions. In contrast, only one ATP-liganded magnesium ion was found to be coordinated by N462 and D475. The space for the second magnesium ion was occupied by the ATP molecule (Figure 3.5D).

Another distinctive feature of AceK is a loop region (Loop-β2αB) that covers the ATP binding site and closes the ATP pocket (Figure 3.5B); this loop is not seen in other ePKs. An ATP molecule in the AceK pocket is constrained to the bottom of the pocket by this loop. Compared to the AMP-PCP molecule in FGFR2, ATP in AceK can be seen to adopt a different orientation. For example, the γ-phosphate of ATP in AceK is 4.2 Å away from that of AMP-PCP in FGFR2. Two negatively charged residues, D477 and E478, interact with ATP γ-phosphate (Figure 3.5B). D477 occupies a similar position to that of D534 in FGFR2 (Figure 3.5C), and so it may play a similar role in coordinating a magnesium ion. E478, on the other hand, forms salt bridges with K346 (located in Loop-β2αB), thereby locking and burying ATP in the binding pocket (Figure 3.5B). This locking mechanism may be the reason that a second magnesium ion is not observed in the pocket, and it could also explain why the ATP molecule was not hydrolyzed, as AceK ATPase activity is magnesium dependent.

3.3.5 AMP binding site and interaction between KD and RD

An AMP molecule is observed with well-defined electron density deep inside a surface pocket at the interface between KD and RD (Figure 3.6A). The adenine moiety sits in a hydrophobic pocket making hydrogen bonds with main-chain groups of E376 and N377 on the β5 strand of KD, and with H113 in RD. In addition, the 3’ hydroxyl group of the ribose interacts
Figure 3.5 A) 2σ 2Fo-Fc electron density of the ATP molecule and the magnesium ion in AceK. B) Interaction of ATP with the side chains of residues in the ATP binding pocket in AceK. C) Interaction of ATP with the side chains of residues in the ATP binding pocket of FGFR2 (PDB accession code: 3cly). D) Side chain alignment of AceK (magenta) with FGFR2 (green) kinase at the ATP binding sites of the two proteins. The side chains of five highly conserved and catalytically important residues are highlighted as stick models. An AMP-PCP molecule is in cyan and the Mg$^{2+}$ ions are shown as yellow spheres in FGFR2. ATP is shown in light grey.
with a threonine (T295) residue located near the N-terminus of αH helix adjacent to the site. The phosphate group interacts with the positive side chains of two other residues (K291 and K294) and with the hydroxyl group of serine (S105) from RD. Importantly, these residues interact with each other and with AMP to form three scaffolding bridges between the two domains: N377|E376--AMP--H113|T295, E374--K294--AMP--S105, and E374--K294--AMP--K291--D55. Thus, AMP acts as a bridge pier that holds KD and RD together (Figure 3.6B). In addition, we observed that residue K361, sitting in between the two domains, forms interactions with KD residues D371 and T372, and RD residues Q290 and Y256, which further tie the two domains together. Interestingly, the KD residues involved in inter-domain interactions are mainly located on β5 strand and its extended loop region (Figure 3.6C). They do not directly interact with the RD residues, but rather use AMP and K361 as intermediaries. This may explain why AMP acts as an allosteric regulator for this protein. As the AMP binding site is 24.8 Å away from the ATP binding site, this raises the questions of how the AMP binding process allosterically affects the ATP binding site. To investigate this, we analyzed the interactions between the AMP binding site and the ATP binding site. Several bridging interactions were noted: both E374--K294--AMP--S105 and E374--K294--AMP--K291--D55 were extended to Y414 via the interaction between E374 and Y414, and Y414 further extends the interaction to Y357 via π-stacking. Y357 together with E416 stabilize K336, which is the key residue for ATP binding. Thus, there is an intricate network of interactions that connect the AMP and ATP binding sites, allowing them to communicate from afar (Figure 3.6C). Our suspicion is that the network connecting the two sites is even more complex and dynamic than we have described here, perhaps involving interactions that open or close the inter-nodal space as AMP is entering or leaving.
While ATPase activity is commonly associated with kinase activity, the unusually strong ATPase activity of AceK as well as the fact that AMP is found at a site remote from the ATP-binding site makes us wonder whether ATPase activity could occur in the AMP-binding site. To test this hypothesis, we generated two mutants, one at the ATP site and one at the AMP site, and measured their ATPase activity by a colorimetric assay used previously in our work. Our preliminary results showed that the D475A mutant, which likely disrupts the binding of a metal cofactor in the ATP-binding site, resulted in the complete loss of ATPase activity. In contrast, D55A in the AMP site did not affect the ATPase function (Figure 3.7). These results demonstrate, unambiguously, that the AMP site is not involved in ATPase activity but represents a bona fide allosteric regulatory site.

Another puzzling aspect of our AceK crystal structure is simply the actual presence of AMP at all. AMP was not part of our crystallization conditions, and AceK could not have hydrolyzed the ATP that we did add in the protein solution to give AMP. Because AMP is an inhibitor of AceK kinase and ATPase activity, and because the protein is produced in a nutrient-rich medium in which AceK kinase activity is expected to be inactivated, we postulate that AMP may have bound to AceK during protein expression in the cell and co-purified with it. Overexpression of cellular AceK will inactivate ICDH and prevent the Krebs’ cycle; to avoid this, the cell needs to inhibit the over-expressed AceK kinase activity, which it does with AMP. The presence of AMP in our crystal structure may thus reflect the physiological state of AceK when cells grow in a nutrient-rich medium, demonstrating the possible allosteric regulation performed by AMP.

As mentioned above, there is an intricate array of regulators that act directly upon AceK (Table 3.1) or indirectly upon the substrate to provide a complex and sophisticated regulatory
Figure 3.6 A) 2σ 2Fo-Fc electron density of the AMP molecule in AceK. B) Electrostatic surface potential analysis shows two deep pockets at the interface of the KD and RD. C) Interaction between AMP and the side chains of residues in the AMP binding pocket in AceK, and the interactions amongst the residues between the ATP binding site and the AMP binding site.
mechanism\textsuperscript{166}. While the complete AceK regulatory mechanism is still unknown, a variety of regulators have been identified, many of which promote phosphatase activity. The large number of regulators that AceK responds to is consistent with its central role as a regulator of metabolic carbon flux. Eight compounds have been shown to activate AceK phosphatase activity while inhibiting kinase activity: isocitrate, 3-phosphoglycerate, pyruvate, AMP, ADP, oxaloacetate, $\alpha$-ketoglutarate, and PEP\textsuperscript{40}. As many of these effectors are derived from the end products of the glyoxylate bypass, they represent negative feedback inhibition, decreasing the amount of isocitrate available to ICL. Four other compounds have been identified that inhibit AceK kinase activity: NADP, citrate, fructose-6-phosphate, and glyoxylate. However, these compounds do not appear to have any effect on AceK phosphatase activity\textsuperscript{40}. Interestingly, no regulators were found that activate AceK kinase activity or inhibit AceK phosphatase activity, which suggest that the glyoxylate bypass is strictly controlled, only being activated when a cell is nutrient deprived.

There are multiple regulators of AceK. How do these effectors work on AceK? Do they share the same binding site or are there multiple allosteric binding sites? We performed electrostatic surface potential analysis of AceK in order to identify other potential allosteric interaction sites. Of particular interest is the interface surface between the KD and RD (Figure 3.6B). In addition to the AMP binding site located therein, two other pockets are formed at the interface. One very positively charged pocket is locate in the upper part of the interface near the AMP site, and the other is a negatively charged site located in a lower part of the interface. These two pockets extend far into the interface. We speculate that these two pockets might represent other allosteric binding sites. Regulator binding in these pockets will no doubt interfere with the
Figure 3.7 The ATPase activity of AceK is harboured in its ATP binding site, not the AMP binding site. The D55A mutant (AMP_DA) of the AMP binding site does not significantly affect the ATPase activity of AceK. In contrast, the D475A (ATP_DA) mutant of the ATP binding site resulted in the total loss of ATPase activity.
interactions between the two domains. However, further experiments are required to investigate this hypothesis.

From an analysis of the ATP and AMP binding sites, we believe that our AceK crystal structure represents its “default” phosphatase form because 1) the protein is produced from a nutrient-rich medium in which there is no need for the glyoxylate bypass (which requires the kinase activity) to conserve the carbon source; 2) AMP, an inhibitor of AceK kinase activity, was found to bind to an allosteric site in the structure; and 3) ATP is completely buried and inaccessible.

3.3.6 Previous mutational study on AceK activities

Despite the lack of structural information, researchers previously carried out mutational studies in attempts to identify residues that are involved in AceK’s kinase and phosphatase function. In addition to K336, which was identified to play the role in binding ATP in both kinase and phosphatase activities, the majority of mutations tested were randomly generated and resulted in a decrease in phosphatase activity with little effect on kinase activity or an increase in kinase activity. Q373R, Y414C, D371 and V472F resulted in near total loss of phosphatase activity \(^{174,175}\). However, two other mutations, D477N and R365C abolished phosphatase activity while elevating kinase activities above wild type levels. It was also found that Q373R and Y414C altered AceK sensitivities to regulators and mutation of D371 demonstrated a decrease in AceK expression and phosphatase activity suggesting it may also be involved in conformation stability of the kinase and possibly the binding of ATP \(^{176}\). It was postulated that because these mutations selectively inhibited phosphatase activity, it is likely that these residues reside in a regulatory domain and affect the catalytic site indirectly \(^{174,175}\). In our structure, R365, D371, Q373 locate at the interface of KD and RD domains near AMP binding site. The mutations of
these residues may interfere the binding of the effector(s), therefore affecting AceK function. As mentioned above, residue Y414 sits between the ATP binding site and AMP binding site, which may act as an interaction node of connection between ATP and AMP. Mutation of this residue perhaps does not affect the AMP binding, but it may block the allosteric effect of AMP on ATP binding site, therefore altering the AceK’s sensitivity to AMP. An interesting observation is D477, which locates at the ATP binding site, when mutated (D477N), phosphatase activity severely decreased and kinase activity increased. Above we postulated that D477 may play the role in coordinating the second magnesium ion, if present, in the ATP binding site. We are still puzzled how it functions, as located at ATP binding site, D477 should decrease both kinase and phosphatase activities when mutated. This unique residue may provide a clue to understanding the mechanism of activity switch between kinase and phosphatase. Some other mutations were also reported, including the mutations of N377, D403 and E439. They locate at unimportant regions and are away from the ATP or AMP binding sites in the structure. Indeed, they did not exhibit a significant decrease in expression or phosphorylating activity. Our structure not only seems consistent with those mutational studies but also gives rise to some plausible explanations.

3.3.7 AceK/ICDH complex structure and interaction between the two proteins

It has been a puzzle for a long time that how ICDH could be phosphorylated by AceK, as the phosphorylation site S113 is buried inside the ICDH structure, making it inaccessible to the enzyme. Previously, two structural modes of ICDH were reported: an open conformation and a closed conformation at active site. The active mode of ICDH is in the closed conformation, for the substrates of ICDH (isocitrate, NADP+) stabilize the closed conformation.
Figure 3.8 Overall complex structure of AceK and ICDH. The phosphorylation loop is coloured in blue, and the substrate reorganization loop (SRL) is coloured in red.
The inactive phosphorylation of S113 of ICDH does not trigger the conformational change at the active site. Then what triggers the conformational transition from the closed to open? The open structure mode of ICDH was obtained when researchers were previously attempting to crystallize ICDH with AceK, even though it is still hard to say that it is AceK that triggered the conformational change. Nevertheless, AceK can only selectively bind to the open conformation of ICDH. However, even within the open conformation of ICDH, S113 still face inside and is inaccessible. It was proposed that the binding of AceK to the open ICDH structure may induce further conformational changes and to position S113 fully accessible. The crystal structure of the AceK-ICDH complex presented here reveals a nonphosphorylated ICDH and AceK with both ATP and AMP bound. ICDH forms a tight homodimer with the two active sites on the opposite sides. Each phosphorylation loop (residues 104-113) in the ICDH dimer binds to an AceK KD domain at the active cleft formed between Loop-β2αB and SRL. The SRL on AceK inserts into the opened active site of ICDH and forms intimate interactions with ICDH dimer (Figure 3.8).

The interface between AceK and ICDH is nearly identical in the two copies of AceK-ICDH present in the asymmetric unit. There are two main AceK-ICDH interaction regions that encompass SRL of AceK from dimeric ICDH: one is from one of dimeric ICDH named M1, and another is from the second molecule of dimeric ICDH named M2 when AceK interacts with the phosphorylation loop on M2 (Figure 3.9). Since the intact interaction “epitope” of ICDH consists of two molecules of dimeric ICDH, AceK needs to associate with both ICDH molecules in order to form a “productive” complex with one substrate site. This is one of the reasons why AceK only recognizes the tertiary structure of the substrate and thus, cannot utilize a peptide substrate.
Figure 3.9 A) The interface between AceK SRL and the ICDH active site cleft. B) The side chain interactions between AceK and ICDH. Residues from ICDH are coloured in blue and residues from AceK are coloured in red.
SRL of AceK extends 32 Å long forming a small α-helix at the forepart. The interaction between SRL and the active site cleft of ICDH includes hydrophobic packing, salt bridge, hydrogen bond. Hydrophobic packing interactions formed at the back end of SRL, including V490\textsuperscript{a}, Y495\textsuperscript{a}, W505\textsuperscript{a}, Y506\textsuperscript{a} and V508\textsuperscript{a} (residues from AceK denoted with \textsuperscript{a}) with V107\textsuperscript{M2}, I111\textsuperscript{M2}, Y296\textsuperscript{M1}, V280\textsuperscript{M1} and W263\textsuperscript{M1} (residues from M1 denoted with \textsuperscript{M1} and residues from M2 denoted with \textsuperscript{M2}). Additional interactions in this region include a salt bridge between S597\textsuperscript{a} and K242\textsuperscript{M1} and a hydrogen bond between S597\textsuperscript{a} and D279\textsuperscript{M1}. At the forepart α-helix of SRL, the interactions mainly comprise of salt bridges and hydrogen bonds, including salt bridges of E499\textsuperscript{a} and K344\textsuperscript{M2}, E497\textsuperscript{a} and R292\textsuperscript{M1}, and a hydrogen bond between E499\textsuperscript{a} and Y345\textsuperscript{M2}, and a hydrogen bond between S502\textsuperscript{a} and the backbone of T338\textsuperscript{M2}, and a hydrogen bond between T105\textsuperscript{M2} and backbone of S502\textsuperscript{a}. These interactions between SRL of AceK and ICDH promote the open conformation of the active site cleft of ICDH, and expose the S113 phosphorylation loop to the catalytic site of AceK (Figure 3.9).

3.3.8 Comparison between the apo and complex ICDH structures

A superposition of the ICDH structure from the AceK-ICDH complex and the apo ICDH structure in the open conformation \textsuperscript{163} (PDB accession code 1sjs) shows that the two structures have a very similar overall conformation, with an r.m.s.d. value across the 828 Ca atoms of 0.62 Å. There are, however, several notable differences between the structures (Figure 3.10). The most conspicuous difference is the behavior of the phosphorylation loop, residues 104-113. In crystal structures of ICDH alone, the phosphorylation loop adopts a more open conformation, while αN helix at the N-terminus of the phosphorylation loop adopts a more closed form (Figure 3.10, pink). Conversely, in the structure of ICDH from the AceK-ICDH complex, the
Figure 3.10 Overall structural comparison between ICDH from the AceK/ICDH complex structure (green, with the phosphorylation loop coloured in red) and the closed form of apo ICDH (PDB accession code: 1sjs; cyan, with the phosphorylation loop coloured in pink).
phosphorylation loop adopts a more closed conformation, while αN helix has shifted, opening up more space on the inside of the active site cleft. This latter shift is caused by the insertion of AceK SRL in this region. Interestingly, with three sequential glycines (residues 108-110) rendering it highly flexible, the phosphorylation loop in the AceK-ICDH complex structure appears to have moved approximately 8.5 Å. Despite the fact that the phosphorylation loop adopts a closed conformation in the complex structure, it is positioned to fit in the AceK active site. Following a closer look at the phosphorylation site, the most significant side chain alterations occur at S113 and R112, both of which have rotated more than 100° so that the S113 oxygen was moved from a buried, inward facing orientation to an exposed, outward facing one. This conformation change renders S113 accessible for phosphorylation (Figure 3.10).

3.3.9 Comparison of the apo and complex AceK structures

A superposition of the AceK structure from the AceK-ICDH complex with apo AceK structure (PDB accession code: 3eps) shows that there is very little conformational change in AceK upon ICDH binding (r.m.s.d. of 0.55 Å over 571 Ca atoms). However, as mentioned above, Acek SRL, whose density was absent in the apo AceK structure is seen clearly in the AceK-ICDH complex. By binding to ICDH, SRL appears to have become stabilized and shifted out of the AceK active site by as much as 16 Å, underscoring the flexibility of this loop. While little overall conformational change was seen in the AMP-binding pockets in the apo and complex AceK structures, there is substantial variance at the ATP binding sites between the two structures. The orientation of the ATP molecule shifts significantly, with the γ-phosphate group moving approximately 4 Å out from the bottom of the ATP binding pocket. Correspondingly, the interactions between the γ-phosphate group and the side chains of the residues in the pocket have also changed. Previous strong interactions between the γ-phosphate group and E478 have been
Figure 3.11 A) Overall structural comparison of AceK from the AceK/ICDH complex structure (green, SRL is coloured in purple) and the apo structure (PDB accession code: 3eps). The Loop-\(\beta2aB\) is coloured in red. B) Comparison of the ATP orientations and interactions with the residues in the ATP binding pockets from two different AceK structures.
severed. Following this conformational shift, the $\gamma$-phosphate group is able to interact with K461, D457, K346, and N462, while still maintaining interactions with D475 and D477. Notably, this last interaction is only maintained because the side chain of D477 is able to re-orient near the $\gamma$-phosphate group, approximately 3.7 Å from its initial position inside the pocket. Moreover, N462, previously coordinating one magnesium ion, now interacts with $\gamma$-phosphate group directly. No magnesium ion was observed in the complex structure, suggesting that it must have been displaced during complex formation. Instead, two positive-charged residues, K461 and K346, interact directly with the $\gamma$-phosphate group, reprising the stability role that the Mg$^{2+}$ ion performs on the $\gamma$-phosphate group in the apo structure. Interestingly, the interaction between K346 and E478, which acted as a lock to close the ATP pocket in the apo structure, was interrupted. Without this K346 “catch”, the side chain of E478 moves out of the pocket. This creates an opening for D457, which fills the vacated position and interacts with the $\gamma$-phosphate group. As previously mentioned, the ability of D457 to move freely into a stabilizing position is enhanced by the lack of a histidine residue in Brenner’s motif. Furthermore, the orientation of ATP in the complex structure is very similar to the orientation of AMP-PCP in the FGFR2 kinase. Because of this, we believe that ATP is bettered positioned for phosphor-transfer. Clearly, it is the interaction with ICDH that triggered this ATP positional change within the ATP binding pocket.

3.3.10 Phosphorylation site

In the AceK-ICDH complex structure, the phosphorylation site, S113, is found to be unphosphorylated, even though the crystallization conditions contained both ATP and Mg$^{2+}$. The reason may be due to the unexpected appearance of AMP, an inhibitor of AceK kinase, in the
RD. However, the phosphorylation loop of ICDH is found at the AceK ATP binding site, even though S113 is still far from ATP. Indeed, the distance between the S113 hydroxyl group and ATP $\gamma$-phosphate group is 13.2 Å. We find it curious that S113 should remain so far from the $\gamma$-phosphate group instead of approaching it to enable the phosphor-transfer reaction. Analysis of the environment around the phosphorylation loop is revealing. The three glycines (residues 108-110) locating at the substrate binding site have no interactions with other residues and offer the loop a high degree of flexibility and mobility. Residue I111 sitting in hydrophobic pocket also provides some mobility. Residue R112, despite its long side chain, projects into an open space and has no interactions with other residues. The only interaction that might account for the distance between S113 and the $\gamma$-phosphate group of ATP is residue Q345 from the Loop-$\beta$2$\alpha$B of the AceK KD. It forms two hydrogen bonds with the backbone of S113. Interruption of these interactions is anticipated to free S113, allowing it to approach ATP for the phospho-transfer reaction. Taken together, because of AMP binding and the resulting inhibition of kinase activity, AceK does not adopt its kinase conformation that would trigger even larger conformational change of ICDH (and perhaps AceK as well) that would allow S113 and ATP interact directly.

3.3.11 Proposed regulatory mechanism for the AceK activity switch

Based on structural analyses of both the apo AceK structure and the complex structure of AceK-ICDH, and with the previous mutational result and the knowledge of AceK regulators in mind, we summarize below the pertinent facts about, and propose a regulatory mechanism for, the AceK activity switch: The AceK structure shows two functional domains: a kinase domain where the kinase, phosphatase and ATPase reactions occur and a regulatory domain which help form regulator binding pockets that that can regulate the catalytic domain’s function. The kinase
Figure 3.12 Interface between the ICDH phosphorylation loop and the AceK active site cleft. ATP and the phosphorylation site S113 are shown as ball and stick.
domain contains an SRL loop that only interacts with the intact, tertiary structure of ICDH, increasing substrate specificity. Although the structure of the AceK kinase domain resembles those from eukaryotic protein kinases, suggesting that they may share a similar mechanism for phosphor-transfer, a unique loop (Loop-β2αB) was found in the AceK kinase domain that is absent in eukaryotic kinases. In the closed conformation, this loop covers the ATP binding pocket, preventing ATP hydrolysis. ATP can adopt two different orientations in the ATP binding pocket of AceK. AceK-ICDH interaction can trigger an ATP positional change even when Loop-β2αB is in the closed conformation. The closed conformation of Loop-β2αB loop prevents unphosphorylated ICDH S113 from approaching the AceK ATP binding site. This is accomplished by interactions between residue Q345 from Loop-β2αB loop of AceK and the backbone of S113, resulting in inhibited kinase activity. The closed Loop-β2αB loop is opened when the regulator AMP is released. In a structure of AceK determined from another crystal form, AMP is absent and ATP was hydrolyzed into ADP. In this AceK structure, the Loop-β2αB was found to adopt an open conformation. However the structure has not been fully refined, with high R and R_free values of 33% and 38%, respectively. It is likely that some conformational changes in the structure have not been identified thus far. Nevertheless the density map clearly showed that there was no AMP density at the AMP binding site, and the presence of an ADP molecule with a Mg^{2+} ion at the ATP binding site. The conformation of the Loop-β2αB is somehow controlled by AMP or other regulators. Mutations in between the AMP and ATP sites compromise the communication between the two sites and therefore affect the allosteric effect of AMP, suggesting that the regulators can remotely control kinase domain activity. In the current complex structure, AMP-bound AceK and unphosphorylated S113 ICDH are an “unmatched” pair. It is noted that unphosphorylated S113 cannot interrupt the interactions between Q345 and
the backbone of S113. However, once phosphorylated, P-S113 of ICDH would be able to enter the active site and undergo dephosphorylation, because the large, negatively charged phosphate group could impair the interactions between Q345 and S113.

3.4 Conclusion

An understanding of *E. coli* metabolism is of tremendous benefit for studying this bacterium’s growth and survival in various nutritional environments, such as in drinking and swimming water. As part of the acetate switch, the glyoxylate cycle appears to increase growth efficiency under glucose-limited conditions. The glyoxylate cycle works only in the presence of acetate and fatty acids, conferring more metabolic flexibility to *E. coli*. AceK, a highly unusual bifunctional protein kinase/phosphatase involved in regulating the glyoxylate cycle, is an integral part of *E. coli* metabolism with a sophisticated regulatory mechanism that controls the reversible phosphorylation/ inactivation of ICDH, a central enzyme in the Krebs’ cycle. AceK plays a vital role in directing the carbon flux to various enzymes during metabolism, and this direction is central to a cell’s ability to adapt its metabolic processes to best manage acetate. The “acetate switch” has received renewed interest following the discovery of an important intermediate, acetyl-phosphate which might act as a global signal in regulating diverse cellular processes. However, prior to this, research on AceK was hindered for many years because of a lack of the important structural information. Our structures of apo-AceK and the AceK-ICDH complex have shed new light on its function, allowing us to correct some previous misunderstanding about AceK “catalytic triad” 172, and to propose a plausible model of the regulatory mechanism of AceK activity. However, many mysteries still remain, such as what triggers AceK kinase activity when cells grow on acetate, and what is the basic enzymatic mechanism of AceK phosphatase
activity. It is our hope that these novel AceK structures will instigate new and exciting research directions in this field.

3.5 Materials and methods

3.5.1 Cloning, expression and purification

The aceK gene in E. coli encodes isocitrate dehydrogenase kinase/phosphatase, a 66 kDa protein containing 578 amino acids. The complete aceK gene (GI number: 944797) was amplified by polymerase chain reaction using a forward primer of 5’-AAAAAGATCTAATAACAGCGCTTTTACTTTCCAGAC-3’ and a reverse primer of 5’-AAGAATTCATAATACATAGGTGTTAATTGCAAAGG-3’. The icdH gene (GI number: 1787381) in E. coli encodes isocitrate dehydrogenase, a 46 kDa protein containing 416 amino acids. Restriction enzyme sites for BglII and EcoRI (underlined in the primer sequences) were engineered in the forward and reverse primers respectively. After digestion of the PCR product by BglII and EcoRI, the purified fragments were cloned into the BglII and EcoRI sites of a modified pET15b vector (Novagen). The resulting fusion protein contained a TEV protease cleavable N-terminal extension (MGSSHHHHHHHHHGS) and was expressed in E. coli BL21(DE3) cells. Both AceK and ICDH were purified with the same procedure using a two-step protocol. The hexa-histidine tag produced a quick and efficient initial purification step, which was followed by size exclusion chromatography to produce a nearly homogenous protein preparation as evaluated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). The frozen cells were suspended in 50 ml of lysis buffer containing 50 mM Tris-HCl (pH 8.0), 300 mM NaCl and 0.1% Triton X-100. Cells were lysed on ice by sonication. Cell debris was removed by centrifugation for 30 min at 15000 rpm in a JA-25.50 rotor in a Beckman
high speed centrifuge. The clarified lysate was applied to Ni\(^{2+}\)-nitrilotriacetic acid-agarose affinity resin (Qiagen) equilibrated with buffer A (30 mM Tris-HCl, pH 8.0 and 300 mM NaCl) followed by a 10-column-volume wash in the lysis buffer containing 20 mM imidazole. The protein was eluted in several fractions by using a step gradient of increasing imidazole concentrations up to 300 mM. The fractions were analyzed using 12.5% SDS-PAGE (Laemmli 1970). The fractions containing protein were concentrated to 10mg/mL using a Centricon-3 (Millipore) and further purified using an AKTA FPLC system (Amersham Biosciences) with a size exclusion Hiload Superdex S-200 16/60 column in a buffer containing 20 mM Hepes pH 7.0, 2 mM DTT, 100 mM NaCl, and 10% glycerol. The major peak of protein was recovered and was concentrated to 10 mg/mL using a Centricon-3 (Millipore). Purity of the concentrated protein was determined by SDS-PAGE (Figure 3.1).

3.5.2 Preparation of the AceK/ICDH complex

The purified dimeric ICDH (46 kDa/subunit) and dimeric AceK (66 kDa/subunit) were mixed in a 1:1 molar ratio in the presence of 1 mM ATP in a buffer consisting of 100 mM NaCl, 20 mM Hepes, 2 mM DTT, and 0.1% Glycerol at pH 7.0. The mixture was stored at 4 °C overnight and concentrated to 5-10 mg/mL using a Centricon-3 (Millipore) and then further purified using an AKTA FPLC system (Amersham Biosciences) with a size exclusion Hiload Superdex S-200 16/60 column in a buffer containing 25 mM Hepes pH 7.5, 2 mM DTT, 100 mM NaCl, and 10% glycerol. The major peak of protein was recovered and was concentrated to 10 mg/mL using a Centricon-3 (Millipore). The components of the peak were determined by SDS-PAGE (Figure 3.1). The final yield of the complex preparation was 5 mg per liter culture.
3.5.3 AceK crystallization

The preliminary crystallization conditions were screened by the sparse matrix method (Jancarik et al. 1991) using standard screening kits. The protein concentration was diluted to 5 mg/mL with the same protein buffer. ATP was added to a final concentration of 1 mM. The method of hanging drop vapor diffusion was used. Hanging drops were set up to contain 2 µL of protein solution mixed with 2 µL of well solution. Crystals appeared in six days and grew to full size within two weeks. Three crystal forms were obtained from very similar crystal growth conditions (Figure 3.2). The primary optimal crystallization condition in the reservoir was 15% glycerol, 2 mM DTT, 0.2 M magnesium chloride, 0.1 M MES buffer (pH ranging from 5.5 to 7.0), and with 12-15% PEG 8000 as the precipitating agent at room temperature. When the pH of the well solution was adjusted to 5.5 and with the addition of 1 mM ZnCl₂, a hexagonal crystal form was obtained. When the well solution pH was adjusted to 6.5 and 1 mM ZnCl₂ was added, a tetragonal crystal form was obtained. When the buffer pH was adjusted to 7.0 and contained 1 mM CaCl₂, an orthorhombic crystal form was obtained.

3.5.4 Complex crystallization

The preliminary crystallization conditions for the AceK/ICDH complex were screened by the sparse matrix method. The protein concentration was diluted to 5 mg/mL with the same protein buffer. ATP was added to a final concentration of 1 mM. Hanging drops were set up to contain 2 µL of protein solution mixed with 2 µL of well solution. Crystals appeared in three days and grew to full size within two weeks. The primary optimal crystallization condition in the reservoir was 10% glycerol, 2 mM DTT, 0.05 M magnesium chloride, 0.1 M MES buffer (pH
and with 25-30% PEG 300 as the precipitating agent at room temperature. A hexagonal crystal form was obtained.

3.5.5 Identification of the components of the complex crystal

The complex crystals of the hexagonal form from one hanging drop were harvested and transferred to an Eppendorf tube and washed with the crystallization mother liquid several times to remove soluble protein from the drop. The components of dissolved crystals were identified by SDS-PAGE (Figure 3.2).

3.5.6 X-ray data collection

Diffraction data for the AceK crystals were collected on the A1 and F1 beamlines at the Cornell High Energy Synchrotron Source (CHESS) (Cornell University, Ithaca, NY) using an ADSC QUANTUM 210 CCD detector at a wavelength of 0.9789 Å at the A1 station, and an ADSC QUANTUM 270 CCD detector at a wavelength of 0.9179 Å at the F1 station. Prior to data collection, crystals were soaked in crystallization buffer containing 25% (v/v) glycerol for 1 min, followed by flash-cooling using liquid nitrogen. Single-wavelength anomalous dispersion datasets were collected with an oscillation angle of 1.0° for hexagonal and tetragonal crystal forms at the A1 station for a total of 120°. High resolution datasets for the tetragonal and orthorhombic crystals were collected at the F1 station. The synchrotron data were indexed and integrated using HKL2000 (Otwinowski & Minor 1997). The tetragonal crystal was determined to be of $P4_12_12$ symmetry and produced unit-cell parameters of $a = b = 124.61$ Å and $c = 267.49$ Å. A total of 49741 reflections were collected in the resolution range of 20-2.9 Å with an $R_{merge}$ of 0.097 and a completeness of 94.4%. Calculation of the Matthew’s Coefficient (CCP4) (Matthews, 1968) indicated that each asymmetric unit contains two molecules of AceK,
corresponding to ~68% solvent content (Table 3.2). The orthorhombic crystal was determined to be of $P2_12_12_1$ symmetry and produced unit-cell parameters of $a = 64.20$ Å, $b = 134.18$ Å, and $c = 187.23$ Å. A total of 44356 reflections were collected in the resolution range of 30-2.7 Å with an $R_{merge}$ of 0.102 and a completeness of 93.2% (Table 3.2). In this crystal form, it was determined that each asymmetric unit contains two molecules of AceK, corresponding to ~59% solvent content. The hexagonal crystal was determined to be of $P3_212_1$ symmetry and produced unit-cell parameters of $a = b = 152.71$ Å and $c = 45.19$ Å. A total of 67225 reflections were collected in the resolution range 50-3.0 Å with an $R_{merge}$ of 0.076 and a completeness of 93.7%. Each asymmetric unit was determined to contain one molecule of AceK as well, corresponding to ~49% solvent content (Table 3.2).

Diffraction data for the complex crystals were collected at the A1 station of CHESS. Single-wavelength anomalous dispersion datasets were collected with an oscillation angle of 1.0° for hexagonal and tetragonal crystal forms at the A1 station for a total of 120°. The hexagonal crystal was determined to be of $P6_3$ symmetry and produced unit-cell parameters of $a = b = 196.80$ Å and $c = 156.46$ Å. A total of 75640 reflections were collected in the resolution range 30-3.0 Å with an $R_{merge}$ of 0.115 and a completeness of 99.6%. Each asymmetric unit was determined to contain two units of AceK/ICDH complex molecules, corresponding to ~67% solvent content (Table 3.2).

### 3.5.7 Structure determination and refinement

The single-wavelength anomalous dispersion (SAD) data at the peak wavelength (0.979 Å) permitted the heavy-atom positions to be located. The positions of 18 selenium atoms and the initial phases were determined by Solve $^{179}$. The Sharp program was used for phasing calculation
and refinement, including solvent flattening, noncrystallographic symmetry identification and averaging. The initial AceK structure was partially traced by RESOLVE \(^{180}\). Additional manual tracing and model building were carried out by using Xfit in XtalView \(^{181}\). CNS (version 1.1) \(^{182}\) was used for refinement.

**Acknowledgements**

The authors would like to thank Drs. M. Cygler and A. Matte for their support as part of the Montreal-Kingston Bacterial Structural Genomics Initiative. This work was supported by the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada to ZJ. ZJ is a Canada Research Chair in Structural Biology.
Chapter 4

Identification of an ITPase/XTPase in Escherichia coli by Structural and Biochemical Analysis

Preface: This chapter was published in Structure

Jimin Zheng was responsible for protein expression, purification, crystallization, data collection, structure solution and refinement, and all biochemical experiments in the investigation of YjjX function. The yjjX gene was cloned by Christian M. Udell. Dr. Vinay K. Singh helped with in vitro pull-down assay and mass spectrometry. The manuscript was written by Jimin Zheng with editorial input from Dr. Zongchao Jia.
4.1 Abstract

Inosine triphosphate (ITP) and xanthosine triphosphate (XTP) are formed upon deamination of ATP and GTP as a result of exposure to chemical mutagens and oxidative damage. Nucleic acid synthesis requires safeguard mechanisms to minimize undesired lethal incorporation of ITP and XTP. Here, we present the crystal structure of YjjX, a protein of hitherto unknown function. The three-dimensional fold of YjjX is similar to those of Mj0226 from *Methanococcus janschii*, which possesses nucleotidase activity, and of Maf from *Bacillus subtilis*, which can bind nucleotides. Biochemical analyses of YjjX revealed it to exhibit specific phosphatase activity for inosine and xanthosine triphosphates and have a possible interaction with elongation factor Tu. The enzymatic activity of YjjX as an inosine/xanthosine triphosphatase provides evidence for a plausible protection mechanism by clearing the noncanonical nucleotides from the cell during oxidative stress in *E. coli*. 
4.2 Introduction

When a cell is under oxidative stress, chemical mutagenesis and oxidative deamination damage DNA nucleotides by the conversion of certain amino groups on bases to keto groups. Specifically, this alteration results in the formation of inosine (I) from adenine and xanthine (X) from guanine. Oxidative deamination of DNA bases occurs not only in duplex DNA strands, but also in the free nucleotide pool, the latter being most susceptible to chemical modifications. Accumulation of these damaged nucleotides in the nucleotide pool can trigger an increased mutation rate in the chromosomal DNA. Mechanisms to prevent incorporation of noncanonical nucleotides into DNA and RNA are therefore required.

Known protective enzymes capable of degrading noncanonical nucleotides are members of the Nudix hydrolase family, which includes versatile, widely distributed “housekeeping” enzymes that can hydrolyze nucleoside di- or triphosphates. These enzymes are involved in the maintenance of physiological homeostasis through modulation of levels of metabolic intermediates, signaling molecules, and potentially toxic substances. For example, the Nudix protein MutT eliminates 8-oxo-dGTP from nucleotide pools. Previously, evidence for the presence of an enzyme hydrolyzing ITP, dITP, and XTP was presented for human, mouse, and rabbit tissues. Recently, a novel NTPase, Mj0226 from Methanococcus jannaschii, was found to hydrolyze dITP, ITP, and XTP with release of pyrophosphate. A potential cellular function for this enzyme is to protect against the incorporation of rogue nucleotides, such as ITP, dITP, and XTP, into DNA and/or RNA.

The yjjX gene in E. coli encodes an uncharacterized protein of 20 kDa molecular mass (184 amino acids). To date, no functional annotation has been provided for its product, YjjX, or for any of its close homologs, though the structure of a closely related hypothetical protein,
UPF0244 from *Salmonella typhimurium*, was deposited into the Protein Data Bank (PDB) (1U14). The *yjjX* gene is positioned in between the genes *trpR* and *gpmB* and is predicted to have an independent promoter. A PSI-BLAST search found 56 bacterial or archaeal homologs with *E* values less than 0.003. Top hits displaying similarity to YjjX can be grouped into two categories: (a) hypothetical proteins where nothing is yet known about function or structure; and (b) nucleotidyl transferases. Previous work has also implicated YjjX in thiamine metabolism, more specifically in the detoxification of bacimethrin and 4-amino-2-trifluoromethyl-5-hydroxymethyl-pyrimidine (CF3-HMP). However, the exact role of YjjX in this process remains unclear.

Here, we present the crystal structure of the YjjX protein at 2.25 Å resolution. The structure displays similarity to the NTPase Mj0226 from *Methanococcus jannaschii* and to Maf from *Bacillus subtilis*, both of which are nucleotide binding proteins. It was suggested that Mj0226 might be involved in the prevention of mutations resulting from the incorporation of noncanonical nucleotides into DNA. We present kinetic evidence showing that YjjX can hydrolyze ITP and XTP almost 100-fold faster than other nucleoside triphosphates. Furthermore, we report possible association of YjjX with elongation factor Tu and discuss the implications of this observation.

4.3 Results

4.3.1 Overall Structural Characterization

The final structure of YjjX was determined to 2.25 Å resolution by the method of single-wavelength anomalous dispersion from selenomethionyl derivative, aided by the native data set. There are eight YjjX molecules in the asymmetric unit, which contains 1,365 residues, 15 sulfate
ions, and 747 solvent molecules. The histidine tag residues are not visible. The R factor is 20.0% for 21,753 unique reflections in the resolution range 30–2.25 Å; $R_{\text{free}}$ is 25.4% for ~10% of the data left out from the refinement. The data and refinement statistics are given in Table 4.1. The Ramachandran plot shows 95% of residues in the most favored regions and the remaining 5% of the residues in allowed areas, as defined in PROCHECK.  

The overall structural architecture of YjjX represents an elongated α/β fold, with five extended mixed β strands (β1, β2, β3, β4, and β5) forming a twisted β sheet at the center of the protein, which connects two lobes containing several short α-helical segments, termed Lobe 1 and Lobe 2 (Figure 4.1). The smaller Lobe 1 contains four short helices, α3, α4, α5 and α6, which are between 4 and ~10 residues long with interlinking loops. These four helices pack and surround one face of the central β sheet with a roughly perpendicular orientation.  

In Lobe 2, the central β sheet is extended by two short β strands with parallel topology (β1 and β2) on the β3 side. These are packed between the helices α1 and α7 (Figure 4.1). A small 5 residue loop (Loop 3) connects β2 and α1. In addition, a 12 residue loop (Loop 2) is inserted between strand β1 and helix α2 and arches over Lobe 1 and connects to the longest helical segment in the structure (α2, 19 residues). The two lobes create an approximately 8 Å wide cleft between them, where three loops (Loops 1, 2, 3) enclose a ~10 Å deep cleft.  

The eight YjjX molecules, which are essentially identical in structure, form four dimers in the asymmetric unit. Each monomer is related to its dimeric partner by a 2-fold axis running approximately vertical to the β sheet. The dimer interface is formed through extensive hydrogen bonding between the β5 strands of each molecule, and it connects the two core 3 strand β sheets of the molecules and forms an extended central β sheet with the clefts facing opposite each other.
Table 4.1 Crystallographic Data and Refinement Statistics

<table>
<thead>
<tr>
<th>Crystal</th>
<th>$\lambda$ (Å)</th>
<th>$D_{\text{min}}$ (Å)</th>
<th>Observed Reflections</th>
<th>Unique Reflections</th>
<th>% Complete All/Last Shell</th>
<th>$I/\sigma(I)$</th>
<th>% $R_{\text{merge}}$ All/Last Shell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>1.54</td>
<td>2.25</td>
<td>62,058</td>
<td>21,753</td>
<td>90.3/82.2</td>
<td>8.1</td>
<td>7.5/41.6</td>
</tr>
<tr>
<td>Se-Met crystal at inflection</td>
<td>0.989</td>
<td>2.5</td>
<td>107,637</td>
<td>35,671</td>
<td>100/100</td>
<td>5.2</td>
<td>8.7/39.8</td>
</tr>
</tbody>
</table>

Refinement Statistics

<table>
<thead>
<tr>
<th>Native P2$_1$ (30–2.25 Å)</th>
<th>Cell dimensions: $a = 71.9$ b = 129.2 c = 85.6 Å, $\beta = 90.27^\circ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique/free reflections</td>
<td>21,753/6,226</td>
</tr>
<tr>
<td>R factor/R$_{\text{free}}$ (%)</td>
<td>20.0/25.4</td>
</tr>
</tbody>
</table>

- R$_{\text{merge}} = \Sigma I(k) - (I)/\Sigma I(k)$, where $I(k)$ and $(I)$ represent the diffraction intensity values of the individual measurements and the corresponding mean values, respectively. The summation is over all measurements. R factor $= \Sigma |F_o| - |F_c|/\Sigma |F_o|$. R$_{\text{free}} = \Sigma |F_{o(\text{free})}| - |F_{c(\text{free})}|/\Sigma |F_{o(\text{free})}|$. $F_o$ is the observed structure factor; $F_c$ is the calculated structure factor based on the model. SAD data were collected in inverse beam mode. No $\sigma$ cutoff was applied to the data, and 10% of reflections were excluded from refinement for calculation of R$_{\text{free}}$. Rmsd, root mean square deviation. The last shell is from 2.33 Å to 2.25 Å for native data, and from 2.59 Å to 2.50 Å for SAD data.
(Figure 4.2). Through gel filtration and dynamic light scattering experiments, we also found that YjjX forms a dimer in solution (data not shown).

4.3.2 Structure Comparison

We carried out YjjX structural homology searches and found YjjX to have high structural similarity with Maf from *Bacillus subtilis* 60 (PDB code 1EX2), and with Mj0226 from *Methanococcus jannaschii* 59 (PDB code 1B78), with Z scores of 8.6 and 7.4, respectively (Figure 4.3). However, sequence alignments between YjjX with either of these proteins exhibit only 19% identity. Most interestingly, Mj0226 and Maf are both nucleotide binding proteins 59,60 and are classified in the HAM1 protein family. Similar to YjjX, these two proteins exhibit a mixed α/β fold, with the central β sheet surrounded by several helices and loops. It is this fold that forms the two lobes and a large cleft 195-197. The rms deviations between the structures of YjjX and Mj0226 and Maf are 2.8 Å and 3.2 Å for 114 and 128 Cα atoms, respectively. However, there are some very obvious differences between YjjX and the two homologous structures. The most notable deviations between these structures are found in the loops and helices of Lobe 1 and sections of Lobe 2 (Figure 4.3). The regions with the largest conformational deviations between Mj0226 and YjjX are also those regions that display high temperature factors in the individual molecules, suggesting flexibility of these segments. Due to the deviations in the flexible region of Lobe 1 in the two structures, the size and shape of the cleft located between Lobe 1 and Lobe 2 of YjjX is different from those of Mj0226 and Maf. The clefts in Mj0226 and Maf are more elongated and wider, whereas the cleft of YjjX is more enclosed and deeper (Figure 4.3).
Figure 4.1 A stereodiagram of the Overall Structure of the YjjX Monomer. The structure has two lobes: Lobe 1 (top) and Lobe 2 (bottom). There is a cleft between the two lobes. Lobe 1 contains α3, α4, α5, and α6 helices. Lobe 2 contains helix α1, helix α7, and stands β1 and β2. Three loops, 1, 2, 3, interact to form an active site cleft. All diagrams of the structure of YjjX were generated by PyMOL (http://www.pymol.org), unless otherwise stated.
Figure 4.2 Dimer Comparison of YjjX, Mj0226, and Maf. YjjX is shown on the top, Mj0226 is shown in the middle, and Maf is shown on the bottom. It is obvious that, despite excellent monomer homology, the dimers are very different for YjjX, Mj0226 (PDB code 1B78), and Maf (PDB code 1EXC).
Figure 4.3 The Stereodiagrams Showing Structural Comparison of the YjjX Monomer with the Mj0226 and the Maf Monomers. A) Superposition of YjjX (purple) with Mj0226 (green) has an rmsd of 2.79 Å. B) Superposition of YjjX (purple) with Maf (cyan) has an rmsd of 3.20 Å.
4.3.3 Active Site

Surface charge analysis for YjjX was performed in order to predict the location of a potential active site. Of particular interest was the large cleft positioned between two structural lobes (Figure 4.1). In this region of the structure, a number of amino acids, including Asn13, Lys16, Arg61, and Glu79 from Lobe 1 and Arg153 from Lobe 2, are highly conserved in all ten YjjX homologs from Archaea and other prokaryotic organisms. The electrostatic surface of YjjX clearly shows that this pocket is positively charged (Figure 4.4), indicating that it could interact with a negatively charged ligand. Additionally, strong and unambiguous electron density was observed for two sulfate ions in the pocket (Figure 4.4). The positioning of these ions is similar to the location of the α- and γ-phosphoryl groups of nucleoside triphosphates as observed in the Maf structure. It is important to note that one of these sulfate ions interacts directly with a number of conserved residues in this site, suggesting that these residues may be important for substrate recognition and stabilization in the cleft. Importantly, the putative active site is also in a location similar to the active sites of Mj0226 and Maf (Figure 4.4), providing further support for the identification of the active site in YjjX.

4.3.4 Identification of YjjX as an ITPase/XTPase

Active site analysis and structure comparison strongly suggested that YjjX is an NTP binding protein. Our experimental results clearly indicate that YjjX has nucleoside triphosphatase activity. Among the nucleoside triphosphates tested, ITP and XTP were the best substrates for YjjX and had apparent $K_m$ values of 0.5 mM and 1.17 mM, respectively, and have a $K_{cat}/K_m$ of 1147 mM$^{-1}$·s$^{-1}$ and 2243 mM$^{-1}$·s$^{-1}$, respectively (Table 4.2). dITP had an apparent $K_m$ of 4.51 mM and a $K_{cat}/K_m$ of 136.4 mM$^{-1}$·s$^{-1}$. No apparent activity was observed for dATP,
Figure 4.4 Electrostatic Surface and the Active Site Geometry Comparison. A) Positive charges are in blue, and negative charges are in red. YjjX possesses a more positively charged active site than Maf, and the YjjX pocket is more enclosed and deeper. Mj0226 exhibits a slightly negatively charged surface at the active site. The difference in charge and geometry at the active sites is assumed to contribute to the different substrate specificity. The diagrams were generated by GRASP \textsuperscript{200}. A stereodiagram for key residues in the active site is shown (right panel). B) Clear density for sulfate ions is shown in the binding pocket that is surrounded by conserved residues, such as Thr11, Asn13, Lys16, Arg61, and Arg153 (left panel). Close-up view of the putative active site with sulfates bound (right panel). The diagrams were generated with Pymol (http://www.pymol.org), Xfit in Xtalview \textsuperscript{201}, and GRASP\textsuperscript{200}.
dGTP, dCTP, or dTTP. For GDP, the $K_m$ value is 0.55 mM and the $K_{cat}/K_m$ is $205.7 \text{ mM}^{-1} \cdot \text{s}^{-1}$.

We observed non-Michaelis-Menten-type behavior in the enzyme kinetics, which is consistent with positive cooperativity. The effect of the addition of other metal ions is reported in Table 4.3. Furthermore, the addition of ATP, GTP, and TPP (thiamine pyrophosphate) resulted in inhibition of the ITPase activity of YjjX (Figure 4.5).

4.3.5 In Vitro Pull-Down Assay and Mass Spectrometry

We performed an in vitro pull-down assay by using His-tagged YjjX as bait. Three apparently distinct bands that were absent in the control lane, in which bait protein was not used, appeared on an SDS gel (Figure 4.6). Peptide fingerprinting with mass spectrometry of the trypsin-digested proteins revealed the most significant band to be an elongation factor Tu protein from *E. coli*. Sensitivity for the identification was very specific ($p < 0.05$), as the digested peptides covered 30% of the protein sequence (Figure 4.6). Other proteins isolated were heat shock protein and $\alpha$-galactosidase, which have no apparent relevance to YjjX and hence are not discussed further.

Mass spectrometric analysis of tryptic peptides is shown. MS data were submitted to the MASCOT search engine and used to match the nonredundant NCBI database. The inset table shows the identified peptide sequences from the EF-Tu protein of *E. coli*, which cover 30% of the protein sequence. The SDS-PAGE inset displays the proteins pulled down in vitro by using His$_6$-tagged YjjX as bait (lane 2). Ni$^{2+}$-nitrilotriacetic acid-agarose resin was employed to pull down interacting proteins. Ni$^{2+}$-nitrilotriacetic acid-agarose resin in the absence of the His-tagged YjjX served as a negative control (lane 1). Protein bands indicated by arrows were
Table 4.2 Kinetic Data of YjjX with Various Substrates

<table>
<thead>
<tr>
<th>Substrates</th>
<th>$V_{\text{max}}$ (µmol·min$^{-1}$·mg$^{-1}$)</th>
<th>$K_{\text{m}}$ (mM)</th>
<th>$K_{\text{cat}}$ (s$^{-1}$)</th>
<th>$K_{\text{cat}}/K_{\text{m}}$ (mM$^{-1}$·s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITP</td>
<td>1720</td>
<td>0.5</td>
<td>573.3</td>
<td>1147</td>
</tr>
<tr>
<td>XTP</td>
<td>7876</td>
<td>1.17</td>
<td>2625</td>
<td>2243</td>
</tr>
<tr>
<td>dITP</td>
<td>1849</td>
<td>4.51</td>
<td>616.5</td>
<td>136.4</td>
</tr>
<tr>
<td>ATP</td>
<td>n/d</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CTP</td>
<td>n/d</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>GTP</td>
<td>67</td>
<td>1.24</td>
<td>22.3</td>
<td>18</td>
</tr>
<tr>
<td>GDP</td>
<td>358</td>
<td>0.58</td>
<td>119</td>
<td>205.7</td>
</tr>
<tr>
<td>UTP</td>
<td>n/d</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>TPP</td>
<td>n/d</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>dATP (8 mM)</td>
<td>102</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>dCTP (8 mM)</td>
<td>65.3</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>dTTP</td>
<td>n/d</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>dGTP</td>
<td>n/d</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Fructose 1, 6 bis phosphate</td>
<td>n/d</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3-PGA (15 mM)</td>
<td>n/d</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

For kinetic measurements, YjjX was assayed in the presence of 10 mM MgCl$_2$ in the standard assay mixture with varying concentrations of substrate. n/d, not detected.
Table 4.3 Metal Cofactor of YjjX

<table>
<thead>
<tr>
<th>Metal Ion Cofactor</th>
<th>Relative Activity (%)</th>
<th>$V_{\text{max}}$ (µmol·min$^{-1}$·mg$^{-1}$)</th>
<th>$K_m$ (mM)</th>
<th>$V_{\text{max}}/K_m$ (µmol·min$^{-1}$·mM$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg$^{2+}$</td>
<td>100</td>
<td>1264</td>
<td>0.91</td>
<td>1389</td>
</tr>
<tr>
<td>Mn$^{2+}$</td>
<td>85</td>
<td>571</td>
<td>1.39</td>
<td>410</td>
</tr>
<tr>
<td>Zn$^{2+}$</td>
<td>28</td>
<td>439</td>
<td>2.44</td>
<td>179</td>
</tr>
<tr>
<td>Ca$^{2+}$</td>
<td>5.37</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cu$^{2+}$</td>
<td>0.66</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Li$^{2+}$</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>K$^+$</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Figure 4.5 Inhibition of ITPase Activity of YjjX. To assay inhibition of its ITPase activity, YjjX was assayed under standard assay conditions (pH 6.75, 25°C), with varying concentrations of ATP, GTP, and TPP.
Figure 4.6 Tryptic Digestion and Peptide Fingerprinting of In Vitro Pulled-Down Proteins Mass spectrometric analysis of tryptic peptides is shown. MS data were submitted to the MASCOT search engine and used to match the nonredundant NCBI database. The inset table shows the identified peptide sequences from the EF-Tu protein of *E. coli*, which cover 30% of the protein sequence. The SDS-PAGE inset displays the proteins pulled down in vitro by using His<sub>6</sub>-tagged YjjX as bait (lane 2). Ni<sup>2+</sup>-nitrilotriacetic acid-agarose resin was employed to pull down interacting proteins. Ni<sup>2+</sup>-nitrilotriacetic acid-agarose resin in the absence of the His-tagged YjjX served as a negative control (lane 1). Protein bands indicated
subjected to tryptic digestion and MS analysis. The elongation factor Tu from *E. coli* is marked by an asterisk.

### 4.4 Discussion

In the absence of a clear functional annotation for YjjX, we have determined the three-dimensional crystal structure of this protein for the purpose of deriving a structure-based function annotation. The overall structure of YjjX exhibits a mixed α/β fold and displays significant structural similarity to Mj0226, a nucleotide phosphatase from *M. jannaschii*, and to the UTP binding protein, Maf, from *B. subtilis*, with 19% sequence identity with either of these proteins. Closer examination of the shape and electrostatic surface potentials of the putative active site clefts of YjjX, Maf, and Mj0226 reveals that YjjX is more similar to Maf (Figure 4.4). In comparison to Maf, while there is some difference in the shape of the putative active site, similar electrostatic potentials are observed, suggesting that both can bind to small, negatively charged ligands. The active site in Mj0226 is more surface exposed and open, in contrast to YjjX, in which the pocket is buried. Furthermore, the putative active site of YjjX is more positively charged than the corresponding region in Mj0226. The better-defined and deeper pocket, together with more characteristic charge distribution in the putative active site of YjjX, suggests higher substrate specificity as compared to Mj0226.

Two negatively charged sulfate ions were found in the YjjX active site, one of which corresponds to the position of the phosphate moiety of the bound ligand in the Maf protein structure. The phosphate ion in Maf is in contact with Ala8, Ser9, Gln10, Ser11, Arg14, and Lys53, whereas in YjjX, one of the sulfate ions is surrounded by Ala10, Thr11, Thr12, Asn13, Lys16, and Arg61. The conservation of these residues suggests that the active site of YjjX is capable of nucleotide binding. In the case of Mj0226, it was predicted that Asp73 may play a key
role in the hydrolysis of PPI from NTP and that the enzyme is involved in the conversion of xanthine or inosine triphosphates to the corresponding monophosphates \(^{59,60}\). Interestingly, in YjjX, this aspartate is not found; however, Glu79 is in a similar position at the bottom of putative active site, and it may play a role similar to that of Asp73 in Mj0226 (Figure 4.4).

The key residues Thr11, Asn13, Lys16, Ser42, Glu46, Arg61, Glu79, Ser89, Try157, and Arg153 of YjjX correspond to the Thr15, Asn17, Lys20, Tyr43, Glu45, Asp73, Ser74, His177, and Arg178 residues in Mj0226 and to Ser9, Ser11, Arg14, Glu31, Glu34, Lys53, Asp70, Thr71, and Lys82 in Maf (Figure 4.4). In view of several lines of evidence, including ITP/XTP activity, the well-defined active site of YjjX, and its similarity to Mj0226 and Maf at monomer levels, we sought to determine a complex structure with ligand bound. Unfortunately, cocrystallization of YjjX with various nucleotides is unsuccessful under the native conditions, and screening for new conditions has not worked out. Moreover, crystals cracked in soaking experiments with ATP, GTP, ITP, AMP, AMPNP, or GMP. Meanwhile, computational ITP docking with AutoDock 3.0 \(^{202}\) was carried out, and it indicates that the putative active pocket has an excellent space fit for ITP, with the phosphate moiety in a similar position to the sulfate ion in the YjjX crystal structure (Figure 4.7).

The ITPase/XTPase activity of YjjX, as reflected by its \(K_m\) and \(K_{cat}\) values (Table 4.2), is comparable to those of human ITPase, \(EcO197\), and Mj0226 \(^{59,192,197}\). ITP is a favored substrate for YjjX because the enzyme has a low apparent \(K_m\) value for ITP; XTP is a more efficient substrate with higher turnover shown by \(K_{cat}/K_m\), though the apparent \(K_m\) value is higher. dITP was found to have comparable \(V_{max}\) and \(K_{cat}/K_m\) to ITP, but significantly higher \(K_m\), suggesting negligible activity toward dITP under physiological conditions. YjjX exhibited no phosphatase activity toward pyrimidines (CTP, TTP, UTP, dCTP, and dTTP), and this is in contrast with what
is seen with the purines GTP and GDP, which were hydrolyzed with modest efficiency. No activity was detected with dGTP, however. It is important to note that inosine has neither an amino group nor a keto group at the 2-position, an observation that may explain the relative specificity of YjjX for XTP or GTP. Another interesting and important observation is that the optimum pH for YjjX activity was 6.75, as compared to 9.6–10 for human ITPases. With regard to the effect of metal ions, similar to human ITPase, YjjX required Mg\(^{2+}\) for activity. We found that ITPase activity was maximal in the presence of Mg\(^{2+}\), while Mn\(^{2+}\) decreased the activity by 20% (Table 4.3). ATP, GTP, and TPP were found to be competitive inhibitors of ITPase activity (Figure 4.5). TPP was a more potent inhibitor (\(I_{50}\) 0.91 mM) than ATP and GTP (\(I_{50}\) 1.91 mM and 1.36 mM, respectively). A recent study with a genetic screen suggested the involvement of YjjX in thiamine metabolism. Specifically, YjjX was implicated in detoxification of bacimethrin and 4-amino-2-trifluoromethyl-5-hydroxymethylpyrimidine (CF3- HMP) in the cell, though the exact role of YjjX remains unknown. Our finding of TPP as an inhibitor of ITPase activity of YjjX may provide the missing link between YjjX and thiamine metabolism, based on which further functional analysis of YjjX in the context of thiamine metabolism can be taken.

As previously mentioned, YjjX forms an intimate dimer in the crystal structure and also in solution. It has been documented that ITPase is catalytically active only as a dimer comprised of two subunits. Our activity assay showed that the two active sites in the YjjX dimer have positive cooperativity (\(n = 1.84\)). Whereas YjjX, Mj0226, and Maf form homodimers, interestingly the way in which the two monomers associate differs for each protein. In YjjX, the interaction between the monomers is dominated by hydrogen bonding interactions between the
**Figure 4.7** Computational Docking of ITP in the Active Site of YjjX. ITP docking to the YjjX was performed by using AutoDock 3.0. ²⁰²
two β5 strands, whereas, in Mj0226, the dimeric interface is formed by the packing of hydrophobic residues 59.

So far the dogma has been that, unlike DNA replication, RNA transcription does not have a proofreading mechanism, although RNA polymerase has been shown to be capable of proofreading to correct undesired base incorporation and to avoid incorrect proteins 206. ITP mimics GTP and forms a very stable Watson-Crick pair 207,208. The incorporation of dITP inhibits further nucleotide addition in human polymerase II to an extent similar to that of mismatched bases. It has also been shown that, in elongation complexes, ITP incorporation decreases the stability of the RNA-DNA hybrid 209,210 have demonstrated that the overall elongation rate of the transcript in the presence of ITP is significantly lower than with the normal nucleotides, and in vitro experiments in which ITP was substituted for GTP reveal Rho-dependent early termination 211. In view of the fact that YjjX has a high specificity and turnover for ITP, dITP, and XTP, and no activity for dATP, dCTP, dGTP, dTTP, or any other nucleosides, we suggest that YjjX is an ITPase/XTPase probably functioning to prevent or minimize incorporation of undesired nucleotides into RNA.

In an attempt to provide further insights, we performed an in vitro pull-down assay that showed that elongation factor Tu was associated with YjjX. EF-Tu, a prokaryotic peptide elongation factor involved in protein translation, promotes codon-directed binding of aminoacyl-tRNA to the A site of ribosomes. When encountering stress conditions, such as oxidative stress and antibiotics, cells can trigger protective mechanisms to slow down or even stop translation to maintain viability. In fact, a novel stress-response protein that binds at the ribosomal subunit interface and arrests translation has been reported212. Given the fact that the elongation factor Tu may associate with YjjX, we postulate that YjjX may be a member of the class of cell-protective
enzymes. In addition to intercepting the incorporation of the dITP, ITP, or XTP into DNA or RNA by hydrolysis activity, YjjX may interact with EF-Tu·GDP or EF-Tu·GTP to hinder the formation of the EF-Tu·GTP--aminoacyl-tRNA complex.

In conclusion, the three-dimensional structure of previously unannotated, to our knowledge, YjjX has permitted the structure-based functional annotation. We have performed biochemical characterization of YjjX and confirmed it to be an ITPase/XTPase in *E. coli*. We postulate that YjjX may have profound implications in oxidative cell stress response, in which it plays a protective role.

4.5 Experimental Procedures

4.5.1 Materials

DNA restriction and modification enzymes were obtained from Fermentas (Burlington, Canada), Promega (Nepean, Canada), and Roche (Laval, Canada). PCR reagents were obtained from Invitrogen (Burlington, Canada). Protease inhibitor cocktail, ATP, ADP, AMP, AMPNP, CTP, GTP, GDP, UTP, TTP, TPP, and other fine chemicals were purchased from Sigma (Oakville, Canada). XTP was purchased from Jena Biosciences (Germany). Ni\(^{2+}\)-nitrilotriacetic acid-agarose affinity resin, the QIAprep plasmid extraction kit, and the QIAquick gel extraction kit were purchased from Qiagen (Mississauga, Canada). Reagents for *E. coli* culturing and protein overexpression were purchased from Bioshop (Canada). Crystallization screening kits were purchased from Hampton Research (Riverside, USA).

4.5.2 Cloning, Expression, and Purification

The open reading frame of *yjjX* was amplified by polymerase chain reaction by using 5′-CGAATTCCATGCTGCTGATTATGCACCAAGT-3′ as the forward primer and 5′-
CGCGGCCGCTATAACGGCATATTGAAACG-3’ as the reverse primer (restriction sites are in bold type). The resulting product was digested with EcoRI and NotI and ligated into the pET21b vector (Novagen). The recombinant YjjX was expressed as C-terminal hexahistidine-tagged protein in BL21 (DE3) E. coli cells under the T7 promoter. YjjX was purified to near-homogeneity by using a two-step protocol. The hexahistidine tag permitted quick and efficient first-step purification by using Ni\textsuperscript{2+}-nitrilotriacetic acid-agarose resin, followed by final purification with a Resource Q column on an Akta FPLC system (Amersham Biosciences). In brief, the frozen cells were resuspended in 50 ml 30 mM Tris-HCl (pH 8.0), 300 mM NaCl, 0.1% Triton X-100. Cells were lysed by sonication, and samples were kept on ice during the procedure. The supernatant was collected by centrifugation for 30 min at 15000 × g. The clarified extract was applied to preequilibrated Ni\textsuperscript{2+}-nitrilotriacetic acid-agarose affinity resin, washed with 20 mM imidazole in buffer containing 30 mM Tris-HCl (pH 8.0) and 300 mM NaCl. Protein was eluted in 50 ml of the same buffer containing 100 mM imidazole. Fractions were analyzed with 12.5% Laemli polyacrylamide gels\textsuperscript{213}. Fractions containing YjjX protein were pooled, and the protein was further purified by using a 6 ml Resource Q column. YjjX was eluted with a linear gradient of 0–500 mM NaCl in 20 mM Tris-HCl (pH 8.0), 30 mM NaCl. A single fraction of protein was recovered, and its purity was determined by SDS-PAGE. The protein was concentrated to 1 ml volume and buffer exchanged into 20 mM Tris-HCl (pH 8.0), 30 mM NaCl by using Centricon-3 (Millipore). Selenomethionine-substituted YjjX was expressed in DL41 (DE3) E. coli cells in LE Master medium supplemented with 50 mg/l D-L-selenomethionine (Fisher-Acros)\textsuperscript{214} and was purified under the same conditions as the native protein. The final yield was ∼55 mg of native protein and ∼10 mg of selenomethionine-labeled protein from 1 l E. coli culture.
4.5.3 Crystallization and Diffraction Data Collection

The preliminary crystallization conditions were screened by the sparse matrix method, by using standard screening kits. The optimal crystallization conditions for the native protein were 0.1 M Tris-HCl buffer (pH 8.0) containing 1–1.2 M lithium sulfate and 2% (v/v) MPD as a precipitating agent at room temperature. The protein concentration was 8 mg/ml, and the method of vapor diffusion by hanging drop was used. The crystallization conditions for the selenomethionine protein were the same as those for the native protein. Crystals appeared in 3 days and grew to full size within 1 week. All diffraction data were collected at cryogenic temperature (100 K). Data from the native YjjX crystal were collected at the home X-ray source equipped with a copper rotating-anode X-ray generator and a 30 cm Mar Research imaging plate. The crystal-to-detector distance was 180 cm, with an oscillation of 1° for a total of 130° of data collection. Data were indexed and integrated with DENZO and SCALEPACK (Otwinowski, 1993). The multiple-wavelength dispersion data were collected at beamline 19BM of the Advanced Photon Source at Argonne National Laboratory in Chicago by using an SBC1 3k × 3k CCD detector. Data of 360° were collected with 0.5° oscillation. The synchrotron data were indexed and integrated with HKL2000. The space group is monoclinic P2\(_1\) with cell dimensions: a = 71.9, b = 129.2, c = 85.6 Å, \(\beta = 90.27^\circ\) for both native and selenomethionine crystals. Cocry stallization and soaking with nucleotide were attempted. Various nucleotides dissolved in 50 mM Tris-HCl (pH 8.0) containing 300 mM NaCl were mixed with purified YjjX protein in a 10:1 ligand:protein molar ratio, tested for the similar crystallization conditions of the native protein, and underwent regular screenings. YjjX crystals were also soaked in crystallization solution containing nucleotides.
4.5.4 Structure Determination and Refinement

The single-wavelength anomalous dispersion (SAD) data at the inflection wavelength in combination with the native data permitted the heavy-atom positions to be located. The positions of eight selenium atoms and the initial phases were determined by SOLVE. The RESOLVE program was used for phasing calculation and refinement, including solvent flattening, noncrystallographic symmetry identification, and averaging. The initial YjjX structure was partially traced by RESOLVE. Additional manual tracing and model building were carried out by using Xfit in XtalView. CNS (version 1.1) was used for refinement. A structural homology search was carried out with program Dali.

4.5.5 Phosphatase Assays

All assays were linear with respect to time and concentration of the enzyme assayed. One unit (U) of activity is defined as the amount of the enzyme required for the hydrolysis of 1 µmol substrate per min at 25°C. Substrates were screened, and kinetic data were determined by measuring the P_i released in an enzymatic assay by coincubation of substrate with YjjX. The P_i released was measured colorimetrically. P_i released in the presence of triphosphate nucleotide and nucleoside substrates, but in the absence of YjjX in control assays, was measured for background correction.

Acid-washed, 96-well microtiter plates were used for all kinetic studies. Each 40 µl assay mixture contained 50 mM Bis-Tris (pH 6.75), 5 mM MgCl_2, and 1.25 nM protein. In the case of low and negligible activity, higher protein concentrations were used to confirm the results. Assays were initiated by the addition of 5 mM substrate (unless otherwise stated), allowed to proceed for 10 min at 25°C, and terminated by the addition of 70 µl color developing reagent,
which was prepared fresh by mixing six parts 0.42% ammonium molybdate in sulfuric acid with one part 10% (w/v) ascorbic acid. Samples were incubated for 20 min at 45°C, and the $A_{660}$ was determined by using a Powerwave XS Microplate spectrophotometer (Bio-Tek Instruments, Inc.). In all cases, controls were run for background correction of $P_i$ present at each substrate concentration tested. To calculate activities, a standard curve over the range of 1–100 nmol $P_i$ was prepared for each set of assays.

### 4.5.6 Kinetic Analysis

Apparent $V_{\text{max}}$ and $K_m$ values were calculated from the Hill equation fitted by using a nonlinear least-squares regression kinetics program $^{221}$. The $I_{50}$ values (concentration of inhibitor producing 50% inhibition of phosphatase activity) were determined by using this program. All kinetic values are the means of three separate measurements and are reproducible within ±10% SE of the mean value. Protein concentrations were determined with the Coomassie blue G-250 dye binding method $^{222}$ by using bovine $\gamma$-globulin as the protein standard.

### 4.5.7 In Vitro Pull-Down Assay

In vitro His-YjjX bait pull-down experiments were carried out at 4°C. Each experiment used cells from 500 ml $E. \text{coli}$ culture grown overnight at 37°C. The cells were lysed by sonication in buffer A (50 mM NaH$_2$PO$_4$/Na$_2$HPO$_4$ [pH 8.0], 150 mM NaCl, 10 mM EDTA, 1 mM DTT, 1 mM PMSF, 0.1% Triton X-100, 20 µl protease inhibitor cocktail [Sigma] per 20 ml lysis buffer) and clarified by centrifugation. The lysate was mixed with 150 µl of 0.5 mg of Histagged YjjX and a 40 µl bed volume of preequilibrated Ni$^{2+}$-nitrilotriacetic acid-agarose resin and incubated for 1 hr. The control experiment included Ni$^{2+}$-nitrilotriacetic acid-agarose resin
only without the bait protein. After collecting the supernatant, the beads were washed three times with 1 ml buffer A and three times with 1 ml buffer A containing 50 mM imidazole. Subsequently, beads were boiled for 5 min in 50 µl loading dye, and 10 µl of each sample was analyzed on a 16% SDS-polyacrylamide gel and visualized with Coomassie brilliant blue dye.$^{213}$

4.5.8 Peptide Fingerprinting by Mass Spectrometry

For MS analyses, Coomassie blue-stained bands were excised from SDS gels and digested with trypsin by using standard protocols. Samples destined for MALDI-TOF MS were mixed 1:1 (v/v) with 5 mg/ml α-cyano-4-hydroxy-cinnamic acid (Sigma) matrix in 50% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid. For calibration purposes, sample spots are near point lock mass corrected by spotting a mixture of synthetic peptides at 1 pM/µl in the appropriate standard wells of the target plate. The peptide standard is composed of Angiotensin I (1296.687 Da), [glu]Fibrinopeptide B (1570.677 Da), Renin (1758.933 Da), and ACTH (2465.199 Da). Lock mass correction results in a mass accuracy of approximately 30 ppm over the range of 1000–3000 Da. Spectra were acquired and analyzed on a Micromass MALDI (Waters Corporation; Milford, MA, USA) by using MassLynx 4.0 software (Waters Corporation). Spectra were acquired in positive-ion reflector mode. A total of 40 laser shots were averaged per spectrum by using an 8%–95% on-scale sample intensity energy setting. The mass range scanned is typically 1000–3000 Da. Raw spectra were processed under ProteinLynx Global Server 2.0 (Waters Corporation) to produce centroid data. The top 50 most intense peaks were submitted for the database search. Search parameters were set at trypsin with up to one missed cleavage, fixed carbamidomethyl modification of cysteine, and variable oxidation of methionine.
Acknowledgments

This work is supported by the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada (NSERC). We thank Dr. Graham Cote for constructive discussions. We are grateful to Melanie Adams for her critical reading of the manuscript. The cloning of \textit{yj}jX by Christian M. Udell is greatly appreciated. We also acknowledge the support of Drs. M. Cygler and A. Matte and the Protein Function Discovery program at Queen’s University. Z.J. is a Canada Research Chair in Structural Biology and a recipient of the NSERC Steacie Fellowship.
Chapter 5

Structural and biochemical characterization of YhdE, a novel dTTP/UTP pyrophosphatase from *Escherichia coli*

*This chapter will be submitted for publication.*

Jimin Zheng and Zongchao Jia*

Department of Biochemistry, Queen's University, Kingston, ON, K7L 3N6, Canada

*Correspondence e-mail: jia@queensu.ca

Jimin Zheng was responsible for protein expression, purification, crystallization, data collection, structure solution and refinement, and all biochemical experiments in the investigation of YhdE function. The *yhdE* and *YhdE* E33A mutant genes were cloned by John Wagner. The manuscript was written by Jimin Zheng with editorial input from Dr. Zongchao Jia and with additional help from Brent Wathen.
5.1 Abstract

A novel dTTP pyrophosphatase, YhdE, from *Escherichia coli* has been characterized. The gene, *yhdE*, is part of the *mre* operon in *E. coli*, which is known as the cell shape operon. The optimal *in vitro* enzymatic conditions for YhdE activity is 25 °C, 5 mM Mn$^{2+}$ in 20 mM HEPES buffer pH 7.0. The apparent $K_m$ of YhdE for dTTP is 0.09 mM and the apparent $k_{cat}/K_m$ is 503 mM$^{-1}$sec$^{-1}$. Some activity was also found using UTP as a substrate, but the $K_m$ was greater at 0.50 mM, and the apparent $k_{cat}/K_m$ was smaller at 191 mM$^{-1}$. Residue E33 of YhdE was found to play a significant catalytic role: when mutated to alanine, the YhdE dTTPase activity was abolished. The structures of the apo YhdE E33Amutant and the YhdE_E33A/dTTP complex were determined for mechanistic analysis. Moreover, overexpression of YhdE in *E. coli* demonstrates a similar phenotype to that of Maf, a YdhE homologue in *B. subtilis*, and the yhdE-null cell was observed to grow much faster when compared to the wild type and to the YhdE overexpression *E. coli* strain. The identification that YhdE is a dTTPase suggests that its cellular function might involve the regulation of the intracellular dTTP pool, thereby affecting DNA replication and cell division.
5.2 Introduction

The *E. coli* mre (murine region) operon consists of five genes, *mreB*, *mreC*, *mreD*, *yhdE* and *rng* \(^{76-78,223-225}\). The genes *mreB*, *mreC* and *mreD* are essential for maintaining the natural rod cell shape of bacteria such as *Bacillus subtilis* (*B. subtilis*) and *E. coli* \(^{80,81}\). When bacteria cells are depleted of MreB, MreC or MreD, either individually or as a complex, the cells display a spherical phenotype \(^{82,226,227}\). MreB is a prokaryotic homolog of actin found beneath the cell surface in a helical array \(^{81}\). In addition to its involvement in regulating cell width, MreB is important for chromosomal DNA segregation and cell division \(^{80,81,83}\). The cellular localization and functions of MreB seem to be coordinated with the recruitment of cell division machinery such as FtsZ \(^{228,229}\). Like actin, MreB can form polar filaments \(^{230}\) and its inherent polarity contributes to the dynamic localization of regulatory proteins and machinery, as well as to the global polarity of the cell \(^{229,231,232,233}\). MreB also forms a membrane-associated complex with MreC and MreD \(^{82}\). Although the cellular and biochemical functions of MreC and MreD are still unknown, there is evidence suggesting that these two proteins act as localization and/or anchoring machinery for MreB \(^{82,226,234}\). How these proteins are mechanistically and functionally coordinated remains unknown.

In *E. coli*, downstream from the *mre* genes are *yhdE* and *rng* (previously called *cafA* or *orfF*). RNaseG, encoded by the *rng* gene, is a non-essential ribonuclease specific for adenine and uracil rich regions \(^{84,85}\). RNaseG is homologous to the amino-terminal part of RNaseE \(^{86}\), a protein involved in the regulation of the FtsZ/FtsA ratio \(^{87}\). FtsZ and FtsA are proteins involved in septum formation and cell division \(^{88,89}\). Overexpression of RNase G in *E. coli* results in the production of cytoplasmic axial filaments that cause the formation of chained cells and minicells, suggesting that RNase G is involved in chromosome segregation and cell division \(^{77}\).
The *yhdE* gene (previous called *orfE*) in the *mre* operon of *E. coli* encodes a 21.5 kDa protein, which is the only gene in the *mre* operon of unknown function. Its orthologous genes are notably conserved in most of the completely sequenced bacterial genomes, as well as in eukaryotes and in some of the archaea. However, to date, no specific function has been assigned to any of the corresponding proteins. Previous studies of the *maf* gene from *B. subtilis*, the counterpart of a *E. coli yhdE* gene, found that overexpression of *maf* in *B. subtilis* cells using a multicopy plasmid led to elongated cell shape, formation of filamentous cells and also appeared to arrest septum formation, whereas insertional inactivation of the *maf* gene suggested that it was not essential for cell viability in either rich or minimal medium. Moreover, functional characterization of the yeast *Saccharomyces cerevisiae* genes by high-throughput gene deletion and parallel analyses found that *yor111*, the *yhdE* homolog in yeast, is not essential for cell viability in either rich or minimal medium. The three-dimensional X-ray crystallographic structure of the Maf protein has been determined in both its apo-form as well as bound to dUTP. There is structural similarity to the *Methanococcus jannaschii* Mj0226 dNTP pyrophosphatase and the *E. coli* YjjX ITPase/XTPase, both of which are nucleotide hydrolases. However, the Maf protein was not found to have the ability to cleave any of the phosphate bonds in either dUTP or dATP.

A comprehensive analysis of previous findings involving the *mre* operon revealed a common phenotypic feature of the *mreB, yhdE*, and *rnG* genes. Overexpression, but not inactivation, of the products of these genes prevents cell division resulting in the formation of filamentous cells. There are also indications that these genes might be involved in cell septum formation or cell division. These observations imply that the genes in the *mre* operon might be cooperatively involved in a common physiological pathway that may play a regulatory role in...
cell division. However, it is not clear how these genes coordinate with each other both mechanistically and functionally. An understanding of YhdE function would not only shed light on YhdE’s role but also provide novel insights into the regulation of cell division. Towards this end, we have in this study identified the biochemical function of YhdE. We established that YhdE is a pyrophosphatase which is highly specific, primarily for the deoxyribonucleotide dTTP and secondly for UTP. From a structural analysis of Maf, we were able to pinpoint a catalytic YhdE residue, Glu33, which, upon mutation to Ala, leads to a complete loss of its dTTPase activity. Further, we determined the structures of the YhdE E33A mutant with and without TTP present in the crystallization conditions. The structure showed a different dimerization mode than is exhibited by Maf, even though the overall structural differences between YhdE E33A and Maf are minimal. Since the YhdE overexpression phenotype comprised slow growth with elongated filaments due to arrested cell division and the \textit{yhdE} gene-knockout \textit{E. coli} was found to have increased cell grow, we speculate that YhdE, being a member of the \textit{mre} operon, might coordinate with other operon members to regulate DNA segregation and cell division by controlling the dTTP pool in the cell.

5.3 RESULTS

5.3.1 Preliminary investigations of YhdE activity

To investigate its activity, YhdE was expressed and purified to near homogeneity on a Ni$^+$ column. The subsequent purification with size-exclusion chromatography indicated that YhdE exists as a dimer in solution with a size of approximately 45 kDa (Figure 5.1). The purified YhdE protein was concentrated to about 20 mg/ml. Previous studies of Maf, the \textit{B. subtilis} homolog of YhdE, found that it has nucleotide binding abilities, but when tested no phosphate bond cleaving activity was demonstrated for either dUTP or dATP. In our nucleotide screening
assay, a variety of NTPs and dNTPs at 100 µM and 500 µM were individually mixed with YhdE to test for hydrolysis activity on these canonical and noncanonical nucleotides; nucleotides tested included ATP, CTP, GTP, UTP, TTP, TDP, TMP, dCTP, dTTP, ITP and XTP. Preliminary tests showed that incubation of YhdE with TTP, UTP and dTTP resulted in the liberation of inorganic phosphate, strongly indicating that these three (d)NTPs are probably YhdE substrates. At a substrate concentration of 100 µM, YhdE demonstrated specific activity for dTTP alone. At a substrate concentration of 500 µM, in addition to dTTP, pyrophosphate activity was observed for TTP and UTP. YhdE TTP pyrophosphatase activity was observed to be less than 50% of that for dTTP, while YhdE UTP pyrophosphatase activity was about equal to that for dTTP (Figure 5.2A). No activity was observed at 100 µM and 500 µM concentrations for the following nucleotides: ATP, CTP, GTP, TDP, TMP, dCTP and dUTP.

5.3.2 Optimal conditions for YhdE pyrophosphatase activity

Optimal conditions for in vitro enzymatic kinetic activity of YhdE were investigated by analyzing the effects of temperature, metal ions and pH on activity. YhdE pyrophosphatase activity with TTP as substrate was analyzed at various temperatures, with maximal activity achieved at 25 °C; at higher temperatures (37 °C, 45 °C), activity decrease by 60% (Figure 5.2C). YhdE nucleotide hydrolysis activity was also analyzed with and without the addition of metal ions. The presence of divalent cations was essential for YhdE pyrophosphatase activity (Figure 5.2B). Optimal activity was achieved with the addition of Mn²⁺, although activity was still observed with the addition of Zn²⁺, Cu²⁺, Mg²⁺ and K⁺, albeit at lower levels. In addition, a pH range of 6.0-11.0 was also examined for its effects on YhdE pyrophosphatase activity. At neutral pH, YhdE demonstrated maximum activity, while in more alkaline conditions activity
Figure 5.1 A) Fast Protein Liquid Chromatography Purification. Absorbance at 280 nm (y-axis) as a function of volume eluted (x-axis). The buffer used was Bis-Tris buffer, pH 6.75. YhdE eluted at about 81 mL, corresponding to a molecular weight of approximately 45 kDa. This corresponds to the dimer of YhdE. B) Purification of YhdE. Shown is the purified YhdE protein sample run on an SDS-PAGE gel after concentrating to about 20 mg/ml. Lane 1: molecular weight ladder; Lane 2: YhdE protein sample. This shows that YhdE has a molecular weight of about 22 kDa, equivalent to that of the monomeric YhdE.
dropped to about 40% of the optimal level (Figure 5.2D). The optimal in vitro condition for YhdE was therefore determined to be 2 mM Mn$^{2+}$, 20 mM HEPES buffer pH 7.0 at 25 °C.

5.3.3 Determination of YhdE products by mass spectrometry

To determine which phosphate bond is cleaved by YhdE, mass spectrometry analysis was utilized. The mass spectra were obtained for control samples containing TTP, TDP and TMP, as well as for a reaction sample containing YhdE with substrate TTP and the resulting products. Analysis of the mass spectra of the reaction sample had a peak corresponding to a molecular weight of 321.239 (Figure 5.3B). This is equivalent to the molecular weight corresponding to the peak found on the mass spectra of TMP (Figure 5.3A).

From the mass spectrometry analysis, we conclude that TMP is the product when YhdE and TTP are incubated, implying that free inorganic pyrophosphate must also be a product. There was no indication of any TDP formation (Figure 5.3B), eliminating the possibility that YhdE was able to cleave the $\gamma$-$\beta$ phosphate bond. The mass spectrometry results confirm that YhdE has phosphate bond cleavage ability and more specifically, cleaves at the $\alpha$-$\beta$ phosphate bond. This result establishes that YhdE is a pyrophosphatase.

5.3.4 Enzymatic kinetics of YhdE pyrophosphatase activity

The enzymatic parameters of YhdE pyrophosphatase activity were determined for UTP and dTTP (Table 5.1). Activity saturation was reached for both UTP and dTTP (Figure 5.5). The apparent $K_m$ of UTP and dTTP was 0.50 mM and 0.09 mM, respectively. Despite the fact that UTP has a greater apparent $k_{cat}$ (95 s$^{-1}$) than dTTP (43 s$^{-1}$), dTTP still displayed a greater
Figure 5.2 A) YhdE Substrate Specificity. Two different concentrations of substrate were tested. 1.5 µM YhdE, 2 mM Mn2+, 20 mM HEPES buffer pH 7.0 mixed with either 100 µM or 500 µM TTP was incubated for 10 minutes at 25 °C. Each experiment was completed in triplicates twice. Shown is the percent activity for each of the different nucleotides tested. Activity is expressed as percentage of the YhdE activity with dTTP as substrate. B) Optimization of Temperature. 6.25 µM YhdE, 5 mM Mn2+, 20 mM HEPES buffer pH 7.0 and 100 µM TTP was incubated for 10 minutes at each temperature. Each experiment was completed in triplicates twice. C) Optimization of Metal Ions. 6.25 µM YhdE, 5 mM metal ion, 20 mM HEPES buffer pH 7.0 and 100 µM TTP was incubated for 10 minutes at 37 °C. Each experiment was completed in triplicates twice. D) Optimization of pH. 6.25 µM YhdE, 5 mM Mn2+, 20 mM buffer and 100 µM TTP was incubated for 10 minutes at 25 °C. Each experiment was completed in triplicates twice. The standard error of the mean was used to calculate the error bars.
Figure 5.3 Mass Spectrometry Analysis.  
A) Control: TTP with a molecular weight of 481.129, TMP with a molecular weight of 402.211, and TMP with a molecular weight of 321.239.  
B) Reaction sample: 6.25 µM YhdE, 5 mM Mn$^{2+}$, 20 mM HEPES, pH 7.0 and 500 µM TTP. Reaction time: 2 hours at 25 °C. Only a peak matching the TMP peak of 321.239 was observed.
apparent catalytic efficiency. This value is represented by the apparent $k_{\text{cat}}/K_m$ values, which were 191 mM$^{-1}$ s$^{-1}$ and 503 mM$^{-1}$ s$^{-1}$ for UTP and dTTP, respectively. Our analysis also indicates that the apparent Hill coefficient for UTP and dTTP is approximately two.

5.3.5 YhdE E33A mutant

Based on an analysis of the structure determined for the Maf-like protein from *Trypanosoma brucei*, residue Glu45 was found to be important for coordinating the metal ions (Figure 5.6). We performed molecular docking simulation of dTTP using AutoDock 3.0$^{202}$. The results showed that the $\gamma$- and $\alpha$-phosphates of dTTP fit the positions of the two sulfates and co-ordinate the manganese ions. This suggests that Glu45 may play a key role in coordinating the metal ions to stabilize nucleotide binding, making it central to the hydrolysis reaction. This residue is conserved in YhdE and is found at position E33. To test the involvement of this residue in enzyme activity, the mutant YhdE E33A was generated, expressed and purified for testing YhdE pyrophosphatase activity with a variety of nucleotides (ATP, CTP, GTP, TTP, TDP, TMP, UTP, dCTP, dTTP, and dUTP). No YhdE pyrophosphatase activity was observed for any of the nucleotides.

5.3.6 The crystal structures of the *E. coli* YhdE E33A mutant

To date, no crystal structure of YhdE from *E. coli* has been reported, although apo-structure of its homolog is available (the Maf-like protein from *Trypanosoma brucei*, PDB accession code: 2AMH). Since the YhdE E33A mutant knocked out YhdE dTTP/UTP pyrophosphatase activity, we reasoned that the dTTP substrate can be trapped in the active site of
Figure 5.4 A) UTPase pyrophosphatase activity of YhdE. For kinetic measurements, 1.5 µM YhdE was assayed in standard conditions (2 mM Mn$^{2+}$, 20 mM HEPES buffer pH 7.0, 25 °C), along with various concentrations of the substrate UTP. B) dTTPase pyrophosphatase activity of YhdE. For kinetic measurements, 1.5 µM YhdE was assayed in standard conditions (2 mM Mn$^{2+}$, 20 mM HEPES buffer pH 7.0, 25 °C), along with various concentrations of the substrate dTTP. Kinetic parameters were calculated using SigmaPlot.
Table 5.2 Kinetic Data of YhdE with dTTP and UTP

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Substrate Structure</th>
<th>(K_m) (mM)</th>
<th>(V_{max}) (mM s(^{-1}))</th>
<th>(k_{cat}) (s(^{-1}))</th>
<th>(k_{cat}/K_m) (mM(^{-1}) s(^{-1}))</th>
<th>(n_H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dTTP</td>
<td><img src="image1" alt="dTTP Structure" /></td>
<td>0.09</td>
<td>0.06</td>
<td>43</td>
<td>503</td>
<td>2.2</td>
</tr>
<tr>
<td>UTP</td>
<td><img src="image2" alt="UTP Structure" /></td>
<td>0.50</td>
<td>0.14</td>
<td>95</td>
<td>191</td>
<td>1.9</td>
</tr>
</tbody>
</table>

For kinetic measurements, 1.5 µM YhdE was assayed in standard conditions (2 mM Mn\(^{2+}\), 20 mM HEPES buffer pH 7.0, 25 °C) with various concentrations of the substrate.

**Figure 5.5** Crystal structure of the Maf-like protein from *Trypanosoma brucei* (PDB accession code: 2AMH). Two sulfates and two manganese ions were found in the crystal structure. The Glu45 residue co-ordinates two manganese ions. A dTTP molecule was docked into the active site devoid of sulfates with the program AutoDock 3.0. \(^{202}\) In the docking, the γ- and α-phosphates of dTTP fit the positions of the two sulfates and co-ordinate the manganese ions. Sequence analysis showed Glu33 in YhdE is the counterpart of Glu45 in Maf-like protein.
YhdE, a beneficial outcome that would allow for the characterization of the YhdE/dTTP interaction. We therefore attempted to crystallize YhdE E33A in the presence and absence of dTTP. Crystals were obtained for both scenarios in very similar crystallization conditions (Figure 5.6), with space groups of $P4_2$ without dTTP and $P2_12_12_1$ with dTTP. Crystal structures of apo YhdE E33A and the YhdE_E33A/dTTP complex were solved at resolutions of 2.0 Å and 2.5 Å, respectively, by molecular replacement using the Maf protein structure as a search template (Table 5.2). In the crystal structures, YhdE forms a homodimer (Figure 5.7A), consistent with the results of our size exclusion experiments ($M_r = 45$ kDa). Similar to the Maf protein structure, the subunit structure contains YhdE E33A monomer with an $\alpha/\beta$ fold that has seven $\alpha$-helices and six extended, mixed $\beta$-strands. The six strands form a twisted $\beta$ sheet at the center of the protein, which connects two lobes (lobe A and B). A large cavity is located between the two protein lobes. Most of the conserved residues of YhdE are located inside this cavity, suggesting that it accommodates the active site of this enzyme. The conserved residues include Ser8, Ser10, Arg13, Val28, Glu31, Glu33, Lys52, Asp70, and Lys82. A similar large cavity contains the active sites of YjjX, MJ0226 and ITPA. Although Maf also forms a dimer in solution, the dimeric mode of YhdE E33A is different from that of Maf. In YhdE, the interface is at a small $\beta$-stand between the two monomers, which renders the A lobe of one monomer next to the B lobe of another monomer, bringing the two active sites close together. In Maf, the interface is at the elongated $\beta$-stand between the two monomers, which render the B lobe of one monomer next to the B lobe of another monomer, separating their active sites on opposite sides of the Maf dimer. This may explain the cooperative nature of the YhdE enzymatic activity: binding of one substrate molecule will affect the binding of another substrate molecule in this extended active site. In contrast, we propose that Maf may not have this cooperative reaction feature.
Figure 5.6  

A) A YhdE E33A crystal ($P4_3$) and its typical diffraction image.  

B) A YhdE_E33A/dTTP complex crystal ($P2_12_12_1$) and its typical diffraction image.
Table 5.2 Statistics of X-ray diffraction data.

<table>
<thead>
<tr>
<th>Protein form</th>
<th>YhdE_E33A</th>
<th>YhdE_E33A+TTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.5418</td>
<td>1.5418</td>
</tr>
<tr>
<td>Space group</td>
<td>$P4_3$</td>
<td>$P2_12_12_1$</td>
</tr>
<tr>
<td>Unit-cell parameters (Å)</td>
<td>$a = b = 70.48, c = 78.83$</td>
<td>$a = 45.974, b = 86.047, c = 88.558$</td>
</tr>
<tr>
<td>Total number of reflections</td>
<td>71565</td>
<td>101687</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>24143 (2135)</td>
<td>17495 (1270)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50-2.0</td>
<td>40-2.5</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.7 (87.6)</td>
<td>96.2 (71.4)</td>
</tr>
<tr>
<td>$&lt;I/\sigma&gt;$</td>
<td>17.7 (4.4)</td>
<td>26.8 (3.7)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.0 (2.5)</td>
<td>5.8/3.0</td>
</tr>
<tr>
<td>Molecules per AU</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$R_{\text{merge}}$ (%)*</td>
<td>6.9 (23.3)</td>
<td>6.4 (28.2)</td>
</tr>
<tr>
<td>$R/R_{\text{free}}$ (%)</td>
<td>16.6/21.0</td>
<td>18.9/23.6</td>
</tr>
<tr>
<td>Solvent content</td>
<td>0.45</td>
<td>0.48</td>
</tr>
</tbody>
</table>

$R_{\text{merge}} = \sum_{hkl} \sum_{i} [I_i(hkl) - \langle I(hkl) \rangle]/\sqrt{\sum_{hkl} \sum_{i} I_i(hkl)}$, where $I_i(hkl)$ is the intensity of the $i^{th}$ observation of reflection $hkl$ and $\langle I(hkl) \rangle$ is the weighted average intensity of $i$ observations of reflection $hkl$. Values in parentheses are for the highest resolution shell.
Comparison of the apo-YhdE E33A structure and the YhdE_E33A/dTTP complex structure show that dTTP binding induces a movement of several structural elements toward the bound substrate, resulting in a “closed” conformation of the enzyme (Figure 5.9). The most prominent structural change includes the downshift of the helices and the loops in the A lobe and the upshift of the loop and side chains in the B lobe by 1.6-11.5 Å, closing the active site and moving a cluster of conserved residues (Ser8, Gly9, Ser10, Lys12, Glu15, Arg26, Val28 and Asp145) close to the substrate (Figure 5.8). In the YhdE_E33A/dTTP complex structure, the side chain of Arg13 is positioned next to the Pγ and Pβ oxygens of the bridging Pγ-O and Pβ-O bonds of dTTP (3.4 and 2.9 Å away), where it appears to stabilize these phosphates during catalysis. The ITP-induced shift of the α-helix in the B lobe moves the side chains of the conserved Ser8 close to the oxygen atoms of the substrates α-phosphate (3.4 Å and 3.6 Å away). Further stabilization of the substrate triphosphate moiety is achieved by the interaction of Lys52 with the α-phosphate oxygen and Lys82 with the γ-phosphate oxygen. Conserved Arg26 is shifted close to interact with the 6-keto oxygen of the base, moving 11.5 Å in the direction of the substrate. In addition, YhdE_E33A-dTTP binding down shifted the loop and the conserved Asp145 by 2.0 Å towards the nucleotide base, closing the substrate in the active site. This closed confirmation is also observed in the Maf-dUPT complex structure (Figure 5.8B), although the dUTP in this Maf complex has a different orientation than the dTTP in the YhdE E33A complex. This may explain why dUTP is not a substrate for YhdE.

In summary, substrate binding and hydrolysis causes YhdE to shift between open and closed conformations because of interactions between several of the static and movable YhdE elements and phosphates of the substrate. The static components include a fixed β-sheet floor accommodating the predicted catalytic residue Asp70 and an associated hydrogen-bonding
Figure 5.7 Dimer comparison of YhdE E33A with the Maf protein. A) The YhdE E33A dimer. B) The Maf protein dimer. The active sites are marked by yellow arrows.
Figure 5.8 A) Structural superposition of apo-YhdE E33A (white) with the YhdE_E33A/dTTP complex (magenta). B) Structural superposition of the YhdE E33A/dTTP complex (magenta) with the Maf/dUTP complex (green). C) Comparison between the open structure of YhdE E33A (white) and the closed dTTP bound conformation (light green). The interactions between dTTP and the residues in the active cleft are shown. D) dTTP in the charged-surface pocket (blue: positively charged; red: negatively charged).
network (Lys52), as well as a group of conserved residues involved in substrate binding (such as Arg13). The movable elements include the α-helix and the flexible loops in the two lobes containing residues involved in substrate coordination and catalysis (Ser8, Ser10, Arg12, Arg 26, Val28, Lys 82 and Asp145). The binding of the dTTP molecule to the open YhdE conformation induces a large shift of movable components toward the bound substrate, transforming the enzyme into a closed conformation, which reverts back to an open conformation after substrate hydrolysis.

5.3.7 YhdE overexpression and cell morphology

Previous studies suggested that Maf overexpression impairs cell septum formation, leading to elongated cell shape and the formation of filamentous cells. As the counterpart gene of the maf gene in B. subtilis, yhdE in E. coli is however in a different gene organization. As mentioned, the yhdE gene is located in the mre+ operon, whereas the maf gene is not in the mre+ operon, but is instead located upstream of mre+ operon. Does YhdE have the same influence for E. coli phenotype as Maf for B. subtilis? To examine whether morphological changes would ensue upon overexpression of YhdE, we performed electron microscopy (EM) of phosphotungstic acid-stained BL21(DE3) cells harboring pET-YhdE-His₆ in which YhdE expression was either endogenously produced or overexpressed during log phase growth induced by IPTG. The morphology of BL21(DE3) YhdE-overexpressing cells appeared quite elongated, with a propensity for clustering (Figure 5.9). This morphological change suggests that the phenotype of YhdE is likely consistent with that of Maf, including the observation that overexpression impairs cell septum formation and blocks cell division.
5.3.8 Cell growth

Interestingly, during the production of YhdE for crystallization studies, we observed that the overexpression of YhdE significantly increased cell doubling time. Further investigation of the growth rate of using wild type *E. coli* K-12 W3110 and a *yhdE*-knockout strain as well as cells overexpressing YhdE showed that the growth of *yhdE*-knockout cells is faster than that of wild type cells. Knockout cells enter the log phase earlier than wild type cells, indicating faster cell division, whereas YhdE overexpression slows cell growth.

5.4 Discussion

The biochemical and structural properties of YhdE have not been extensively studied in *E. coli*. Initial results from Minsov *et al.* suggested that the YhdE homolog in *B. subtilis*, Maf, has nucleotide binding abilities, but it is undetermined if Maf is able to specifically cleave any of the nucleotide phosphate bonds. In this study we determined the structure of the YhdE E33A mutant alone and complexed with dTTP. The overall structure of YhdE is very similar to that of Maf, and both are also similar to the structures of YjjX and Mj0226 ITPase. ITPases are pyrophosphatase proteins highly specific for the non-canonical nucleotides ITP and XTP. YhdE contains the same fold found in ITPases and conserves some of the same residues found around their triphosphate binding site, which suggests that it shares a similar nucleotide binding and catalytic mechanism with these enzymes. The structural features strongly suggest that YhdE and Maf are pyrophosphatase proteins. Based on this premise, we screened canonical and noncanonical nucleotides to find YhdE substrates. Initially, TTP, UTP and dTTP were found to be hydrolyzed by YhdE. By using mass spectrometry to analyze an incubated solution containing YhdE and its substrate TTP, it was determined that YhdE cleaves the α-β phosphate bond and is
Figure 5.9 EM of cells overexpressing YhdE. Electron micrographs of phosphotungstic acid-stained BL21(DE3) cells containing pET-YhdE-His<sub>6</sub> in which YhdE expression was not induced (A) or induced (B) with IPTG. Scale bars represent 2.0 µm.
**Figure 5.10** Growth curve of the wild type *E. coli* strain (WT), cells overexpressing (OE) and the *yhdE*-deleted *E. coli* strain (KO), in an LB medium under aerobic conditions at 30 °C.
a pyrophosphatase. We found no mass spectrometry evidence that YhdE cleaves the β-γ phosphate bond. This result suggests that YhdE is a novel pyrophosphatase.

The use of a pyrophosphate detection kit found that YhdE has varying degrees of pyrophosphatase activity for the following nucleotides: TTP, UTP and dTTP. Consistent enzymatic kinetics was obtained for UTP and dTTP. The $K_m$ for dTTP was about 5.5-fold smaller than the $K_m$ for UTP. A much smaller concentration of dTTP than UTP is required to reach half the maximal velocity for enzymatic activity. The catalytic efficiency, $k_{cat}/K_m$, also differs between the two substrates. It is about 2.5-fold greater for dTTP than UTP. Taken together, the $K_m$ and catalytic efficiency strongly suggest that dTTP is the preferred substrate of YhdE.

The kinetic results indicated that both substrates follow a non-Michaelis-Menten model and participate in positive cooperative binding to YhdE, indicated by Hill’s coefficient (n) greater than one (the Hill’s coefficients for both dTTP and UTP are approximately two). YhdE have one binding sites per subunit and YhdE exists as a dimer in solution (Figure 5.1). This suggests that YhdE could bind up to two substrates at any one time. In optimal conditions, dTTP binds at two sites of the YhdE dimer. The structural studies showed that the dimeric mode of YhdE make two active sites close to each other, therefore making cooperativity possible.

Substrate recognition is an important concept to consider for mechanistic studies of YhdE. Comparison of the structures of the two substrates, UTP and dTTP, suggests that YhdE is able to display higher specificity for dTTP based on either the methyl group of dTTP or the lack of the 2’ hydroxide group on dTTP (Table 5.1). Furthermore, YhdE does not display significant activity for dUTP; therefore, this suggests that the methyl group on dTTP is a distinguishing feature for substrate recognition. It is possible that YhdE does not need to have any apparent enzymatic
activity for dUTP because there are highly specific pyrophosphatases for dUTP. The charged surface analysis revealed that the methyl group sits in a small hydrophobic cleft, in which Val28 forming hydrophobic interact with the methyl group (Figure 5.8D, yellow circle). This feature makes dTTP a more attractive substrate than UTP.

When tested for pyrophosphatase activity with a variety of nucleotides, the YhdE E33A mutant did not show any activity. In the Maf-like protein structure, the corresponding E45 residue is shown to make contact with the metal ion, and earlier results demonstrated that the presence of metal ions is essential for YhdE pyrophosphatase activity. The charged surface analysis showed that E33 is located in a very negatively charge region, and therefore it is very possible that this is the region which adopts the metal ion; in contrast, the dTTP binding site is positively charged. The loss of YhdE pyrophosphatase activity indicates that E33 is important in coordinating the metal ion.

Here, we have proposed that YhdE is a novel pyrophosphatase protein specific for dTTP. YhdE displays some structural similarity to several housecleaning nucleotide hydrolase enzymes, which are responsible for “cleans[ing] the cell of potentially deleterious endogenous metabolites and to modulate the accumulation of intermediates in biochemical pathways” If YhdE was in fact part of the housecleaning family of proteins, one would expect that YhdE deficiency would result in significant observable changes in the cell. Contrary to this, knockout studies of the maf gene suggest that Maf and YhdE are not essential for cell viability; instead, it is overexpression of YhdE that results in a detrimental phenotype. The high specificity that YhdE has for dTTP, a nucleotide important for DNA replication, and the fact that excess amounts of YhdE may hydrolyze dTTP in the cell pool (resulting in inhibition of DNA replication) suggests that YhdE could play a role in regulating DNA replication by controlling the intracellular dTTP nucleotide
pool. The pyrophosphatase activity of YhdE suggests that YhdE is important for the metabolism of dTTP. Metabolic degradation of dTTP has been discussed in other studies\textsuperscript{237-240}, but this is the first report of a protein that specifically hydrolyzes dTTP to dTMP in bacteria.

Thymine has been shown to play a critical role in the viability of cells. Prokaryote and eukaryote cells that are starved of thymine are not viable, and cell death from thymine deficiencies are known as thymineless death (TLD). The specific pathways resulting in TLD are unknown, but Morganroth and Hanawalt\textsuperscript{241} have eliminated unbalanced growth, the SOS response and nucleotide excision repair as causal factors of TLD, and suggest that active DNA synthesis is not required for TLD. Interestingly, this phenomenon is not observed when cells are starved of other growth factors, highlighting the importance of thymine for cell functionality\textsuperscript{242}. Thymine deficiencies also result in changes to the pools of other nucleotides. Similar to the overexpression of Maf, thymine starvation of \textit{E. coli} cells results in a filamentous phenotype that is observed as part of TLD\textsuperscript{242}. The similarity in phenotype between these two processes suggests there is an overlap or coordination of mechanisms that results in the inhibition of separation and elongation of cells. TLD demonstrates the importance of nucleotide regulation for synchronizing cellular processes, including DNA synthesis and cell division that are necessary for cellular viability. It is still not clear how thymine deficiency results in TLD, but as it results in a similar phenotype to that observed with dTTP deficiency, it may be that reduced amounts of thymine leads to reduced amount of dTTP, and thus a reduction in DNA replication. The loss of colony-forming ability that accompanies the filamentous phenotype of overexpressed Maf in \textit{B. subtilis} cells provides evidence that dTTP regulation not only affects DNA replication, but also cell growth and morphology. Further studies into the effects of knockout and overexpressed \textit{yhdE} are expected to provide additional insights into the mechanism by which YhdE functions to regulate
dTTP and other intracellular nucleotide pools. Such studies could also provide insight into the observed correlation between dTTP pools and cell growth and morphology changes.

To further understand how yhdE expression is synchronized with DNA replication and cell growth, localization studies of yhdE during the cell cycle should be attempted. As suggested by Mathews and Sinha \textsuperscript{243}, dTTP levels are higher at replication sites, and so it would be interesting to see if nucleotide regulators are also concentrated at such sites. As discussed earlier, yhdE is located in the mre operon in \textit{E. coli}, which contains other genes involved in achieving the rod-shape of \textit{E. coli} cells. Notably, the actin homologue in bacterial cells, \textit{mreB}, is also contained in this operon. MreB has many of the characteristics of actin including inherent polarity and involvement in the localization of other cellular molecules. In particular, MreB is a prime candidate for facilitating the movement and localization of YhdE throughout the cell. Moreover, it would be of interest to see if YhdE has any potential cellular binding partners. It is clear that the \textit{mre} operon, consisting of \textit{mreB, mreC, mreD, yhde, rng} genes, is involved in regulating DNA replication, chromosomal segregation, cell division, and the overall synchronization of cell morphology, but the coordination of these mechanisms and their implications for other cellular processes including TLD remain unknown. The identification of YhdE as a dTTP pyrophosphatase protein is an important step towards unraveling the functions associated with the \textit{mre} operon in \textit{E. coli}.

\textbf{5.5 Experimental Procedures}

\textbf{5.5.1 Materials}

DNA restriction and modification enzymes were obtained from Fermentas (Burlington, Canada), Promega (Nepean, Canada), and Roche (Laval, Canada). PCR reagents were obtained
from Invitrogen (Burlington, Canada). ATP, CTP, GTP, UTP, TTP, TPP, TMP and other fine chemicals were purchased from Sigma (Oakville, Canada). dCTP, dTTP, and dUTP, and reagents for *E. coli* culturing were purchased from Bioshop (Canada). Ni\(^{2+}\)-nitrilotriacetic acid agarose affinity resin, the QIAprep plasmid extraction kit, and the QIAquick gel extraction kit were purchased from Qiagen (Mississauga, Canada). A pyrophosphate detection kit was purchased from Invitrogen (Burlington, Ontario).

5.5.2 Cloning, expression and purification of wild type YhdE and the YhdE E33A mutant

The cloning of YhdE and mutagenesis (E33A mutant) were carried out using standard procedures. Recombinant YhdE and the E33A mutant were expressed in BL21(DE3) *E. coli* cells. Expression and purification of YhdE were completed as previously reported for YjjX \(^75\).

5.5.3 Preliminary investigation of YhdE activity

The standard solution consisted of 6.25 \(\mu\)M YhdE, 5 mM Mg\(^{2+}\), 20 mM Bis-Tris buffer pH 6.75 and 100 \(\mu\)M substrate in a 30\(\mu\)L volume incubated for 10 minutes at 37 °C. The reactions were terminated by the addition of 70 \(\mu\)l of a color developing reagent, which consisted of six parts 0.42% ammonium molybdate in sulfuric acid with one part 10% (w/v) ascorbic acid. The samples were incubated for 20 minutes at 45 °C. Any inorganic phosphate liberated by YhdE was detected by absorbance at 660 nm. The absorbance was determined with a Powerwave XS Microplate spectrophotometer (Bio-Tek Instruments, Inc.).

5.5.4 Determination of YhdE products by mass spectrometry
6.25 µM YhdE was added to 100 µM TTP in 2 mM Mn\textsuperscript{2+}, 20 mM HEPES pH 7.0 in a 100 µL volume. The mass spectrometry analysis was assisted by the Mass Spectrometry Laboratory, Department of Chemistry, Queen’s University.

5.5.5 Optimal conditions for the pyrophosphatase activity of YhdE

The standard reaction for optimization of YhdE pyrophosphatase activity utilized 6.25 µM YhdE, 100 µM TTP, 20 mM buffer, and 2 mM metal ion (unless otherwise indicated). The reactions were allowed to proceed for 10 minutes at 37 °C for all optimizations except for the temperature optimization. The effects of a variety of metal ions on YhdE activity were probed by the addition of YhdE to the substrate and 5 mM metal ion in Bis-Tris buffer pH 6.75. YhdE TTP pyrophosphatase activity in 2 mM Mn\textsuperscript{2+} and 20 mM Bis-Tris buffer pH 6.75 was also analyzed at the following temperatures: 4, 16, 25, 37 and 45 °C. YhdE activity was investigated in a variety of buffers within a pH range of 6.0-11.0 by combining YhdE with substrate and 2 mM Mn\textsuperscript{2+}. The buffers used were Bis-Tris buffer for pH 6.0-6.75, HEPES buffer for pH 7.0-7.6, Tris-Cl buffer for pH 8.0-9.5 and CAPS buffer for pH 10.0-11.0. Inorganic pyrophosphate was detected by using the P\textsubscript{i}Per Pyrophosphatate Assay Kit from Molecular Probes, Invitrogen. The absorbance at 565 nm was determined by using a Powerwave XS Microplate spectrophotometer (Bio-Tek Instruments, Inc.), and a control lacking enzyme was used to correct for background readings.

5.5.6 Substrate specificity of YhdE (wild type and mutant)

The substrate specificity of YhdE was analyzed by comparing YhdE enzymatic activity in the presence of a variety of concentrations of different nucleotides. The standard reaction to test
for YhdE (d)NTP pyrophosphatase activity contained: 1.5 µM YhdE, 100 or 500 µM substrate, 2 mM Mn²⁺ and 20 mM HEPES buffer pH 7.0, all in a 50 µl volume. The samples were incubated at 25 °C for ten minutes. These assays were repeated for the analysis of the YhdE E33A mutant. All reactions were analyzed by the P_iPer Pyrophosphatate Assay Kit from Molecular Probes, Invitrogen. The absorbance at 565 nm was determined by using a Powerwave XS Microplate spectrophotometer (Bio-Tek Instruments, Inc.), and a control lacking enzyme was used to correct for background readings.

**5.5.7 Enzymatic kinetic analysis of YhdE pyrophosphatase activity**

The standard reaction for the kinetic analysis of YhdE dTTP and UTP pyrophosphatase activity contained: 1.5 µM YhdE, 0-300 µM substrate, 2 mM Mn²⁺ and 20 mM HEPES buffer pH 7.0 in a 50 µl volume. The samples were incubated at 25 °C for ten minutes. One unit of enzyme is defined as the amount of the enzyme required for the hydrolysis of 1 µmol of substrate per minute at 25 °C. All reactions were analyzed by the P_iPer Pyrophosphatate Assay Kit from Molecular Probes, Invitrogen. The absorbance at 565 nm was determined by using a Powerwave XS Microplate spectrophotometer (Bio-Tek Instruments, Inc.), and a control lacking enzyme was used to correct for background readings.

**5.5.8 Crystallization**

The preliminary crystallization conditions for the YhdE E33A mutant were screened by the sparse matrix method (Jancarik et al. 1991) using standard screening kits. The protein concentration was diluted to 10 mg/mL in the same buffer used for size-exclusion purification. The hanging drop vapor diffusion method was used. Hanging drops were set up with 2 µL of
protein solution mixed with 2 µL of well solution. The primary optimal crystallization condition in the reservoir was 0.1 M magnesium sulfate, 0.1 M MES buffer (pH 6.5), and with 10-15% PEG 8000 as the precipitating agent at room temperature. A hexagonal crystal form appeared in six days and grew to full size within two weeks. For the YhdEE33A/dTTP complex crystallization, dTTP was added to a final concentration of 1 mM of YhdE E33A protein solution. The crystals were obtained in the same growth condition, though a different crystal form developed.

5.5.9 Structure determination

The diffraction data sets for the YhdE E33A crystal and the YhdE E33A/dTTP complex crystal were collected using in-house facilities, including a Rigaku rotating anode generator, and diffraction intensities were recorded on MarResearch Mar345 imaging plate. All data were processed with HKL2000. Molecular replacement solutions were found with the aid of the Phaser program using the structure of Maf protein as a search template. The remainder of the model was built and all side chains were corrected manually using the program Xtalview. All subsequent refinements were carried out with REFMAC5 within the CCP4 suite of programs.

5.5.10 Electron microscopic analysis

Bacterial cultures (1 L) grown at 37 °C in Terrific Broth–ampicillin were grown to an OD$_{600}$ of 0.4 and induced with 100 µM isopropyl-β-thiogalactoside (IPTG) (Bioshop). Cultures were then grown for 3 hours, and 1.5 mL of uninduced cultures or 2 × 1.5 mL of induced cultures was harvested by centrifugation on a tabletop centrifuge (5000g, 1 minutes) in Eppendorf tubes. Pellets were resuspended in 1.5 mL of 50 mM NaH$_2$PO$_4$ (pH 7.0), from which
50 µL volumes were mounted onto copper grids. After 1 minute of incubation, the copper grids were subsequently dried with filter paper and treated with 50 µL of 1% (w/v) sodium tungstophosphoric acid (pH 7.4) for 1 minute and dried again with filter paper. Copper grid-mounted bacteria were then examined on a Hitachi 7000 EM operated at 75 kV.

5.5.11 Cell growth

*E. coli* K-12 W3110, the *yhdE*-knockout strain, and the plasmid of the *yhdE* full-length construct were kind gifts from the Genome Analysis Project in Japan ([http://ecoli.aist-nara.ac.jp](http://ecoli.aist-nara.ac.jp)). Overnight Luria–Bertani (LB) broth cultures of three strains were diluted to the same optical density at 600 nm (OD$_{600}$) of 0.10 and then left to grow for two and a half hours. The cultures were induced with 100 µM IPTG. The cultures were then incubated and shaken at 37 °C for 12 hours; OD$_{600}$ measurements of the three cultures were measured every 30 minutes.

ACKNOWLEDGMENTS

The authors would like to thank Drs. M. Cygler and A. Matte for their assistance in cloning. This work was supported by the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada to ZJ. ZJ is a Canada Research Chair in Structural Biology.
Chapter 6

General discussion and summary
6.1 From structure to function by superfamily structural comparison

Sequence similarity comparison is a traditional way of categorizing or classifying related proteins to identify and analyze their functions. However, amino acid degeneracy complicates the detection of proteins that are structurally or functionally related. A protein superfamily is defined as a group of proteins that have evolved from a common ancestor protein, but have diverged over time in sequence, structure and function. Protein families usually display clear sequence similarity and highly similar structures, whereas superfamilies group protein families that share distant common ancestry with other protein families. Proteins in a superfamily usually maintain similar overall structure due to evolutionary pressure to retain their functions, and proteins with similar structures are usually related and often have similar function, owing to their shared genetic heritage. Therefore protein structure is more conserved than sequence and is thus much more useful for determining protein function at the superfamily level, where sequence similarity has degraded. There are, however, several cases where proteins that have similar structures have completely unrelated functions, and there are also cases of proteins with different structures having similar functions. In those latter cases, related functional subdomains or motifs are generally found to have been conserved. As mentioned in Chapter 1, for example, numerous kinases all contain a similar ATP binding domain for the purpose of phospho-transfer despite having substantially different overall structures. These proteins warrant a detailed analysis of subtle changes in their sub-structures.

Protein function typically relies on a constellation of a few amino acid residues that are embedded in a distinct globular domain. The rest of the globular domain functions as a structural scaffold supporting the functional centre and also contributes to interactions with the substrates, co-factors and other proteins. The key functional residues are often regularly distributed in
sequence to form a signature sequence motif. Generally, even though the overall protein sequence will vary over the course of evolution, the catalytic or cofactor-binding residues and the local sequence motif remain conserved. Variations tend to occur in the substrate recognition sites. This principle had been extensively exploited by natural selection to generate the diversity of enzymes found in the living world from a relatively small number of ancestral enzymes. Identification of small, signature functional motifs is therefore useful as a clue for the functional annotation of a protein. Bioinformatics methods have been developed, for example, to computationally identify modular signaling domains within protein sequences with a high degree of accuracy. In contrast, little success has been achieved in predicting short linear sequence motifs or key functional residues within proteins.

The Structural Classification of Proteins (SCOP) database collects the protein structures from the PDB and classifies protein hierarchy based on a tiered system, with class, fold, superfamily, and family hierarchical levels. Utilizing the SCOP database of protein superfamilies, and protein-substrate complexes found in the PDB that elucidate conserved structural features, one may define specific signature motifs. The combination of genomic sequencing with structural genomics has provided a wealth of new individual protein structures and/or complex structure of protein with its binding partners or substrates. This accumulated structural information has the potential to unlock the sequence motifs responsible for specific protein functions, shedding light on the structure-function relationship and providing further insight into catalytic mechanisms.

6.2 Structure-based functional annotation of four proteins in practice

While the preceding section outlined the theme underlying homology-based functional annotation, in many cases such procedures provide only rudimentary models at best of protein function.
functionality. This is due to the complex nature of protein function, the difficulties in determining function precisely by systematic experiments, and, at a more fundamental level, uncertainties in the definition of protein function itself. “Function” can be described at many levels, and thus proteins are often annotated with varying degrees of specificity. This can range from the classification of the protein into a domain superfamily to the definition of a specific and measurable enzymatic activity. At the same time, despite highly conserved macromolecular structures, many proteins in structural families often exhibit a variety of biological functions. As functional annotation places a protein into a structural class for which a likely cellular role of the protein can be clearly predicted, it is imperative that annotation be as specific as possible. It should be based not only on overall structural homologies, but also should consider any and all other sources of information as well, such as operon prediction, loose functional annotation of additional operon members, local structural analysis focused on the conservation of catalytic residues, as well as co-crystallization trials and virtual ligand screening. In the present case, a combination of information from a variety of experimental and theoretical perspectives has permitted the functional annotation of four proteins described herein.

By comparing protein structures with members of their superfamilies, we investigated four proteins (YihE, Acek, YjjX and YhdE), which belong to two different protein superfamilies: eukaryotic protein kinases and Ham1/Maf pyrophosphatases. In the case of YjjX, the structure has a clear nucleotide binding fold. Moreover, a local or regional comparison between YjjX and members of its superfamily exhibited remarkable similarity in the putative ligand-binding pocket, with most residues being either conserved, or replaced by functionally equivalent amino acids among YjjX, Maf, and Mj0226 (Figure 6.2A). Two sulfates were bound in this region of YjjX, both positioned in similar conformations to bound phosphates identified in the Maf structure. As
previous studies of the Mj0226 protein suggested that it played a role in the conversion of xanthine or inosine triphosphates to the corresponding monophosphates, the conservation of key structural features in the putative catalytic site of YjjX strongly suggested that this protein retained NTP hydrolase activity. However, granting that YjjX is a novel NTP hydrolase, its substrate specificity still remained unknown. We therefore performed nucleotide library screening to identify its binding partner(s). Because Mj0226 is a house-keeping enzyme, we suspected that YjjX might also possess house-keeping functions. Thus we screened not only the canonical nucleotides, such as ATP, GTP, but also noncanonical ones as well, which ultimately resulted in a positive substrate match. In this case, the high similarity in both their overall folds and the conserved active site between YjjX and Mj0226 provided strong evidence for YjjX functional annotated as a house-keeping nucleotide phosphatase. Further biochemical analysis permitted us to identify YjjX as a novel ITPase/XTPase that functions as a house-keeping enzyme to protect the E. coli from the formation of noncanonical nucleotides during oxidative stress.

It is often the case that the initial structural analysis only indicates the functional family for a given protein, and not its ligand or substrate. The identification of this latter information can be impeded by the diversity of functions within a particular structural classification of proteins family, by a highly selective and specific ligand-binding site, or by the identification of a novel protein fold. For example, although the structure of Maf from B. subtilis was determined ten years ago, and was subsequently classified into a nucleotide binding protein family, its substrate specificity still remained unknown. In other cases, a protein can exhibit poor overall secondary structural alignment even though a specific subdomain or the geometry of specific catalytic residues is superimposable upon a functionally homologous protein. It is therefore extremely
important that structural details of new protein families be scrutinized for the smallest profiles, which may provide functional clues. To date, much of this work still requires substantial manual inspection and is thus labour intensive. In the case of AceK, for example, when we used the full structure as the template for structural homology search, no matches were found. However when we searched for homologues of the kinase domain on its own, more than 100 matches to members of the receptor Tyr kinase superfamily were found, with an r.m.s.d. of 3.5~4.0 Å. Moreover, when searching for homologues for just the ATP binding region on its own, the resulting hits had an r.s.m.d. of just 1.5 Å, indicating that AceK shared a conserved phosphor-transferring mechanism with the receptor Tyr kinase superfamily. Furthermore, this comparison of ATP binding uncovered other residues that contributed to AceK ATP binding, which proved instrumental for explaining AceK activity. Overall, it is clear that improvements in methods for rapidly identifying and comparing locally conserved structural features (i.e., structure-based local phylogenetic relationships) have vastly improved and will continue to improve the functional annotation process.

6.3 The role of sequential and structural alignment in identifying a nucleotide binding motif and determining substrate specificity in the Ham1/Maf superfamily

Full sequence alignments between Ham1 ITPase, Maf protein and YjjX-like protein exhibit only 19% identity. However, the overall structures of these three protein families show high levels of structural similarity and are thus classified in the Maf/HAM1 superfamily. Most interestingly, although Mj0226, Maf and YjjX are all nucleotide binding proteins, their substrates are distinct. Thus we were interested in identifying structural features responsible for their shared nucleotide binding abilities, such as the P-loop that is found in many ATPases and GTPases. In this case, the P-loop is not found in the Maf/Ham1 superfamily. Instead, a different loop is found
between a β-strand and an α-helix in the putative active site of these three proteins, one with a set of conserved residues that are similarly positioned in space. In Maf, the corresponding loop’s sequence is \textbf{LILASQSPRRKELL}; in Mj0226 the sequence is \textbf{IYFATGNPNKIEA}; and in YjjX the loop is \textbf{QVVCATTNPARIQAI}. Local sequence alignment of this loop in all Maf/Ham1 superfamily member proteins clearly reveals a sequence motif in this loop: 
\[(L/V/I)X(L/V/I)A(T/S)X(N/S)XX(K/R)XXX(L/V/I)] \] (Figure 6.1). Structure alignment of the loops in YjjX, Maf, Mj0226 and human ITPase also shows high conformational similarity and specific positioning of the conserved residues (Figure 6.1). Notably, the Maf crystal structure, grown in the presence of dUTP, clearly shows that S9 and R14 on this loop interact with the UTP β- and γ-phosphates.

Sequence and structural alignments indicate that the putative nucleotide binding loop plays an important functional role in Maf/Ham1 superfamily proteins. We investigated whether this conserved motif is present in other proteins or is unique to the Maf/Ham1 superfamily. Results from a BLAST search for homologues of the YjjX loop sequence found that GTP Cyclohydrolase II also contains a similar loop sequence. GTP cyclohydrolase II converts GTP to 2,5-diamino-6-β-ribosyl-4(3H)-pyrimidinone 5'-phosphate, formate and pyrophosphate, which is the first step in riboflavin biosynthesis. Judging from the structure of GTP cyclohydrolase II, the corresponding motif of GTP cyclohydrolase II forms a similar loop that also resides between an α-helix and a β-sheet and contains a threonine and a lysine in similar locations to those in the active site of YjjX (Figure 6.2, B right). This occurs even though GTP cyclohydrolase II is twice the size of YjjX and has very little overall similarity to YjjX. Interestingly, the loop motif occurs at the C-terminus of GTP cyclohydrolase II, whereas it is at the N-terminus in Maf/Ham1. The complex structure of GTP cyclohydrolase II with GTP clearly shows that the threonine and
Figure 6.1 Sequence alignment of Maf/Ham1 proteins in the putative nucleotide binding loop clearly shows a conserved motif pattern of (L/V/I)XLA(T/S)XX(K/R)XXX(L/V/I). Structure alignment of the loops in YjjX (green), Maf protein (cyan), Mj0226 (magenta), and human ITPase (yellow) is shown on the right.
lysine residues on the loop interact with \(\beta\)- and \(\gamma\)-phosphates of GTP. This observation demonstrates that this novel nucleotide binding motif is not unique to the Maf/Ham1 superfamily.

Both the identification of a novel nucleotide binding loop in the Maf/Ham1 superfamily, and the determination of the Maf/dUTP complex structure, are illuminating. In the Maf/dUTP complex structure, S9 and R14 on the loop interact with the \(\gamma\)- and \(\beta\)-phosphates of dUTP (Figure 6.2 A, middle), thereby stabilizing and positioning the dUTP in the correct orientation for \(\alpha\)-\(\gamma\) phosphate bond cleavage by D70. The structure of Mj0226 complexed with ATP has also been determined (Hwang et al., 1999). However, it does not show ATP interacting with the nucleotide binding loop as Maf/dUTP does (Figure 6.2 A, left). This discrepancy can be explained by the fact that ATP is not the true substrate of Mj0226, and therefore it may not be able to accurately interact with the active site for hydrolysis. Indeed, the interaction between Mj0226 and ATP may indicate that ATP is a competitive inhibitor. Thus we postulate that this novel nucleotide binding loop is an essential feature of pyrophosphatases. The evidence that GTP cyclohydrolase II also contains this loop at its C-terminus, is a pyrophosphatase, and forms a complex structure with GTP that is similar to the Maf/dUTP complex (Figure 6.2A, right), further supports this hypothesis.

Because the structures of ITP and ATP are very similar, it raises the question as to how Mj0226 and YjjX recognize ITP in a cellular nucleotide pool that contains ATP at concentrations that are three orders of magnitude higher than ITP concentrations. The only difference between these nucleotides is that the amino group on the ATP base ring is replaced by oxygen on the ITP base ring. Since no structure of either Mj0226 or YjjX in complex with ITP presently exists, we performed computational docking of ITP into the active sites of Mj0226 and YjjX. Results show
strikingly similar conformations of ITP in the active sites of both the Mj0226 and YjjX: the ITP phosphates interacted with the nucleotide binding loop and positioned $\alpha$-$\beta$ phosphate bond next to D73 of Mj0226 and E79 of YjjX (Figure 6.2 B). We were surprised to find that the ITP base ring interacted with H177 and R178 in Mj0226, and with R153 in YjjX, which are all highly conserved in the two protein families. Upon reflection, these interactions may account for the ITP specificity of these two proteins: these residues form positively charged regions that can interact with the oxygen atom on the ITP base ring, but which will repel the amino group on the ATP base ring. We therefore postulate that H177 and R178 in Mj0226, as well as R153 in YjjX, play key roles in differentiating ITP from ATP. To investigate this, we plan to mutate these residues and expect to see exchanges in substrate specificity of the mutants.

6.4 The importance of the protein-substrate complex structure for identifying the kinase substrate-reorganization specificity

Although most eukaryotic protein kinases share a common structural fold, a similar ATP binding cleft, and a similar substrate binding mode that is located in the cleft between the two lobes of the kinase structure, they differ in terms of the charge and hydrophobicity of surface residues in their regulatory loops. These differences affect kinase specificity. Thus far, the identification of kinase specificity is still a substantial challenge, largely because multiple mechanisms that contribute to this exquisite specificity have evolved, including the structure of the catalytic site, the local and distal interactions between the kinase and substrate, the formation of complexes with scaffolding and adaptor proteins that spatially regulate the kinase, systems-level competition between substrates, and error correction mechanisms. In most cases, the
Figure 6.2 A) Mj0226/ATP complex (left), Maf/dUTP complex (middle), and GTP cyclohydrolase II/GTP complex (right). B) Computational docking of ITP into the active site of Mj0226 (left) and YjjX (right).
consensus phosphorylation site and Ser/Thr or Tyr specificity is determined by the structure of the catalytic cleft of the kinase and by local interactions between the kinase cleft and the substrate phosphorylation site. Distal binding interactions between the kinase and the substrate often provide additional binding interactions and specificity, and sometimes provide allosteric regulation and localization to specific cellular compartments or structures. Although not all kinases make use of all of these mechanisms, many protein kinases use at least some of them to achieve their specificity. The most direct way to investigate the kinase substrate specificity is to determine the complex structure of the kinase with its substrate. In this study, the complex structure of AceK with ICDH provides a clear example for studying the mechanism of kinase specificity.

As the first phosphorylation system found in *E. coli*, ICDH has several unique features that make AceK an interesting target for studying kinase specificity. Firstly, AceK controls the activity of ICDH in *E. coli* by a cyclic phosphorylation mechanism where it is responsible for both ICDH phosphorylation and dephosphorylation. Thus, AceK must be able to recognize both species at the same active site. Secondly, the crystal structure shows that the phosphorylation site in ICDH, S113, is positioned in the concave active site and barely accessible to the kinase/phosphatase, raising the question of how AceK is able to approach this site to perform its function. Thirdly, the substrate specificity of AceK does not depend solely on the sequence motif around the phosphorylation site of ICDH, but instead on the overall three-dimensional structural interaction between AceK and ICDH. All these features suggest that conformational changes occur in both AceK and ICDH that may be triggered upon interaction. However, any such conformational changes are not evident in the apo-AceK structure. Thus, it appears necessary to determine the complex structure of AceK and ICDH, ideally in both phosphorylated and
dephosphorylated form, to clarify the nature of the protein-protein interactions between these proteins, as well as to answer the questions of how AceK is able to access the buried S113 residue of ICDH for (de)phosphorylation, and how the conformations change upon complex formation. In most cases, enzymes and substrates form weak, transient interactions to facilitate a reaction, the completion of which typically results in enzyme-substrate dissociation that allows the enzyme to engage in another reaction. However, due to the highly unusual bifunctional abilities of AceK, both ICDH and phosphorylated-ICDH are the substrate and product for AceK. It has been postulated that there must be an extensive interface between AceK and ICDH that allows these proteins to form a stable complex in solution. However, all previous attempts to co-crystallize AceK and ICDH have been unsuccessful. We postulated that AceK would have a stronger association with ICDH if we could limit the dissociation reaction from happening. To test this, we mixed AceK and ICDH with 1 mM ATP, but in the absence of an important cofactor for the reaction, Mg$^{2+}$, to prevent the phosphorylation reaction. After incubating overnight at 4 ºC, the mixture was purified by size-exclusion chromatography. Analysis by chromatogram and SDS-PAGE gel clearly showed that AceK and ICDH formed a stable complex in the presence of 1 mM ATP in solution. Interestingly, we found that the protein complex is more stable when compared to the native AceK protein, which gradually precipitates out of solution after a week of storage at 4 ºC. From an analysis of the complex structure, we find that the AceK SRL, known to be very flexible in AceK monomers, is stabilized by the ICDH dimer. For this reason, we found that the AceK/ICDH complex was more readily crystallizable than apo-AceK. From the complex structure, we identified that it is this AceK SRL that interacts with ICDH at the active cleft, reorganizing it so that the (de)phosphorylation site becomes exposed. Thus, AceK specificity appears to hinge on the distal binding interactions between the kinase and the substrate.
6.5 Conclusions

*E. coli* has evolved intricate systems that enable it to protect itself from death in harsh environments, such as high temperature, limited nutrients or oxidative stress. In this work, four *E. coli* proteins, a protein kinase and a protein kinase/phosphatase (YihE, AceK) and two pyrophosphatase (YjjX, YhdE), which are involved in three different response systems (Cpx response system, acetate switch and housekeeping system), were studied for structural-based functional characterization. *yhdE*, a Cpx responsive gene, encodes a novel Ser/Thr protein kinase, which is involved in the regulation of bacterial adhesion in association with the Cpx stress response. YihE participates in a new type of bacterial phosphorylation mechanism that combines His/Asp phosphorylation in sensing stress (TCS) and the transcriptional upregulation of a Ser/Thr protein kinase to maintain cellular function. AceK, the first Ser/Thr protein kinase/phosphatase found in *E. coli*, has a highly unusual ability to regulate the activation of ICDH by reversible phosphorylation, and thus it is able to control the flow of carbon into either the glyoxylate bypass or the full Krebs’ cycle. Structure determinations of both apo-AceK and the AceK/ICDH complex have greatly facilitated our efforts to explain the regulatory mechanism and substrate reorganization of the AceK/ICDH complex interaction. Interestingly, both YihE and AceK show eukaryotic protein kinase folds with very low sequence identity. The structural study of these two kinases demonstrates that prokaryotic kinases have conserved both their fold and catalytic mechanism(s) in the presence of tremendous sequence variation to allow for substrate binding and activity. Since the initial discovery of ICDH phosphorylation, the field of prokaryotic protein phosphorylation has grown both in size and in stature. Large numbers of phosphor-proteins have been detected but only a few have been identified. Our study contributes to the understanding of phosphorylation mechanisms of bacterial metabolic and regulatory
processes. YjjX and YhdE were function unknown proteins prior to our structural characterization. Via structure determination, both YjjX and YhdE were found to be novel nucleotide hydrolyses. YjjX is a novel ITPase/XTPase. ITP and XTP are produced by deamination of ATP and GTP in response to exposure to chemical mutagens, and they reduce both the rate and efficiency of transcription. As a consequence, mechanisms are required in nucleic acid synthesis to safeguard against undesired incorporation of ITP and XTP. The role of YjjX, a housekeeping enzyme which specifically cleans up ITP/XTP, might therefore be to prevent or minimize the incorporation of undesired nucleotides ITP and XTP into RNA. YhdE was determined to be a novel dTTPase/UTPase, which is involved in septum formation and cell division. Both studies provide clear examples of the power of structural biology as a tool for determining the function of proteins.

Future work will involve:

a) Characterization of YihE kinase substrates and its implication in the Cpx stress response pathway in collaboration with Dr. N. Martin in the Department of Microbiology and Immunology,

b) Complete structure refinement of the structure in its opened conformation form.

c) Investigation of binding of allosteric regulators and establishment of the mechanism of the molecular switch between the kinase and phosphatase activities of AceK.

d) Determination of the complex structure of AceK with phosphorylated ICDH to explain the AceK phosphatase mechanism.

e) Structure-based engineering and directed evolution of the AceK kinase form.

f) determination of the mechanism of these house-cleaning enzymes to target specifically
their natural substrates from a nucleotide pool containing a 1000-fold excess of canonical nucleotides such as ATP and GTP, which are very close structural analogues of ITP and XTP.

g) Characterization of YhdE In vivo function in regulating the cell division as dTTPase.
Reference List


27 G. Sapkota, et al., "Dephosphorylation of the linker regions of Smad1 and Smad2/3 by small C-terminal domain phosphatases has distinct outcomes for bone morphogenetic protein and transforming growth factor-beta pathways," J. Biol. Chem. 281(52), 40412 (2006).


Z. Li, S. Pandit, and M. P. Deutscher, "RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA," EMBO J 18(10), 2878 (1999).


P. N. Danese and T. J. Silhavy, "The sigma(E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in Escherichia coli,"


B. Myrnes, P. H. Guddal, and H. Krokan, "Metabolism of dITP in HeLa cell extracts, incorporation into DNA by isolated nuclei and release of hypoxanthine from DNA by a hypoxanthine-DNA glycosylase activity," Nucleic Acids Res 10(12), 3693 (1982).


