

Structural Studies on HPrK/P, an Enzyme Involved in the Control  
of Carbon Catabolite Repression in Gram positive Bacteria

by

Gregory Scott Allen

A DISSERTATION

Presented to the Department of Biochemistry & Molecular Biology  
and the Oregon Health & Science University

School of Medicine

in partial fulfillment of  
the requirements for the degree of  
Doctor of Philosophy

July 2<sup>nd</sup>, 2002

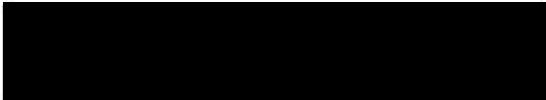
School of Medicine  
Oregon Health & Science University


---

Certificate of Approval

---

This is to certify that the Ph.D. thesis of  
Gregory Scott Allen  
has been approved

  
\_\_\_\_\_  
Professor in charge of thesis

  
\_\_\_\_\_  
Member

  
\_\_\_\_\_  
Member

  
\_\_\_\_\_  
Member

  
\_\_\_\_\_  
Member

## Table of Contents

Tables & Figures.....	iii
List of Abbreviations.....	v
Acknowledgments.....	vi
Abstract.....	1
Chapter I.	
Introduction.....	3
Carbon Catabolite Repression.....	6
Mechanisms of Ser/Thr Kinases & Phosphatases.....	31
Chapter II.	
Protein Crystallography Methods.....	61
Chapter III.	
Crystallization, preliminary X-ray analyses, and biophysical characterization of HPr kinase/phosphatase of <i>Mycoplasma     pneumoniae</i> .....	116
Chapter IV.	
Crystal structure of HPr kinase/phosphatase from <i>Mycoplasma pneumoniae</i> .....	134
Chapter V.	
Conclusion.....	182

## Appendix

Crystallization and preliminary crystallographic analysis of a complex of CcpA, P~Serine46-HPr, and Cre DNA from <i>Bacillus megaterium</i> .....	188
Bibliography.....	204



## Tables & Figures

- 1-1 Comparison of the major mechanisms regulating carbohydrate uptake and metabolism in *E. coli* and *B. subtilis*
- 1-2 CcpA sequences from various gram-positive bacteria aligned with LacI/GalR family members from *E. coli*
- 1-3 Schematic view of the PTS from *E. coli*
- 1-4 Alignment of HPr sequences from organisms that encode HPrK/P with Crh from *B. subtilis* and HPr & NPr from *E. coli*
- 1-5 Schematic depiction of the four major types of transporters found in living organisms
- 1-6 *B. subtilis* non-PTS proteins phosphorylated and regulated by PTS proteins
- 1-7 The major mechanism of CCR in *B. subtilis*
- 1-8 Model of CcpA from *B. megaterium*
- 1-9 Alignment of HPrK/P from various organisms
- 1-10 Proposed mechanisms regulating expression of the *B. subtilis glpFK* operon at the DNA and RNA level
- 1-11 Proposed catalytic mechanism for PKA
- 1-12 Proposed VHR catalytic mechanism
- 1-13 Catalytic mechanism of the PPP family member, PP2B
  
- 2-1 The seven crystal systems
- 2-2 The 14 Bravais lattices
- 2-3 A 3 fold screw axis
- 2-4 Vector diagram of parallel scattering from two dipoles
- 2-5 Vector geometry of scattering
- 2-6 The carbon scattering factor
- 2-7 Bragg's Law
- 2-8 Real- Reciprocal space lattice comparison for orthogonal unit cells
- 2-9 Real- Reciprocal space lattice comparison for triclinic unit cells
- 2-10 The Ewald sphere
- 2-11 Harmonic analysis of a step function
- 2-12 A simple Patterson map
- 2-13 The atomic scattering factor in the complex plane
- 2-14 Structure factors in the complex plane (Argand diagram)
- 2-15 Bijvoet pairs mirrored about the real axis

- 2-16 The anomalous scattering factor of Bromouridine
- 2-17 Harker construction for a SIRAS experiment
- 2-18 A graph of a typical  $P(\alpha_p)$  distribution for isomorphous data
- 2-19 The total probability for a reflection in the SIRAS method
- 2-20 Two different probability curves for a SIRAS experiment

Table 3-1 Data collection statistics for HPrK/P from *M. pneumoniae*

- 3-1 Enzyme activity of HPrK/P following DTNB-labelling
- 3-2 Crystal of HPrK/P of *Mycoplasma pneumoniae*

Table 4-1 Selected Crystallographic Data

- 4-1 The structure of *M. pneumoniae* HPrK/P
- 4-2 View of the NTD of the F subunit
- 4-3 View of the CTD from the E subunit
- 4-4 The *M. pneumoniae* HPrK/P dimer
- 4-5 The open and closed conformations of the P-loop
- 4-6 Superimposition of structural homologues of *M. pneumoniae* HPrK/P
- 4-7 View of the putative ATP binding pocket of the *M. pneumoniae* HPrK/P
- 4-8 View down the molecular 3-fold axis into the putative ATP binding site
- 4-9 Superimposition of the N-terminal domain of the *M. pneumoniae* HPrK/P onto structural homologues

Table A-1 MIR Data For CcpA Complex

Table A-2 MAD (Bromouridine) Data For CcpA Complex

Table A-3 Data from MAD (Selenomethionine) experiment

## List of Abbreviations

CCR – carbon catabolite repression  
ATP – adenosine triphosphate  
PTS – phosphoenol-pyruvate phosphotransfer system  
HPr – histidine phosphocarrier protein  
CcpA – catabolite control protein A  
HPrK/P – HPr kinase/phosphatase  
NTD – N-terminal domain of HPrK/P  
CTD – C-terminal domain of HPrK/P  
PEPCK – phospho-enol pyruvate carboxy kinase  
MAD – multiple anomalous dispersion  
MIR – multiple isomorphous replacement  
AS – anomalous scatterer  
SIRAS – single isomorphous replacement with anomalous scattering  
ML – maximum likelihood  
DM – density modification  
cAMP – cyclic adenosine monophosphate  
PRD – PTS response domain

# Acknowledgments

To Ron and Jo.

## Abstract

Carbon catabolite repression (CCR) in bacteria is a global response to the availability of preferred carbohydrates in the growth medium. Bacterial metabolism accomplishes a great number of tasks in parallel and feedback loops exist to coregulate these tasks. Specifically, bacteria couple the uptake of carbohydrates with the uptake of important molecules such as nitrogen, phosphorous and sulfur such that a lack of these molecules will significantly reduce active transport of carbohydrates. They also negatively regulate transcription and translation of genes encoding enzymes involved in the metabolism of carbohydrates either not present in the culture media or less preferable to those that are present. Bacteria also negatively regulate, through phosphorylation or allosteric inhibition of activity, enzymes involved in the metabolism of carbohydrates either absent or less preferable to those that are present. And they must simultaneously activate or increase the expression of enzymes involved in the metabolism of carbohydrates that are either present in the medium or preferable to those present.

These metabolic regulatory circuits are well documented in the Gram-negative bacteria *Escherichia coli*. cAMP and cAMP receptor protein (Crp) produce the major mechanism of CCR in *E. coli*. But the absence, in the

Gram positive *Bacillus subtilis*, of cAMP and Crp alludes to the vast differences in the details of the CCR mechanisms of the two bacteria. In *B. subtilis* the HPr kinase/phosphatase (HPrK/P), histidine phosphocarrier protein (HPr), and catabolite control protein A (CcpA) are central components of a very different mechanism of CCR. Therefore we expected that the structures of close relatives of these *B. subtilis* proteins would provide some insight on the mechanistic details of CCR in *B. subtilis* as well as elucidate novel structural features. To that end the x-ray crystal structure of the HPrK/P from the Gram positive bacterium *Mycoplasma pneumoniae* was determined to 2.5 Å resolution. In addition, initial efforts to define the structure of CcpA bound to its corepressor, phospho-HPr, which has been phosphorylated on residue Ser46, and DNA has yielded data quality crystals. The structural analysis of HPrK/P has implicated residues that are involved in catalysis and identified the ATP binding site. we report the structure of the Gram positive *Mycoplasma pneumoniae* HPrK/P and an attempt to define the structure of a CcpA:P-Ser46-HPr:DNA complex from the Gram positive *Bacillus megaterium*. This work represents a significant advance towards the goal of a structure-function understanding of CCR in Gram positive bacteria.

# Chapter I

## Introduction

### Introduction

Glucose is the preferred carbon source for organisms from Man to yeast to bacteria. It follows then, that these organisms possess elaborate mechanisms to recognize and utilize their favorite carbohydrate. The naïve expectation might be that mechanisms common to so many organisms might also be similar in detail, yet this is probably not true in any strict sense. Mechanistic differences are not well understood though, as researchers who study human glucose metabolism have only recently begun to compare their system with other simpler systems in the hope of further insight (Rutter *et al.*, 2000).

However, two major mechanistic themes have been identified for regulation of glucose metabolism: glucose or a glycolytic product acts directly to effect metabolic change (e.g. inhibition of enzyme); or may regulate transcription or translation of metabolic enzymes (e.g. repressor-ligand binding). Single cell organisms use both themes concurrently, in contrast to multi-cellular organisms. In mammals only one theme is involved in any given cell (Rutter *et al.*, 2000). In the Beta cells of the islets

of langerhans of the mammalian pancreas, for example, glucose sets off a signaling cascade, which leads to transcription of insulin, a hormone that triggers glucose uptake (Rutter *et al.*, 2000). In yeast, however, glucose is recognized by a G-protein coupled receptor transiently activating cAMP production leading to protein kinase A (PKA) activation. PKA then activates a phosphorylation cascade resulting in a switch from a non-fermentative metabolism to the fermentation of glucose to the bacteriocide ethanol (Versele *et al.*, 2001).

## **Bacteria**

Many bacteria also have elaborate mechanisms for regulation of glucose metabolism, which is intimately related to carbon metabolism. Although the full range of carbon sources utilized by a given bacterial species is highly individualized, sugars and related carbohydrates are major sources of energy and carbon for most well studied bacteria. However, many bacteria have distinct preferences for some carbon sources. As Monod discovered for *E. coli* and *B. subtilis*, glucose is consumed exclusively before lactose in a culture that contains both sugars.<sup>1</sup> This phenomenon, called diauxie, also occurs in many other species of bacteria.

---

<sup>1</sup> Monod J 1942 *Recherches sur la croissance des cultures bacteriennes*. PhD thesis as cited in (Stulke, J. and W. Hillen (2000). "Regulation of carbon catabolism in *Bacillus* species." *Annu Rev Microbiol* 54: 849-80; and Deutscher, J., A. Galinier, et al.. (2002).



Diauxie is merely the outward manifestation of a complex metabolic regulatory scheme called carbon catabolite repression (CCR). The 'glucose effect', a term that subsumes the phenomena of diauxie, and hints at the importance of glucose, also hides the true complexity of CCR:

"Early analyses of the glucose effect led to the postulate that it occurs whenever growth conditions are such that degradation (catabolism) exceeds biosynthesis (anabolism). This postulate in turn led to a second notion, namely, that it was the accumulation of one or more cytoplasmic catabolites, derived from the repressing carbohydrate, that gave rise to the glucose effect. This concept, now known to be in error in some instances, but not in others, caused the term 'catabolite repression' to be coined." **The glucose effect in *E. coli* and other bacteria consists of at least four kinetically and mechanistically distinguishable, but physiologically related, phenomena termed permanent repression, transient repression, inducer exclusion, and inducer expulsion.** Permanent repression and transient repression refer to long and short-lived forms of repression that occur in response to glucose addition to the growth medium of a microorganism." (Saier *et al.*, 1995)

In the following discussion we will use the terms CCR and 'glucose effect' interchangeably.

---

*Carbohydrate uptake and metabolism. Bacillus subtilis and its closest relatives: from genes to cells.* A. L. Sonenshein, J. A. Hoch and R. Losick. Washington, D. C., ASM Press: 129-150.

## Carbon Catabolite Repression

As the title of this dissertation suggests we are primarily interested in CCR of a specific family of Gram positive bacteria but it is necessary to first discuss CCR in *E. coli* for several reasons. Historically, diauxie or biphasic growth, in *E. coli* represents one of the earliest examples of CCR in bacteria. Perhaps because of this, CCR in *E. coli* has been one of the most studied metabolic systems of bacteria in general. Consequently, the high level of understanding of CCR in *E. coli* lends itself to comparison with *B. subtilis* and other Gram positive bacteria. Also this research, in *E. coli*, specifically with regard to lactose repression and its use in protein expression, forms the basis for protein overproduction methods and therefore provides a good reference point for most biochemists. As we will show, the overall effects of CCR are similar in both families of bacteria but the mechanistic details, of at least the major mechanism of CCR, are quite different (Fig. 1-1). It is this difference in detail that has motivated our work.

### *E. coli*

Continued study of diauxie in *E. coli* led to the first level of understanding of CCR. Aboud *et al.* showed that expression of the genes of the *lac* operon was repressed by the lactose repressor (LacI) in the presence of lactose. Furthermore, beta-galactosidase mRNA, encoded in *lacZ*, did not

appear until glucose disappeared from culture media containing both glucose and lactose (Aboud *et al.*, 1970). LacI, encoded by *lacI* (also part of the *lac* operon), possesses a helix-turn-helix DNA binding motif and was shown to bind to a pseudo-palindromic site in the *lac* operator in the absence of the inducer lactose thereby repressing transcription of the *lac* operon (Nick *et al.*, 1985). The bound Lac repressor may interfere with the binding of RNA polymerase to DNA due to overlapping binding sites (Schlax *et al.*, 1995). LacI has also been shown to interrupt ongoing transcription (Deuschle *et al.*, 1986) or keep the polymerase from escaping the promoter (Lee, J. *et al.*, 1991). But in the presence of lactose, LacI binds the inducer and releases its grip on the lactose operator de-repressing transcription (Matthews *et al.*, 1998).

LacI was the founding member of what is now a large family of DNA binding proteins known as the LacI/GalR family (Weickert *et al.*, 1992) (Fig. 1-2). The first crystal structure of a protein in this family, PurR (Schumacher *et al.*, 1994), showed how these proteins likely bind DNA and how binding of inducer leads to conformational changes, which must unfavorably realign the DNA binding N-terminal domains and lower the proteins affinity for its operator site (Schumacher *et al.*, 1995). LacI provides a good example of how transcription may be regulated by the

presence or absence of small molecule effectors (i.e. inducers). However, the activated transcriptional state is not necessarily equal to de-repression or the presence of inducer, as in PurR. That is, the presence of lactose will lead to an increase in transcription of the *lac* operon only if the medium does not contain glucose.

In order to achieve activation the global activator, catabolite repressor protein (Crp, encoded by *crp*), bound to cyclic AMP (cAMP), in complex (cAMP:Crp) binds to an operator site upstream of the start site and interacts with RNA polymerase in such a way that it is transcriptionally competent (Nick *et al.*, 1985). A portion of Crp was also shown to interact with the C-terminal domain of RNA polymerase, possibly recruiting it to the operator and upregulating transcription of the *lac* operon (Savery *et al.*, 2002). Subsequently, a large number of genes were identified that contained cAMP:Crp binding sites giving the appearance of generality to the mechanism of cAMP:Crp activation.

Early work on the *lac* operon indicated that the glucose effect arose from fluctuations in the level of cAMP produced by adenylate cyclase. Adenylate cyclase is activated by the histidine phosphorylated form of EIIA<sup>glc</sup> (*crr*), the cytoplasmic subunit of EII<sup>glc</sup>, and a soluble component of the sugar phosphotransferase system (PTS) (Fig. 1-3). The PTS transfers

phosphates from phosphoenolpyruvate (PEP) to incoming PTS sugars through the intermediaries of histidine phosphoproteins HPr, Enzyme I (EI) and sugar specific Enzymes IIABC (EIIs), where EIIBC forms a membrane bound complex and EIAs are cytoplasmic (Saier *et al.*, 2002). Glucose specific EIIA (EIIA<sup>glc</sup>) is histidine phosphorylated in the absence of glucose due to the fact that no glucose is transported to receive the phosphate. The interaction of EIIA<sup>glc</sup> with adenylate cyclase begins the signal cascade to activate transcription at *lac*, if lactose was present to de-repress LacI.

The cAMP:Crp complex can repress or activate transcription depending on where in the operator the Crp site is located; if its upstream of the -35 site it may act as an activator but if its downstream or within the gene it may act as a repressor (Kolb *et al.*, 1993). This allows for a more general response to the absence of glucose or low energy state; in some cases genes need to be repressed, but activated in others depending upon their ultimate role in CCR. Crp binds DNA with low affinity but upon binding of cAMP the complex (cAMP:Crp) binds with high affinity to a 22 base pair palindromic DNA sequence (Stormo *et al.*, 1989) as a dimer. Crystal structures of cAMP:Crp bound to DNA show that the DNA is sharply bent around the dimer, which may be one mechanism of transcriptional regulation (Parkinson *et al.*, 1996). Usually, when

cAMP:Crp activates transcription its binding site is either close to RNAPol and interacting with the polymerase or further upstream interacting with other proteins in a complex activation mechanism. For example, in genes that are regulated by the cytidine repressor (CytR , LacI/GalR family member) cAMP:Crp increases the binding affinity of CytR 100-1000 fold *in vitro* repressing a number of genes involved in nucleoside and deoxynucleoside uptake and metabolism (Pedersen *et al.*, 1991). And in some cases cAMP:Crp may have operator-binding sites overlapping the site of another regulatory protein so that regulation is achieved by the number of active DNA binding proteins and may allow for coregulation (Kolb *et al.*, 1993).

The elucidation of the opposing effects of cAMP and lactose on the regulation of the *lac* operon has led many to assume, probably correctly, the existence of similar control features throughout *E. coli*. But, as the above quote suggests, CCR is more complex; the exclusion of inducers and the expulsion of cAMP from the inside of the cell also play important roles in *E. coli* CCR. In fact glucose not only inhibits production of cAMP it promotes its expulsion (Saier *et al.*, 1975). Also, EIIA<sup>glc</sup> inhibits uptake of lactose through its interaction with lactose permease (LacY), and this phenomenon is called inducer exclusion. EIIA<sup>glc</sup> is special in that it is an allosteric

activator in both its phosphorylated and unphosphorylated states. When glucose is not available EIIA<sup>glc</sup> is phosphorylated and acts to reverse the glucose effect (e.g. activates adenylate cyclase); and when glucose is available EIIA<sup>glc</sup> is unphosphorylated and acts to repress transcription of enzymes of non-glucose metabolism (Fig 1-1; A&B).

As previously noted, inducer exclusion is a major component of the glucose effect. It is tied to the PTS through the allosteric interactions of unphosphorylated EIIA<sup>glc</sup> with other, non-PTS, carbon-source permeases specific for maltose (MalK), lactose (LacY), melibiose (MelB) and raffinose (RafB). EIIA<sup>glc</sup> interacts with a consensus binding sequence found in these permeases between the central loop connecting transmembrane helices 6 and 7 in LacY. This interaction inhibits uptake of their respective carbohydrates (Titgemeyer *et al.*, 1994). Recent work suggests that it is the ratio of PEP to pyruvate, a telltale sign of energy level, which determines the phosphorylation state of EIIA<sup>glc</sup> (Hogema *et al.*, 1998). In fact the locus *ptsG*, which encodes the PTS glucose transporter EIIBC<sup>glc</sup>, is upregulated by cAMP:Crp and its increased expression may be the origin of diauxie (rather than the reduction of cAMP levels and concomitant removal of the global activator cAMP:Crp). The increased production of EIIBC<sup>glc</sup> also leads to a reduction in phosphorylated EIIA<sup>glc</sup> by providing a phosphate acceptor sink

(Fig. 1-1B). Lactose is excluded from the cell if EIIA<sup>glc</sup> shuts down lactose permease (LacY) making it impossible for RNAPol to transcribe through bound lactose repressor (LacI) (Kimata *et al.*, 1997). This explains the preference for PTS sugars over non-PTS sugars but not the preference for glucose over other PTS sugars, the reasons for which remain to be discovered.

The regulation of glycerol metabolism reveals a slightly different method of inducer exclusion. Glycerol enters the cell by facilitated diffusion mediated by GlpF. In order to enter the glycolytic pathway, however, glycerol must be phosphorylated by glycerol kinase (GlpK) to produce glycerol-3-phosphate. GlpK is allosterically inhibited by the glycolytic intermediate FBP and unphosphorylated EIIA<sup>glc</sup> (Fig 1-1; A&B). FBP inhibits formation of the active tetramer (Pettigrew *et al.*, 1996) while EIIA<sup>glc</sup> interacts with a helix far from the active site (Hurley *et al.*, 1993) possibly effecting long range conformational changes that lead to inhibition of the kinase.

In many instances cAMP:Crp provides a global activation signal in response to phosphorylated EIIA<sup>glc</sup> (production of cAMP) while individual repressors (e.g. LacI) or activators (e.g. PhoB see below) control the transcriptional response to inducers. But *E. coli* also uses antitermination as



a means to control transcription. Genes controlled by antitermination have a specific upstream RNA sequence recognized by antiterminator proteins that will adopt a conformation that releases RNAPol unless the antiterminator binds.

The *bgl* operon in *E. coli*, which encodes genes involved in uptake and assimilation of  $\beta$ -glucosidic sugars, is controlled by the antiterminator BglG. BglG is a member of the PTS regulation domain (PRD; composed of two subdomains 1&2) (van Tilbeurgh *et al.*, 2001a) family of proteins whose activity is regulated by phosphorylation of their PRD domains by members of the PTS (Gorke *et al.*, 1999). The first two gene products of the operon, BglG and EII<sup>bgl</sup>, are sufficient for its antitermination (Gorke *et al.*, 1999). When EII<sup>bgl</sup> is phosphorylated, due to low glycolytic flux or the absence of  $\beta$ -glucosidic sugars, this phosphate is transferred to the negative regulation domain of BglG and it is unable to antiterminate leaving the *bgl* operon untranscribed. But phosphorylated-His15-HPr (p-his15-HPr) can transfer its phosphate to the activation domain of BglG, which then undergoes a conformational change allowing it to bind to two  $\rho$ -independent antiterminators in the operon thus allowing transcription of *bgl* (van Tilbeurgh *et al.*, 2001a).

Interestingly, it seems that nitrogen metabolism is tied to carbon metabolism through the PTS. Apparently *rpoN*, the operon which encodes an alternate sigma factor,  $\sigma^N$ , required for transcription of genes necessary for nitrogen fixation, also encodes two genes homologous to EIAs and to HPr of the PTS, called EIA<sup>nt</sup> and NPr respectively ( Fig. 1-4) (Powell *et al.*, 1995). Powell *et al.* showed that EIA<sup>nt</sup> could be phosphorylated by HPr and that NPr could be phosphorylated by EI, thus establishing a mechanism by which nitrogen assimilation is linked to carbon uptake.

Similarly, two other important cellular components, sulfur and phosphate are coordinately regulated with carbon. Adenylate cyclase is responsive to sulfur deprivation suggesting a pathway for regulation linkage (Quan *et al.*, 2002) and the phosphate starvation inducible gene E (*psiE*) is under the control of PhoB (a response regulator sensitive to the histidine kinase PhoR) and cAMP:Crp (Kim, S. K. *et al.*, 2000). The promoter of *psiE* contains two Pho boxes, one upstream and one at the -35 site where a cAMP:Crp binding site overlaps the PhoB site. cAMP:Crp represses transcription of *psiE* in low glucose conditions presumably because it overlaps the RNAPol binding site. However, in low phosphate conditions, when PhoB is phosphorylated and DNA binding competent, *psiE* is transcribed. Apparently gene products for growth under phosphate

starvation are important enough that they need to be transcribed even at low energy levels.

Not only are the activities of carbohydrate transporters regulated but their expression and type are also regulated in response to glucose levels. Upon passing cultures of *E. coli* from high to low glucose concentrations there are changes in the overall transporter profile; at high glucose concentrations (mM) EII<sup>glc</sup> is the principal transporter while at lower glucose concentrations (optimal at  $\mu$ M glucose) the galactose transporter (and periplasmic binding protein), which also transports glucose, (MglA) high affinity transporter predominates (Ferenci, 1996). This occurs through endoinduction where intracellular concentrations of galactose and maltotriose, manufactured during glucose famine, activate expression of the genes of the *mgl* (MglA) and *mal/lamB* (outer membrane porin, LamB) regulons, which are required for high affinity transport (Manch *et al.*, 1999). cAMP also plays a role in induction of high affinity transport, however *crp* mutants do not induce *mgl* or *mal* under limiting glucose (0.1 – 300 $\mu$ M glucose), suggesting endoinduction is the dominant regulatory scheme for the switch to high affinity glucose transport.

The mechanisms of the glucose effect in *E. coli*, described above, are all mediated by cAMP:Crp. That is, cAMP:Crp is somehow involved in

making the glucose effect palpable for these processes. But there is at least one subsystem of CCR, namely the *fru* regulon, mostly unaffected by the noted disruptions, and therefore presumed to occur outside of cAMP:Crp and PTS regulation.

The *fru* operon encodes an alternative PTS system called the fructose specific PTS, which allows *E. coli* with a deleted HPr gene (*ptsH*), to perform gluconeogenesis if fructose, but no other PTS sugar, is in the medium. The *fru* operon has an EIIBC<sup>fru</sup> and a protein DTP homologous to both HPr and EIAs capable of being phosphorylated by EI of the PTS (Saier *et al.*, 1996). The similarities between the two phosphoryl transfer systems and the fact that no other sugars have their own PTS may indicate that fructose is a special sugar and that it may be a primordial PTS; possibly this is why glucose is epimerized to fructose in the glycolytic pathway. Yet the *fru* operon also encodes a LacI/GalR family member, catabolite repressor/activator (Cra; formerly FruR), which has been shown to be important for regulating transcription of general metabolic enzymes; positive regulation of PEP synthase (*ppsA*), PEP carboxykinase (*pckA*), and cytochrome d oxidase (*cyd*) *et al.*; negative regulation of fructose and mannitol catabolic enzymes (*fruBKA* and *mtlADR*), HPr and Enzyme I of the PTS (*ptsHI*) (Ryu *et al.*, 1995), as well as enzymes of glycolysis (*edd eda*).

Some of these operons (e.g. *ppsA*) are only regulated by this mechanism as they have no cAMP:Crp binding sites in their operators and are therefore not regulated by the cAMP levels in the cell. The significance of metabolic genes regulated by Cra and its indifference to cAMP suggests the existence of a cAMP independent control circuit for carbon metabolism, which would affect glycolytic flux (Saier *et al.*, 1996).

### ***B. subtilis***

The work presented in the body of this dissertation involves a dual activity enzyme, HPr Kinase/Phosphatase (HPrK/P) from *Mycoplasma pneumoniae*. *Mycoplasma pneumoniae* belongs to a class of bacteria called the low guanine-cytosine (G+C), Gram positive bacteria. However, little work has been done to elucidate the mechanisms of CCR in most of these organisms. Rather, low G+C Gram positive bacteria are assumed to have CCR mechanisms similar in kind to *B. subtilis* due to their overall relatedness as well as extensive homologies between orthologous proteins. In contrast to the dearth of information available for *M. pneumoniae* a great deal of information exists bearing on the CCR mechanisms of *B. subtilis*. In fact, unlike *E. coli* and other Gram-negative enteric bacteria, *B. subtilis* have neither a Crp polypeptide (Kunst *et al.*, 1997) nor does addition of cAMP

lead to metabolic changes (Nihashi *et al.*, 1984). So, although *E. coli* and *B. subtilis* have many similarities, *B. subtilis*, and presumably the other low G+C Gram positive bacteria, lack the major mechanism of transcriptional control found in *E. coli*.

### Similarity to *E. coli*

Transport phenomena in *B. subtilis* are intimately tied to CCR through transcriptional control of expression of permeases. Generally permeases in both *E. coli* and *B. subtilis* fall into four classes (Fig. 1-5): 1) channels (e.g. GlpF), 2) Major Facilitator Superfamily of secondary transporters ~15 of which are thought to uptake sugars through sugar-proton symport, 3) ATP Binding Cassette (ABC) family of primary transporters, ~10 of which are thought to transport sugars (e.g. MglA in *E. coli*), which use ATP to pump against a gradient, and finally 4) the group translocators of which there are ~17 (4 families: glucose, fructose, lactose, mannose) that transport PTS sugars. The EII transporters are composed of 3 or 4 polypeptides (EIIB, EIIC, and possibly EIID for mannose), which couple phosphorylation and transport in a poorly characterized mechanism that consumes one ATP per one molecule of sugar transported (Saier *et al.*, 2002). *B. subtilis* also encode

HPr and EI, however there is an additional HPr homolog, Crh, unique to bacilli (Reizer *et al.*, 1999; Deutscher *et al.*, 2002)

As previously noted in the case of *E. coli* the relative mixture of permeases expressed and operative at the cell surface depends on a signal transduction network. At low external concentrations of sugars the ABC transporter family proteins, which have high affinity for specific sugars, become important. However, these permeases require a helper protein that is soluble and periplasmic in *E. coli* but is tethered to the membrane in *B. subtilis*. The high affinity glucose permease in *E. coli* MglA has a homolog YufO in *B. subtilis* (Quentin *et al.*, 1999).

As in *E. coli* transcriptional repressors of *B. subtilis* negatively control expression of operons encoding enzymes of carbohydrate catabolism and transport. And in some cases these repressors are related to the LacI/GalR family of repressors. AraR a negative regulator of L-arabinose metabolism controls the *araABDLMNPQ-abfA* metabolic operon for arabinose catabolism and *araE* gene for arabinose permease (Mota *et al.*, 1999). AraR bears some semblance to LacI/GalR at its C-terminal end, which is the inducer-binding region of the Lac repressor family; however, at its N-terminal, DNA binding domain it resembles the non-LacI family member GntR, the repressor of the *gntRKPZ* operon required for gluconate

metabolism. GntR is the founding member of a large family of DNA binding proteins found in bacteria that have been shown to bind to DNA in the absence of their ligands but not in its presence (Yoshida *et al.*, 1995). Recently a crystal structure of a GntR family protein, FadR, was reported and it is reminiscent of a shortened PurR (van Aalten *et al.*, 2000).

The PRD proteins of *B. subtilis* illustrate how the PTS regulates the energy level in the cell and transduces the signal to a transcriptional response. All eight PRD containing antiterminators found in the *B. subtilis* genome regulate the expression of PTS permeases and their catabolic enzymes (Reizer *et al.*, 1999). Significantly, BglG is the only PRD containing protein that has been characterized in *E.coli*. In contrast the PRD domains are more numerous and varied in *B. subtilis*; transcriptional activators with PRDs have been characterized as well. PRD control of permease expression may be a major mechanism of inducer control in *B. subtilis* since they have no global allosteric regulator like EIIA<sup>glc</sup> from *E. coli*.

The PTS regulation domain is used for transcriptional control. PRD domains are divided in two; PRD1 and PRD2. Each domain has two histidine phosphorylation sites, which are largely conserved throughout the PRD family (Fig 1-6). Phosphorylation at some sites leads to dimer



formation while at other sites dimer destabilization (Amster-Choder *et al.*, 1992). Obviously the dimer state will have transcriptional consequences.

In the *lic* operon, which encodes genes that are important for the metabolic use of oligomeric beta-glucosides, LicT and LicR play key roles in the regulation of this operon. LicT acts as an antiterminator while LicR is an operon activator. The RNA forms cruciform structures, where RNAPol stalls in the absence of LicT, but if LicT is activated it binds the RNA inhibiting formation of secondary structure and allowing RNAPol to transcribe through. The crystal structure of the LicT-PRD, a fragment of LicT containing both PRDs, gives some indication of a general mechanism (van Tilbeurgh *et al.*, 2001b). Since PRD1 is believed phosphorylated by EII<sup>bgl</sup> in the absence of inducer (inactivating) and PRD2 is phosphorylated by EI~P and HPr~pH15 in the absence of glucose (Tortosa *et al.*, 1997) perhaps phosphorylation of PRD1 by EII<sup>bgl</sup> disrupts RNA binding dimer formation while EI-p and HPr-p phosphorylation of PRD2 favor RNA binding dimer formation (Gorke *et al.*, 1999).

Based on the LicT insights into the PRD mechanism, other antiterminator and activator proteins, which contain PRDs, are postulated to have a similar mechanism (van Tilbeurgh *et al.*, 2001a). Also LicR has an EIIA-like domain known to be a site for phosphorylation by EII<sup>lic</sup> (Tobisch

*et al.*, 1999a). However, they both have N-terminal DNA (LicR; helix turn helix) or RNA (LicT) binding domains and PRD1 and PRD2 (Deutscher *et al.*, 2002). No general rule exists for the role of any given histidine; their roles seem protein specific but the type and number of histidine phosphorylations will affect the overall conformation of the protein and thus its nucleic acid binding affinities. Perhaps even the dimerization state is effected by phosphorylation (Amster-Choder *et al.*, 1992). It has also been postulated that sequestration of PRD proteins at the membrane by EII may be one mechanism for deactivation (van Tilbeurgh *et al.*, 2001a).

Thus van Tilbeurgh and Declerk (van Tilbeurgh *et al.*, 2001a) propose a mechanism like LicT in which the phosphorylation state of EII<sup>lic</sup> effects the phosphorylation state of LicR. That is, if inducer is present phosphate flows to it from PRDs and HPr dephosphorylating, and activating LicR, which leads to transcription of the *lic* operon. But if glucose is present *lic* is only 5-fold induced. If no inducer is present EII<sup>lic</sup> phosphorylates LicR in its EIIA-like domain inactivating LicR regardless of the phosphorylation state of PRDs.

*Lactobacillus brevis*, a low GC Gram positive bacterium, uses modified HPr that has been phosphorylated on residue serine 46 (P-Ser46-HPr) by HPrK/P to regulate the galactose:H<sup>+</sup> symporter permease (galP) by

binding to this transporter and uncoupling sugar-proton symport. This allosteric interaction, reminiscent of the EIIA<sup>glc</sup> interaction with LacY in *E. coli*, allows galactose to diffuse out of the cell (Fig 1-1 D). Radioactive galactose diffusion was demonstrated in *B. subtilis* by simply expressing *galP* from *L. brevis* in a mutant aspartate 46 HPr (S46D HPr) background thus reconstituting inducer regulation in *B. subtilis*, which does not normally display this type of inducer control (Djordjevic *et al.*, 2001). In a related experiment, inactive mutants of HPrK/P in *Lactobacillus casei* destroy this type of inducer control showing the importance of HPrK/P in CCR (Dossonnet *et al.*, 2000).

Expulsion of inducers (like cAMP) apparently does not occur in *B. subtilis*, although multi-drug efflux (Saier *et al.*, 2002) and other transporters with outward polarity do exist in the genome. It seems prevention of induction (modulated by PRDs) is the preferred method of controlling the expression of many secondary catabolic operons.

### **Differences with *E. coli***

Clearly the extent and utilization of the PRD proteins in *B. subtilis* is different from *E. coli* but many similarities do exist with the BglG mechanism of antitermination found in *E. coli*. However, the major

mechanism of CCR in *B. subtilis* has no parallel in *E. coli*. In *B. subtilis* HPr kinase/phosphatase (HPrK/P) phosphorylates the HPr (and Crh) proteins on serine 46, which act as corepressors of catabolite control protein A (CcpA; LacI/GalR member) to effect global metabolic control (Fig. 1-1 C&D).

Historically, the mechanism of CCR in *B. subtilis* was delineated by identifying the action of CcpA at the alpha amylase operon.

The alpha amylase operon (*amyE*), was known from previous work to experience negative regulation under glucose repression (Weickert *et al.*, 1990). Transposon inactivation of the cis acting element of negative regulation at the *amyE* operon identified the *ccpA* gene, which encodes the LacI/GalR family member CcpA (Henkin *et al.*, 1991) (Fig. 1-2). *amyE* was found to contain an operator sequence overlapping the translational start site for the operon (Weickert *et al.*, 1990). Kim *et al.* showed that purified CcpA binds to *amyE* DNA *in vitro* (Kim, J. H. *et al.*, 1995). This operator sequence was found to fit a consensus sequence 14 base pair pseudo-palindromic subsequently named the catabolite response element (CRE) (Hueck *et al.*, 1994). Using this consensus sequence to probe the genome of *B. subtilis* revealed that CREs were found near the promoter for many genes encoding extra-cellular enzymes, intracellular carbohydrate metabolism enzymes, some glycolytic enzymes, and sporulation enzymes (Hueck *et al.*,

1994; Tobisch *et al.*, 1999b; Yoshida *et al.*, 2001). Moreno *et al.* discovered that in excess of 250 genes show greater than 2-fold glucose repression or activation directly attributable to CcpA control (Moreno *et al.*, 2001).

Although the identification of CREs and their binding partner CcpA may seem to explain repression of many genes there is another level of complexity. A serine 46 to alanine mutation in HPr, which was known to be phosphorylated at serine 46 and histidine 15 (Eisermann *et al.*, 1988), led to significant relief of CCR (Deutscher *et al.*, 1994; Reizer *et al.*, 1996); and when a closely related protein, Crh (85aa and 45% identity to HPr) was mutated to alanine at serine 46, in addition to the same mutation in HPr, complete CCR relief was achieved (Galinier *et al.*, 1997). Furthermore, the *in vitro* binding of CcpA to CRE sites was significantly enhanced in the presence of phospho-S46-HPr (Deutscher *et al.*, 1995) or phospho-S46-Crh, (Galinier *et al.*, 1999). These results suggested that HPr, Crh and their modifying enzyme HPr kinase/phosphatase, are critical components of CCR (Fig. 1-7).

Further *in vitro* binding studies showed that phospho-S46-HPr binds to CcpA only if HPr is not phosphorylated at His15 (Deutscher *et al.*, 1995). Kraus *et al.* noticed that the various CcpAs found in the low G+C Gram positive bacteria contained additional regions of conservation with respect to

other LacI/GalR family members (Fig. 1-2). Five mutations of the CcpA regulator (R47S, Y89E, Y295R, A299E and R303D) from *Bacillus megaterium* were identified that abrogated both catabolite repression *in vivo* and P-Ser46-HPr binding *in vitro* (Kraus *et al.*, 1998) (Fig. 1-8).

As we noted previously HPrK/P can both phosphorylate and dephosphorylate residue Serine 46 of HPr and Crh. The activities of HPrK/P are regulated by the allosteric effector fructose-1,6-bisphosphate (FBP; favors kinase activity), and the concentration differential between ATP/GTP and inorganic phosphate ( $P_i$ ; favors phosphatase activity). To date some 32 HPrK/P sequences from various bacteria have been described. These genes are found primarily in low G+C gram positive bacteria and purple bacteria, but the spirochete *Treponema pallidum* also encodes the enzyme. An alignment of representative HPrK/P sequences shows homology and conservation throughout the coding region, albeit with stricter conservation in the C-terminal half of the predicted proteins (Fig. 1-9). Recently crystal structures were reported of a fragment of the *Lactobacillus casei* HPrK/P (Fieulaine *et al.*, 2001) and the full-length *Staphylococcus xylosus* HPrK/P (Marquez *et al.*, 2002), which are discussed at length in Chapter 4.

In another interesting parallel with *E. coli*, glycerol metabolism is also allosterically regulated in *B. subtilis*, however the regulator is p-His15-HPr

rather than EIIA<sup>glc</sup>. Glycerol is the only known carbohydrate that is transported by facilitated diffusion, which is dependent on GlpF, an aquaporin family member (Deutscher *et al.*, 2002). In order to accumulate glycerol within the cell and degrade it through glycolysis glycerol must be phosphorylated by glycerol kinase (GlpK). GlpK that is phosphorylated on a histidine by p-His15-HPr is 10-fold more active (Darbon *et al.*, 2002). The product, glycerol-3-phosphate, then induces the *glp* regulon through its interaction with the glycerol antiterminator (GlpP). Interestingly, the *glp* regulon is also controlled by a CRE element thereby providing a means to repress transcription in the presence of glucose (Fig 1-10). This system nicely illustrates how *B. subtilis* achieves a very different signaling pathway with proteins that it shares with *E. coli*.

## Conclusion of Comparison

The PTS system plays a central role in the metabolic cycle of both *E. coli* and *B. subtilis*. The proteins of these PTS systems are highly homologous and appear to function in like manner. Furthermore, the enzymes of the PTS exert regulatory control over much of CCR. However, in *E. coli* the locus of regulation appears to be EIIA<sup>glc</sup> through inhibition of non-PTS permeases in its unphosphorylated state and activation of adenylate

cyclase in its phosphorylated state. The lack of inducers and the presence of cAMP:Crp, respectively, have widespread effects on transcription of genes important in various metabolic cycles of *E. coli*.

The absence of Crp and cAMP in low G+C Gram positive bacteria is the most striking difference with respect to *E. coli* and other enteric bacteria. The PTS is still central to CCR in *B. subtilis* and other low G+C Gram positive bacteria; however here global regulation of gene expression is controlled by P-Ser46-HPr (and pP-Ser46-Crh in bacilli); and induction at some operons encoding sugar permeases is coordinately controlled by P-His15-HPr and their corresponding EIIBCs through PRD proteins. HPrK/P controls the serine phosphorylation state of HPr (and Crh in bacilli), which leads to control of transcription of a large portion of the *B. subtilis* genome.

### ***Mycoplasma pneumoniae***

*M. pneumoniae* is a human pathogen of the respiratory tract and the causative agent in atypical pneumoniae (Krause, 1996). Although *M. pneumoniae* is a member of the low G+C gram positive bacteria family, the most well known member of which is *B. subtilis*, at 41% G+C it is on the high end of the low GC spectrum. Microbiologists now believe that *Mycoplasma pneumoniae* and its closest known relative *Mycoplasma genitalium* evolved from a *B. subtilis*-like (average sequence identity with



Bs. is 65%) precursor through gene reduction facilitated by a parasitic lifestyle. The *M. pneumoniae* genome, first sequenced in 1996 and recently re-annotated and updated, has one of the smallest genomes of free living organisms (Dandekar *et al.*, 2000). At 816 Kbp and 688 open reading frames some researchers believe *M. pneumoniae* may be approaching a theoretical limit to genome reduction. However, global transposon mutagenesis suggests that only 350 of the orfs may be indispensable (Hutchison *et al.*, 1999).

Perhaps as a result of gene reductions the mycoplasmas display numerous anomalies. Conspicuously, they have no cell wall. Also, transcription in *M. pneumoniae* does not seem to have a strong consensus -35 box; and transcripts display heterogeneity in their start sites possibly related to the fact that *M. pneumoniae* has only one, rather large sigma factor (Weiner *et al.*, 2000). Also, *M. pneumoniae* have no TCA cycle genes nor do they respire; apparently all ATP is generated from glycolysis, ribose phosphate, and other pathways. Interestingly, genes encoding glycolytic enzymes and proton motive force enzymes were found undisrupted in a transposon mutagenesis study implying that their gene products are required for viability. The identification of relatively few orthologous proteins with known functions suggests that the Mycoplasmas have only a handful of

regulatory genes (Hutchison *et al.*, 1999). However, while *M. pneumoniae* encodes homologues of the PTS (including HPr), homologues of CcpA and Crh are conspicuously absent. Consequently, to the extent CCR might occur in *M. pneumoniae*, its mechanism would likely be different from that of *B. subtilis*.

The HPrK/P enzyme from *M. pneumoniae*, at 312 amino acid residues, is only 2 residues longer and displays significant homology to the *B. subtilis* enzyme (33% identity and 54% similarity) (Fig. 1-9). Although maintained throughout the protein, this homology is more striking in the C-terminal domain, which contains the canonical ATP binding P-loop (Saraste *et al.*, 1990) and the HPrK/P family consensus sequences (Reizer *et al.*, 1998). Both *B. subtilis* HPrK/P and *M. pneumoniae* HPrK/P display *in vitro* phosphatase activity with concentrations of ATP and  $P_i$  of 200  $\mu$ M and 5 mM, respectively. If 10 mM FBP is added both enzymes regain kinase activity (Steinhauer, 2002). *In vitro* kinase activity assays using *B. subtilis* HPrK/P show ATP stimulates phosphorylation cooperatively (half maximal activity at 1 mM ATP), while FBP can stimulate phosphorylation at lower concentrations of ATP (half maximal activity at 1 mM FBP, 25  $\mu$ M ATP) (Jault *et al.*, 2000). *M. pneumoniae* HPrK/P kinase activity assays suggest that it is also activated cooperatively by ATP, although at a much lower

concentration (half maximal activity at 10  $\mu$ M). However, FBP has no stimulatory effect on phosphorylation; it merely overcomes inhibition due to  $P_i$  (half maximal kinase activity 50  $\mu$ M ATP, 5 mM  $P_i$ , and 10  $\mu$ M FBP) (Steinhauer, 2002).

## **Mechanisms of Ser/Thr Kinases and Phosphatases**

### **Histidine Kinases**

The histidine kinase (HK) – response regulator (RR) system is perhaps the most well known protein kinase system found in bacteria. Most HKs are homodimers localized to the membrane where they have a periplasmic sensory domain at the N-terminus, which transduces a signal to the C-terminal kinase domain in the cytoplasm (Stock *et al.*, 2000). The HK catalytic core of about 350 residues is unlike any serine/threonine or tyrosine protein kinase in sequence or structure (Tanaka *et al.*, 1998). Upon binding of ATP the catalytic core autophosphorylates on a conserved histidine. The phosphate is then transferred to a conserved aspartate on the RR leading to conformational changes, which presumably facilitate DNA binding of the winged helix-turn-helix domains of the RR (Lee, S. Y. *et al.*, 2001). The DNA binding signal is quenched by the eventual auto-de-phosphorylation

activity intrinsic to the RR (West *et al.*, 2001). However, the HK catalytic domain has no obvious sequence similarities to HPrK/P.

### **Protein Kinases**

Protein kinases initiate and perpetuate the complex intracellular signaling cascades, found predominantly in eukaryotes, by phosphorylating specific signaling proteins on serine, threonine, or tyrosine residues. Most Ser/Thr/Tyr protein kinases in eukaryotes belong to the cAMP dependent kinase (cAPK  $\alpha\beta\gamma$ ) catalytic subunit superfamily based on homologies found in their various catalytic domains. An alignment of 117 eukaryotic Ser/Thr protein kinases delineated 12 short sequence motifs conserved throughout the catalytic domains (Hanks *et al.*, 1991). In many cases these sequence motifs are known to be involved in the mechanism of phosphorylation. In fact small but correlated variations in these domains is diagnostic for Ser/Thr or Tyr protein kinases so that the type of kinase is easily predicted by its primary sequence. Shi *et al.* in a bioinformatic approach used the protein kinase superfamily sequence motifs from eukaryotes to search bacterial genomes (Shi *et al.*, 1998). They found 28 open reading frames of putative kinases including 2 in *B. subtilis* and 1 each in *M. genitalium* and *M. pneumoniae*; these sequences share no homology with HPrK/P from *M. pneumoniae*. So, PKA-like protein kinases do not

appear to be confined to eukaryotes. However, definitive evidence of a bacterial PKA awaits biochemical verification.

The first crystal structure of PKA revealed a dual lobed, single domain monomer (Knighton *et al.*, 1991). The conserved core of the PKA structure is composed of a smaller N-terminal subdomain composed primarily of a 5 stranded antiparallel  $\beta$  sheet, which binds nucleotide, and a larger C-terminal domain of mostly  $\alpha$  helical structure, which contains most of the conserved catalytic residues. A single  $\beta$  strand links the two lobes. The active site is on one side of a deep fissure between the two lobes facilitating the proximal binding of the peptide substrate.

Figure 1-11 shows a model of a putative transition state structure of PKA with ADP,  $AlF_3$ , and a peptide substrate that suggests a catalytic mechanism (Madhusudan *et al.*, 2002). The proposed mechanism is also depicted in Figure 1-11. Residue Asp166 has been suggested as the likely catalytic base. However, quantum mechanical calculations suggest that Asp166 may not be the general base but rather simply orients the hydroxyl group for nucleophilic attack (Hutter *et al.*, 1999). The Lys168 side chain probably also plays an important role in catalysis by stabilizing the planar charge formation of the putative transition state of the phosphate thus

facilitating nucleophilic attack by the substrate hydroxyl (Madhusudan *et al.*, 2002).

### **P-loop kinases**

The phosphate binding loop, the so named P-Loop, has the canonical sequence GxxxxGK(T/S) and was originally identified by Walker in 1982 (Walker *et al.*, 1982). The P-loop is also called the Walker A box. The P-loop is found in many dissimilar proteins that bind ATP or GTP, HPrK/P among them. Adenylate kinases, ras proteins, elongation factors, ATP synthase  $\beta$ s, myosin heavy chains, thymidine kinases, phosphoglycerate kinases and the HPrK/Ps from many disparate organisms can all be aligned with the above consensus sequence; and individual families (e.g. adenylate kinases) may have extensions and variations on the canonical GxxxxGK(T/S) (Saraste *et al.*, 1990). Numerous X-ray crystal structures of proteins containing the P-loop have been reported. These structures show that the P-loop occurs as an intervening loop between a helix and a beta strand that hydrogen bond with each other. Each residue is important for nucleotide binding. For example, in p21-ras the invariant lysine, which is found at the end of the loop and the beginning of the helix, hydrogen bonds with both the  $\beta$  and  $\gamma$  phosphate of the bound nucleotide. The invariant

Ser/Thr residue following the lysine is involved in coordinating the bound magnesium ion (Milburn *et al.*, 1990).

A survey of crystal structures of proteins with bound nucleotides reveals most kinases that transfer the  $\gamma$ -P of ATP or GTP to an acceptor have 3 recognizable structural motifs (Traut, 1994): the kinase-1, kinase-2, and kinase-3 motifs. Kinase-1a, a subgroup of kinase-1 having an invariant lysine in the motif, resembles the P-loop and interacts with the nucleotide phosphates. Kinase-2 contains an invariant aspartate involved in coordinating the divalent cation. Kinase-3, less often involved in binding, interacts with the base of the nucleotide.

### **Isocitrate Dehydrogenase Kinase/Phosphatase**

Like HPrK/P, isocitrate dehydrogenase kinase/phosphatase (IDHK/P) is a bifunctional enzyme, which regulates the activity of isocitrate dehydrogenase (IDH) through reversible phosphorylation on a serine found in the active site of (LaPorte, 1993). Phosphorylated IDH is inactive and allows isocitrate to flow through the glyoxalate shunt required for growth on acetate. Miller *et al.* showed that wild type IDHK/P displays micromolar  $K_d$ s for both phosphorylated and unphosphorylated IDH (Miller *et al.*, 1996). The maximum velocities of both kinase and phosphatase activities are similar but the intrinsic ATPase activity is 10-fold higher than either and

only partially inhibited by phosphatase and kinase activity. ATP or ADP is required for both activities but a non-hydrolyzable ATP analogue does not support phosphatase activity. The model developed as a result of this information maintains that IDHK/P has a single active site existing in two different catalytically competent conformations - phosphatase ready or kinase ready - and that the phosphatase uses the kinase back-reaction, while the kinase alone is dependent on ATP hydrolysis.

A more recent study reveals that IDHK/P is responsive to two classes of allosteric effectors probably acting at different sites: AMP, 3-phosphoglycerate, and pyruvate activate the phosphatase activity and inhibit the kinase and ATPase activities; NADPH and isocitrate inhibit kinase activity without activating phosphatase activity and NADPH also inhibits phosphatase activity (apparently it binds to both phosphorylation states of IDH, whereas isocitrate only binds to dephosphorylated IDH). Isocitrate and NADPH bind to IDH, possibly inhibiting binding of IDH to IDHK/P, and do not affect the ATPase activity of IDHK/P (Miller *et al.*, 2000).

IDHK/P has a number of interesting parallels with HPrK/P and may provide some clues to future investigators of the kinetics of HPrK/P. Clearly, the most interesting is the presence of the same dual catalytic activities of



these two classes of enzymes and allosteric control by metabolically important small molecules.

## **Phosphatases**

### **Dual specificity phosphatases**

Protein phosphatases halt the intracellular signaling cascades found in eukaryotes by dephosphorylating specific signaling proteins phosphorylated on serine, threonine, or tyrosine residues. Phosphatases are classified according to the type of phosphoryl bond they hydrolyze; phosphoprotein phosphatases (PPs) are specific for the serine/threonine-phosphate bond while phosphotyrosine phosphatases (PTPs) are specific for the tyrosine-phosphate bond. Although PTPs generally are outside the scope of this discussion one sub-class, called dual specificity PTP (DS-PTP), can dephosphorylate both phospho-serine/threonine and phospho-tyrosine. The eukaryotic tyrosine specific protein phosphatases (PTPs) all contain the C(X)<sub>5</sub>R(S/T) consensus sequence involved in phosphate binding (Jackson *et al.*, 2001) however the DS-PTPs have an expanded consensus sequence (VXVHCXXGXXRS(X)<sub>5</sub>AY(L/I)M). Vaccinia H1 Related (VHR) phosphatase, is composed of only the ~250 residues of the catalytic domain common to all DS-PTPs. The X-ray crystal structure of VHR (Yuvaniyama *et al.*, 1996) reveals the presence of a phosphate binding loop containing the

CXXGXXRS portion of the consensus sequence, which is structurally reminiscent of P-loops but has a sequence more akin to the kinase-1b motif (Traut, 1994).

Another defining feature of the PTPs shared by VHR and other DS-PTPs is the mechanism of dephosphorylation. PTPs use the conserved cysteine to carry out a nucleophilic attack on the phosphorylated residue. The cysteine is then temporarily phosphorylated as a covalent intermediate, which is then hydrolyzed thus completing the enzymatic cycle. Figure 1-12 gives a schematic view of the PTP mechanism as it occurs in VHR. Specifically, residue Asp92 is thought to act as the general acid in the development of the phosphoenzyme intermediate and the general base responsible for the renewal of the enzyme (Jackson *et al.*, 2001; Kim, J. H. *et al.*, 2001).

### **Serine/threonine specific phosphatases**

Jackson & Denu point out that the serine/threonine specific phosphoprotein phosphatases (PP) use catalytic mechanisms very different from those of the PTPs. PPs use a “dinuclear metal ion center proposed to activate a water molecule for direct hydrolysis of the Ser/Thr phosphoester bond” (Jackson *et al.*, 2001). The PPs are categorized as PPP or PPM based

on homology, vulnerability to inhibitors, and metal requirements among other criteria.

PPPs have an identifiable consensus motif [DXH(X)<sub>23</sub>-<sub>26</sub>GDXXDR(X)<sub>20-26</sub>GNH(E/D)]. Some well known members of this class of phosphatases are PP1 and PP2B; X-ray crystal structures of PP1 (Goldberg *et al.*, 1995) and PP2B (Kissinger *et al.*, 1995) reveal some common structural features. The catalytic domains have two subdomains connected by one strand; each subdomain has a beta sheet and the two sheets are juxtaposed to make a double sheet. Many of the conserved residues of the PPP family consensus motif (Fig. 1-13) are found in intervening loops at the edge of the double sheet where they coordinate two metal ions: two Mn ions in the case of the PP1; a Fe and a Zn in the case of PP2B.

Figure 1-13 illustrates a proposed mechanism for the PPPs based on the PP2B structure. At the start of the cycle conserved residue Asp118 and Asn150 help coordinate the metals and a water molecule, perhaps a hydroxide ion, interacting with both metals. Binding of the phosphoenzyme orients the phosphoester for nucleophilic attack by the hydroxide ion. His151 acts as a general acid by providing a proton to the enzyme leaving group while the hydroxide attacks the phosphate. Phosphoenzyme

intermediates have not been observed for the PPPs thus water replacement of the phosphate ion regenerates the enzyme (Jackson *et al.*, 2001).

The crystal structure of the only representative PPM family member, PP2C, was determined by Das *et al.* (Das *et al.*, 1996). PP2C has some similarities with the PPP structures described above. However, the double stranded sheet seen in PP1 and PP2B now appears as more of a beta barrel in PP2C; the active site lies along the axis of the 'barrel' and above the opening. The loops between the beta strands of the barrel provide 4 conserved aspartates and a glutamate that partially coordinate two Mn ions. The two hexa-coordinated Mn ions are bridged by a water and an aspartate while four other waters hydrogen bond with a bound phosphate. The bound phosphate is also coordinated by both metal ions and an arginine appears to position the bound phosphate above the metal ions. Although little work has been done to delineate the mechanisms of members of the PPM family, Das *et al.* have proposed a catalytic mechanism similar to those of PP1 and PP2B (Das *et al.*, 1996).

Interestingly, of the three types of phosphatases known to dephosphorylate phospho-serine/threonine residues only the DS-PTPs have any sequence homology to HPrK/P. This homology is limited but there are obvious structural similarities between the VHR phosphate binding site and

some known structures of P-loops. Ironically, the more specific serine/threonine phosphatases, the PPs, have no structural features that resemble the P-loop found in HPrK/P.

## **Conclusion**

Histidine kinases and response regulators comprise well known bacterial signalling pathways that are in some respects analogous to the HPrK/P-HPr system. IDHK/P also displays a number of similarities with HPrK/P and is well established as a distinctly prokaryotic system of signalling.

Most protein kinases can perhaps trace their lineages to a common ancestor that resembles PKA; and bioinformatics has identified numerous candidate protein kinases in bacteria but fails to identify HPrK/P among them. Many kinases not related to the PKA family and not normally associated with ser/thr protein kinase activity (e.g., ras) do show ser/thr protein kinase activity in a signalling context and also have a nucleotide binding motif called the P-loop, which is found in HPrK/P.

Much of the work on phosphatases involves eukaryotic proteins but recent genomic searches have discovered that each of the five phosphatase classes may have members in some bacteria (Kennelly, 2002). Three of the five classes of protein phosphatases dephosphorylate proteins

phosphorylated on serine and threonine. The DS-PTPs can also dephosphorylate phospho-tyrosine. The DS-PTPs have a structural element that is similar to the P-loop, which is involved in binding of the phospho-tyrosine. The strictly ser/thr phosphatases of the PPP and PPM classes have some structural similarities to each other but not to the DS-PTPs; they bind two metals at the opening of a beta-barrel, which interacts with the phosphate

It is clear that *E. coli* and *B. subtilis* undergo significant metabolic changes in response to available carbon source. These changes have similar outcomes on the transport and utilization of sugars but they are accomplished with very different regulatory mechanisms. In *E. coli* cAMP and Crp perform the role of global transcriptional activators; operon specific repressors (e.g. LacI) repress transcription in the absence of inducer; and the intracellular availability of inducer is controlled by its exclusion according to a hierarchy of preferred carbon sources. In contrast *B. subtilis* do not make cAMP nor do they encode a Crp homolog, rather CcpA and P-Ser46-HPr (and P-Ser46-Crh) appear to act as global regulators; there is wider and more varied use of the PRD proteins to regulate the antitermination and activation of operons important for carbon metabolism. HPrK/P is central to *B. subtilis* CCR; phosphorylation of residue serine 46 of HPr and Crh adds a

further level of regulation of the PTS not apparent in *E. coli*; and these phosphorylations lead to transcriptional regulation of a large portion of the genes involved in carbon metabolism of *B. subtilis*.

With regard to *M. pneumoniae* it is not clear what role HPrK/P fulfills in the metabolism of these parasites. Perhaps the differences between *M. pneumoniae* HPrK/P and *B. subtilis* HPrK/P will provide some insight to the role of HPrK/P in these parasites. However, the sequence similarity between the two HPrK/Ps (33% identical/ 54% homologous) virtually guarantees that their overall fold and structure will be similar.

The studies described in this thesis are meant to begin the delineation of a key aspect of CCR, namely the structural mechanism of HPrK/P catalytic function.

**Figure 1-1** Comparison of the major mechanisms regulating carbohydrate uptake and metabolism in *E. coli* and *B. subtilis*. **A.** The stimulating effect of P~EII<sup>glc</sup> on adenylate cyclase activity in *E. coli*. **B.** Inducer exclusion in *E. coli*, inhibition of GlpK and non-PTS permeases by EII<sup>glc</sup>, which prevails over P~EII<sup>glc</sup> in the presence of a PTS sugar. **C.** P~His-HPr, predominant when PTS sugars are absent, phosphorylates and activates GlpK in *B. subtilis*. **D.** P~Ser46-HPr, produced by HPrK/P when FBP and ATP are present, binds to CcpA to exert CCR in *B. subtilis*; also inhibits non-PTS permeases in other low GC gram-positive bacteria. (Text and figure adapted from (Deutscher *et al.*, 2002)).



The diagram illustrates the cAMP-dependent transcriptional activation pathway in *E. coli*. It begins with the entry of Lactose into the cell via a permease. Lactose is then converted to Lactose-1-phosphate by the enzyme Active lactose permease. This step is coupled with the conversion of ATP to ADP. Lactose-1-phosphate is then converted to Glycerol 3-P by the enzyme Glycerol 3-P dehydrogenase. Glycerol 3-P enters the Glycolysis pathway. Simultaneously, Glycerol is converted to Glycerol 3-P by the enzyme Glycerol 3-P dehydrogenase. Glycerol 3-P is then converted to Glycerol by the enzyme Glycerol 3-P dehydrogenase. The resulting Glycerol is then converted to PEP by the enzyme PEP carboxylase. PEP is then converted to Pyruvate by the enzyme Pyruvate kinase. Pyruvate is then converted to P-EI by the enzyme P-EI. P-EI is then converted to P-HPF by the enzyme P-HPF. P-HPF is then converted to HPF by the enzyme HPF. HPF is then converted to EIAGlc (low concentration) by the enzyme EIAGlc. EIAGlc (low concentration) is then converted to EIAGlc (high concentration) by the enzyme EIAGlc. EIAGlc (high concentration) is then converted to Inactive adenylate cyclase by the enzyme Inactive adenylate cyclase. Inactive adenylate cyclase is then converted to Active adenylate cyclase by the enzyme Active adenylate cyclase. Active adenylate cyclase is then converted to 2 ATP by the enzyme 2 ATP. 2 ATP is then converted to cAMP by the enzyme cAMP. cAMP is then converted to Active CAP by the enzyme Active CAP. Active CAP is then converted to Transcriptional Activation by the enzyme Transcriptional Activation.

The diagram illustrates the catabolite repression mechanism in *E. coli*. It shows the cell membrane with transporters for Glycerol (GlpF), Lactose (Inactive lactose permease), and Glucose (IIIGlc). Inducer exclusion occurs when Glc-6-P is present. Glycerol enters the cell and is converted to P-HPPr, which is then converted to P-EHA/Glc. Lactose enters the cell and is converted to P-EHA/Glc. Glucose enters the cell and is converted to Glc-6-P, which enters glycolysis. P-EHA/Glc (low concentration) is converted to P-EHA/Glc (high concentration), which then inhibits the Inactive CAP protein. The Inactive CAP protein is shown binding to DNA, leading to the Absence of Transcriptional Activation and CATABOLITE REPRESSION.

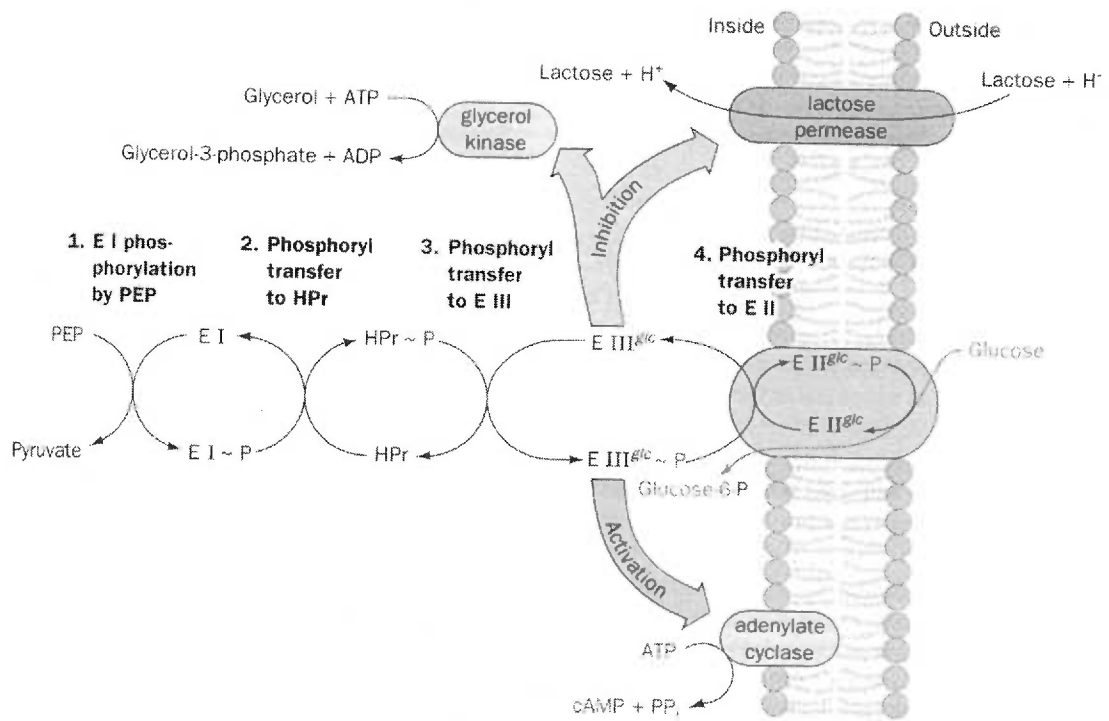
The diagram illustrates the metabolic regulation of the CcpA repressor. Key components and processes include:

- Cell Membrane:** Separates the "Out" (extracellular) and "In" (intracellular) environments.
- Transporters:**
  - GlpF:** A channel for Glycerol transport.
  - Active sugar permease:** A transporter for Sugar transport, which is also involved in the conversion of Sugar to Glycerol 3-P.
- Intracellular Metabolism:**
  - Glycerol:** Enters the cell via GlpF and is phosphorylated to **Glycerol 3-P** by **Glycerol kinase (GlpK)** using **ATP**.
  - Sugar:** Enters the cell via the active sugar permease and is converted to **Glycerol 3-P** using **ADP**.
  - Glycerol 3-P:** Enters the **Glycolysis** pathway.
- Regulatory Pathway:**
  - P-His-HPr (high concentration):** The active form of the repressor, shown in a dark box.
  - HPr kinase/Phosphatase:** A protein that converts **P-His-HPr** to **P-Ser-HPr (low concentration)** using **ATP** and releasing **ADP**.
  - P-Ser-HPr (low concentration):** The inactive form of the repressor, shown in a light box.
- Transcriptional Regulation:**
  - Inactive CcpA:** Represented by a hexagon.
  - Absence of Transcriptional Repression:** Indicated by a large arrow pointing to the right.

The diagram illustrates the catabolite repression mechanism in the *lac* operon. It shows the cell membrane with transporters for Glycerol, Lactose, and Glucose. Glycerol enters via an inducer exclusion mechanism involving Inactive GpK. Lactose enters via an Inactive sugar permease. Glucose enters via a transporter (II<sup>Glc</sup>) and is converted to Glc-6-P, which enters glycolysis to produce FBP. FBP is phosphorylated by HPr Kinase/phosphatase to P-Ser-HPr (high concentration). P-Ser-HPr inhibits Inactive CcpA, converting it to Active CcpA. Active CcpA binds to the *cre* site, leading to transcriptional repression. The diagram also shows the PEP to Pyruvate conversion and the EI to F-EI cycle.

**Figure 1-2** CcpA sequences from various gram-positive bacteria aligned with LacI/GalR family members from *E. coli*. Accession numbers: sp|P46828|CCPA - *Bacillus megaterium*; sp|P25144|CCPA - *Bacillus subtilis*; sp|Q56194|CCPA - *Staphylococcus xylosus*; tr|Q48518 CCPA - *Lactobacillus casei*; tr|O87746 - *Enterococcus faecalis*; tr|Q99TC8 CcpA - *Staphylococcus aureus*; tr|Q8Y6T3 CcpA - *Listeria monocytogenes*; sp|P06964|CYTR - *Escherichia coli*; sp|P03024|GALR - *Escherichia coli*; sp|P03023|LACI - *Escherichia coli*; sp|P15039|PURR - *Escherichia coli*. (Prepared with ClustalW and Boxshade).





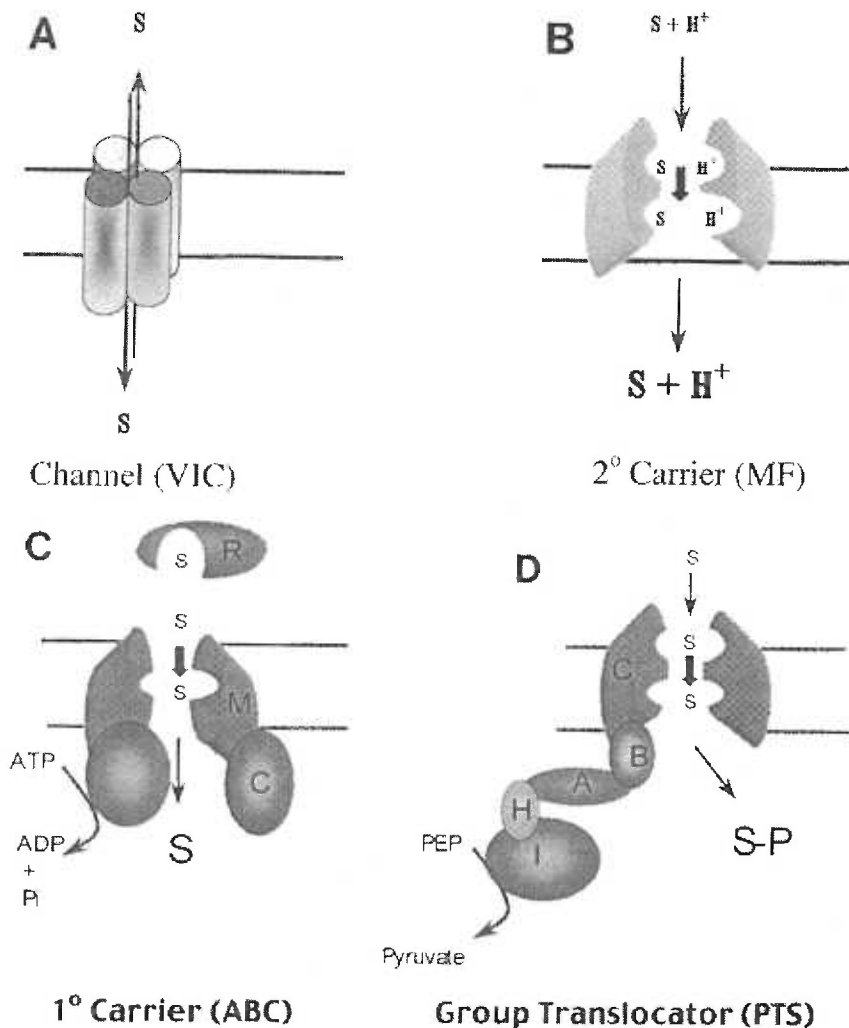
**Figure 1-3** Schematic view of the PTS from *E. coli*. [Adapted from (Voet & Voet 2000)]

B_subtilis	1	--AQKTFKVTADSGIHARPATVVLVQTASKYDADVNLEYN-GKTVNLKSLMGVMSLGIKAG
Crh_B_subtilis	1	-MVQQKVEVRLKTGLQARPAALVQEAANRFTSDVFLEKD-GKKVNAKSLMGVMSLAVSTG
B_megaterium	1	-MAQKTFTVTADSGIHARPATVVLVQAASKFSDINLEFN-GKTVNLKSLMGVMSLGIQKG
M_pneumoniae	1	-MKKIQVVVKDPVGIHARPASIIAGEANKFKSELKLVSPSGVEGNIKSIINLMSLGIKQN
E_faecalis	1	-MEKKFPHVAETGIHARPATVVLVQTASKFNSDINLEYK-GKSVNLKSLMGVMSLGVGQG
L_casei	1	-MEKREFNIIAETGIHARPATVVLVQAASKFNSDINLEYK-GKSVNLKSLMGVMSLGVGQG
S_aureus	1	-MEQNSYVITIDETGIHARPATVVLVQTASKFSDIQLEYN-GKSVNLKSLMGVMSLGVGQG
N_meningitidis	1	-MLKQSIETINKLGLHARASNKFTQTASQFKSEVWVTKNDSRVNGKSLMGVMSLGLAAAKG
T_pallidum	1	-MVVKTVRVLNRAGVHARPAALIVQAASREDSKIMLVRD-TIRVNAKSLMGVMSLGLAAAGCG
E_coli	1	-MFQQEVTITAPNGLHTRPAAQFVKEAKGSETSEITVTSN-GKSASAKSLFKLQTLGLTQG
NPr_E_coli	1	MTVKQTVETINKLGMHARPAMKLFELMQGSDAEVLLRNDEGTEAEANSVIALLMIDSAKG


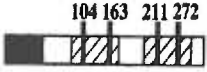

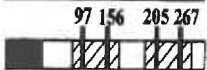
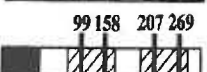

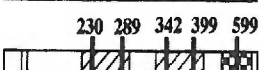
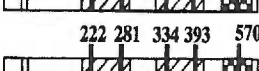
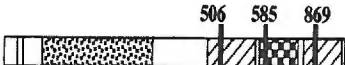
  

B_subtilis	58	AEITISFSGADENDALNALBETMKSEGLGE-
Crh_B_subtilis	59	TEVTLIAQGEDEQEAEKLAAYVQEEV----
B_megaterium	59	ATITISAEGSDEADALAALEDTMSKEGLGE-
M_pneumoniae	60	DHITIKAEGTDEEEALNAIKAVLEKHQVI--
E_faecalis	59	SDVTITVDGADEAEGMAAIVETLQKEGLAE-
L_casei	59	ADVITISAEGADEADATAAITDTMKKEGLAE-
S_aureus	59	AEITIVADGSDESDAIQAISDVLSKEGLTK-
N_meningitidis	59	TVIELETDGADEAEAMRALTDLINGYFGEGE
T_pallidum	59	SELELVVEGPDVAALSAIERLFQNKFEED-
E_coli	59	TVVTISAEGEDEQKAVEHLVKLMAELE----
NPr_E_coli	61	RQHEVEATGPQEEALAAVIALFNSGFDED-

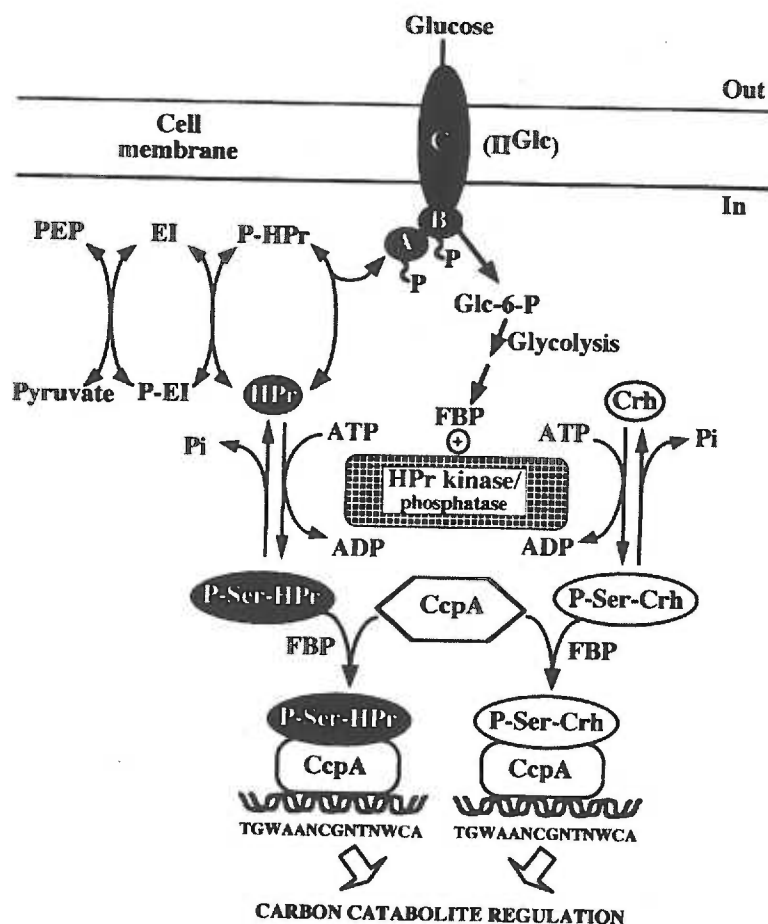
**Figure 1-4** Alignment of HPr sequences from organisms that encode HPrK/P with Crh from *B. subtilis* and HPr & NPr from *E. coli*. Accession numbers: sp|P08877| HPr - *Bacillus subtilis*; tr|O06976 CRH - *Bacillus subtilis*; sp|O69250| HPr - *Bacillus megaterium*; sp|P75061| HPr - *Mycoplasma pneumoniae*; sp|P07515| HPr - *Enterococcus faecalis*; sp|Q9KJV3| HPr - *Lactobacillus casei*; sp|P02907| HPr - *Staphylococcus aureus*; sp|Q9JQN1| HPr - *Neisseria meningitidis*; sp|O83598| HPr - *Treponema pallidum*; sp|P07006| HPr - *Escherichia coli*; sp|P33996| NPr (Nitrogen related HPr) - *Escherichia coli* (prepared with ClustalW and Boxshade).



**Figure 1-5** Schematic depiction of the four major types of transporters found in living organisms. **A.** Channels (usually oligomeric) a voltage gated ion channel (VIC) is depicted. **B.** Secondary carriers (usually mono- or dimeric) A major facilitator (MF) superfamily porter with solute-proton symport is depicted. **C.** Primary active transporters An ATP binding cassette (ABC) uptake permease is shown here. **D.** Group translocators A PTS family member is represented. The functionally homo-dimeric transporter EIIC is energized by a series of phosphoryl transfer reactions in the other enzymes of the PTS (EI, HPr, and EIIA) to phosphorylate the sugar upon import. (Text and figure adapted from (Saier *et al.*, 2002))

Protein		Number of residues	Domain organisation and conserved histidine residues	Phosphorylation or regulation by HPr	Negative regulation by the PTS
<b>Metabolic enzyme</b>	<b>Enzyme</b>				
GlpK	Glycerol kinase	496		Yes (SDM and ivP)	No
<b>Antiterminator</b>	<b>Controlled operon</b>				
GlcT	<i>ptsGHI</i>	285		Yes (SDM)	Yes IICBA (PtsG)
LicT	<i>bglPH, licS</i>	277		Yes (SDM and ivP)	Yes IICBA (BglP)
SacT	<i>sacPA, sacXY</i>	276		Yes (SDM)	Yes IIBC (SacP)
SacY	<i>sacB, sacXY</i>	280		Yes (ivP)	Yes IIBC (SacX)
<b>Transcriptional activator</b>					
<b>DeoR family</b>					
LicR	<i>licBCAH</i>	641		Yes (SDM)	Yes IIA, IIB (LicA, LicB)
YdaA (MtlR)	<i>mtlAD ?</i>	694		?	MtlA ?
YjdC (ManR)	<i>yjdDEF ?</i>	648		?	YjdD?
<b>NifA/NtrC family</b>					
LevR	<i>levDEFG sacC</i>	938		Yes (SDM and ivP)	Yes IIA, IIB (LevD, LevE)

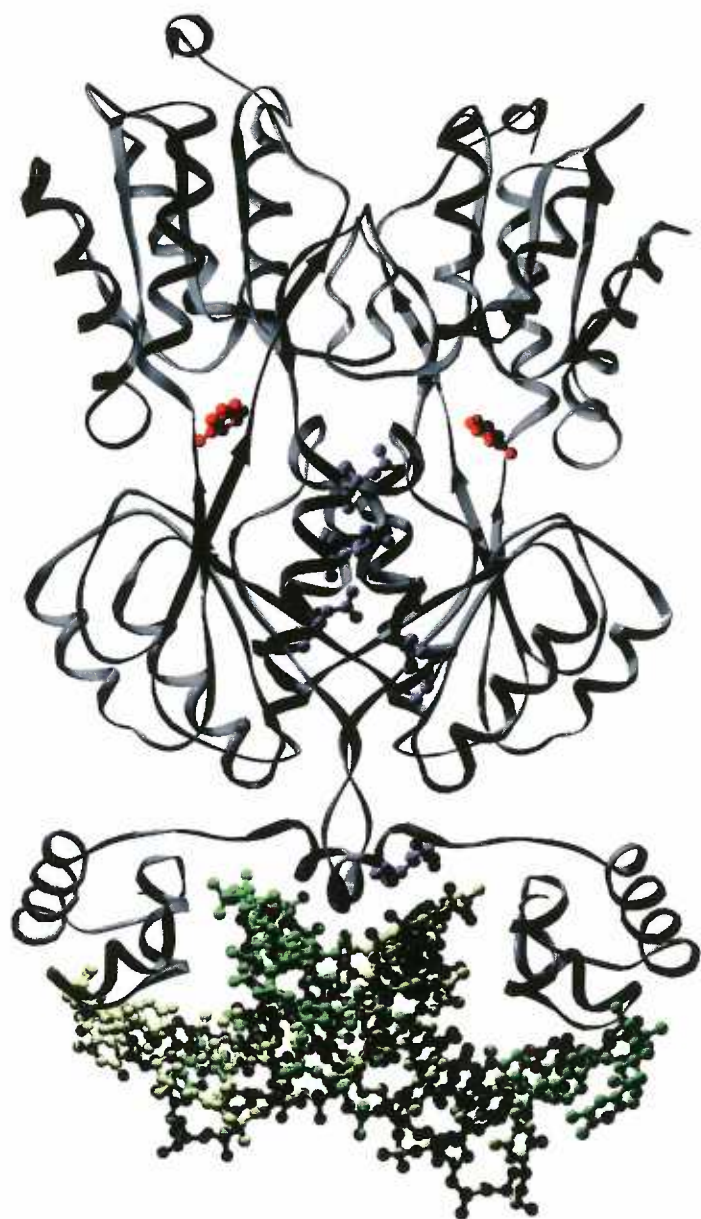
**Figure 1-6** *B. subtilis* non-PTS proteins phosphorylated and regulated by PTS proteins. Presented are GlpK and PRD containing transcriptional regulators; with a black box for the RNA binding domain of antiterminators; or a small white box for an N-terminal DNA binding motif. Bold bars indicate conserved histidyl phosphorylation sites in PRDs (striped box) or EIIA-like domains (checkered box). Phosphorylation sites determined by site directed mutagenesis (SDM) or *in vitro* phosphorylation experiments (ivP). SacY and GlcT are phosphorylated by P~His-HPr but are active in absence of functional HPr. All PRD regulators are probably negatively controlled by P~EIIB mediated phosphorylation; the corresponding EIIB is listed. (Text and figure adapted from (Deutscher *et al.*, 2002))



**Figure 1-7** The major mechanism of CCR in *B. subtilis*. The uptake of PTS sugars leads to an increase in the FBP concentration in the cell, which stimulates the ATP dependent HPrK/P catalyzed phosphorylation of HPr and Crh at serine 46. Only the seryl-phosphorylated forms of HPr and Crh are capable of binding to CcpA, an interaction possibly stimulated by FBP. The CcpA complexes can then bind to CRE sites found in catabolite regulated operons. Similar mechanisms are probably operative in most other gram-positive bacteria, however Crh has been detected only in bacilli. (Text and figure adapted from (Deutscher *et al.*, 2002))



**Figure 1-8** Model of CcpA from *B. megaterium*. The model is based on the PurR dimer from *E. coli* bound to duplex DNA, which is shown in green and gold (Schumacher *et al.*, 1995). Mutations of the residues shown in purple (on monomer A) abrogate binding of p-Ser46-HPr to *B. megaterium* CcpA(Kraus *et al.*, 1998); these residues suggest the location and extent of the surface of CcpA contacted by p-Ser46-HPr. Hypoxanthine molecules, shown in red, induce PurR to bind specifically to the *purF* promoter; small molecule inducer may bind in a homologous site on CcpA in addition to P-Ser46-HPr or as part of a separate mechanism. (prepared with SwissPDB and Povray)



```

M_pneumoniae 1 -MKKELVKELEIQFQDCVNLIDGHTNTSN--VTRVPGLRKRVFEMGLGFSSQIG-SVAI
B_subtilis 1 --AKVRTKDVMEQFN--LELISG-EGEIN--RPTMSDLRPGIETAGYFTYVPRRVOL
S_xylosus 1 ---MLTKSLVERFE--LEMIAG-EAGLN--KQKNTDLSRPGLEAGYFHHYASDRIQL
L_casei 1 MADSVTVRQLVKATK--LEVYSGEYLDQ--RQVVLSDLSRPGLETTGYFNYPHERIQL
E_faecalis 1 -MEVVKIYQLVENLS--LEVYVDEESLN--RTTKTGETSRPGLETTGYFNYPHSHDRQL
M_genitalium. 1 -MKHLTVKALVLQFNDCTQLIDGKNNIDN--VITPGLKRSVFELLGLFCKPIG-SVAI
S_aureus 1 --MLTTEKLVETLK--LDLIAG-EGEIS--KPIKNADLSRPGLEAGYFHHYASDRIQL
L_monocytogenes 1 MTKSVTVKDLKERLN--LELIC-ETGLE--RPTSTDLRPGLETTGFFSYYPEDRVOL
T_pallidum 1 -MLKLDLKE-RDSDL--LRCLAG-HHGLA--NPHTISDLNRPGLVLSGFDFVAYRIQL
N_meningitidis 1 -MPSISVRRFDDNQ--YKQLAWAAGNSGADNRIGVEADKPVLAIVGHLNFIHPNQIQV

M_pneumoniae 56 LGKRFFGFLSQKTLVQQQILHNLKLNPPAILTKSFTDPTVLLQVNQTYQVPIKTD
B_subtilis 54 LGKTELSFFEQPPEEKKQRMDSLCTDVTPAILSRDMPIPOETIDASEKNGVPVLRSP
S_xylosus 53 LGTTELSFYNLPPDEERKGRMRKLCRPETPAIVTRDEPPEELIEAAKEHETPLITSKI
L_casei 57 FGRTEISFARNMSSEERLLILKRMATEDTPAFIVSRGLEAPAEITATAAHIPVLSRL
E_faecalis 56 FGSKEITFAERMMPERLLVMRRLCAKDTPAFIVSRGLEPPEELITAAKENGVSVDSP
M_genitalium. 56 LGKREFIFLNQKPVQQKKIIANLLKLPKPAVILTKSFLDCGVLLAVNQTYQVPIKTN
S_aureus 53 LGTTELSFYNLPPDKDRAGMRKLCRPETPAIVTRGTPPEELVEAAKELNTPPIVAKD
L_monocytogenes 56 FGMTSEISEGMEPBERLKRYKQMCTKRTPAFVISRNLEVPKELVAAAKADPIVRSRL
T_pallidum 54 FGRGEHAYLLALLEQGRYGAIEKMTFDLPCCIFSHGTPPEKFLHLAEPSSCPILVTR
N_meningitidis 58 VGLAESEYLNRLLESCTGYQPGDFDLSMSLVVANGLPVSPGRDYCHKNDIPILTSKL

M_pneumoniae 116 FSTELSFVTETVINEQFATVAQIHGVLLEVFVGVLITGRSGIGKSECALDLINKNHLFV
B_subtilis 114 KTELSSRRTNFLESRLAPTTAIGHVLVDIYGVGVLTGKSGVKGKSETALELVKRGHRLV
S_xylosus 113 ATTQLMSRLTTPLEHLEARITSLHGVLVDVYGVGVLTGDSGIGKSETALELVKRGHRLV
L_casei 117 PTRLSSLTTEYLDSQLAERRSMHGVLDVYGVGVLTGDSGVKGKSETALELVKRGHRLV
E_faecalis 116 STERLLGELSSYLDGRLAERTSVHGVLDVYGVGVLTGDSGIGKSETALELVKRGHRLV
M_genitalium. 116 FSTELSFVTETVINEQFATVQKLVHGVLEIFGVGVLECKSGIGKSECALDLINKNHLFV
S_aureus 113 ATTLMSRLTTPLEHALAKTTSLHGVLVDVYGVGVLTGDSGIGKSETALELVKRGHRLV
L_monocytogenes 116 KTRLSSVITNYLESRLAPVISMHGVLDVYGVGVLTGDSGIGKSETALELVKRGHRLV
T_pallidum 114 TSSELSLRLMRVLSNIFAPTIALHGVLEVEYGVGLISGDSGVKGKSETALELVKRGHRLV
N_meningitidis 118 ESPYLMVDVLRITLQRTLAASSVKHGVFLDVFEIGVLITGHSGLGKSETALELVKRGHRLV

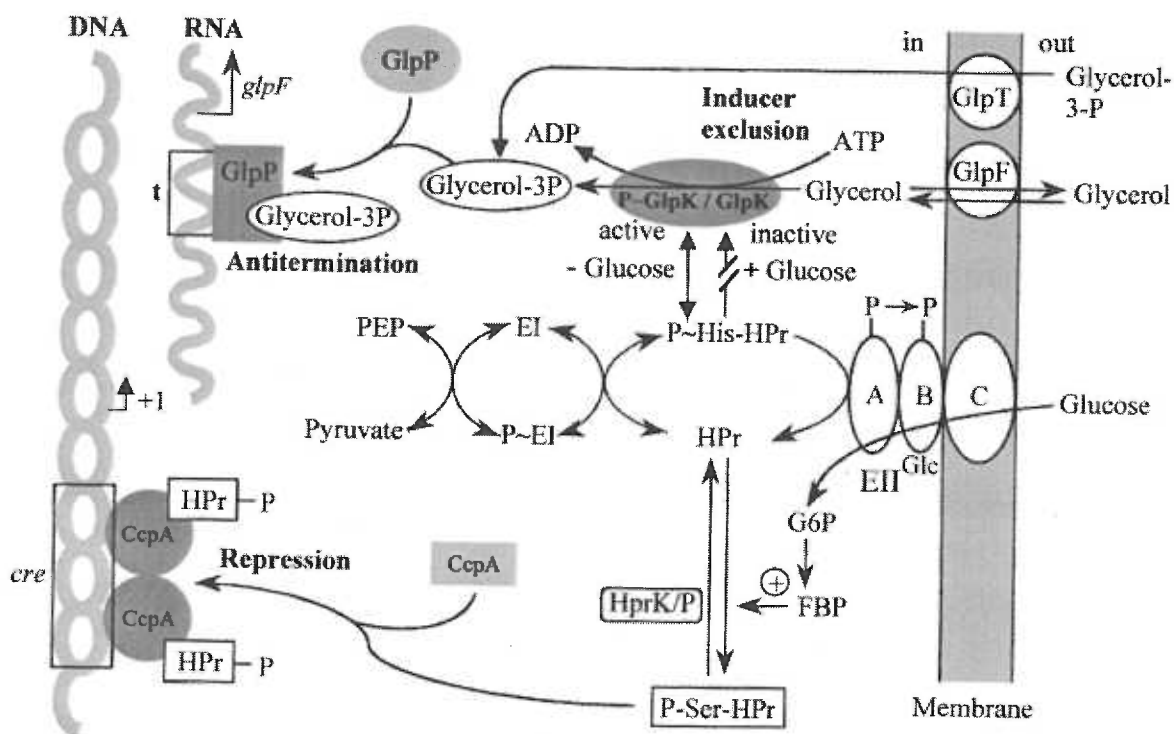
M_pneumoniae 176 GDDAIEIYRLG-NRLFGR-AQEVAKKFMERGLGIINVERFYGLQITKORTEQLMVNVL
B_subtilis 174 ADDCVETIRQEDQDTLVGN-APELTEHLLERGLGIINVMTLFGAGAVRSNKRRTIVMNL
S_xylosus 173 ADDNVEIREISKDELIGR-APKLTEHLLERGLGIINVMTLFGAGSILTEKRLRLNILE
L_casei 177 ADDRVDVYQDEQTVGA-APPILSHLEIRGLGIIDVMNLFAGAGAVREDDTSLIVHLE
E_faecalis 176 ADDRVDVYQDELTIVGE-PPKILQHLERIGLGIIDVMNLFAGASAVRGFMQVQLVYLE
M_genitalium. 176 GDDAIEIYRLG-NRLFGR-AQALAKGFMEIRGLGIINIERAYGLQITKEQTEQLAISIL
S_aureus 173 ADDNVEIROINKDELIGR-PPKLEHLLERGLGIINVMTLFGAGSILTEKRLRLNINLE
L_monocytogenes 176 ADDNVEIROEDEMTLIGS-SPATIEHLLERGLGIINVMTLFGAGAVRSSKKTIVVHLE
T_pallidum 174 ADDLVEISCVNGNSLIGRVHKSIGHHMEIRGLGIINITQLYGVGSIRERKELQMVVLE
N_meningitidis 178 ADDAVELEFRIGPETLEGR-CSPMLRDFLEVRGLGIINIRHTFGETSIRPKKILQLIINLV

M_pneumoniae 234 SLEKQTTVTFERLGTELKQRLGVDLSFYETPISFGRKTSIEIESAVIDEFLKHSYGNS
B_subtilis 233 LWEQG--KOYDRGLGLEETMKIETDEITKLTIPVRPGRNLAVIIEVAAMNRLKRMGINA
S_xylosus 232 NWHKE--KLYDRVGLNEETLRLDTEITKLTIPVRPGRNLAVIIEVAAMNRLNIMGINA
L_casei 236 NWTPD--KTDRGLGSGEQTLQIFDVPVPKLTVPVKVGRNLAVIIEVAAMNFRKSMGYDA
E_faecalis 235 AWEKD--KKYDRGLGSDAMVELANVDVPQIRIPVKTRGRNLAVIIEVAAMNFRKSMGYDA
M_genitalium. 234 SLEEKNNASFERLGSCLKLKNLGVKISYYQIPSSGRKTSIEIESAVIDEFLKHSYGNS
S_aureus 232 NWNKO--KLYDRVGLNEETLSLDTEITKLTIPVRPGRNLAVIIEVAAMNRLNIMGINA
L_monocytogenes 235 NWDPD--KHYDRVGLDQEKTKIFDMDIPKLTIPVRPGRNLAVIIEVAAMNRLKRMGYNS
T_pallidum 234 EWNSS--KAYDRGLGTOELNTLTDVSVPLIEIPVRPGRNIPITIDETAMNRLKRMGYNS
N_meningitidis 237 EADDEYMKQLDRLSIRTETESLLNVNRSVTLFVAVGRNLAVLVEAAVRNVLQLRGKDS

M_pneumoniae 294 ALDEIENQKAILKRR-----KDES--
B_subtilis 291 AEQSTNKIADVIEDG-----EQEE--
S_xylosus 290 AEEFNDRNLNABILRNG--NNGNNGEEK--
L_casei 294 TKTEKKNLNLHLEHNEETDQNSSGDK--
E_faecalis 293 TKTEERLTRLIE-----NSGE--
M_genitalium. 294 ANEEILKQRAMLEEQ-----TDE--
S_aureus 290 AEEFSERLNEETIKN-----SHKSE--
L_monocytogenes 293 AEQSTQDLNNLIGHN-----SSMND--
T_pallidum 292 AKEFNQSVLKLMEONAAHAPYRPDDTY
N_meningitidis 297 TREELERHQTQLKENE----QHNEDRPD

```

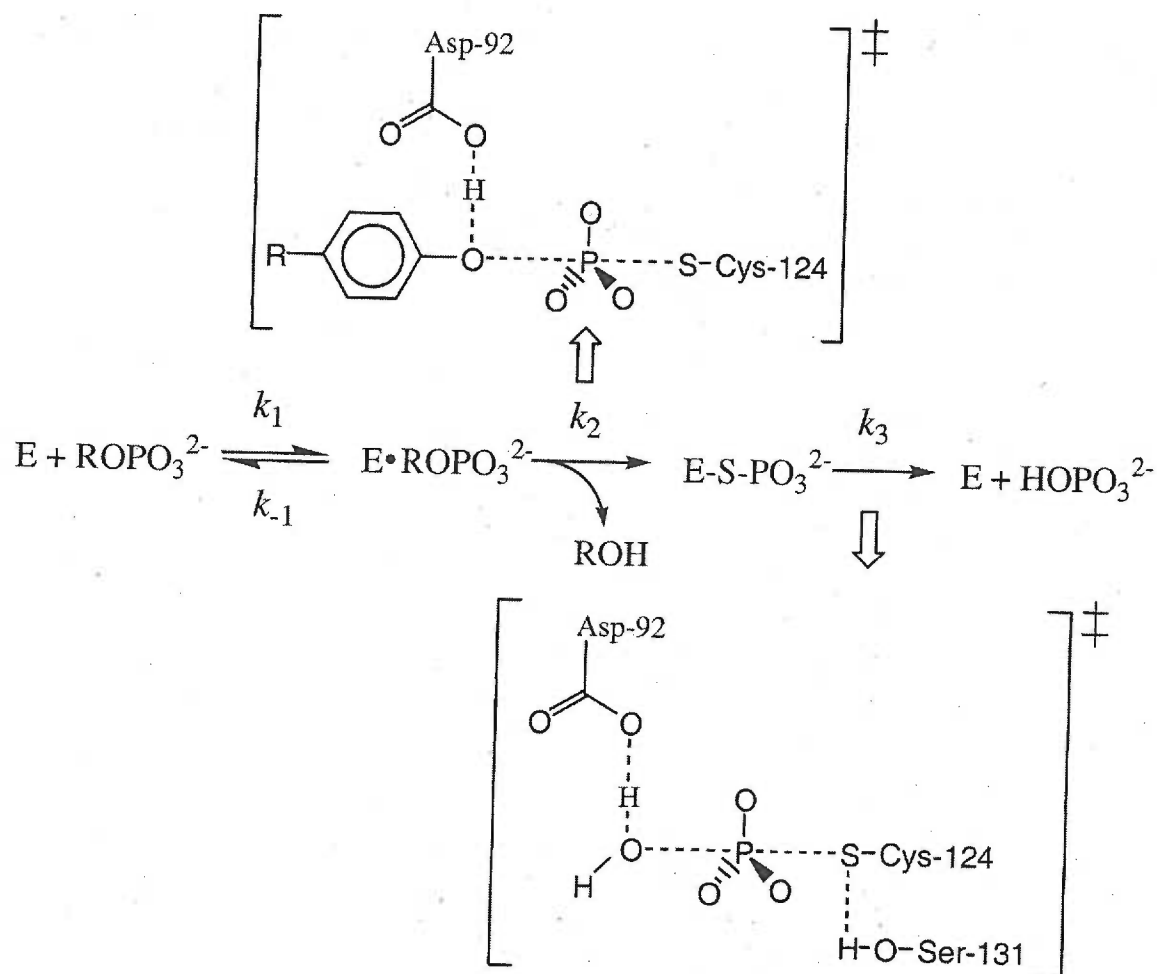
Figure 1-9 Alignment of HPrK/P from various organisms (prepared with ClustalW and Boxshade).



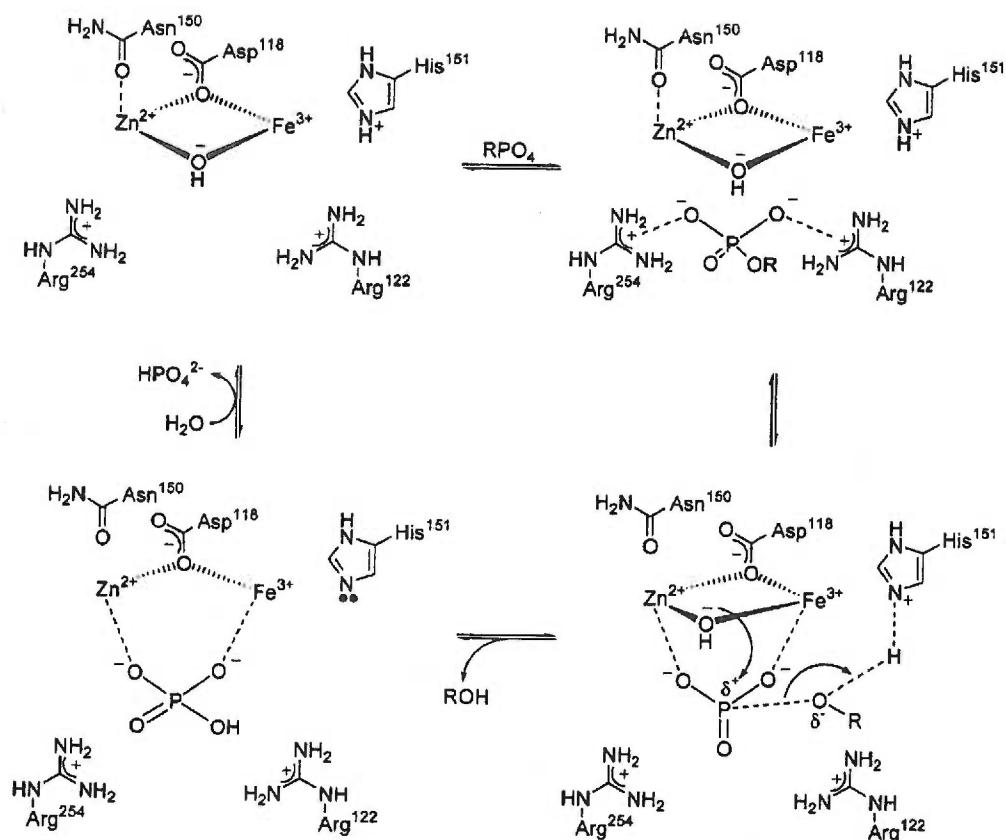
**Figure 1-10** Proposed mechanisms regulating expression of the *B. subtilis* *glpFK* operon at the DNA and RNA level. Similar to P~Ser46-HPr, P~Ser46-Crh can also interact with CcpA to effect CCR of the *glpFK* operon. (Text and figure adapted from (Darbon *et al.*, 2002))

**Figure 1-11.** Proposed catalytic mechanism for PKA. **A.** Schematic showing the hydrogen bonding pattern of PKA,  $\text{AlF}_3$ , ADP,  $\text{Mg}^{2+}$ , solvent, and a serine on P-site, a peptide substrate. The  $\text{AlF}_3$  is supposed to represent a static picture of the planar phosphate intermediate (Figure adapted from (Madhusudan *et al.*, 2002)). **B.** The former proposed function of Asp166 as the catalytic base is uncertain; it may simply orient the hydroxyl group of the substrate. Lys168, on the other hand, appears to be important in stabilizing a developing charge on the phosphate ion (Figure adapted from (Zhou *et al.*, 1997)).





**Figure 1-12.** Proposed VHR catalytic mechanism. In the production of the thiophosphate enzyme intermediate residue Asp92 acts as the general acid, contributing a proton to the tyrosine while Cys124 makes a nucleophilic attack on the phosphotyrosine (or serine/threonine). Residue Asp92 also acts as the general base, activating water to attack the thiophosphate, for the hydrolysis of the phosphoester. Conserved residue Ser131 provides a hydrogen bond to the sulfur of Cys124, which may stabilize the developing negative charge on the sulfur. (Figure adapted from (Kim *et al.*, 2001))



**Figure 1-13.** Catalytic mechanism of the PPP family member, PP2B. Proposed mechanism based on PP2B numbering. The metal ions coordinate a hydroxide ion thought to be the nucleophile, which attacks the bound phosphorylated protein substrate. His151 donates a proton to the developing negative charge on the side chain oxygen. (Figure adapted from (Jackson *et al.*, 2001))



with plasmid pGP204 in DH5 $\alpha$  were grown in M9 minimal medium (128g/L Na<sub>2</sub>HPO<sub>4</sub>\*7H<sub>2</sub>O, 30g/L KH<sub>2</sub>PO<sub>4</sub>, 10g/L NH<sub>4</sub>Cl, 5g/L NaCl) supplemented with: D, L-selenomethionine (100 mg/L), the other 19 amino acids (100 mg/L), thiamine (3 mg/L), biotin (3 mg/L), guanosine (1 mg/L), thymine (1 mg/L), uracil (1 mg/L), adenine (1 mg/L), MgCl<sub>2</sub> (0.5 mM), glucose (50 mM), and ampicillin (100 mg/L). Expression of the recombinant selenomethionyl protein was induced by the addition of IPTG (1 mM) to logarithmically growing cultures (OD<sub>600</sub> of 0.6) at 37 °C. Cells were harvested at 2hrs. The selenomethionine substituted protein, used for MAD experiment, was purified as described above except for the addition of 15 mM  $\beta$ -mercaptoethanol to the purification buffer throughout purification. Mass spectroscopy confirmed ~100% substitution of methionine with selenomethionine in purified HPrK/P. The final yield of the selenomethionine substituted HPrK/P was ~8 mg/L. Native HPrK/P was prepared in Luria Broth (5gm/L yeast extract, 10gm/L NaCl, and 10gm/L pancreatic digest of casein) using the same overproduction protocol as described above.

Harvested cells were resuspended in running buffer: 50 mM Tris/HCl pH 7.5, 10 mM  $\beta$ -mercaptoethanol (15 mM for selenomethionyl HPrK/P), and 0.6 M NaCl. The cell suspension was lysed by sonication on ice. The

clarified lysate was run over a nickel chelate resin (1 mL columns of prepacked Pharmacia HiTrap<sup>TM</sup> Chelating HP; Iminodiacetic acid functional group – 7 carbon linker- crosslinked agarose; charged with 3 mL of 50 mM NiSO<sub>4</sub>) and washed extensively with running buffer. HPrK/P was then eluted with a gradient of imidazole (0-500 mM) in running buffer. In order to remove the imidazole the collected fractions were dialyzed against running buffer. The dialyzate was then concentrated to 15 mg/mL, as determined by absorbance ( $\epsilon_{275} = 1.15 \text{ mL/mg}\cdot\text{cm}$ ), using Amicon filters with a 30 kilodalton cutoff.

### **Protein Crystals**

Crystals of HPrK/P were grown using the hanging drop method. Briefly, the hanging drop method subjects the protein to buffer conditions approximating the desired endpoint; the resultant solution is sealed in a single well of a 48 well plate in vapor contact with a reservoir of the endpoint buffer. The protein solution reaches vapor equilibrium with the reservoir over the course of several days.

Theoretically we can say very little about the crystallization process other than that is assumed that the crystalline state represents a free energy minimum for the protein. The protein solution at a general state of temperature, pressure, and buffer conditions is assumed to have a

concentration of maximum solubility  $x$ ; if the system is not at equilibrium it can attain conditions of supersaturation transiently greater than  $x$ , where protein will precipitate in disordered aggregates or microcrystals depending upon the system. To obtain the largest crystals we attempt to slowly reduce the solubility of the protein over the course of the experiment such that supersaturation is infinitesimal; and, as the crystals grow and reduce the concentration of protein, infinitesimal supersaturation is maintained.

Practically, the following considerations have been shown to be important for crystallizing some proteins: 1) monodispersity, or a diffusion coefficient of small variation, is indicative of high solubility and minimal aggregation (Ferre-D'Amare *et al.*, 1994); 2) 'pure' protein generally increases the chance that crystals obtained will contain the protein of interest; 3) flexible regions, such as large linker peptides, increase overall conformational entropy disfavoring the ordered state of a crystal.

We used the screen method of crystallization, which seeks to widely vary pH, salt, precipitating agent, temperature, the protein concentration in an attempt to find a combination which will produce some crystals (Ollis *et al.*, 1990; Lindwall *et al.*, 2000). Once crystals are obtained parameters are varied in a systematic fashion to obtain larger crystals of data quality.

HPrK/P from *M. pneumoniae* crystallized in a hanging drop with reservoir conditions of 20 mM MgCl<sub>2</sub>, 50 mM Tris-HCl, pH 7.6, 4% PEG-8,000, and 0.8 M NaCl. However, to allow for inevitable variations in the quality and concentration of each purification experiment, the precipitant and protein concentration were varied over a small range in each hanging drop experiment (2-5% PEG 8,000 and 4-8 OD<sub>275</sub>). Freshly purified HPrK/P [ $< 2$  freeze-thaw cycles (-80 C) or  $\sim 72$ hrs] at  $\sim 15$  OD<sub>275</sub> was diluted to the appropriate concentration with 50 mM Tris-HCl pH 7.6, and 0.6 M NaCl and brought to 1 mM ATP analogue {adenosine 5'-[ $\beta,\gamma$ -nitro] triphosphate (AMPPNP)}. Single drops of 5  $\mu$ L of protein-ANP dilution were placed on silanized cover slips and 5  $\mu$ L of reservoir solution added. The coverslips were then sealed to the corresponding reservoir. Crystals appear in 1-2 weeks and grow to 0.5 X 0.5 X 1.0 mm.

### **Lattices, Unit Cells, Symmetry, and Space Groups**

Webster defines a lattice as a regular geometrical arrangement of points or objects over an area or in space; *specifically* it is the arrangement of atoms [proteins] in a crystal. We can provide a mathematical description of (protein) crystals in terms of their 3-dimensional coordinates. The number of coordinate descriptions of the theoretically possible geometries of lattices

(Euclidean) can be narrowed to seven: Triclinic, Monoclinic, Orthorhombic, Tetragonal, Trigonal, Hexagonal, and Cubic (Fig. 2-1).

We define a unit cell as the smallest repeating unit of the crystal along one of the lattice basis vectors - one cell edge is the repeat distance directly along one of the 3 basis vectors. These crystal systems may be envisioned as having one lattice point per unit cell ( $1/8$  of a point at each corner), which are defined as primitive lattices. Non-primitive lattices have 2 or more lattice points per unit cell, e.g., 1 point in the center or half a point in each of two parallel faces. The number of primitive and non-primitive lattices is also theoretically limited; these are the 14 Bravais lattices shown in Figure 2-2.

Symmetry is defined as a rigid motion of a geometric figure that determines a one to one mapping onto itself. The only types of symmetry possible for proteins are 2, 3, 4, and 6 fold screw axes and simple rotation axes, due to the chiral nature of biological macromolecules; mirror symmetry and inversion centers are disallowed. Where a screw axes is a rotation and translation parallel to a basis vector for the crystal system, e.g., a  $3_1$  screw axes is generated by a 120 degree rotation with each  $1/3$  unit cell translation along a cell edge (Fig 2-3). These symmetry elements, (screw and rotation axes) when combined with the Bravais lattices, produce the allowed

space groups, which are also theoretically limited. Therefore, a space group is the set of finite elements  $\{n\}$  closed [equivalent to a general point  $(x, y, z)$ ] under the symmetry operations of rotation and screw axes in a lattice of a given Bravais type.

Real macro-crystals are usually not perfectly geometric but are made up of small regions of perfect micro-crystals called grains (McPherson, 1999). When the same crystal planes in adjacent grains are slightly misaligned with each other we say that the crystal is a mosaic and this leads to reflection broadening, which is preferable since a perfect crystal (e.g. diamond) would lead to secondary extinction or multiple scattering of rays reducing the measured intensity. Each grain gives reflections of small angular width because it is not an infinite lattice, which would give a point reflection (Drenth, 1995). The total intensity of a reflection from a mosaic crystal is just the sum of reflections from each grain since the phase relationships between them are random in general (James, 1962). Thus a sum of reflections of finite width from grains of a mosaic crystal leads to measured reflections of finite width.

HPrK/P crystallizes in the  $P2_12_12_1$  space group; this is a primitive lattice with 2-fold screw axes parallel to each principal axis of the unit cell. A 2-fold (and no higher) symmetry axes about each principal axes is only

consistent with the Orthorhombic crystal system. The 2-fold screw along the **a** axes requires  $(x,y,z) = (x+1/2, -y, -z)$ . And the same is true for **b**  $(x,y,z) = (-x, y+1/2, -z)$  and **c**  $(x,y,z) = (-x, -y, z+1/2)$ . For these crystals there are four equivalent divisions of the unit cell called asymmetric units. Where an asymmetric unit is defined as the unique portion of the unit cell from which the remaining portions of the unit cell may be reconstructed by symmetry operations.

### **Data Collection and Processing**

Even before synchrotron radiation was widely used in macromolecular crystallography radiation damage was a significant problem for crystal structure analysis. The high photon flux at a typical synchrotron beam line today indicates some method of cooling (Burmeister, 2000; Weik *et al.*, 2000) - to collect MAD data sets requires cryocooling. Usually this means vapor phase of liquid nitrogen (~100 Kelvin). Unless cryo-protectants are used most mother liquors will freeze and ice rings would dominate the diffraction pattern at ~3.7 angstroms. Cryo-protectants preclude the formation of ice structure in or near the crystal, forming a glass instead. At concentrations in excess of 20% in the mother liquor glycerol, ethylene glycol, polyethylene glycols (of < 400 dalton), and sugars have been

successful cryo-protectants. Also highly soluble salts may be used at high concentration (Rubinson *et al.*, 2000).

Mounting a crystal for collecting data at 100 K is very simple. Nylon (10  $\mu\text{m}$ ) loops of  $\sim 0.5$  mm diameter are mounted on pins set in a magnetic base that locks to the goniometer head. The HPrK/P crystals (see Chap. III) were scooped out of the drops they grew in with the nylon loops and serially transferred to a final cryoprotectant containing 20 mM  $\text{MgCl}_2$ , 50 mM Tris/HCl pH7.6, 0.2 M NaCl, and 30% glycerol. The whole assembly is quickly placed on the goniometer head such that the loop is completely immersed in a stream of nitrogen gas at 100 K. The crystal is then aligned in the x-ray beam and indexing and data collection proceeds.

It is helpful to classify the type of space group a crystal takes as early as possible in a crystal structure analysis as greater symmetry can simplify all aspects of the X-ray crystal structure determination. If the crystal diffracts well the determination of the crystal system, called indexing, is usually straightforward and requires only one diffraction image (Steller *et al.*, 1997). Indexing is performed for each crystal before data collection to identify the best data collection strategy for the crystal given its crystal system and orientation with respect to the X-ray beam.



In most cases data collection consists of rotating the crystal about one axis ( $\phi$ ) under X-ray irradiation. The beginning and ending points of the rotation have been determined, by the determination of the best strategy, to collect the theoretically greatest possible data, which may not be 100% if the crystal orientation is unfavorable. This data is obtained by rotating the crystal through  $\sim 1^\circ$  wedges during the course of an exposure lasting on the order of minutes. After the exposure the charge coupled device (CCD), located at some distance from the crystal, is read, returning dark current subtracted intensity data for each (0.1mm) pixel.

The purpose of processing the data is to reconstruct a 3-dimensional lattice composed of 'reflections' from the set of wedges (oscillations). However, the wedges are not simply added together due to the following problems among others: radiation decay between the first and later appearances of a given reflection; the amount of time a crystal experiences diffraction for different reflections will vary; absorption of diffracted rays is greater in some directions; and many reflections will be measured only partially during several successive oscillations. When the data set of all oscillations is considered as a whole many of the systematic errors, including those just enumerated, introduced as a consequence of the collection scheme may be corrected. The output is usually in the form of a

list of reflections indexed by their coordinates,  $hkl$ , in the 3-dimensional array called reciprocal space (see below), the corrected intensity  $I(hkl)$ , and the error associated with the intensity  $\sigma(I)$  (Stout *et al.*, 1989).

## Diffraction of X-rays

X-rays, electromagnetic radiation of wavelength in the range of  $\sim 100 - 10^{-4}$  were first discovered by Rntgen in 1895. Quantum theory predicts that light interacting with matter, electrons in the case of X-rays, can be absorbed and reradiated with an energy less than or equal to the incident light. If the energy of the reradiated light is lessened the discrepancy leads to an increase in internal energy of the electron system; this is called inelastic scattering. If the energy of the reradiated light is not reduced we call it elastically scattered light. For the purposes of this discussion we assume that, even in the case of anomalous scattering, we are dealing with elastic scattering.

For elastic scattering then, classical electromagnetic theory predicts that X-rays will interact with matter in the same way as visible light. Generally a plane wave of light with frequency  $\omega$  whose electric vector has amplitude  $E_0$  is described as follows:

$$2.1 \quad \vec{E}(t) = \vec{E}_0 e^{i\omega t}$$

When an X-ray plane wave impinges on a dipole oscillator whose axis is parallel to the electric vector of the X-ray and of charge  $e$  and mass  $m$  the charge undergoes simple harmonic motion. If the displacement of the charge from the center of the dipole is  $\mathbf{x}$  at time  $t$  then the equation of motion of the charge is

$$2.2 \quad \mathbf{x}'' + k\mathbf{x}' + \omega_s^2 \mathbf{x} = e/m \mathbf{E}_0 \mathbf{e}^{i\omega t}$$

where  $k$  is the damping factor and  $\omega_s$  is the natural frequency of the dipole;

$$2.3 \quad \mathbf{x} = (e^2/m) [\mathbf{E}_0 \mathbf{e}^{i\omega t} / (\omega_s^2 - \omega^2 + ik\omega)]$$

and equation 2.3 is the solution to equation 2.2 (James, 1962). The dipole produces a spherical wave of the same frequency as the incident X-ray with amplitude a function of the frequency. Then at time  $t$  at a point  $\mathbf{R}$  on the spherical wavefront whose polar distance is  $\phi$  the magnitude of the electric vector tangent to the sphere  $F_s$  is given by

$$2.4 \quad F_s = - (e^2/mc^2) \mathbf{E}_0 \mathbf{e}^{i\omega t} [(t - |\mathbf{R}|/c)] \sin \phi / |\mathbf{R}| = f(|\mathbf{R}|) \mathbf{e}^{i\omega t}$$

where  $\mathbf{R}$  is very large compared to the wavelength and  $c$  is the speed of light in free space and the negative sign denotes that the phase is opposed to the incident light. At a great distance from the dipole in the plane perpendicular to the dipole and passing through the center of the dipole the amplitude  $F_s$  of the scattered radiation at time  $t$  and distance  $\mathbf{R}$  is

$$2.5 \quad F_s = (e^2/mc^2) [\omega^2 E_0 / (\omega_s^2 - \omega^2 + ik\omega)].$$

We define the scattering factor,  $f$ , of a dipole as the ratio of the scattering due to a dipole to the scattering of a free electron ( $\omega_s = k = 0$  in eq. 2.5)

$$2.6 \quad f = [\omega^2 / (\omega_s^2 - \omega^2 + ik\omega)].$$

Equation 2.6 clarifies the notion that the amplitude is a complex number, which arises from the nature of the equation of motion of dipole motion. If the incident radiation is very different from the natural frequency of the dipole the complex aspect of the scattering factor is negligible. But if the incident frequency is comparable to the natural frequency the complex nature of the scattering factor becomes evident; this gives rise to scattering called anomalous due to the fact that both the magnitude and the phase of the scattered wave are sensitive functions of the incident wavelength.

In the above discussion of dipole radiation it was assumed that the point of zero displacement of the dipole was at the origin of a coordinate system. To describe the scattered wave from a dipole shifted by a vector  $\mathbf{r}$  from the origin let A1 in Figure 2-4 be the origin and A2 the displacement of the dipole. The incident and scattered radiation follow parallel paths but the path length of the scattered radiation differs by

$$2.7 \quad d = (\mathbf{r} \cdot \mathbf{s}) - (\mathbf{r} \cdot \mathbf{s}_0) = \mathbf{r} \cdot (\mathbf{s} - \mathbf{s}_0) = \mathbf{r} \cdot \mathbf{S}$$

and therefore it will have a phase difference  $\phi$

$$2.8 \quad \phi = (2\pi/\lambda) \mathbf{r} \cdot \mathbf{S}$$

where  $\lambda$  is the wavelength of the radiation and  $\mathbf{S}$  has a simple geometrical interpretation shown in Figure 2-5. Assuming that  $\mathbf{R}$  is so great compared to  $\mathbf{r}$  that the amplitude of the scattered wave is essentially the same as that of dipole at the origin we now have the equation of scattering of a dipole at a distance  $\mathbf{r}$  from the origin of

$$2.9 \quad F_s = f(|\mathbf{R}|) e^{i\omega t'} e^{i\phi}.$$

In a real crystal random vibrations of individual scatterers modulate the scattered wave. This is called the temperature effect due to its presumed origin in the random thermal vibrations of scatterer. The temperature effect may be modeled in the lattice of dipoles by superimposing a random vibration on the oscillation of each dipole. For a given dipole this amounts to an equation similar to that of equation 2.9 where  $f$  is now a function of time as well. In practice the time dependence is removed by making the assumption that each dipole vibrates independently and in the same way as the others. The amplitude of the scattering can then be described by

$$2.10 \quad F_s = f(|\mathbf{R}|) e^{i\omega t'} \exp(-8\pi^2 U_s^2 \sin^2\theta/\lambda^2)$$

where  $U_s^2$  is the mean square displacement of the dipole in a direction parallel to  $\mathbf{S}$  and  $\theta$  is the scattering angle (James, 1962). Historically, the quantity  $8\pi^2 U_s^2$  has been called the B factor and is usually folded into the

expression for the scattering factor  $f = f(|R|, B)$ . The ultimate effect, then, of random thermal vibration is a reduction in the scattering factor of the form shown in Figure 2-6, as well as a concomitant reduction in the measured intensity, however it does not affect the phase.

Ultimately, we would like to describe the form of the total scattered wave from a crystal. But it is easier to arrive at that form in a series of steps. To a good first approximation a lattice of dipoles exhibits a mathematical similarity to a crystal (however, a real crystal has quadrupole and, in general, multipole terms as well). Let  $\mathbf{r}$  be the vector describing the coordinates of a dipole in a 3-dimensional lattice having period  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$  in the  $x$ ,  $y$ , and  $z$  direction respectively. Then

$$2.11 \quad \mathbf{r} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$$

where  $u$ ,  $v$ , and  $w$  are integers. The total scattering  $F$  from the lattice of dipoles is just the sum of the amplitudes from each scatterer

$$2.12 \quad F = \sum F_s = F_s(\mathbf{r}_1) + F_s(\mathbf{r}_2) + F_s(\mathbf{r}_3) + \dots$$

But as previously noted, the form of  $F_s$  changes if the dipole is not at the origin of the lattice. There is a phase change relative to the origin simply due to the position of the dipole in the lattice

$$2.13 \quad F = \sum F_s = F_s(e^{2\pi i \bar{S} \cdot \bar{r}_1} + e^{2\pi i \bar{S} \cdot \bar{r}_2} + e^{2\pi i \bar{S} \cdot \bar{r}_3} + \dots).$$

From wave theory the intensity  $I$  at time  $t$  at a point  $\mathbf{R}$  on the spherical wavefront is equal to the amplitude, a complex number, multiplied by its complex conjugate and thus

$$2.14 \quad I(\mathbf{R}) = |\mathbf{F}| \cdot |\mathbf{F}|^* = P |f(\mathbf{R})|^2 [e^{i\omega t'} (e^{2\pi i \bar{\mathbf{S}} \cdot \bar{\mathbf{r}}_1} + e^{2\pi i \bar{\mathbf{S}} \cdot \bar{\mathbf{r}}_2} + e^{2\pi i \bar{\mathbf{S}} \cdot \bar{\mathbf{r}}_3} + \dots)] [e^{-i\omega t'} (e^{-2\pi i \bar{\mathbf{S}} \cdot \bar{\mathbf{r}}_1} + e^{-2\pi i \bar{\mathbf{S}} \cdot \bar{\mathbf{r}}_2} + e^{-2\pi i \bar{\mathbf{S}} \cdot \bar{\mathbf{r}}_3} + \dots)]$$

where  $P$  is defined as the polarization factor, a number which varies between 1 and 1/2 depending on the angle between  $\mathbf{S}$  and the direction of polarization of the incident wave. Equation 2.14 describes the far field intensity of elastic scattering from a lattice of equivalent dipoles but it can be shown that it also describes a 3-dimensional version of the diffraction grating. For large numbers of dipoles, such as would be found in a real crystal,  $I(\mathbf{R})$  is negligible unless the geometry of the scattering satisfies Bragg's condition (James, 1962).

By comparison to Figure 2-4 let  $A_1$  and  $A_2$  be a distance  $d$  apart where now  $d$  corresponds exactly to the distance between two parallel planes of dipoles (Fig. 2-7). Scattering occurs in such a way that it may be considered a reflection from each plane. Constructive interference occurs when the reflected rays from parallel planes have path differences an integral multiple,  $n$  of the wavelength. The path difference is given by

$$2.15 \quad \Delta = 2 d \sin \theta = n\lambda,$$

which is the familiar Bragg formulation of scattering.

The conditions for interference may be described geometrically by way of the concept of the reciprocal lattice. If we imagine a lattice of points  $L$  described by equation 2.11 where a point is given by  $(u, v, w)$  we can define a reciprocal lattice  $R$  described by

$$2.16 \quad \mathbf{r}^* = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$$

where a point in  $R$  is given by  $(h, k, l)$ .  $R$  has the property that

$$2.17 \quad \mathbf{a}^* \cdot \mathbf{b} = \mathbf{a}^* \cdot \mathbf{c} = \mathbf{b}^* \cdot \mathbf{a} = \mathbf{b}^* \cdot \mathbf{c} = \mathbf{c}^* \cdot \mathbf{a} = \mathbf{c}^* \cdot \mathbf{b} = 0.$$

That is  $\mathbf{a}^*$  is perpendicular to the  $a$  plane of  $L$ ,  $\mathbf{b}^*$  is perpendicular to the  $b$  plane of  $L$ , and  $\mathbf{c}^*$  is perpendicular to  $\mathbf{a}$  and  $\mathbf{b}$  (or the  $c$  plane) of  $L$ .

Furthermore, we define the dot product of  $\mathbf{a}$  and  $\mathbf{a}^*$  is equal to one and

$$2.18 \quad \mathbf{a}^* \cdot \mathbf{a} = \mathbf{b}^* \cdot \mathbf{b} = \mathbf{c}^* \cdot \mathbf{c} = 1.$$

If  $L$  is an orthogonal lattice then  $R$  is orthogonal and the lengths of the  $R$  lattice vectors are the reciprocal of the  $L$  lattice vectors as shown in Figure 2-8. However, if  $L$  is a triclinic lattice then the lengths of the lattice vectors are as shown in Figure 2-9.

Using the reciprocal space relations we can interpret the conditions for interference in a simple geometric manner. The reciprocal space vector  $(h, k, l)$  is normal to the reflecting plane in the Bragg relation and has the magnitude of the inverse of the plane spacing. We define the Ewald sphere



as a sphere in reciprocal space with radius  $1/\lambda$  and the origin is a point on its surface. If the incident X-ray beam passes through both the center of the Ewald sphere and the origin of reciprocal space (Fig 2-10) any other point on the Ewald sphere coincident with a reciprocal lattice point P satisfies the Bragg relation (James, 1962).

### Fourier methods

Fourier showed that any periodic function  $f(x)$  could be represented by a series of sinusoidal waves each having a period, an integral fraction of that of the original function, and empirically determined amplitude. It can be shown that the function  $f(x)$  can be approximated, to an arbitrary accuracy related to the number of terms in the series, by the Fourier series

$$2.19 \quad f(x) = A_0/2 + a_n \cos(n\pi x/p) + b_n \sin(n\pi x/p)$$

Where  $n$  is an integer,  $a_n$  and  $b_n$  are amplitudes, and period of  $2p$  (Wylie, 1966) (Fig. 2-11). The infinite series is an exact series representation of the function  $f(x)$

$$2.20 \quad f(x) = A_0/2 + \sum_{\infty} [a_n \cos(n\pi x/p) + b_n \sin(n\pi x/p)] = \sum_{\infty} c_n e^{in\pi x/p}$$

Due to the periodic nature of the crystal lattice we can represent the electron density  $\rho$  at any point (xyz) in a crystal of unit cell volume  $V$  by a triple Fourier series summed over reciprocal space

$$2.21 \quad \rho(xyz) = 1/V \sum_h \sum_k \sum_l F(hkl) e^{-2\pi i(hx + ky + lz)}$$

This equation shows that the electron density in real space at  $\rho(xyz)$  is the Fourier transform of the set of all structure factors  $F(hkl)$  (Stout *et al.*, 1989). Equation 2.21 is also the inverse analog to equation 2.13, which describes the structure factor of a lattice of dipoles as the sum of the amplitudes of individual dipoles. In this way the Fourier transform relates real space to reciprocal space and *vice versa*.

For mosaic crystals the intensity measured on the CCD is assumed to be the square of the modulus of the structure factor amplitude  $F(hkl)$  and it is clear that the phase information of the structure factor has been lost. However, the phase information of the structure factor carries all the information regarding the relative spacing of the scattering masses. The discussion of the Fourier series provides a recipe for reconstructing the electron density in the unit cell only if we know the structure factor. This is the so-called ‘phase problem’ of crystallography; the various solutions of the phase problem involve regaining the lost phase information.

For protein crystallography the most fruitful methods of ‘phasing’ a protein begin with the Patterson synthesis. The Patterson function,

$$2.22 \quad P(\mathbf{u}) = 1/V \sum_h \sum_k \sum_l |F(hkl)|^2 \cos [2\pi(hu + kv + lw)]$$

where  $u$ ,  $v$ , and  $w$  are relative coordinates, is analogous to the Fourier series representation of electron density where  $F(hkl)$  is replaced with  $I(hkl)$  and the phase information is omitted. The Patterson function is computed on the unit cell of the crystal for which the data were collected. In fact it can be shown, using the convolution theorem, that equation 2.22 is equivalent to

$$2.23 \quad P(\mathbf{u}) = \int_V \rho(\mathbf{r}_1) \rho(\mathbf{r}_1 + \mathbf{u}) dV$$

where the integral is taken over the unit cell (Drenth, 1995).

Equation 2.23 shows that the weight of each patterson vector is proportional to the scattering factors of the two volume elements. The scattering factor is proportional to the number of electrons and therefore the weight is proportional to the product of the number of electrons in each atom. The interpretation of the vector function  $P(\mathbf{u})$  is that peaks in the patterson map correspond to inter-atomic vectors (i.e.  $u = x_1 - x_2$ ,  $v = y_1 - y_2$ , and  $w = z_1 - z_2$ ) (Fig. 2-12). Unfortunately, all possible inter-atomic vectors are represented; for  $N$  atoms there are  $N^2$  peaks – ( $N$  of them are length 0 self vectors making the origin a huge peak ameliorated by subtracting the average value of  $|F|^2$ , so called origin removal).

## SIRAS

Single isomorphous replacement with anomalous scattering (SIRAS) is a general Patterson method for determining the location of heavy atoms in

a unit cell, determined by the heavy metal structure factors  $F_h$ , using isomorphous and anomalous data in the Patterson summation. The method requires a native data set and a data set with specifically located heavy metals, which have not perturbed the crystal structure (isomorphism). Because of its generality SIRAS may be used in any experiment where there is significant anomalous scattering. We have used the SIRAS method for our multiple anomalous dispersion (MAD) data.

In general heavy metals have natural frequencies, called absorption edges, in the X-ray domain so that the complex nature (see eq 2.5 and Fig. 2-13) of the scattering factor can no longer be neglected. Near an absorption edge the scattering factor is complex and both the magnitude and the phase of scattered radiation depend on the wavelength of incident radiation (Fig. 2-14). The SIRAS method finds the heavy metal structure factors using a Patterson synthesis of the isomorphous  $(\Delta iso)^1$  and  $(\Delta ano)^1$  anomalous differences between  $F_p$  and  $F_{ph}$  (Fig 2.15). Therefore the Patterson with SIRAS data

$$2.24 \quad P(u) = 1/V \sum_h \sum_k \sum_l (|\Delta iso|^2 + |\Delta ano|^2) \cos [2\pi(hu + kv + lw)]$$

---

<sup>1</sup>  $|\Delta iso| = |F_{ph}| - |F_p| \sim |F_h| \cos(\alpha_{ph} - \alpha_h)$  component of  $F_h$  along  $F_{ph}$   
 $|\Delta ano| = |F_h| \sin(\alpha_{ph} - \alpha_h) \sim f'/2f'' (|F_{ph}(+)| - |F_{ph}(-)|)$  where  $f_{anom} = f' + if''$  component of  $F_h$  perpendicular to  $F_{ph}$   $|F_h|^2 = |\Delta ano|^2 + |\Delta iso|^2$

will be a Patterson map of the heavy metal structure in the unit cell. The heavy atom positions are then refined using the method of least squares to minimize the closure error between observed and calculated  $F_o$ .

In a typical MAD experiment one begins by measuring the fluorescence from the crystal as the wavelength is scanned around an accessible absorption edge of the anomalous scatterer (AS) (selenium  $\sim 0.98$  Å). In this way the total scattering factor for the AS is experimentally measured. The experimental values are then fit to a theoretical absorption curve using the Kramers-Kronig transform (Gonzalez *et al.*, 1999). This allows a functional representation of both the real and complex components of the scattering factor as a function of wavelength (Fig. 2-16).

In order to maximize the isomorphous and anomalous differences between data sets (Hendrickson *et al.*, 1997; Gonzalez *et al.*, 1999) peak (maximum of scattering function), inflection point (maximum of derivative of scattering function), and a high energy remote wavelengths are chosen for data collection. The data sets are collected in order of importance peak, inflection point, and high energy remote wavelengths to reduce the effect of radiation decay on the experiment. The Solve program MADMRG then combines the data into a single native and isomorphous data set with anomalous scattering (SIRAS) (Terwilliger, 1997).

Specifically, let  $F_h(\lambda_j)$  be the structure factor due to anomalous scattering atoms at  $\lambda_j$  and  $F_o$  be the structure factor due to non-anomalous scatterers, which should be independent of wavelength, then the amplitude of the total structure factor

$$2.25 \quad F_t(\lambda_j) = |F_o + F_h(\lambda_j)|.$$

One can then solve algebraically for the structure factor amplitudes of  $F_o$  and  $F'(\lambda_o)$ , (the structure factor of the whole structure at a standard wavelength  $\lambda_o$ ) as well as the anomalous difference at a standard wavelength

$$2.26 \quad \Delta_{ano}(\lambda_o) \sim [\Delta f''(\lambda_o) / \Delta f''(\lambda_j)] \Delta_{ano}(\lambda_j)$$

where  $\Delta f''$  is the complex scattering factor. The quantities  $F_o$ ,  $F'(\lambda_o)$ , and  $\Delta_{ano}(\lambda_o)$  are used as  $F_p$ ,  $F_{ph}$ , and  $\Delta_{ano}$  in the SIRAS methods described elsewhere but implemented in Solve HEAVY. The Solve program HEAVY conducts automated peak searches of origin removed Patterson maps to locate heavy atom positions (Terwilliger *et al.*, 1999).

### **Protein phase angle determination**

Once the positions of the heavy metals have been refined their phases may be calculated in a straightforward manner. Following the reconstruction of the full set of structure factors  $F_h$  the protein structure factors may be determined probabilistically. Figure 2-17 shows a Harker construction of a general structure factor with anomalous scattering. The

magnitude and direction of the AS scatterer structure factor are known while only the magnitudes are known for  $F_p$  and  $F_{ph}$ ; this is depicted in Figure 2-17, which shows three circles of radius  $F_p$ ,  $F_{ph}(+)$ , and  $F_{ph}(-)$  where  $F_p$  and  $F_h$  have the same origin. The intersection of the three circles lies at the correct phase angle  $\alpha$ . All other phase angles will have an associated error of closure  $\epsilon$

$$2.27 \quad \epsilon = |F_p + F_h| - |F_{ph}|_{obs},$$

where  $F_{ph}$  is assumed to contain all the error. When  $\epsilon$  is calculated for all protein phase angles  $\alpha_p$  (phase angle for  $F_p$ ) it describes a distribution of  $\epsilon(\alpha_p)$ .  $\epsilon(\alpha_p)$  divided by its mean square error  $E$  is assumed to be normally distributed and therefore a probability distribution  $P(\alpha_p)$  may be constructed

$$2-28 \quad P(\alpha_p) = N \exp(-\epsilon(\alpha_p)^2 / 2E^2)$$

where  $N$  is chosen to normalize the function such that the total probability is equal to one (Beyer, 1987; Drenth, 1995). A graph of a typical  $P(\alpha_p)$  distribution is depicted in Figure 2-18. A similar equation may be derived for the closure error due to the anomalous differences where

$$2-29 \quad \epsilon(\alpha_p)_{ano} = (|F_{ph}(+)| - |F_{ph}(-)|)_{obs} - (|F_{ph}(+)| - |F_{ph}(-)|)_{calc}$$

The total probability for a reflection in the SIRAS method is the product

$$2-30 \quad P(\alpha_p) = P(\alpha_p)_{iso} P(\alpha_p)_{ano}$$

illustrated in Figure 2-19.

Based on this analysis one might think that  $\alpha_{p\max}(hkl)$ , where P is at its maximum value, should be assigned to  $F(hkl)$  and the Fourier series of the electron density built up from the resulting structure factors

$$2-31 \quad F(hkl) = F(hkl)\exp[i\alpha_{p\max}(hkl)].$$

But an electron density map constructed from the most probable phases does not give the best Fourier map; when the  $\alpha_p$  are chosen as the phase of the centroid (the weighted mean) of their individual joint probability distributions  $P(\alpha_p)$  the resulting Fourier series will have the minimum mean square error in electron density (Drenth, 1995). In general  $|F_{\text{best}}|$  is less than  $|F_{\alpha_{p\max}}|$  and the ratio  $|F_{\text{best}}|/|F_{\alpha_{p\max}}| = m$  is called the figure of merit. The figure of merit provides a single number description of the quality of the joint probability distribution  $P(\alpha_p)$ . If a given probability distribution has a single well defined peak its centroid will have phase  $\alpha_{p\max}$  and  $|F_{\text{best}}|/|F_{\alpha_{p\max}}| = m \sim 1$  as shown in Figure 2-20a. The probability distribution for another structure factor may have two peaks of similar magnitude well separated in phase, which would describe the case of  $m \sim 0$  illustrated in Figure 2-20b.

So the best Fourier map is actually calculated with

$$2-32 \quad F_{\text{best}}(hkl) = m |F(hkl)| \exp [i\alpha_{p\text{centroid}}(hkl)]$$



as the Fourier series terms, where  $|F(hkl)|$  is the measured structure factor amplitude,  $m$  is the figure of merit and  $\alpha_{p,centroid}$  is the best phase. The Fourier summation

$$2-33 \quad \rho(xyz) = 1/V \sum_h \sum_k \sum_l (m|F(hkl)| \exp [i\alpha_{p,centroid}(hkl)]) e^{-2\pi i(hx + ky + lz)}$$

then provides the initial electron density map within which the first structural model is built (Drenth, 1995).

### **Density modification**

Once the best phases have been calculated by the methods described above it may be possible to further improve the calculated electron density of the protein using the techniques of density modification (DM). Generally these techniques make assumptions about the scattering due to solvent, which is part of the total scattering from the unit cell.

$$2-34 \quad F_{total} = F_{protein} + F_{solv}$$

But the primary goal of protein structure determination is to find the structure factor amplitudes and phases due to the protein since these describe the protein structure.

Solvent flattening is one of the simplest and most successful techniques of reducing the effect of bulk solvent on the calculated protein structure factors. Solvent flattening capitalizes on the fact that scattering due to bulk solvent is lower in magnitude and noisier than that due to protein or

bound solvent. Wang created an algorithm by which bulk solvent regions are identified and given an average electron density (flattened) (Wang, 1985). The Fourier transform of the flattened density function of the unit cell provides structure factors, which are used to construct a probability distribution  $P_{\text{flat}}$  analogous to equation 2-28

$$2-35 \quad P_{\text{flat}} = N \exp [ X \cos (\alpha_p - \alpha_{\text{calc}})]$$

where  $\alpha_p$  is the protein phase angle,  $\alpha_{\text{calc}}$  is the phase angle calculated from the flattened density function,  $N$  is a normalization constant and  $X$  is the mean intensity of the unknown part of the structure (Drenth, 1995). A new joint probability distribution is then computed using

$$2-36 \quad P(\alpha_p) = P(\alpha_p)_{\text{iso}} P(\alpha_p)_{\text{ano}} P(\alpha_p)_{\text{flat}}$$

which is analogous to equation 2-30 and from this a new figure of merit calculated for each structure factor. From this best structure factor a new electron density function is computed and the whole procedure repeated until the average figure of merit of the protein converges. For crystal structures containing 50% or more solvent the figure of merit can improve drastically (Wang, 1985).

Maximum likelihood (ML) density modification, as performed in the Resolve program of the Solve suite, is a much more sophisticated approach than solvent flattening (Terwilliger, 2000). Solvent flattening and other non-

ML methods do not provide for the appropriate weighting of the flattened phases with respect to the observed phases since they both contain similar information (the assumption made in 2-36, that the probabilities are independent is false in general). ML constructs a log likelihood function (LL) of a set  $\{F_h\}$  of structure factors that fit the data, which incorporates observed data with foreknowledge of the map (e.g., bulk solvent is flatter than protein)

$$2-37 \quad LL(\{F_h\}) = LL^O(\{F_h\}) + LL^{OBS}(\{F_h\}) + LL^{MAP}(\{F_h\})$$

where  $LL^O(\{F_h\})$  denotes other structure factor information such as the Wilson distribution of intensities. The maximum of  $LL(\{F_h\})$  then provides the most likely set of structure factors. This formalism has the advantage that it can incorporate other previously successful DM techniques and yet avoid the problem of determining the correct relative weighting. When Resolve was used on the HPrK/P structure it improved the average figure of merit from .357 before DM to .567 (Table 4-1).

### **Model Building and Refinement**

The primary goal of DM is to improve the interpretability of the electron density function. Inspection of the resulting density maps following the determination of the protein phases and DM is the best way to decide if anything was gained in the attempt. Usually a map is calculated at

intermediate resolution, e.g., data from 15-3 Å, and searched manually for the telltale secondary structure of an alpha helix. If DM is required to find helices or improves the appearance of a previously identified helix density it is considered a worthwhile procedure.

The ability to examine electron density maps presumes the ability to display them. The Xfit graphical display program from the XtalView suite of programs was used throughout the structure determination of HPrK/P (McRae, 1999). In addition to the ability to display protein structure coordinates and electron density maps Xfit has numerous features that facilitate the process of building a protein model from the observed data, i.e., the electron density map. XtalView and its programs run on computers with UNIX and X11. Some of the XtalView programs can be run from the command line but most use a graphical user interface started from `xtalmgr`, the first window opened from the command line.

After the best structure factors have been found through SIRAS and possibly DM, the observed data is in the form of an  $hkl$  indexed structure factor  $F_{\text{obs}}(hkl)$ ,  $\sigma_{\text{obs}}[F(hkl)]$ ,  $\alpha_{\text{best}}(hkl)$ , and figure of merit ( $\text{fom}(hkl)$ ). When these data are in the right format Xfit uses a Fast Fourier Transform (FFT) to calculate the electron density ‘on the fly’ at the center of the graphical display. Xfit also has a semiautomated fitting feature, which when triggered

places a 'mrk' dummy residue at the center of the display and then adds consecutive dummy residues at distance intervals appropriate for the  $C_{\alpha}$  backbone when the ">" key is punched. The program considers the electron density at several possible locations and then considers a second dummy residue placement for each first location to minimize sidechain solutions; each solution is given a probability and the most probable solution is frequently correct. This allows the rapid tracing of the polypeptide backbone in regions of unambiguous electron density.

When a break in the electron density appears the other end of the chain is elongated until it also reaches a break. The resulting  $C_{\alpha}$  backbone trace is then transformed into a polyalanine chain using the Poly-Ala Fragment command, which uses a database of protein backbone pentamers to find a good fit to 5 successive  $C_{\alpha}$ s. The pentamer is fit by least-squares and the middle three alanines are retained; this process is repeated until the entire  $C_{\alpha}$  backbone is alanine.

Finally, the polyalanine chain is assigned a sequence based on identifiable stretches of the electron density. The protein sequence and the location of selenomethionine residues, the anomalous scatterers, are already known; if Xfit is given the protein sequence and a choice for a given residue in that sequence is identified on the polyalanine chain it will change the

alanines to the given sequence and attempt a best fit for the sidechain. If the electron density is readily interpretable a 300 residue protein model can be built in one sitting. Poor density and frequent breaks in the chain slow the process considerably, so it is well worth producing the best electron density map possible before building the model.

The first round of model building ceases when the easily interpretable portions of the electron density map have been fitted with a reasonable model, at which point refinement of the model commences. Refinement of protein models can be formulated as a mathematical problem whereby the model is improved by minimization of some functional representation of the model and the observed data. The R-factors are defined as

$$2-38 \quad R_{\text{work}} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

and

$$2-39 \quad R_{\text{free}} = \frac{\sum ||F_{\text{obs}}(\text{test})| - |F_{\text{calc}}(\text{test})||}{\sum |F_{\text{obs}}(\text{test})|}$$

where the test set comprises ~ 5% of observed reflections selected randomly and subsequently not used in refinement. R factors provide a simple means of comparison of the observed structure factor magnitudes with the model structure factor magnitudes.  $R_{\text{free}}$  is most useful as a means of comparing various refinement strategies; the superior refinement strategy minimizes  $R_{\text{free}}$ .

In addition to the observed data and the model there are a number of other known quantities to be considered in refinement of proteins. Most importantly there exists now a large database of well defined, high resolution protein structures; these structures provide a great deal of information on allowed bond lengths, bond angles, dihedral angles, planarity of aromatic rings, van der Waals radii and torsion angles. In a conventional least squares refinement scheme as implemented in the program TNT (Tronrud, 1997) a composite function is formulated, which represents each differential between the model and the known database parameters. The method of least squares is used to find the coordinates and B-factors of the atoms of the model, which minimize the composite function.

However, a complicated n-dimensional function such as that found in the least squares implementation of TNT will have numerous local minima in which refinement may be trapped. The method of simulated annealing as implemented in Crystallography and NMR System (CNS) provides a convenient method for avoiding local minima (Brunger *et al.*, 1998). In simulated annealing an n-dimensional composite function of the total energy of the model is formulated,

$$2-40 \quad E_{\text{total}} = E_{\text{chem}} + E_{\text{xray}}$$

$$2-41 \quad E_{\text{chem}} = \sum_{\text{bonds}} k_b(r-r_o)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \text{dihedral terms} + \text{planar} + \dots$$

$$2-42 \quad E_{\text{xray}} = \sum_{\text{hkl}} [ |F_{\text{obs}}(\text{hkl})| - k |F_{\text{calc}}(\text{hkl})| ]^2$$

where the  $k$ s are weighting functions, and  $r_0$  and  $\theta_0$  represent database values.

But here the total energy is an explicit function of the temperature, and the energy minimum is presumed to represent the best fit to the observed data.

Simulated annealing avoids local minima by refining at high temperatures for very short periods of time, which favors minima with large radii of convergence (Brunger *et al.*, 1987; Rice *et al.*, 1994). High temperature refinement is followed by room temperature refinement and then the geometry of the model is regularized.

Once the first round of model building and refinement, is completed the new model must be scrutinized to determine if the refinement procedure produced chemically reasonable results. Usually the new model will require extensive manual modifications before a new round of refinements begins. This process of rebuilding and refinement continues until no further improvements in the R-factors and geometry are realized. At that point the model may have useful phase information in its own right; an electron density map is calculated with observed structure factor amplitudes and model phases, which may be combined with the experimental phases. The model is updated, if the combined density map identifies previously



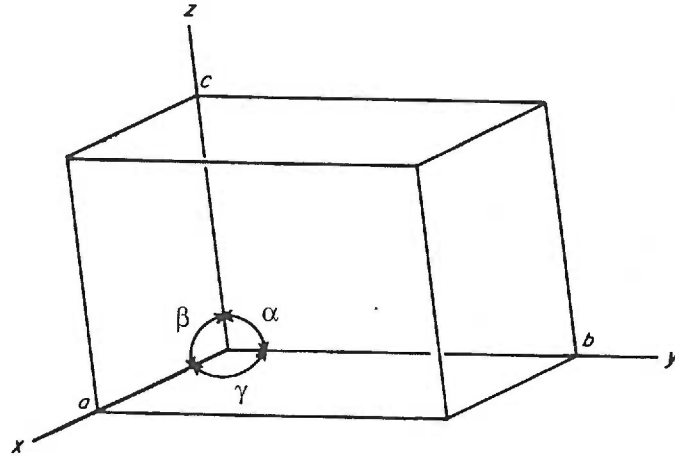
unknown portions of the crystal structure, and the process of rebuilding and refinement commences anew.

Under ideal circumstances the iterative process of model building and refinement would converge rapidly to the correct protein model, but in reality the correct model may be difficult to recognize, so refinement could continue indefinitely. Therefore, we have followed some rules of thumb in deciding when to finish refinement: the R-factor should be less than 25%, the  $R_{\text{free}}$  should be less than 30%, the average root mean square deviation from database values for bond lengths should be less than 0.01 , and the average rmsd for angles should be less than 2. These values are somewhat arbitrary, however, and they represent more of a convention based on what is achievable with current practices, software, and apparatus.

It would be myopic, though to terminate refinement solely on the basis of achieving acceptable cutoff values for refinement parameters. While the overall structure would be acceptable it may contain regions obviously in error based on a Ramachandran analysis for example. We used the program PROCHECK (Laskowski *et al.*, 1996), a considerable improvement on a simple Ramachandran plot, to help finalize the structure of *M. pneumoniae* HPrK/P. In addition to a Ramachandran plot PROCHECK provides graphical output of the  $\phi$ - $\psi$  and sidechain bond angle

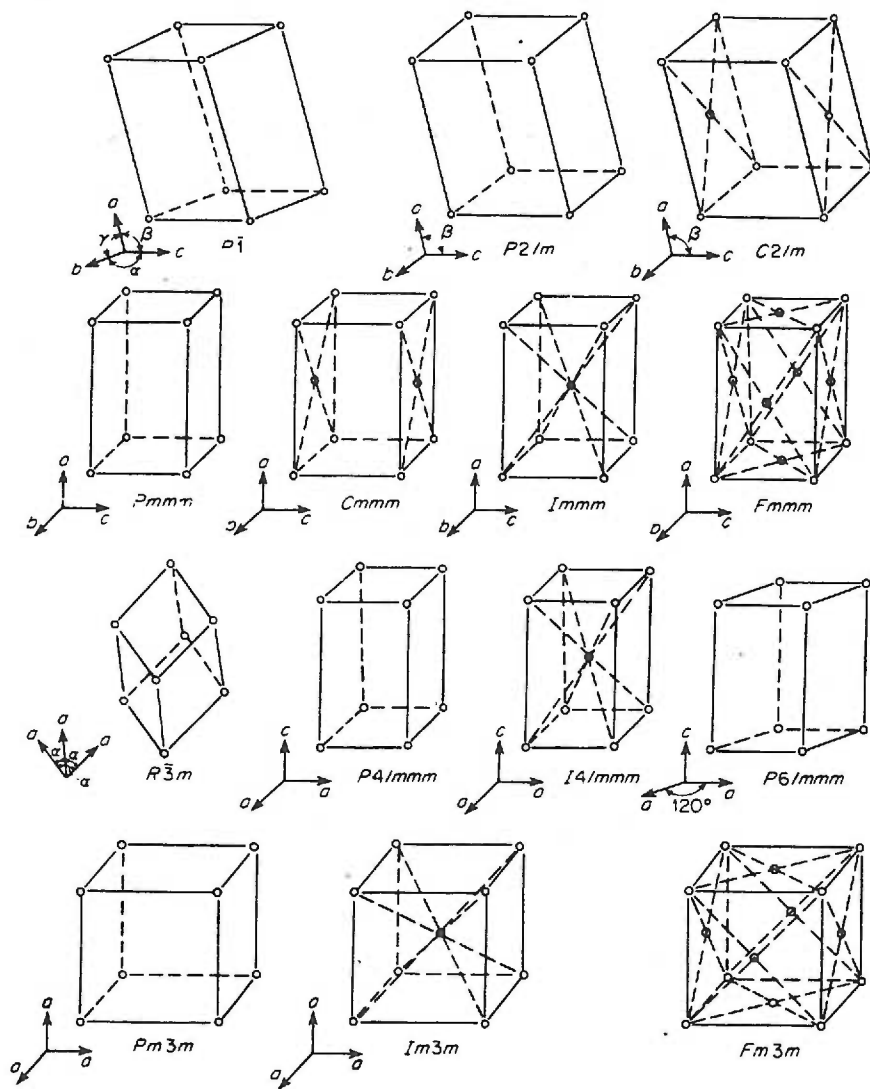
deviations from database values for individual residue types; this is useful because different types of residues have different optimum values for these angles. PROCHECK also provides extensive information on the individual variations of planarity, bond length, chirality, and Luzzatti coordinate error of residues considered to be in error or at least worth further scrutiny.

Armed with the output from PROCHECK the refinement and rebuilding process continues until either all identified errors are fixed or the errors, which had been previously fixed but will not refine, are accepted as unrefinable. It is also possible that the data appear to uphold conformations normally considered to be in error. For example, as we show in chapter 4, the  $\phi$ - $\varphi$  angles of the Asp178 residue of each subunit of *M. pneumoniae* HPrK/P lie in the disallowed region of Ramachandran space. But this example is somewhat unusual and after fixing the errors that were rectifiable the model and coordinates were submitted to the RCSB protein database.

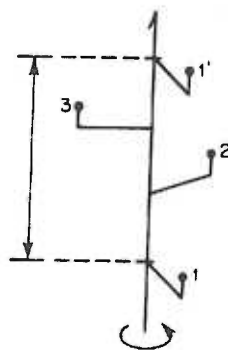


Crystal System	Independent Parameters	Parameters	Lattice Symmetry
Triclinic	6	$a \neq b \neq c; \alpha \neq \beta \neq \gamma$	$\bar{1}$
Monoclinic	4	$a \neq b \neq c; \alpha = \gamma = 90^\circ; \beta > 90^\circ$	$2/m$
Orthorhombic	3	$a \neq b \neq c; \alpha = \beta = \gamma = 90^\circ$	$mmm$
Tetragonal	2	$a = b \neq c; \alpha = \beta = \gamma = 90^\circ$	$4/mmm$
Trigonal rhombohedral lattice	2	$a = b = c; \alpha = \beta = \gamma \neq 90^\circ$	$\bar{3}m$
hexagonal lattice	2	$a = b \neq c; \alpha = \beta = 90^\circ; \gamma = 120^\circ$	$6/mmm$
Hexagonal	2	$a = b \neq c; \alpha = \beta = 90^\circ; \gamma = 120^\circ$	$6/mmm$
Cubic	1	$a = b = c; \alpha = \beta = \gamma = 90^\circ$	$m\bar{3}m$

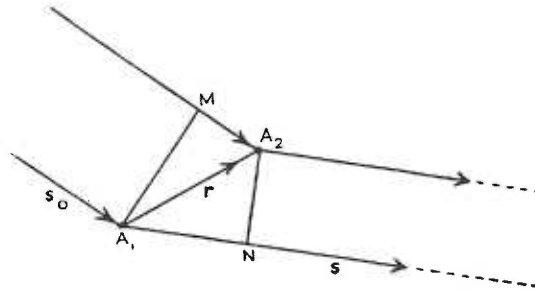
**Figure 2-1** The seven crystal systems. The definitions of the axes and angles between them are drawn in the unit cell box. Note that rhombohedral and hexagonal lattices are both considered Trigonal systems. (Adapted from (Stout *et al.*, 1989))



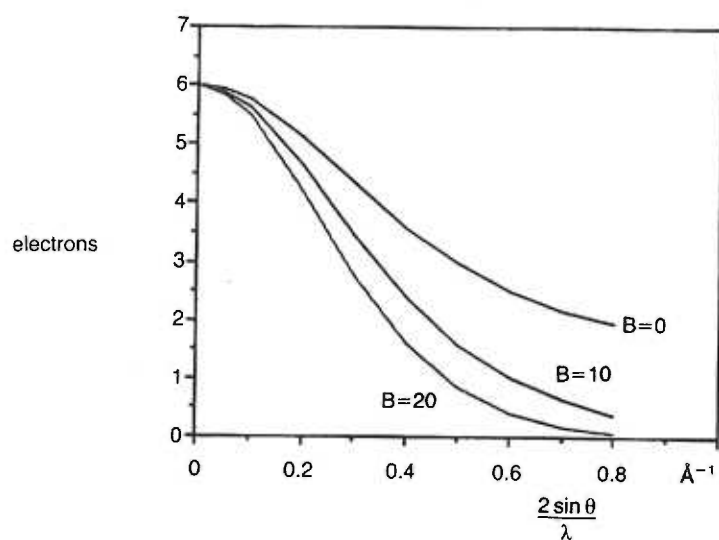
**Figure 2-2** The 14 Bravais lattices. The Bravais lattices include the 7 crystal systems but they are designated Primitive lattices because they have only one lattice point per unit cell ( $1/8^{\text{th}}$  point per corner). The non-Primitive lattices, C, I, and F contain 2 or more lattice points per unit cell. (Adapted from (Stout *et al.*, 1989))



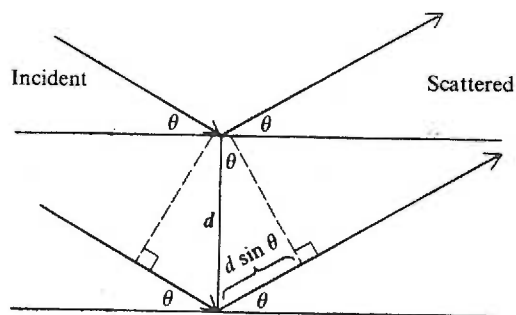
**Figure 2-3** A 3 fold screw axis.  $3_1$  screw axes are generated by a 120 degree rotation with each  $1/3$  unit cell translation parallel to a cell edge (figure adapted from (Drenth, 1995)).



**Figure 2-4** Vector diagram of parallel scattering from two dipoles.  $A_1$  is taken to be the origin.  $A_2$  is a distance  $|r|$  away from  $A_1$  in the direction of  $r$ .  $s_0$  is the incident wave and  $s$  is the scattered wave. The distance  $M-A_2$  is the same as  $r \cdot s_0$ . The distance  $A_1-N$  is the same as  $r \cdot s$ . The difference in the path length traveled is  $d = r \cdot s - r \cdot s_0 = r \cdot (s - s_0) = r \cdot S$  (Text and figure adapted from (James, 1962)).

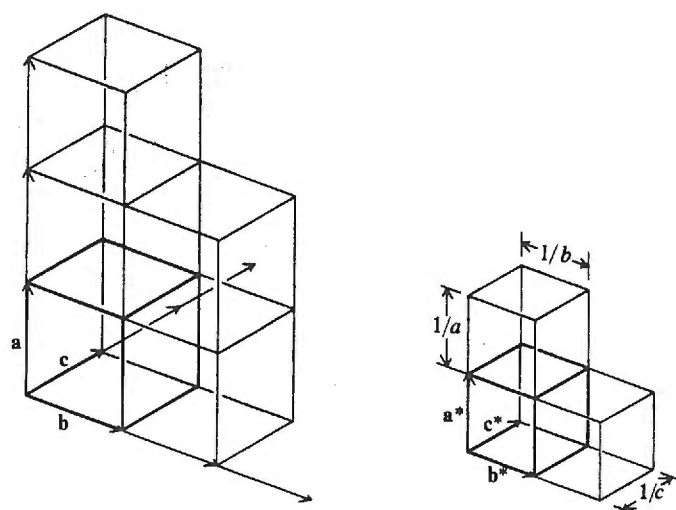


**Figure 2-6** The carbon atom scattering factor. The real component of equation 2.10 displayed as a function of  $2\sin\theta/\lambda$  and constant B. As the scattering angle increases the scattered amplitude decreases more rapidly for a carbon atom with a greater B value. (Text and figure adapted from (Drenth, 1995))



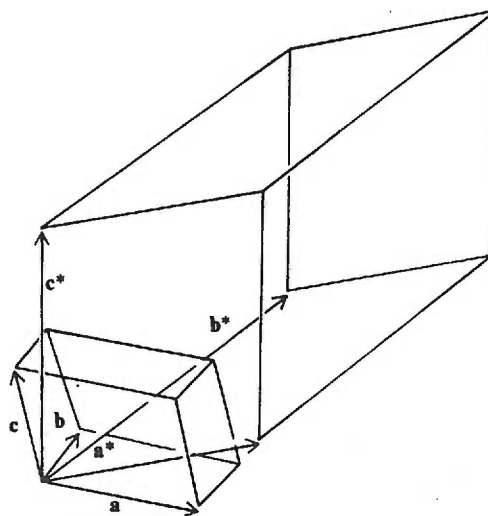
**Figure 2-7** Bragg's Law. This is a special case of the geometry described in Figure 2-4. The incident wave strikes two parallel crystal planes in such a way that it may be considered reflection. The path difference of scattered radiation  $\Delta = 2 d \sin \theta = \lambda$ . This formulation resembles the one dimensional diffraction grating. (Adapted from (Drenth, 1995))



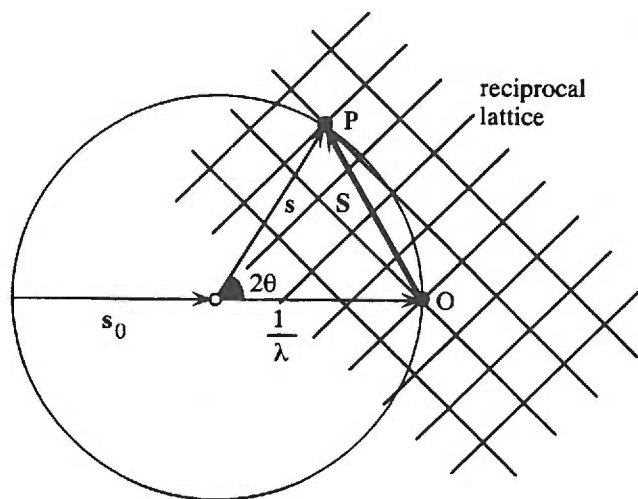


**Figure 2-8** Real and Reciprocal space lattice comparison for orthogonal unit cells. (Adapted from (Cantor *et al.*, 1980))

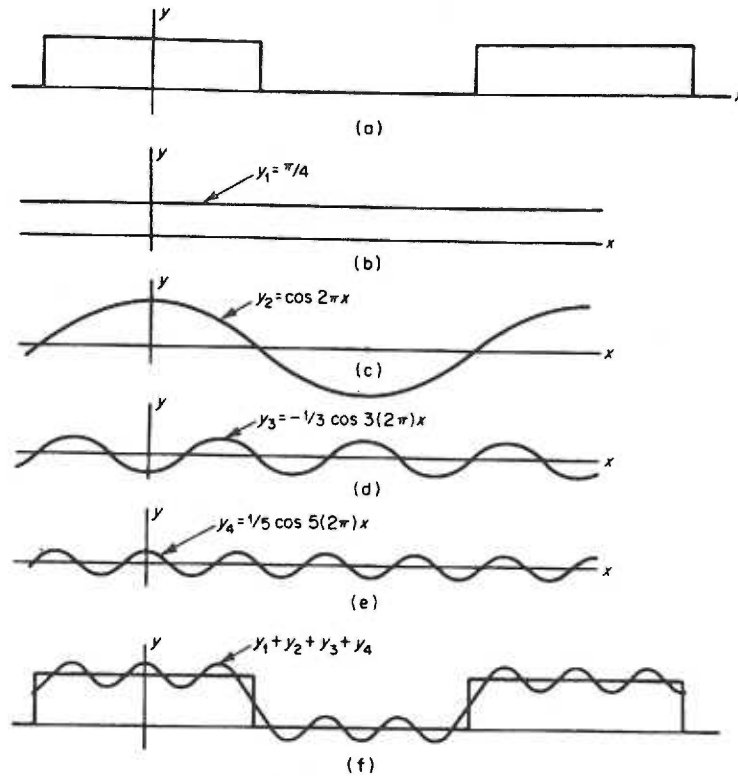
$$\begin{aligned} \mathbf{a}^* &= (1/V)(\mathbf{b} \times \mathbf{c}) \\ \mathbf{b}^* &= (1/V)(\mathbf{c} \times \mathbf{a}) \\ \mathbf{c}^* &= (1/V)(\mathbf{a} \times \mathbf{b}) \end{aligned}$$



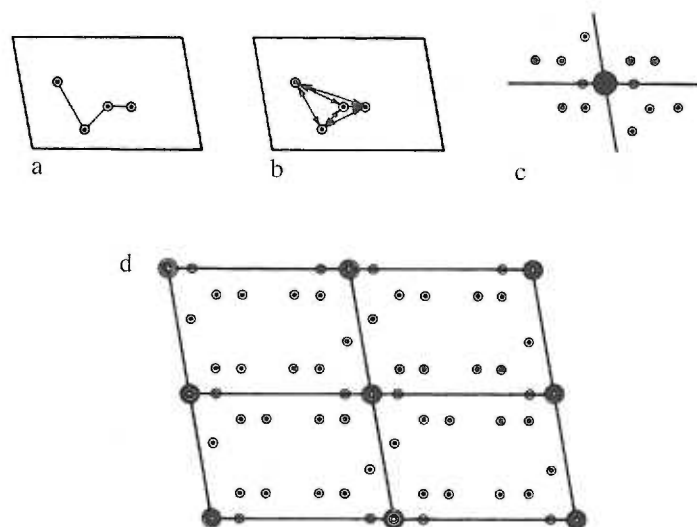
**Figure 2-9** Real and Reciprocal space lattice comparison for triclinic unit cells. The lengths of the reciprocal space lattice vectors are reciprocally related to the real space lattice vector as shown in the box. (Adapted from (Cantor *et al.*, 1980))



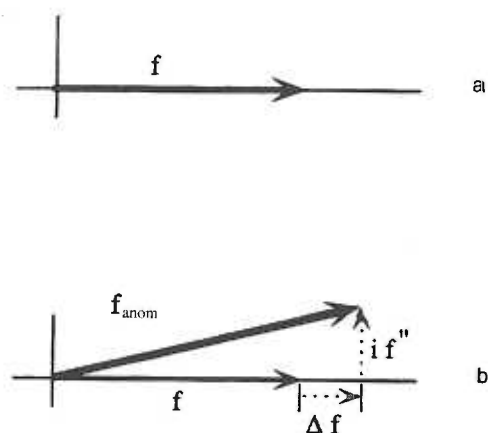
**Figure 2-10** The Ewald sphere. The sphere has radius  $1/\lambda$ . The center of the reciprocal lattice is at  $O$ .  $s_0$  is the incident beam,  $s$  is the scattered beam, and  $S$  is the scattering vector. If the incident X-ray beam passes through both the center of the Ewald sphere and the origin of reciprocal space any other point on the Ewald sphere coincident with a reciprocal lattice point  $P$  satisfies the Bragg relation. (Adapted from (Drenth, 1995))



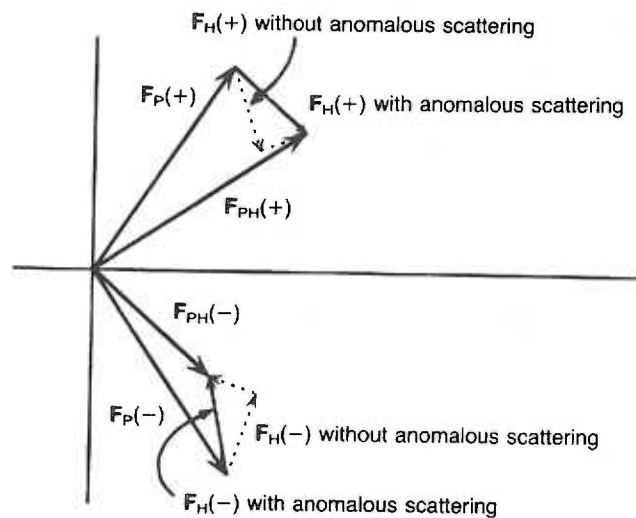
**Figure 2-11** Harmonic analysis of a step function. (a) A periodic step-function. (b) – (e) The first four terms of the cosine form of the Fourier series for the step function. (f) The sum of the first four terms of the cosine series representation of the step function. (Text and figure adapted from (Stout *et al.*, 1989))



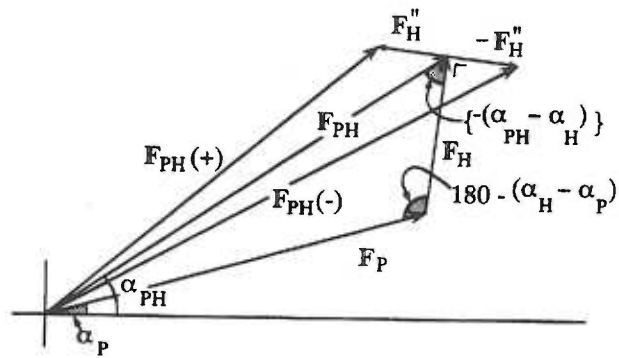
**Figure 2-12** A simple Patterson map. (a) A hypothetical four atom model. (b) The interatomic vectors for the model in (a). (c) The superimposition of four models of (a); each model has a different atom at the origin. (d) The Patterson peaks in four adjacent unit cells. (Text and figure adapted from (Stout *et al.*, 1989))



**Figure 2-13** The atomic scattering factor in the complex plane. (a) A free electron scatters x-rays with no complex component (see equation 2.5). (b) A bound electron scatters x-rays with some, perhaps negligible, complex component. The anomalous scattering factor is formulated as  $f_{\text{anom}} = f + \Delta f + i f'' = f' + i f''$  where  $\Delta f$  denotes the real component of the change in the scattering factor when the frequency of the radiation approaches an absorption edge. (Text and figure adapted from (Drenth, 1995))

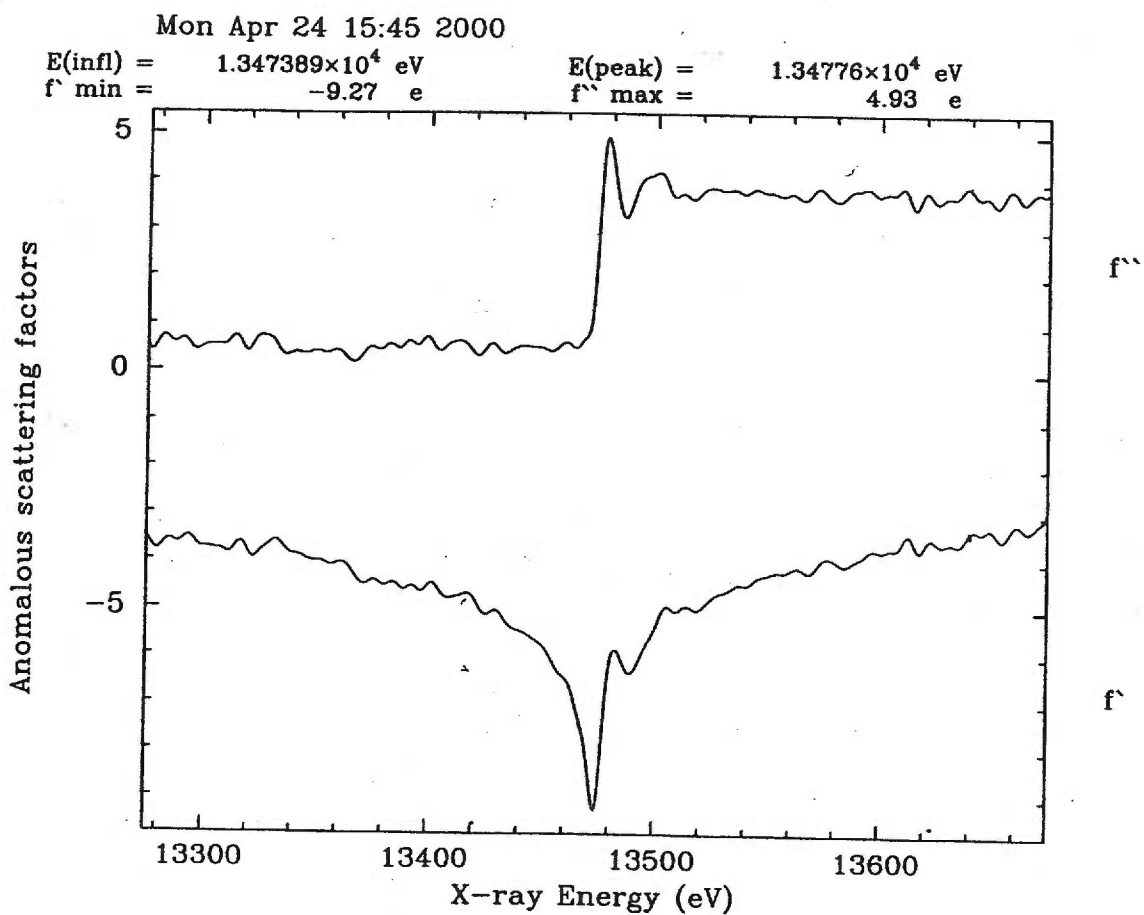


**Figure 2-14** Structure factors in the complex plane (Argand diagram). The protein-heavy metal structure factor  $F_{ph}$  is the sum of the protein and heavy metal structure factors,  $F_p + F_h = F_{ph}$ . In the absence of anomalous scattering  $F_h(+)$ ,  $F_h(-)$ ,  $F_p(+)$ , and  $F_p(-)$  are symmetric about the real (x) axis. In the presence of anomalous scattering the pairs of structure factors are anti-symmetric with respect to the real axis. (Text and figure adapted from (Drenth, 1995))

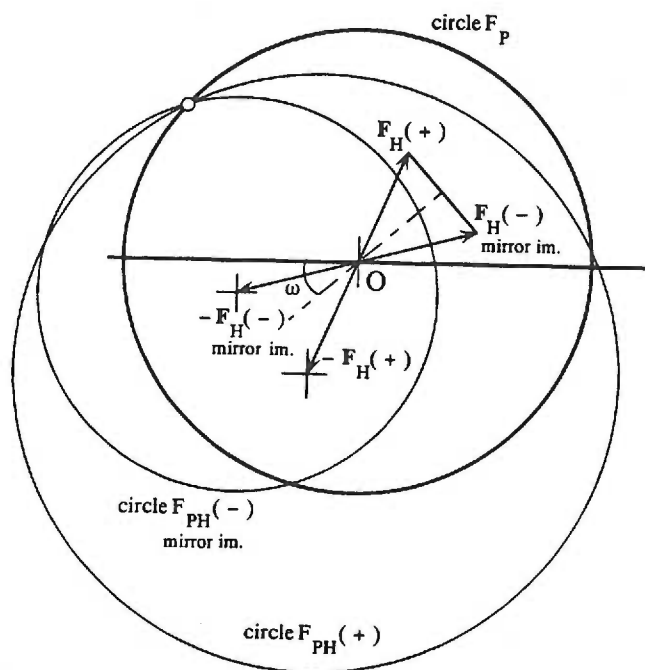


**Figure 2-15** Bijvoet pairs  $[F(hkl)$  and  $F(-h-k-l)]$  mirrored about the real axis. In this view the the complex conjugates  $F_{ph}(-)$ ,  $F_p(-)$ ,  $F_h(-)$ , and  $F_h''(-)$  are superimposed.  $\alpha_h$  is the phase angle for the nonanomalous part of  $F_h$ . (Text and figure adapted from (Drenth, 1995))

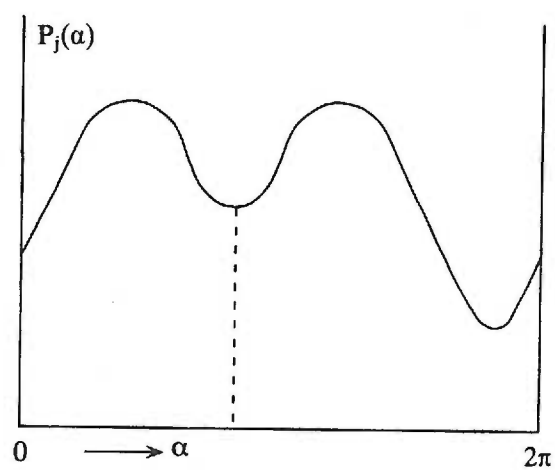




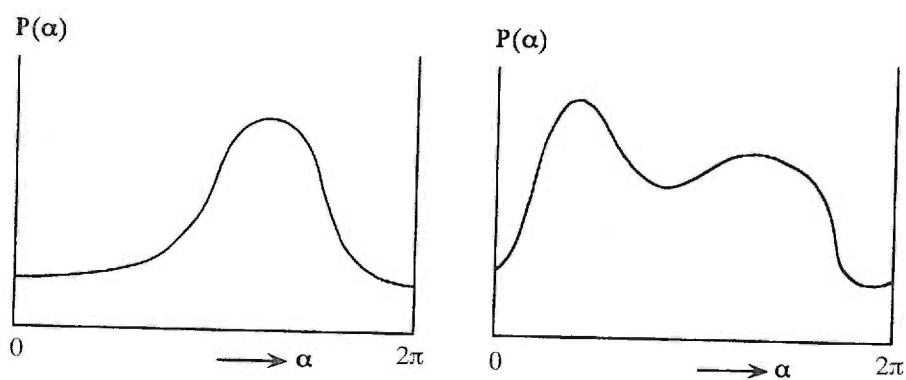
**Figure 2-16** The anomalous scattering factor of Bromouridine. These data were collected as part of the *B. megaterium* CcpA structure determination described in the Appendix.



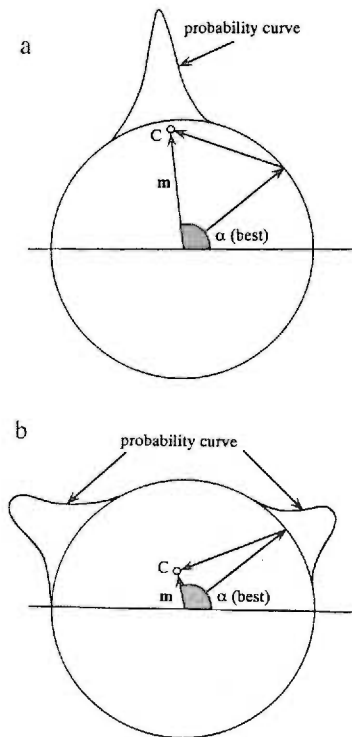
**Figure 2-17** Harker construction for a SIRAS experiment. The magnitude and direction of the AS scatterer structure factor is known while only the magnitudes are known for  $F_p$  and  $F_{ph}$ . Note that  $-F_h(-)$  is depicted as its complex conjugate to simplify the geometry. (Figure adapted from (Drenth, 1995))



**Figure 2-18** A graph of a typical  $P(\alpha_p)$  distribution for isomorphous data.  
(Figure adapted from (Drenth, 1995))



**Figure 2-19** The total probability for a reflection in the SIRAS method is shown for two typical wavelengths. The distributions are clearly more unimodal than for the isomorphous component alone. (Figure adapted from (Drenth, 1995))



**Figure 2-20** Two different probability curves for a SIRAS experiment. The probability curves are mapped onto an argand diagram to illustrate how the best phases appear at high (a) and low (b) figure of merit.  $C$  is the centroid of the distribution and  $m$  is the figure of merit. The best phases are not those that have the highest probability. (Figure adapted from [Drenth, 1995 #387])

## Chapter III

# **Crystallization, preliminary X-ray analyses, and biophysical characterization of HPr kinase/phosphatase of *Mycoplasma pneumoniae***

**Katrin Steinhauer,<sup>a#</sup> Gregory S. Allen,<sup>b#</sup> Wolfgang Hillen,<sup>a</sup>**

**Jörg Stülke<sup>a</sup> and**

**Richard G. Brennan<sup>b\*</sup>**

<sup>a</sup> *Lehrstuhl für Mikrobiologie, Friedrich-Alexander-Universität  
Erlangen-Nürnberg, Staudtstr. 5, D-91058 Erlangen, Germany*

<sup>b</sup> *Oregon Health & Science University, Biochemistry and  
Molecular Biology L224, 3181 SW Sam Jackson Park Road,*

*Portland, OR 97201, USA*

<sup>#</sup> These authors contributed equally to this work.

*Acta Cryst.* (2002) **D58**:515-518

## Synopsis

The *Mycoplasma pneumoniae* HPr kinase/phosphatase (HPrK/P) was crystallized in the space group  $P2_12_12_1$  with cell edges  $a = 117.1 \text{ \AA}$ ,  $b = 127.7 \text{ \AA}$  and  $c = 170.7 \text{ \AA}$ . Both X-ray and biophysical data indicate that HPrK/P is a functional hexamer that undergoes an ATP binding induced conformational change.

## ABSTRACT

The *Mycoplasma pneumoniae* HPr kinase/phosphatase (HPrK/P) is a member of a large family of enzymes, which is central to carbon regulation in Gram positive bacteria. The full-length, *M. pneumoniae* HPrK/P was crystallized from solutions of polyethylene glycol 8000, and KCl or NaCl, which also contained the non-hydrolysable ATP analogue, adenosine 5'-[ $\beta,\gamma$ -methylene] triphosphate (AMPPCP). The crystals take the orthorhombic space group  $P2_12_12_1$  with unit cell edges  $a = 117.1 \text{ \AA}$ ,  $b = 127.7 \text{ \AA}$  and  $c = 170.7 \text{ \AA}$ . A complete X-ray intensity data set has been collected and processed to  $2.50 \text{ \AA}$  resolution. The slow self-rotation function revealed the presence of a 6-fold axis. Dynamic light scattering (DLS) experiments indicated a molecular weight of 197 kDa for HPrK/P, in the absence of AMPPCP, and of 217 kDa in the presence of the ATP analogue. Thus, the biophysical and crystallographic data suggest that HPrK/P is a functional hexamer that undergoes an ATP binding induced conformational change.

## 1. Introduction

*Mycoplasma* species belong to the low Guanine-Cytosine (GC) family of Gram-positive bacteria and, with a genome size which ranges from 0.58-1.35 Mb, are the smallest self replicating organisms (Razin *et al.*, 1998). The fastidious growth requirements of *Mycoplasma* species result from a streamlined genome and dictate their parasitic lifestyle. Known hosts include plants, fish, birds, insects and mammals (Razin *et al.*, 1998). *Mycoplasma* species in humans and animals are usually found attached to the surfaces of epithelial cells, nevertheless they also seem to be capable of entering the intracellular space. The latter may enable them to escape the immune response as well as selective antibiotic treatment (Rosengarten *et al.*, 2000). Their impact on human health and the somewhat mysterious molecular basis of their pathogenicity makes them of particular interest, especially in light of the available genomic information (Rosengarten *et al.*, 2000).

The genomes of *M. genitalium* and the closely related *M. pneumoniae* have been sequenced (Fraser *et al.*, 1995; Himmelreich *et al.*, 1996). The genome sequencing projects of the two *Mycoplasma* species revealed that these organisms possess few genes encoding regulatory proteins. For example, they possess only a single sigma factor and lack two-component regulatory systems (Fraser *et al.*, 1995; Dybvig *et al.*, 1996; Himmelreich *et al.*, 1996; Razin *et al.*, 1998). Also, the genomic sequences of *M. pneumoniae* and *M. genitalium* revealed the existence of orthologues of HPr



kinase/phosphatase (HPrK/P) in these organisms. HPrK/P is the key enzyme in the regulation of carbon catabolism in other, more complex, low GC Gram-positive bacteria such as *Bacillus subtilis*, *Enterococcus faecalis* or *Lactobacillus casei* (Galinier *et al.*, 1998; Reizer *et al.*, 1998; Kravanja *et al.*, 1999; Dossonnet *et al.*, 2000). In these organisms HPrK/P responds to the metabolic state of the cell by phosphorylating or dephosphorylating its substrate, the heat stable protein (HPr) of the bacterial phosphoenolpyruvate:sugar phosphotransferase system (PTS). The phosphorylation state of HPr in turn signals the metabolic state of the cell; the different forms of HPr are recognized by proteins in several distinct regulatory systems (Stulke *et al.*, 1998). Presumably the HPrK/P orthologues found in *Mycoplasmas* also play an important role in the regulation of energy metabolism in these organisms.

*In vitro* biochemical analysis of *M. pneumoniae* HPrK/P revealed that the enzyme exhibits kinase activity at low ATP concentrations ( $[ATP] > 1\mu M$ ), and requires inorganic phosphate ( $[P_i] > 1mM$ ) for phosphatase activity (Steinhauer, 2002). In contrast, HPrK/P from *B. subtilis* needs high ATP concentrations ( $[ATP] > 100\mu M$ ) for kinase activity but low inorganic phosphate ( $[P_i] > 200\mu M$ ) to exhibit phosphatase activity (Jault *et al.*, 2000; Monedero *et al.*, 2001; Hanson *et al.*, 2002). Therefore, the intrinsic activities of the HPrK/P from *M. pneumoniae* and *B. subtilis* differ: HPrK/P

from *M. pneumoniae* is primarily a kinase, while the *B. subtilis* enzyme is by default a phosphatase. Besides having different intrinsic activities, both enzymes are regulated by ATP, inorganic phosphate, and the glycolytic intermediate fructose-1,6-bisphosphate (Reizer *et al.*, 1998; Kravanja *et al.*, 1999; Jault *et al.*, 2000; Hanson *et al.*, 2002; Steinhauer, 2002). This regulation of HPrK/P by low molecular weight effectors is an indication of its role as a sensor of the metabolic state of the cell and points to a mechanism for the regulation of carbon catabolism.

In addition to their enzymatic differences gel filtration experiments suggest that *M. pneumoniae* HPrK/P is active as a hexamer, while the enzyme from *B. subtilis* is active as an octamer (Steinhauer, 2002). Recently, the crystal structure of an N-terminal truncation mutant of the *Lactobacillus casei* HPrK/P, which is missing residues 1 through 127, was determined and found to be hexameric (Fieulaine *et al.*, 2001). However, no three-dimensional structural data are yet available for any full-length HPrK/P enzyme. In order to obtain a more detailed understanding of the structural organisation and mechanism of HPrK/P we carried out further biophysical characterisation as well as the crystallisation of the *M. pneumoniae* enzyme. Here, we describe the crystallization of HPrK/P and experiments designed to address possible structural changes of the protein upon ATP binding.

## **2. Materials and Methods**

### **2.1. Purification of HPrK/P**

HPrK/P from *M. pneumoniae* carrying an N-terminal hexahistidine sequence was purified in buffer containing 10 mM Tris/HCl pH 7.5, 10 mM  $\beta$ -mercaptoethanol, and 0.6 M NaCl. The protein was bound to a nickel chelate resin and eluted with a gradient of imidazole in purification buffer as described earlier (Steinhauer, 2002). In order to remove the imidazole the eluate was dialyzed against purification buffer. The dialyzed protein was then concentrated to 7 mg/ml as determined by Bradford assay.

### **2.2. Preparation of selenomethionine substituted HPrK/P**

A selenomethionine substituted HPrK/P was prepared by transforming the methionine auxotrophic strain *E. coli* B834 (Novagen) with plasmid pGP204 carrying the *M. pneumoniae hprK* gene (Steinhauer, 2002). Cells were grown in supplemented M9 minimal medium containing selenomethionine at a final concentration of 80 mg/l (Doublié, 1997). Cells were grown at 37 °C in the dark to prevent light induced oxidation of selenomethionine. Expression of the recombinant protein was induced by the addition of IPTG (1 mM) to logarithmically growing cultures (OD<sub>600</sub> of 0.6). The selenomethionine substituted protein was purified as described above except for the addition of 15 mM  $\beta$ -mercaptoethanol to the purification buffer throughout purification. The final yield of the selenomethionine substituted HPrK/P was 12 mg/l.

### 2.3. DTNB-titration of HPrK/P

Titration of the thiols of HPrK/P with 5,5'-dithiobis (2-nitrobenzoic acid) (DTNB) was carried out in a buffer containing 10 mM Tris pH 7.2 and 0.6 M NaCl at room temperature; DTNB was dissolved in 0.1 M phosphate buffer, pH 7.2 (Ellman, 1959). The reaction was followed spectrophotometrically at 410 nm and was complete after approximately 100 minutes. The amount of labelled thiol groups per HPrK/P monomer was calculated using the extinction coefficient of  $\text{TNB}^{2-}$  at  $\lambda = 410$  nm (Riddles *et al.*, 1983). Only freshly purified HPrK/P devoid of reducing agent was used for the DTNB titration. Following the reaction with DTNB, labelled HPrK/P was used for activity assays, which were carried out as described (Steinhauer, 2002). Briefly, 20  $\mu\text{M}$   $(\text{His}_6)\text{HPr}$  or  $(\text{His}_6)\text{HPr}(\text{Ser-P})$  were incubated with 350 nM HPrK/P in assay buffer (10 mM  $\text{MgCl}_2$ , 25 mM Tris/HCl, pH 7.6, 1 mM dithiothreitol) in a final volume of 20  $\mu\text{l}$ . After 15 min incubation at 37 °C, the enzyme was inactivated. The proteins were analyzed using 10% native PAGE.

### 2.4. Dynamic light scattering experiments

Dynamic light scattering studies on HPrK/P were done using a DynaPro-801 Dynamic Light Scattering Instrument (Protein Solutions Inc.). Protein samples at 1.5 mg/ml or 1 mg/ml were prepared in solutions of 25 mM Tris/HCl, pH 7.6, containing either 600 mM, 360 mM, or 200 mM

NaCl, and 6 mM  $\beta$ -mercaptoethanol in the presence or absence of 1 mM adenosine 5'-[ $\beta,\gamma$ -methylene] triphosphate, AMPPCP, a non hydrolysable ATP analogue, and 10 mM  $\text{MgCl}_2$ . Prior to the experiment, all protein samples were filtered through 0.1 micron Anotop 10 filters (Whatman) to eliminate any large aggregates. All data were analyzed using AutoPro PC (Protein Solutions Inc.) software. The reported values are averages of 5 scans of 2 minutes each.

## **2.5. Crystallization of HPrK/P**

Crystals of HPrK/P were grown at room temperature in hanging drops by the vapor diffusion method (McPherson, 1999). The crystallization solution was 4% polyethyleneglycol-8,000, 20 mM  $\text{MgCl}_2$ , 50 mM Tris/HCl, pH7.6, and 0.8 M KCl or NaCl. In a typical hanging drop experiment, 3  $\mu\text{l}$  HPrK/P (7 mg/ml) in 10 mM Tris/HCl, pH 7.5, 0.6 M NaCl, and 1 mM AMPPCP were mixed with 3  $\mu\text{l}$  of the crystallization solution and equilibrated over a 0.8 ml reservoir at room temperature. Crystals took about a week to appear and grew for three additional weeks. Crystals of the selenomethionine substituted protein were obtained under the same conditions as described for the native HPrK/P, except that the crystallization solution contained 0.9 M NaCl and the protein concentration used for crystallization was 5 mg/ml.

## 2.6. Cryoprotection and x-ray intensity data collection

Cryo-cooling conditions for crystals of both the native and selenomethionyl substituted HPrK/P proteins were obtained by modifying a method that significantly changes the ionic strength of the crystallization liquor while attempting to maintain its osmolality (McRee, 1999). The scheme involves passing the crystal through 4 drops (5 $\mu$ l each) of increasing glycerol concentration and decreasing salt concentration. This stepwise approach is necessary because the crystals are destroyed if placed directly in the final cryo-protectant solution. Specifically, the first transfer from the crystallization solution (20 mM MgCl<sub>2</sub>, 50 mM Tris/HCl pH7.6, 4% PEG-8,000, and 0.8 M NaCl; ~2.34 Os/kg) removes the PEG-8,000, decreases the NaCl to 0.5 M and adds 10% glycerol (~2.35 Os/kg). The second transfer further decreases the salt concentration to 0.2 M and increases the glycerol to 20% (~3.45 Os/kg). Incubation in the third and fourth transfer drops simply increases the glycerol in 5% increments to a final concentration of 30% (~4.55 and ~5.65 Os/kg, respectively). Transfer drop four preempts formation of ice. The crystal is then placed in the nitrogen cryo-stream. Depending upon the size of the crystal it is incubated for 1-3 minutes per drop; larger crystals are soaked for 3 minutes. X-ray intensity data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) on beamline BL 9.2 using a Quantum 4 CCD and  $\lambda = 0.88560\text{\AA}$ . All data were

collected at 100 K. The intensity data were processed with d\*TREK v5.5i (Pflugrath, 1999).

### **3. Results and Discussion**

#### **3.1. DTNB-titration of HPrK/P**

In order to determine the solvent accessible thiol content of the enzyme, which might provide possible heavy atom derivative sites for mercurial, titration of (His)<sub>6</sub>HPrK/P with 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) was carried out and monitored spectrophotometrically (Ellman, 1959). Repeated titration of freshly purified (His)<sub>6</sub>HPrK/P with DTNB revealed a thiol content of 1.8 per monomer, indicating that both cysteines of the *M. pneumoniae* enzyme are solvent accessible thiols rather than disulfide bonded or buried.

To investigate whether the cysteines are crucial for enzyme activity, we carried out enzyme activity assays following DTNB-titration of HPrK/P. Kinase activity of HPrK/P was tested in the presence of 10  $\mu$ M ATP and in the presence or absence of 10 mM fructose-1,6-bisphosphate (FBP); FBP has been shown to stimulate kinase activity of some HPrK/P proteins (Galinier *et al.*, 1998; Reizer *et al.*, 1998; Kravanja *et al.*, 1999; Dossonnet *et al.*, 2000; Jault *et al.*, 2000), but to exhibit no stimulatory effect on the kinase activity of *M. pneumoniae* HPrK/P (Steinhauer, 2002). The results obtained for the DTNB-labelled HPrK/P also did not show any stimulatory

effect of FBP on kinase activity (Fig. 3-1A). Furthermore, the phosphorylation efficiencies of the non-labelled HPrK/P protein and HPrK/P following DTNB-titration were similar. However, in the presence of 50  $\mu$ M ATP and 5 mM  $P_i$ , addition of 1 mM FBP to the reaction mixture prevented kinase inhibition by  $P_i$  (Steinhauer, 2002). There was also no difference in phosphatase activity or its regulation upon labeling of *M. pneumoniae* HPrK/P when compared to the unlabeled enzyme (Fig. 3-1B).

These results indicate that DTNB treated *M. pneumoniae* HPrK/P neither reduces the enzymatic activity nor its regulation. Therefore, it is very likely that the cysteines in HPrK/P are not essential for the catalytic activity of the enzyme and their ready accessibility makes them good candidates for heavy atom modification by mercurials.

### **3.2. Dynamic light scattering experiments**

Previous size exclusion chromatography experiments indicated that *M. pneumoniae* HPrK/P is an hexamer in solution (Steinhauer, 2002). Dynamic light scattering (DLS) experiments were carried out to further characterize the oligomeric properties of the enzyme in the presence and absence of the AMPPCP. DLS offers the advantage of studying a macromolecule in its true solution state; chromatography or gel matrices, which can confound interpretations of data by their nonspecific interaction with protein, are not present. DLS also indicates the aggregation state of a



protein under a given set of experimental conditions. This information was useful for designing crystallization experiments.

A 1 mg/ml (27  $\mu$ M on a monomer basis) solution of HPrK/P in the presence of 0.6 M NaCl was monodisperse and had an estimated molecular weight of 197 kDa. A reduction of the salt concentration to 360 mM or 200 mM and an increase in the protein concentration to 1.5 mg/ml (41  $\mu$ M) gave similar results. Addition of 1 mM AMPPCP resulted in an increase of the estimated apparent molecular weight of HPrK/P to 217 kDa. Therefore, the data obtained by dynamic light scattering indicate that the oligomeric state of apo (His)<sub>6</sub>HPrK/P is either a pentamer (184 kDa) or a hexamer (221 kDa), whereas the ATP-bound enzyme is clearly a hexamer.

The change in the apparent molecular weight detected upon nucleotide binding has two possible explanations. The first is that in the absence of nucleotides HPrK/P exists as a pentamer, which upon binding ATP adds an additional monomer to the complex resulting in the formation of a hexamer. The second, and more plausible interpretation, is that HPrK/P takes a more compact conformation in the absence of ATP, thereby leading to a decreased hydrodynamic radius, which is detected as a smaller molecular weight. Thus, it is very likely that the *M. pneumoniae* HPrK/P forms hexamers even

in the absence of nucleotides and can undergo significant nucleotide-induced conformational changes.

### 3.3. x-ray Data Collection and Analysis

Orthorhombic crystals of native HPrK/P and its selenomethionine derivative were obtained as described in Materials and Methods (Fig. 3-2). X-ray intensity data were collected at cryogenic temperatures to a resolution greater than 2.5 Å. The data were processed with d\*TREK version 5.5i (Table 3-1) (Pflugrath, 1999).

The diffraction pattern displayed 2/m 2/m 2/m Laue symmetry. Analysis of the  $hk0$ ,  $h0l$  and  $0kl$  zones revealed the systematic absence of all  $2n+1$  reflections along the  $a^*$ ,  $b^*$  and  $c^*$  axes. Thus, the  $(\text{His})_6\text{HPrK/P}$  crystal takes the space group  $P2_12_12_1$  and the unit cell edges are  $a = 117.1$  Å,  $b = 127.7$  Å,  $c = 170.7$  Å and  $\alpha = \beta = \gamma = 90.0^\circ$ . On the basis of the unit cell volume ( $2.564 \times 10^6$  Å<sup>3</sup>) and the assumption of six monomers per asymmetric unit, the calculated  $V_M$  is 2.9 Å<sup>3</sup>/Da (Matthews, 1968). The presence of a hexamer is supported further by the results of the Slow Rotation Function as implemented in *GLRF* (Tong *et al.*, 1997) with resolution and radius cutoffs of 10-3.5 Å and 20 Å respectively, which reveals a 6-fold rotation axis at a peak height that is 8.6 above sigma.

Searches for heavy atom derivatives are underway as are multiple wavelength anomalous diffraction experiments using crystals of selenomethionyl substituted HPrK/P in complex with AMPPCP (Hendrickson, 1991).

**Acknowledgements:** This work was supported by the Deutsche Forschungsgemeinschaft through SFB 473, the Fonds der Chemischen Industrie (to J.S.), NIH training grant AI07472 (to G.S.A.) and NIH grant GM49244 (to R.G.B.). K.S. designed and performed the experiments summarized in Figure 3-1.

**Table 3-1**

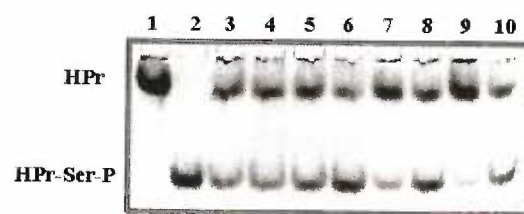
Data collection statistics for HPrK/P from *M. pneumoniae*.

<b>Resolution</b>	<b>30.0 - 2.49</b>	<b>2.59 - 2.49</b>
<b>Observations (#)</b>	336892	34222
<b>Unique Observations</b>	86543	9077
<b>Multiplicity</b>	3.9	3.8
<b>Completion (%)</b>	95.6	91.0
<b>Mean I/σ(I)</b>	16.8	4.0
<b>R<sub>iso</sub>#</b>	0.068	0.433

$$\# R_{iso} = \sum |I_i - I_{avg}| / I_{avg}$$

**Figure 3-1** Enzyme activity of HPrK/P following DTNB-labelling. **A.** Kinase activity of HPrK/P. Lanes 1 and 2 contain as controls, (His<sub>6</sub>)HPr and (His<sub>6</sub>)HPr(PhosphoSer46). Lanes 3 to 6 show HPrK/P kinase activity in the presence of 10  $\mu$ M ATP and in the presence (even numbers) or absence (odd numbers) of 10 mM fructose-1,6-bisphosphate (FBP). In lanes 3 and 4 the result of the kinase reaction in the presence of DTNB-labelled HPrK/P is shown, whereas lanes 5 and 6 represent the reference reaction including unlabelled HPrK/P. Lanes 7 to 10 show HPrK/P kinase activity in the presence of 50  $\mu$ M ATP and 5 mM P<sub>i</sub>, and in the presence (even numbers) or absence (odd numbers) of FBP (1 mM). Lanes 7 and 8 show the reaction with the DTNB-labelled HPrK/P; lanes 9 and 10 are the reference experiments. **B.** Phosphatase activity of HPrK/P. Lanes 1 to 4 show HPrK/P phosphatase activity in the presence (even numbers) or absence (odd numbers) of 5 mM P<sub>i</sub>. In lanes 1 and 2 the reaction mixture included HPrK/P following DTNB-titration. Lanes 3 and 4 show the respective reactions for the unlabelled HPrK/P. Lanes 5 and 6 are controls, containing (His<sub>6</sub>)HPr and (His<sub>6</sub>)HPr(PhosphoSer46).

A



B



## Chapter IV

### **Crystal Structure of HPr Kinase/Phosphatase from *Mycoplasma pneumoniae***

**Gregory S. Allen,<sup>1</sup> Katrin Steinhauer,<sup>2</sup>**

**Wolfgang Hillen,<sup>2</sup> Jörg Stülke,<sup>2</sup>**

**and Richard G. Brennan<sup>1\*</sup>**

<sup>1</sup> Oregon Health & Science University, Department of Biochemistry and Molecular  
Biology, L224

3181 SW Sam Jackson Park Road, Portland, OR 97201, USA

<sup>2</sup> Lehrstuhl für Mikrobiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg,  
Staudtstr. 5, D-91058 Erlangen, Germany

(Journal of Molecular Biology, in press)

## Summary

HPr kinase/phosphatase (HPrK/P) modifies Serine 46 of Histidine containing protein (HPr), the phosphorylation state of which is the control point of carbon catabolite repression in low G+C Gram-positive bacteria. To understand the structural mechanism by which HPrK/P carries out its dual, competing activities we determined the structure of full length HPrK/P from *Mycoplasma pneumoniae* (pdb Id: 1KNX) to 2.5 Å resolution. The enzyme forms a homo-hexamer with each subunit containing 2 domains connected by a short loop. The C-terminal domain contains the well-described P-loop (Walker A box) ATP binding motif and takes a fold similar to phosphoenolpyruvate carboxykinase (PEPCK) from *Escherichia coli* as recently described in other HPrK/P structures. As expected, the C-terminal domain is very similar to the C-terminal fragment of *Lactobacillus casei* HPrK/P and the C-terminal domain of *Staphylococcus xylosus* HPrK/P; and the N-terminal domain is very similar to the N-terminal domain of *Staphylococcus xylosus* HPrK/P. Unexpectedly, the N-terminal domain resembles UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-diaminopimelate ligase (MurE), yet the function of this domain is unclear. We discuss these



observations as well as the structural significance of mutations in the P-loop and HPrK/P family sequence motif.

**Keywords:** HPr kinase, catabolite repression, Ser/Thr kinase, Ser/Thr phosphatase, *Mycoplasma pneumoniae*

## **Introduction**

HPr kinase/phosphatase (HPrK/P) can both phosphorylate and dephosphorylate residue Serine 46 of the Histidine containing protein (HPr), a phosphate shuttle protein of the phosphoenolpyruvate:sugar phosphotransferase system (PTS). The activities of HPrK/P are regulated by the allosteric effector fructose-1,6-bisphosphate (FBP; favors kinase activity), and the concentration differential between ATP/GTP and inorganic phosphate ( $P_i$ ; favors phosphatase activity). To date some 32 HPrK/P sequences from various bacteria have been described. These genes are found primarily in low G+C Gram-positive bacteria and purple bacteria, but the spirochete *Treponema pallidum* also encodes the enzyme. An alignment of representative HPrK/P sequences shows homology and conservation throughout the coding region, albeit with stricter conservation in the C-terminal half of the predicted proteins (Fieulaine *et al.*, 2001). Given these

high homologies (e.g., *M. pneumoniae* HPrK/P, an outlier, is 53% similar to the predicted amino acid sequence of *Bacillus subtilis* HPrK/P) we expected the various HPrK/P orthologues might fulfill similar roles in the organisms in which they are found.

The role of HPrK/P in the metabolism of *B. subtilis*, which is the archetype of low G+C gram-positive bacteria, has been well defined. *B. subtilis*, like most bacteria, has distinct hierarchical preferences for carbon source (Saier *et al.*, 1996; Stulke *et al.*, 2000; Deutscher *et al.*, 2002); this phenomenon, known as carbon catabolite repression (CCR), is the end result of several pathways working in concert. In *B. subtilis* HPrK/P is a critical component of CCR; it senses the energy state of the cell through its interactions with ATP/GTP,  $P_i$ , and FBP (the concentration of FBP is higher in the presence of glucose than with other sugars) (Fujita *et al.*, 1979; Deutscher *et al.*, 1985). A high-energy state leads to repression or activation of transcription of genes encoding a plethora of extracellular enzymes, intracellular carbohydrate metabolic enzymes, some glycolytic enzymes and sporulation enzymes (Hueck *et al.*, 1994; Tobisch *et al.*, 1999; Moreno *et al.*, 2001; Yoshida *et al.*, 2001).

Repression of these genes is effected by catabolite control protein A (CcpA), (Henkin *et al.*, 1991) a LacI/GalR family member (Weickert *et al.*,

1992). CcpA binds to a 14 base-pair pseudo-palindromic consensus sequence, the catabolite response element (CRE),(Kim *et al.*, 1995) found in or near the promoters of genes known to be glucose repressed (Weickert *et al.*, 1990; Hueck *et al.*, 1994). The binding of CcpA to CRE sites is significantly enhanced in the presence of phospho-Ser46-HPr (Deutscher *et al.*, 1995) or phospho-Ser46-Crh, a non-PTS protein with 45% identity to HPr (Galinier *et al.*, 1999) HPrK/P either phosphorylates or dephosphorylates HPr or Crh at Ser46 in response to the ratio of  $P_i$ :[ATP/GTP]; where relatively high concentrations of ATP/GTP favor phosphorylation, while dephosphorylation occurs at relatively high concentrations of  $P_i$  (Jault *et al.*, 2000).

In contrast to *B. subtilis*, the role of HPrK/P in *Mycoplasma pneumoniae* is not as well understood. Although both *B. subtilis* HPrK/P and *M. pneumoniae* HPrK/P display *in vitro* phosphatase activity with concentrations of ATP and  $P_i$  of 200  $\mu$ M and 5 mM, respectively, if 10 mM FBP is added both enzymes regain kinase activity (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002); yet the enzymes diverge significantly in regard to their respective kinase activities and their regulation. *in vitro* kinase activity assays using *B. subtilis* HPrK/P show ATP stimulates phosphorylation cooperatively (half maximal activity at 1 mM ATP), while

FBP can stimulate phosphorylation at lower concentrations of ATP (half maximal activity at 1 mM FBP, 25  $\mu$ M ATP) (Jault *et al.*, 2000). *M. pneumoniae* HPrK/P kinase activity assays suggest that it is also activated cooperatively by ATP (half maximal activity at 10  $\mu$ M), but that FBP has no stimulatory effect on phosphorylation; it merely overcomes inhibition due to  $P_i$  (half maximal kinase activity 50  $\mu$ M ATP, 5 mM  $P_i$ , and 10  $\mu$ M FBP) (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002).

This shift to activation of the kinase activity of HPrK/P, and perhaps CCR, at a lower concentration of ATP in *M. pneumoniae* may correlate with its lifestyle as a parasite of the mammalian respiratory tract. *M. pneumoniae* is a member of the low G+C Gram-positive bacteria family, like *B. subtilis*, and probably evolved from a *B. subtilis*-like precursor through gene reduction driven by its parasitic lifestyle; indeed these bacteria have no TCA cycle genes nor do they respire (Hutchison *et al.*, 1999) However, while *M. pneumoniae* encodes homologues of the PTS (including HPr), homologues of CcpA and Crh are conspicuously absent. Consequently, to the extent that CCR might occur in *M. pneumoniae*, would require its mechanism to be different from that of *B. subtilis*.

We expected, therefore, that the differences in both lifestyle and metabolism between *B. subtilis* and *M. pneumoniae* inferred from a genomic

comparison were related, at least in part, to the different activities of their respective HPrK/Ps. Furthermore, the different activities of the two HPrK/Ps should have a structural basis for understanding them; and that any structural information should provide insight into the general enzyme mechanism of the HPrK/P family. In order to advance our state of knowledge of these wider issues we have determined the three-dimensional structure of *M. pneumoniae* HPrK/P.

Specifically we describe here the 2.5 Å resolution crystal structure of the HPrK/P from *Mycoplasma pneumoniae*, a 312 residue homo-hexamer. As anticipated the C-terminal domain of *M. pneumoniae* HPrK/P is virtually identical to the same domain of the *Lactobacillus casei* HPrK/P fragment (Fieulaine *et al.*, 2001) and to the full-length *Staphylococcus xylosus* HPrK/P (Marquez *et al.*, 2002). Structural homology searches using the C-terminal domain from *M. pneumoniae* HPrK/P also reveal a significant similarity to *E. coli* phosphoenolpyruvate carboxykinase (PEPCK) (Russell *et al.*, 2002). Structural homology searches using the N-terminal domain from *M. pneumoniae* HPrK/P suggest that it is related to the N-terminal, uridyl-binding domain of *E. coli* MurE, an enzyme which catalyzes the addition of meso-diaminopimelic acid to the nucleotide precursor UDPMurNAc-L-Ala-D-Glu, an intermediate step in cell wall peptidoglycan

synthesis (Gordon *et al.*, 2001). Finally, mutational analyses of the C-terminal domain of *M. pneumoniae* HPrK/P, combined with results from structural homology searches, indicate that ATP interacts with the enzyme at the P-loop.

## Results

### Hexameric Structure

The asymmetric unit contains a hexamer of the *Mycoplasma pneumoniae* HPrK/P (Fig. 4-1). The individual subunits of the hexamer have two domains: an N-terminal domain (residues 1-132) and C-terminal domain (residues 136-312) that are connected by a 3-residue stretch (residues 133-135). The topology of the N-terminal domain (NTD) is  $\beta 1(4-6)$ ,  $\alpha A(7-12)$ ,  $\beta 2(17-21)$ ,  $\alpha B(23-27)$ ,  $\beta 3(28-30)$ ,  $\alpha C(39-43)$ ,  $\beta 4(53-56)$ ,  $\alpha D(58-66)$ ,  $\alpha E(69-81)$ ,  $\beta 5(86-90)$ ,  $\alpha F(96-106)$ ,  $\beta 6(109-113)$ , and  $\alpha G(117-132)$  (Fig. 4-2), where the residues of each element are enclosed within parentheses. The NTD has a central 4-stranded beta sheet ( $\beta 2\beta 6\beta 5\beta 4$ ) where  $\beta 2$  is anti-parallel to the other strands. In addition strands  $\beta 1$  and  $\beta 3$  form a second, anti-parallel, beta sheet. Helix  $\alpha F$  packs against the 'top' of the central  $\beta$  sheet. Helices  $\alpha E$  and  $\alpha D$  form a triangle with the sheet, the vertex of which

points away from the sheet. This triangle is stabilized by interactions with  $\alpha F$  and  $\alpha C$  on either side. Helix  $\alpha B$  and strand  $\beta 3$  of the secondary sheet ( $\beta 1\beta 3$ ), pack against the 'bottom' of the main sheet with strand  $\beta 3$  nearly perpendicular to those of the main sheet. Helix  $\alpha A$  interacts with  $\alpha G$  as a quasi-coiled coil and these too are found on the 'bottom' of the central  $\beta$  sheet.

To compare the structures of the N-terminal domains of each subunit we computed the pair-wise root mean square deviations (rmsd) of their corresponding  $C_{\alpha}$  carbons (residues 1–132) using the F subunit as the probe. The F polypeptide was used because its electron density was the most well determined throughout its NTD. The NTD of F on E gave an rmsd = 0.31 Å, F on C an rmsd = 0.36 Å, F on B an rmsd = 0.50 Å, F on D an rmsd = 0.58 Å, and F on A an rmsd = 0.70 Å. The rmsd values show that the NTDs are virtually identical with only small variations

The NTD is connected to the C-terminal domain (CTD) by a short loop consisting of residues 133–135. The structure of the *M. pneumoniae* CTD, comprising residues 136 to 312, is essentially identical to the CTDs of the *Lactobacillus casei* and *Staphylococcus xylosus* HPrK/P; (Fieulaine *et al.*, 2001; Marquez *et al.*, 2002) where the rmsd is 1.7 Å for 158 corresponding  $C_{\alpha}$  atoms when comparing the B subunit to the *L. casei* HPrK/P fragment

and 1.2 Å for 152 corresponding C<sub>α</sub> atoms when comparing B subunit residues 133-289 to the *S. xylosus* CTD. Using the nomenclature of the *L. casei* structure, the CTD topology is βA(136-138), βB(140-145), βC(148-153), α1(157-170), βD(173-178), βE(180-185), βF(188-194), η1(195-197), βG(199-203), βH(207-210), α2(211-215), η2(216-218), βI' (219-221), βJ(223-226), βK(228-233), Eα(242-245), βK' (251-255), βK(258-267), α3(275-289), and α4(293-311) (Fig. 4-3) where the residues of each element are enclosed in parentheses. Residues 242-245 are either missing (subunit C), in a loop (subunit A, B, D, and F), or in a helix (subunit E), where they form Eα. Helix Eα (residues 242-245) contacts αE (residues 69-81) from the NTD of subunit B of a symmetry mate. The *L. casei* structure does not contain βI', Eα or βK'; and the *S. xylosus* structure does not contain Eα.

The CTD has a central 7-stranded mixed beta sheet (I'DBCJKK', where strands B and K' are anti-parallel), a 4-stranded antiparallel sheet (AEFI, with a beta hairpin between strands E and F), and a 2-stranded antiparallel sheet (GH). Strands βG and βH contain part of the HPrK/P family sequence motif, which is thought to mediate interaction with HPr (Reizer *et al.*, 1998; Hanson *et al.*, 2002). The P-loop, residues G<sub>154</sub>RSGIGKSE<sub>162</sub>, appears on the surface of the protein at the C-terminus of βC and extends into α1. There are 3 additional beta hairpins, connecting



strands A and B, K and K', and G and H and a poorly defined loop between J and K' (or E $\alpha$ ).

Additional structural details include the involvement of helix  $\alpha_4$ , the C-terminal tail of *M. pneumoniae* HPrK/P, in extensive contact with  $\beta G$  and  $\beta H$  of a neighboring CTD, which is not the 'dimer' partner. Also, residues 12-15, 126, and 127 from the NTD contact residues 184 –190 (the beta hairpin between  $\beta E$  and  $\beta F$  as well as part of  $\beta F$ ) of the CTD of its dimer partner, with a salt bridge formed between Asp15 and Arg184 of the dimer partner.

The six CTDs show slightly different conformations and overlaying the CTD of the B subunit, which appears to be in an ATP-binding competent conformation (see below), onto the five remaining CTDs reveals rmsds of 0.58 Å (A on B), 0.90 Å (C on B), 1.02 Å (F on B), 1.07 Å (D on B), and 1.77 Å (E on B). Evidently the CTDs of subunits A and B are identical within experimental error, while those of B and E are significantly different. However, the calculated rmsd variations are directly related to P-loop conformational variations (discussed below) and the core 150 residues are virtually identical. If the P-loop and the unstructured loop between J and K' are not considered, i.e., only residues 136-153:163-234:246-305, the rmsd for B on A becomes 0.48 Å, but that of B on E drops to 0.79 Å.

## Dimer

HPrK/P from *Enterococcus faecalis* has been shown to be functional as a dimer (Kravanja *et al.*, 1999). Hence, although *M. pneumoniae* HPrK/P is functional as a hexamer (Steinhauer, K. *et al.*, 2002) it may be viewed as a trimer of dimers, i.e., dimers AB, CD, and EF (Fig. 4-1). The subunits, which comprise the dimer, bury  $\sim 3600 \text{ \AA}^2$  of accessible surface area (Fig. 4-4). The interface between the subunits contains a 2-fold axis parallel to a radial line originating at the center of the hexamer. Nearly all inter-subunit NTD contacts are made with its dimer partner NTD and bury  $\sim 1800 \text{ \AA}^2$  per dimer.

The N-terminal interactions involved in dimerization include a water mediated hydrogen bond between the side-chain carboxyls of Glu125 and Glu125' (where prime indicates residues of the other subunit) and hydrogen bonds between the gamma hydroxyl of Ser121 and the amide nitrogen of Asn129' found in the quasi coiled coil,  $\alpha G - \alpha G$ , formed at the dimer interface. A hydrophobic patch, composed of residues Phe40, Leu45, and Phe46 from helix  $\alpha C$ , interacts with Leu43 and Phe63 from its dimer mate. The dimerization interface also contains four salt bridges (Glu60:Arg37'; Glu41:Arg59' plus symmetric cognates).

The C-terminal interactions involved in dimerization include hydrophobic contacts as well as multiple electrostatic interactions and hydrogen bonds. The N $\epsilon$  of Lys285 hydrogen bonds variously with the backbone oxygens of Phe146', Leu217', Gln218', and Thr220'. Side-chains of residues Glu144 and Phe146 in the loop between  $\beta$ B and  $\beta$ C of one subunit contact side-chains of residues Lys221 and Gln218 respectively, in  $\eta$ 2 and  $\beta$ I' of the other subunit; Lys221 and Glu144 form a four cornered salt bridge with their dimer mates (CD and EF dimers). Electrostatic interactions between dimer mates also include a salt bridge between Lys272 and Glu166 (CD and EF dimers). The disruption of some bonding patterns is evidence for long-range effects of conformational changes at the P-loop (see below).

## Hexamer

If the N-terminal domains are disregarded, the CTDs alone can be visualized as two plates of trimers with 32 point group symmetry, as was observed for the corresponding CTDs of *L. casei* and *S. xylosus* (Fig. 4-1). The C-terminal domains of each subunit contact each of the other CTDs to varying extents but interact primarily with their dimer partners and the CTDs of their trimer. The CTD interface of two trimer partners includes extensive,

primarily electrostatic, contacts between the C-terminal tail of one subunit and the ‘capping motif’ (the GH sheet,  $\alpha 2$ ,  $\eta 1$ , and  $\eta 2$ ; see Fig. 4-3)(Fieulaine *et al.*, 2001) of its neighbor (not the dimer partner). Also, the ‘K3 loop’, the loop between  $\beta K$  and  $\alpha 3$ , of one subunit closely approaches both the ‘capping motif’ and the P-loop helix,  $\alpha 1$ , of its neighbor (see below). The trimer CTD interface buries  $\sim 2,500 \text{ \AA}^2$  per subunit for a total of  $\sim 15,000 \text{ \AA}^2$  for both trimers of CTDs. The dimerization of the two trimers (with NTD contributions included) buries an additional  $\sim 11,000 \text{ \AA}^2$ . Therefore, in all the hexamer buries  $\sim 26,000 \text{ \AA}^2$  of surface area.

## Discussion

### P-loop

Crystallization of *M. pneumoniae* HPrK/P requires the presence of AMPPNP,(Steinhauer, K. *et al.*, 2002) and we expected to find this non-hydrolysable ATP analogue bound to the P-loop (residues  $G_{154}RSGIGKSE_{162}$ ). Although we find no evidence for the nucleotide in any of the six independent subunits, we do see evidence for an anion, most likely chloride, which is present as a crystallization reagent at 1 mM, in subunits A and B. Alternatively, but less likely, this anion could be a tightly bound

phosphate that has copurified with the protein. Regardless, the anion interacts with the main chain nitrogens of P-loop residues Gly157, Ile158, Gly159 and Lys160 (Fig. 4-5). The four remaining P-loops differ in conformation, where helix-like phi-psi angles occur variously at residues 159(D), 158(C and F), and 157(E), defining the beginning of the helix  $\alpha 1$ , which follows the P-loop. In the A and B subunits this helix begins at residue 160. Thus, the various conformations of the loop are: 'open' and thus able to bind ATP or  $P_i$  (subunits A and B) and 'closed' and folded over the putative nucleotide-binding pocket (subunits C, D, E, and F) (Fig. 4-5). The B factors for P-loop backbone atoms range from 45 – 65 Å<sup>2</sup> in subunits A, B, C, and E but in subunits D and F the B factors for these atoms range from 65 – 100 Å<sup>2</sup> indicating that the P-loop is flexible.

In comparing the most open nucleotide binding pocket, that of the B subunit, to the most closed pocket, that of the E subunit, a pathway for the movements undertaken by the P-loop to bind to the ATP can be envisioned. In both subunits, residues Ala154 and Lys160 are virtually superimposable, while the intervening residues undergo large changes, underscored by the 5 Å displacement of the C $_{\alpha}$  atom of Ile158 (Fig. 4-5). In the open form, the hydroxyl group of Ser156(B) interacts with the chloride ion, which also contacts Lys160(B). In its closed form the Ser156(E) hydroxyl group

hydrogen bonds with the backbone nitrogen of Gly159(E), thereby capping helix  $\alpha 1$  (Aurora *et al.*, 1998). Interestingly, if Ser156, which is conserved throughout the HPrK/P family, is mutated to alanine we see a four-fold increase in phosphatase activity while kinase activity is unaffected (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). However, a threonine substitution of Ser156 abrogates phosphatase activity and diminishes kinase activity four-fold (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). This suggests that Ser156 is an important regulator of the catalytic activities of HPrK/Ps. Perhaps threonine is better than serine at capping helix  $\alpha 1$ , which closes the binding pocket and diminishes ATP or phosphate binding and enzyme activity. In contrast alanine might not be effective at capping the helix, the result of which favors the open conformation and thereby ATP or phosphate binding, and an increase in phosphatase activity.

### **Structural Homologues of the CTD**

In order to gain insight into the kinase and phosphatase activities of *M. pneumoniae* HPrK/P, DALI (Holm *et al.*, 1995) searches were performed using the CTD and full-length protein. Since using the full-length protein and the CTDs of the B or E subunit gave similar results we report only the

results of a typical search. Using the CTD of the B subunit DALI returned the HPrK/P fragment from *L. casei* (1jb1-A) ( $Z = 23.2$ ; rmsd = 1.8 Å for 158 residues) and PEPCK from *E. coli* (1ayl) ( $Z = 7.3$ ; rmsd = 2.8 Å for 126 residues), (where a self-alignment gives  $Z = 32.2$  and a  $Z < 2.0$  suggests no structural similarity). Structural homology with PEPCK was previously reported for both the *L. casei* HPrK/P fragment (Russell *et al.*, 2002) and the *S. xylosus* HPrK/P (Marquez *et al.*, 2002) structures.

As expected, visual inspection of the *L. casei* HPrK/P overlay of *M. pneumoniae* HPrK/P shows that the structures of the CTD of each enzyme are virtually identical (Fig. 4-6A). The structural alignment of subunit B of *M. pneumoniae* HPrK/P with *L. casei* HPrK/P, optimized with LSQMAN, (Kleywegt, 1999) gives an rmsd of 1.3 Å for 153 of the 161 corresponding C $\alpha$  atoms. Notable differences include the ~5 Å movement of the end of the loop between  $\beta$ E and  $\beta$ F, which contacts the N-terminal domain. Also the entire GH sheet, corresponding to residues 199 to 210 in our structure, has shifted, en masse, ~5 Å. The 'K3 loop', residues 265-273, and the C-terminal tail,  $\alpha$ 4, have also shifted but in all these differences the actual secondary structure elements are homologous in both models albeit with some small translation or rotation. Finally, in *M. pneumoniae* HPrK/P  $\beta$ K is four residues longer at its N-terminus where it forms part of a beta

hairpin and continues the sheet with  $\beta K'$ , which is not seen in *L. casei* HPrK/P.

The *E. coli* PEPCK overlays the *M. pneumoniae* CTD with an rmsd = 1.6 Å for 119 non-contiguous C $\alpha$  atoms, as determined by LSQMAN. The structural alignment includes the P-loop and the HPrK/P family sequence motif of *M. pneumoniae* HPrK/P ( $\beta$  strands G and H). Essentially, the entire central beta sheet of the CTD of *M. pneumoniae* HPrK/P aligns well with the corresponding structure of PEPCK. The two enzymes have the kinase-1a, kinase-2, and possibly, kinase-3 motifs found in nucleotide binding sites (see below) (Traut, 1994).

The full length *S. xylosus* HPrK/P structure, which was reported while this manuscript was in preparation, is also structurally homologous to *M. pneumoniae* HPrK/P (Marquez *et al.*, 2002). In fact the CTD of HPrK/P from *S. xylosus* (B subunit, residues 133-298) is virtually identical to *M. pneumoniae* HPrK/P (B subunit, residues 136-301). The beta-sheet structures of both polypeptides superimpose exactly, the P-loops of both enzymes have the same conformation, and the GH loops are identical; the rmsd = 1.2 Å for 152 corresponding C $\alpha$  atoms when the CTDs alone are compared (Fig. 4-6B). Interestingly the K3 loops take slightly different conformations, whereby loop residues 262-264 are about 3 Å further away



from the putative ATP binding site in *S. xylosus* HPrK/P. Figure 4-6C shows the alignment of the three HPrK/P structures solved to date.

### **The Kinase Motifs**

Extensive mutational studies on the *M. pneumoniae* HPrK/P have been done (Hanson *et al.*, 2002; Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). To understand the structural basis of their effects on the kinase and phosphatase activities of the enzyme, we have modelled the ATP and magnesium from the PEPCK structure into the putative nucleotide binding site of the B subunit of *M. pneumoniae* HPrK/P (Figs. 4-7 & 4-8). We used knowledge of the ATP and magnesium binding network and water structure of PEPCK (Matte *et al.*, 1997) to infer homologous interactions in *M. pneumoniae* HPrK/P. With respect to the kinase-1a motif (P-loop; residues 154-162) and the kinase-2 motif (Asp176 and Asp177) functional assignment is straightforward due to sequence and structural homology between the HPrK/P and PEPCK enzymes. Also His139 and the HPrK/P family sequence motif (the GH loop) have homologues in PEPCK. But the putative kinase-3 motif (K3 loop of HPrK/P), although similar to that of PEPCK, is less obvious and does not fit the canonical kinase-3 motif.

### **Kinase-1a**

In the kinase-1a motif only the lysine is invariant. However, the three glycines of the *M. pneumoniae* HPrK/P P-loop (G<sub>154</sub>RSGIGKSE<sub>162</sub>) are conserved throughout the HPrK/P family (Reizer *et al.*, 1998; Fieulaine *et al.*, 2001) as well as in PEPCK. We would expect that the glycines of the P-loop are catalytically important probably for conformational reasons; (Matte *et al.*, 1997) the presence of three glycines would facilitate the necessary hinge motion of the P-loop upon binding ATP (Figs. 4-5 & 4-7). In accord with this inherent flexibility, there are large changes in the phi-psi angles of two of these glycines, Gly157 and Gly159, in their open form (see below for open form values) as compared to their values in the closed form  $\{(-67^\circ, -32^\circ) \text{ and } (-61^\circ, -29^\circ) \text{ respectively}\}$ . The critical importance of each glycine is highlighted by mutational analysis of the *M. pneumoniae* HPrK/P as substitution of any one of them dramatically effects both kinase and phosphatase activities (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002).

Specifically, replacement of Gly154 by alanine abolishes phosphatase activity and reduces kinase activity four-fold (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). Noting the phi-psi angles of Gly154 in the B subunit ( $141^\circ, 178^\circ$ ), which are similar in all six independent *M. pneumoniae* HPrK/P Gly154 residues, we conclude that glycine is required

because only glycine can occupy this part of Ramachandran space. Likewise, replacement of Gly157 by alanine resulted in an enzyme with a four-fold reduction in kinase activity and very weak phosphatase activity (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). The amide nitrogen of Gly157 of subunits A and B interacts with the bound anion, which occupies the position of the beta phosphate of ATP. The phi-psi angles of Gly157 of the B subunit ( $93^\circ$ ,  $1^\circ$ ) probably also require glycine as these dihedral angles are not favored by other residues. Finally, the alanine mutation of Gly159 completely inactivates both catalytic functions of the enzyme (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). The corresponding mutation of *B. subtilis* HPrK/P has similar effects: it has no phosphatase activity while the kinase activity is drastically reduced (Hanson *et al.*, 2002). Again, the phi-psi angles of Gly159 in the open form of subunit B ( $130^\circ$ ,  $12^\circ$ ) are energetically unfavorable for alanine or any other residue.

The similarities of the P-loops of the *M. pneumoniae* HPrK/P ( $G_{154}RSGIGKSE_{162}$ ) and *E. coli* PEPCK ( $G_{248}LSGTGKTT_{256}$ ) are evident. The lysine found in these sequences is invariant in kinase-1a motifs, but in the related kinase-1b motif, which is found in phosphofructokinase (PFru)(Shirakihara *et al.*, 1988), an arginine is substituted for the lysine. The guanidinium moiety interacts with oxygens of both the beta and gamma

phosphates of ATP. Interestingly, an arginine (or alanine) substitution of the lysine in *M. pneumoniae* HPrK/P destroys all catalytic activity (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). The *in silico* mutation of K160R places the guanidinium group of arginine about 2 Å closer to the gamma phosphate of ATP, implying that arginine is simply too bulky to fit into the nucleotide binding pocket of HPrK/P.

Additional insight into the functionality of the P-loop of *M. pneumoniae* HPrK/P was gained by substitutions of Ser161. The homologous residue in PEPCK, Thr255, directly coordinates a bound magnesium. In accord with a role in cation binding, substitution of Ser161 in *M. pneumoniae* HPrK/P with an alanine reduces kinase activity 10-fold (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). This substitution also abolishes phosphatase activity. The Ser161Thr substitution reduces kinase activity only 2-fold. However, phosphatase activity is very weak suggesting an unbranched hydroxyl group is key for this activity. In sum these mutational data indicate that Ser161 directly coordinates the magnesium ion necessary for the catalytic activity of *M. pneumoniae* HPrK/P (Fig. 4-7).

Residue Glu162, which is conserved in HPrK/P P-loops, interacts with the putative anion found at the ATP binding site in our A and B

subunits as does the carboxylate of Glu163 of *L. casei* HPrK/P (Fieulaine *et al.*, 2001). This implies that these glutamine residues are protonated at neutral pH or that they share a proton with the anion. Modelling reveals that the Glu162 side chain must move upon ATP binding as the carboxylate clashes with the beta phosphate of the ATP (Fig. 4-7). In PEPCK the homologous residue is a threonine, which does not clash with the beta phosphate of ATP. When Glu162 is substituted with an aspartate, the kinase activity is unaffected but phosphatase activity is abolished (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). These data indicate that Glu162 is important for kinase activity but plays a critical role in the phosphatase catalytic mechanism.

### **Kinase-2 motif**

Residues Asp268 and Asp269 in PEPCK form another important part of the ATP binding pocket, the kinase-2 motif, (Traut, 1994) which interacts with magnesium through a water intermediary. These residues superimpose upon Asp177 and Asp178 in *M. pneumoniae* HPrK/P. Alanine mutations of either of the corresponding residues in the *B. subtilis* HPrK/P virtually abolishes both activities (Galinier *et al.*, 2002). The sidechain carboxyl group of Asp177 hydrogen bonds to the hydroxyl of Ser161 and a water in the

binding pocket similar to residues Asp268 and Thr255 in PEPCK (Fig. 4-7), which are involved in cation binding (Matte *et al.*, 1997).

However, in the *M. pneumoniae* HPrK/P structure, Asp178 forms a salt bridge with the N $\delta$  ring nitrogen of His139 contorting the Asp178 backbone and placing its phi-psi dihedral angles in the disallowed region of Ramachandran space. Breaking the salt bridge between Asp178 and His139 would allow the imidazole ring of His139 to rotate freely and form a hydrogen bond to the gamma phosphate of ATP as observed for the corresponding histidine (His232) in PEPCK.

Residues His139 and Gly140 do not belong to a kinase motif, yet we note here that they are conserved in the HPrK/P family, whereas in PEPCK a cysteine has replaced the glycine. Interestingly, substitution of Gly140 with alanine increases *M. pneumoniae* HPrK/P phosphatase activity 6-fold without affecting kinase activity (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). Since we see a significant rotation about the His-Gly peptide, with respect to the His-Cys peptide of PEPCK, we postulate that a glycine at position 140 may be required for salt-bridge formation between Asp178 and His139.

#### **HPrK/P family sequence motif**

A stretch of conserved sequence found throughout the HPrK/P family, is found between residues 200 and 220 in *M. pneumoniae* HPrK/P (Fig. 4-6C) (Reizer *et al.*, 1998). In the *M. pneumoniae* structure this corresponds to  $\beta$ G- $\beta$ H- $\alpha$ 2 (Fig. 4-3 and 4-6C) and represents most of the so-called ‘capping motif’, which was defined for the *L. casei* HPrK/P (Fieulaine *et al.*, 2001). The  $\beta$ G- $\beta$ H sheet is partially buried but the hairpin, residues 203–206, is solvent exposed and close to the P-loop (Fig. 4-7). Substitution of  $\beta$ H residue Gly207 with alanine completely inactivates both catalytic functions of the enzyme (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002); the phi-psi angles of Gly207 of subunit B, which are typical of the other subunits, are (109°, -150°) and therefore, energetically accommodate only glycine at this location.

In the PEPCK structure the guanidinium group of Arg333 hydrogen bonds with an oxygen of the gamma phosphate of ATP. Although PEPCK Arg333 does not superimpose directly upon any *M. pneumoniae* HPrK/P residue, Arg204 of the  $\beta$  hairpin between  $\beta$ G and  $\beta$ H is proximal (4.4 Å C $\beta$  to C $\beta$ ) and would be capable of contacting the gamma phosphate of the superimposed ATP from PEPCK (Fig. 4-7). This arginine seems a likely functional homologue of Arg333. Mutation of the conserved Arg204, to a lysine in *M. pneumoniae* HPrK/P, significantly reduces phosphatase activity

without affecting kinase activity (Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002). While a lysine at this site could provide the salt bridge with the gamma phosphate of ATP, the planar guanidinium group of arginine may be necessary to provide stabilization of the transition state complex of phospho-Ser46-HPr during its dephosphorylation by HPrK/P, in a manner analogous to the tyrosine phosphatase from *Yersinia pestis* (Zhang *et al.*, 1994; Jackson *et al.*, 2001).

### **Kinase-3**

The adenine and ribose interactions made by PEPCK are provided by secondary structure elements far removed in sequence from the P-loop: the guanidinium group of Arg449 provides direct contacts to the ribose ring oxygen; the main chain carbonyl of Ile450 and the side chain oxygen of Thr455 contact the N6 nitrogen of the adenine base. These residues are in a loop and the beginning of an  $\alpha$  helix. The strand preceding the loop (residues 435-441), the loop (residues 442-449), and the helix (residues 452-464) of PEPCK are homologous to strand  $\beta$ K (residues 260-265), the K3 loop (residues 266-271), and helix  $\alpha$ 3 (residues 272-284) of *M. pneumoniae* HPrK/P, respectively (Fig. 4-8). In the HPrK/P structure the loop is significantly shortened, and the N-terminal portion of the helix has moved about 3 Å, but the resemblance of the homologous portions is obvious. More



specifically we suggest that residues Ser268 and Thr273 of *M. pneumoniae* HPrK/P are homologues of Ile450 and Thr455, respectively, and are part of the kinase-3 motif (Fig. 4-8).

On a given side of the HPrK/P hexamer three CTDs assemble in such a way that lines connecting the putative ATP binding pockets would form an equilateral triangle, the sides of which ( $\sim 30$  Å in length) pass through the K3 loops. In our model of ATP bound to the B subunit of *M. pneumoniae* HPrK/P (Fig. 4-8), the ATP clashes with the K3 loop of the neighboring E subunit indicating that the binding of ATP to HPrK/P necessarily induces structural changes in this loop, requires this loop to move, perhaps by HPr binding, or both. After loop movement residues 268 and 273 from the K3 loop of subunit E could contact the adenine ring. The movement of the K3 loop of monomer E would also change the shape of the nucleobase binding site in the E subunit thus providing a mechanism for presumed cooperativity of ATP binding (Jault *et al.*, 2000; Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J., 2002).

### **Structural Homologues of the NTD**

The functional relevance of the 132 residue N-terminal domain of *M. pneumoniae* HPrK/P is unknown. In an attempt to gain insight into the

activity of this domain a second DALI search was carried out. Using the NTD of the F subunit, which had the best experimental electron density for this domain, a good match was found with the N-terminal domain of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-diaminopimelate ligase (MurE) from *E.coli* ( $Z=6.7$ ; pdb ID 1e8c-A) (Gordon *et al.*, 2001).

Specifically, the N-terminal 101 amino acid residues of MurE align with the NTD of HPrK/P with an rmsd = 2.8 Å for 86 residues. The overlay of MurE on the NTD of *M. pneumoniae* HPrK/P (Fig. 4-9A) indicates general agreement of their topology and connectivity, although there are several insertions of one or more residues in each secondary structure element of the NTD with respect to MurE. Three of the four  $\beta$  strands ( $\beta_6\beta_5\beta_4$ ) of the central  $\beta$  sheet of the NTD are identical in MurE.

Residues 1-58 of the NTD have comparable secondary structure elements in MurE with the exception of the insertions of  $\alpha_B$  and  $\alpha_C$  in the NTD. The NTD also has a significant insertion, the triangle of  $\alpha_D$  and  $\alpha_E$ , which represents a ten residue lengthening and unfolding of the center of what is a single helix in MurE. The remainder of the NTD (residues 81 to 132) is essentially homologous to MurE with the exception of  $\alpha_F$  of the NTD, which is a long loop in MurE; a portion of this loop, residues 72-76, is a strand of the central  $\beta$  sheet of MurE.

Gordon *et al.* describe how the UDP moiety of the substrate binds in a cleft between the N-terminal and central domains of MurE (Gordon *et al.*, 2001). Residues H<sub>43</sub> QA<sub>45</sub>, found at the end of  $\beta$ 5, interact directly with the two phosphates of the UDP moiety (Fig. 4-9A). Intriguingly, the corresponding region of the NTD, G<sub>57</sub> KRE<sub>60</sub> is highly conserved in the HPrK/P family (Reizer *et al.*, 1998). This is a part of the dimerization interface between two NTDs, which contains four salt bridges {Glu60:Arg37 (internal), Glu41:Arg59' and their symmetric cognates} and numerous charged sidechains that surround a large solvent filled pocket. Several solvent molecules display low B-factors (half the solvent average). Given its role in ligand binding in MurE, the corresponding pocket of *M. pneumoniae* HPrK/P might well serve as a binding site for a small molecule allosteric regulator.

As expected the NTD of the HPrK/P of *S. xylosus* (residues 1-129) was most similar to the NTD of the *M. pneumoniae* HPrK/P (residues 1-132) with an rmsd = 1.9 Å for 127 residues (Marquez *et al.*, 2002). In overlay it appears that, although the two structures have identical secondary structural elements in all cases, several helices are out of register. In fact, helices  $\alpha$ C,  $\alpha$ E, and  $\alpha$ F are in disagreement by greater than 2 Å. The portions of helices  $\alpha$ E and  $\alpha$ F that are out of register are on the side opposite the common

dimer interface and phosphate binding site of *S. xylosus* HPrK/P; but  $\alpha$ C is involved in dimer-dimer contacts.

Sequence alignment of HPrK/Ps suggests that a region of the NTD encompassing  $\alpha$ C,  $\beta$ 4, and  $\alpha$ D, is conserved throughout the family (Fieulaine *et al.*, 2001); glycines 44 and 57 and Glu60 of *M. pneumoniae* HPrK/P are strictly conserved (Fig 4-6C). Marquez *et al.* show that this is the region of *S. xylosus* NTD inter-subunit contacts and protein-phosphate contacts (Marquez *et al.*, 2002). The overlay of the two HPrK/P structures demonstrates that phosphate binding moves the C-terminus of  $\alpha$ C by 5 Å and rearranges the following loop (residues 45 – 52 of *M. pneumoniae* HPrK/P) suggesting that the first conserved glycine, Gly44, facilitates this restructuring (Fig. 4-9B). The *S. xylosus* structure also reveals the second conserved glycine, Gly57, is close to the phosphate binding site and the conserved glutamate forms a salt bridge with Arg33, which interacts directly with the phosphate. These observations, together with our earlier comparison of the *M. pneumoniae* NTD to the MurE ligand binding site, support our conjecture that this is the binding site of an allosteric effector that acts here *in vivo* to modify the activity of the HPrK/P enzyme.

## Mechanistic Implications

In summary, we have determined the full-length structure of HPr kinase/phosphatase from *Mycoplasma pneumoniae* to 2.5 Å resolution. The asymmetric unit of the crystal contains a hexamer, which is the functional unit of the enzyme (Steinhauer, K. *et al.*, 2002). Each subunit is composed of two domains: the NTD and CTD. Using the DALI search routine we have identified proteins that show strong structural homology with each domain of *M. pneumoniae* HPrK/P. As we had expected *M. pneumoniae* HPrK/P is virtually identical to the C-terminal fragment of *L. casei* HPrK/P and the full-length *S. xylosus* HPrK/P. Using the CTD alone, the DALI search has revealed another related structure, PEPCK from *E. coli*. In fact the ATP binding region of PEPCK can be superimposed upon the putative ATP binding pocket of the 'open' subunits of *M. pneumoniae* HPrK/P; the beta phosphate overlays, nearly perfectly onto a putative chloride ion found in the A and B subunits at the P-loop. We therefore expect that HprK/P would bind ATP in the same manner as PEPCK binds ATP (Figs. 4-7 & 4-8), likely even taking the *syn* conformation of the adenine base that is found in the PEPCK structure. In either the *syn* or *anti* conformation the base would certainly clash with the K3 loop of the neighboring subunit suggesting that movement of this loop either accompanies, or precedes binding of ATP. Furthermore, the proximity of the K3 loop to two

neighboring ATP binding sites provides a clear linkage between sites, which may be used to attain cooperative ATP binding as seen with *B. subtilis* HPrK/P (Jault *et al.*, 2000).

A series of kinase defective mutations further support our model of ATP binding to the *M. pneumoniae* HPrK/P and begins to suggest a mechanistic interpretation. Alanine mutations of strictly conserved glycines in the P-loop, Gly154, Gly157 and Gly159, likely decrease kinase activity by decreasing the conformational flexibility of the loop and thereby increasing the energetic cost associated with binding ATP. Alanine substitution of Lys160 produces a dead enzyme because the  $\epsilon$  amino group may stabilize the developing charge on the  $\beta$  phosphate of ATP. Substitution of Lys160 with arginine also produces a dead enzyme possibly because the guanidinium group is too bulky to allow productive ATP binding.

Finally, a search of the structural database using the NTD returned the N-terminal, uridyl-binding portion of MurE as a likely homologue. Furthermore, the overlay of the NTD of *S. xylosus* HPrK/P on the NTD of *M. pneumoniae* HPrK/P is virtually identical, with the exception of the phosphate-binding region of *S. xylosus* HPrK/P; this phosphate-binding region is conserved among HPrK/Ps and overlays the uridyl-binding region of MurE. We therefore speculate that the bound phosphates in *S. xylosus*

HPrK/P may not be artifacts of crystallization but are the binding sites for anionic allosteric effectors. We do see some large scale structural differences that would be expected upon binding of allosteric effector. As noted above, when the CTDs are superimposed the rmsd is 1.2 Å (153 residues) and when NTDs are superimposed the rmsd is 1.9 Å (127 residues); but when we superimpose the *S. xylosus* HPrK/P (dimer in asymmetric unit) on the AB dimer of our structure the rmsd is 2.7 Å (540 residues).

## Experimental Procedures

Overexpression of selenomethionyl *M. pneumoniae* HPrK/P, its purification, crystallization, cryo-cooling procedure and intensity data collection have been described in detail elsewhere (Steinhauer, K. *et al.*, 2002). Briefly, isomorphous selenomethionyl substituted HPrK/P crystals were grown in 0.9 M NaCl, 50 mM Tris pH 7.6, 3% PEG 8000, and 1 mM AMPPNP. The space group is P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. Intensity data were collected at SSRL on BL 9-2 using Blu-Ice data acquisition software and a Quantum 4 CCD detector (ADSC). Three wavelengths were collected in 180° swaths, using 1° oscillations, before the energy was changed. d\*Trek v7.0 was used to process the MAD data while v7.2 was used for the native data (see Table 1) (Pflugrath, 1999). Solve v1.18 was used to determine the selenium atom

structure and Resolve v1.05 to implement maximum likelihood density modification (Terwilliger *et al.*, 1999; Terwilliger, 2000). The resulting electron density map was of excellent quality; greater than 90% of the C $\alpha$  backbone was traced using the auto-fit feature in xfit from XtalView v4.1 (McRee, 1999). The four selenomethionines in each subunit allowed unambiguous sequence assignment.

CNS v1.1 was used to refine the resulting hexamer model (Brunger *et al.*, 1998). First, multiple rounds of simulated annealing and model rebuilding were performed. Composite omit maps were generated in CNS to help trace ambiguous regions of electron density of the protein structure and place solvent molecules. Simultaneous positional and B-factor refinement was used to complete the model. The current  $R_{\text{work}}$  is 22.7% and  $R_{\text{free}}$  is 28.0% to 2.5 Å resolution. The final model contains the following residues: subunit A (1-235, 242-309); subunit B (1-236, 241-308); subunit C (1-234, 246-311); subunit D (1-236, 240-307); subunit E (1-238, 240-311); subunit F (1-237, 239-307); and 14,435 protein and 282 solvent atoms. None of the polypeptide chains of the six subunits are contiguous due to a region of uniformly poor electron density between  $\beta$ J and  $\beta$ K' of the CTD. Residues Glu236(D), Glu236(E), Gln238(E), Glu236(F), Lys237(F), and Thr239(F) (subunits in parentheses) were modelled as alanines. The N-termini also had



regions of poorer density and the following residues were modeled as alanines: Lys3(A); Lys2(B); Lys2(C); Lys2(E) & Lys3(E). Additionally the electron density maps at the C-termini became poor after residue 305 in some subunits; the following residues were modelled as alanines: Glu311(C); Lys308(E) & Lys309(E); Lys306(F) & Arg307(F). Chains A and D had relatively high B factors ( $> 100 \text{ \AA}^2$  for A;  $> 80 \text{ \AA}^2$  for D) for residues 60-120 ( $B_{\text{ave}} \sim 125 \text{ \AA}^2$ ); some side-chains could not be placed and the following residues were modelled as alanines: Val70(A), Gln73(A), Gln74(A); Lys67(D), Glu71(D), Val98(D), Leu99(D). Finally, the side-chain of Arginine 155 in the P-loop of chain F had such poor density it was also modelled as an alanine. Buried surface areas were calculated using a 1.4 Å probe radius as implemented in CNS (Brunger *et al.*, 1998).

All figures were generated in SwissPDB viewer v. 3.5,(Peitsch *et al.*, 1995) and rendered in POV-Ray<sup>TM</sup> v. 3.1 (<http://www.povray.org>) except Figure 4-5, which was generated with O(Jones *et al.*, 1991) and Figure 4-6C, which was generated with Alscript (Barton, 1993).

**Acknowledgements:** This work was supported by the Deutsche Forschungsgemeinschaft through SFB 473, the Fonds der Chemischen

Industrie (to J.S.); NIH training grant AI07472 and N. L. Tartar Fellowship (to G.S.A.); and NIH grant GM49244 (to R.G.B.).

**Table 4-1: Selected Crystallographic Data**

Data Sets (Å)	1	2	3	Native
	( $\lambda = 0.93218$ )	$\lambda = (0.97919)$	( $\lambda = 0.97903$ )	( $\lambda = 0.8856$ )
a	117.50	117.52	117.58	117.73
b	128.93	128.96	129.01	127.84
c	172.40	172.51	172.59	170.77
Resolution (Å)	30.0 – 3.03	30.0 – 3.01	30.0 – 3.04	30.0 – 2.49
Completeness (%)	98.5	98.5	98.5	97.1
Overall $R_{\text{merge}}$	6.7	6.9	5.5	6.8
Overall $\langle I/\sigma(I) \rangle$	7.9	7.9	9.5	22.0
# Reflections (#)	191393	195937	189457	347151
# Unique Refs. (#) *	96847	99160	96085	91137
High res. Shell (Å)	3.19 – 3.03	3.16 – 3.01	3.2 – 3.04	2.59 – 2.49
Completeness (%)	98.7	98.7	98.7	94.8
$R_{\text{sym}}$	25.4	26.5	21.2	41.8
$\langle I/\sigma(I) \rangle$	2.4	2.2	2.9	3.0
Solve figure of merit		0.36		
Resolve figure or merit		0.57		
Refinement				
$R_{\text{free}}$ (%)				28.0
$R_{\text{work}}$ (%)				22.7
atoms (#)				14,728
rmsd bond (Å)				0.007
rmsd angle (deg.)				1.2
rmsd B (Å <sup>2</sup> )				4.09
$\langle B \rangle$ (Å <sup>2</sup> )				53.8

\* Friedel pairs treated separately for MAD data

$$R_{\text{merge}} = \sum_h \sum_l |I_{hl} - \langle I_h \rangle| / \sum_h \sum_l \langle I_h \rangle$$

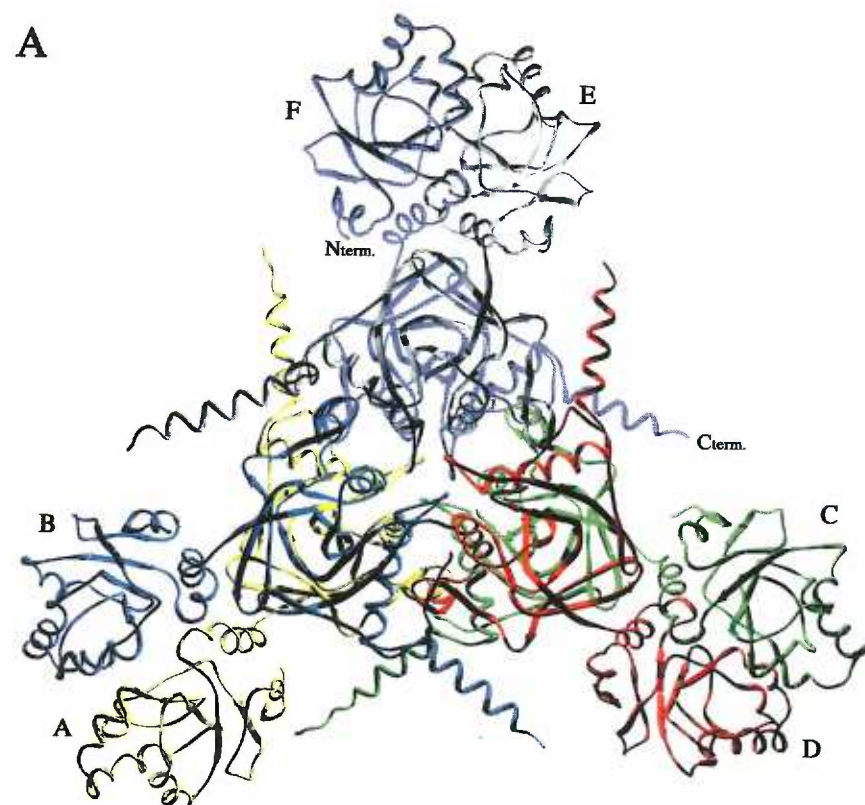
$$R_{\text{sym}} = \sum_h \sum_l |I_{hkl} - \langle I_{hk} \rangle| / \sum_h \sum_l \langle I_{hk} \rangle$$

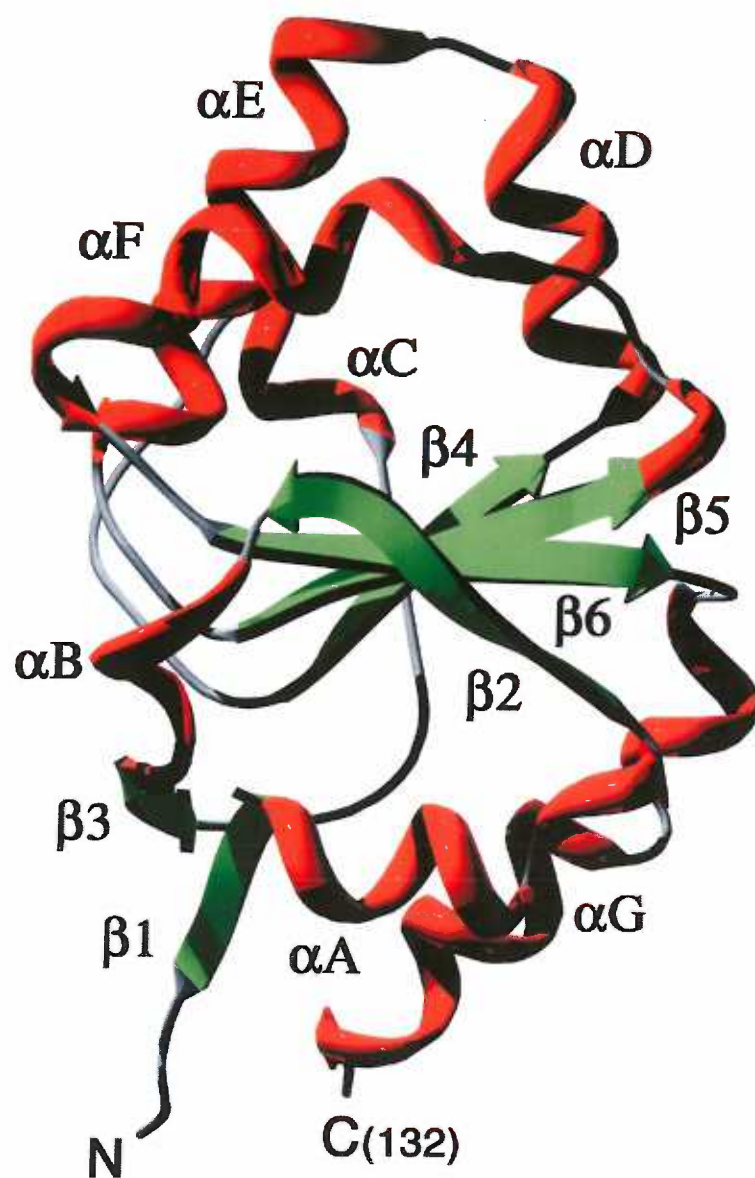
Figure of merit =  $\langle |\sum P(\alpha) e^{i\alpha} / P(\alpha)| \rangle$  where  $\alpha$  is the phase and  $P(\alpha)$  is the phase probability distribution

$$R_{\text{work}} = \sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$$

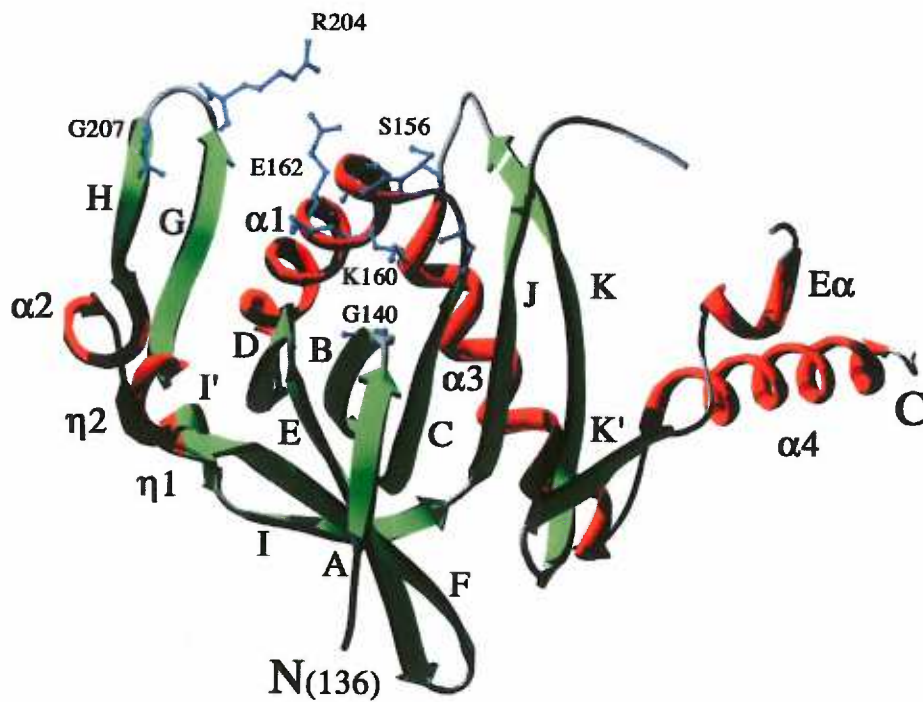
$$R_{\text{free}} = \sum |F_{\text{obs}}(\text{test})| - |F_{\text{calc}}(\text{test})| / \sum |F_{\text{obs}}(\text{test})| \text{ where test set comprises 5\% of observed reflections selected randomly}$$

**Figure 4-1:** The structure of *M. pneumoniae* HPrK/P. **A.** View of the hexamer along the molecular 3-fold axis. The color code is as follows: A subunit in yellow; B subunit in blue; C subunit is green; D subunit is red; E subunit is grey; F subunit is purple. The N and C termini of the F subunit are labelled. **B.** View of the hexamer along a 2-fold axis perpendicular to the view in A. The D and E subunits have been omitted for clarity.

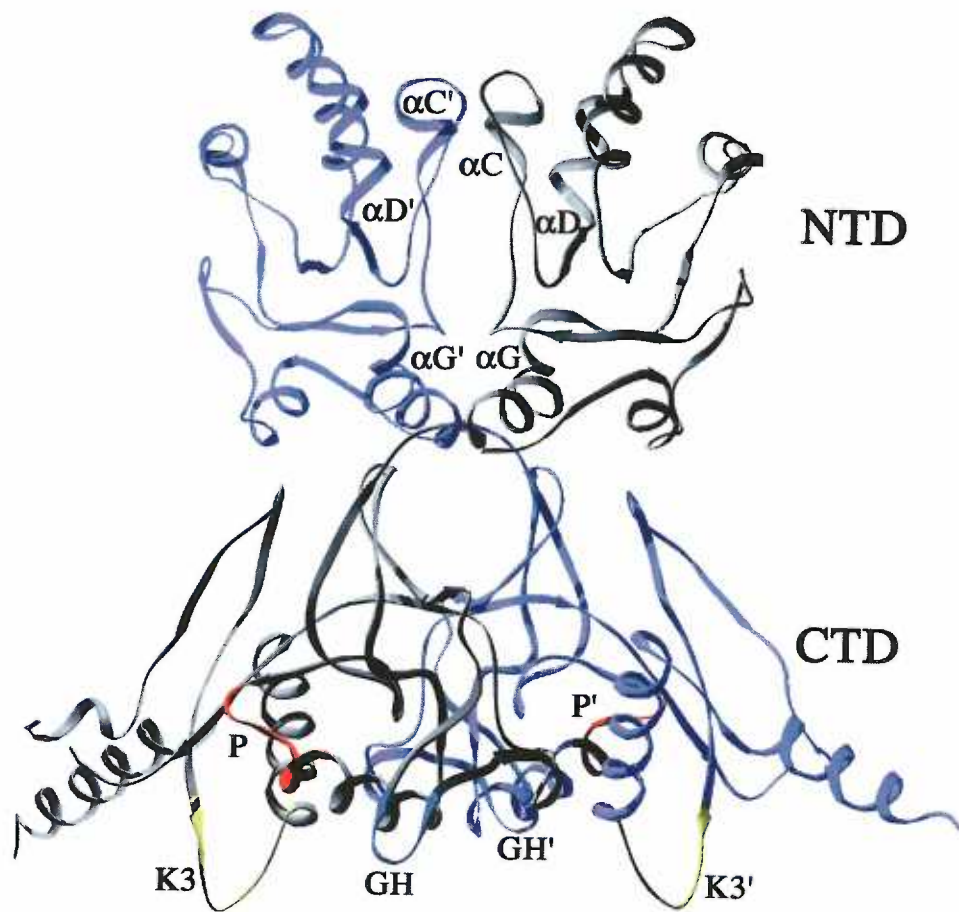




**Figure 4-2:** View of the NTD of the F subunit. Beta sheets are colored green and helices are colored red. Residue 132, the last residue of the NTD, is labelled as C(132).

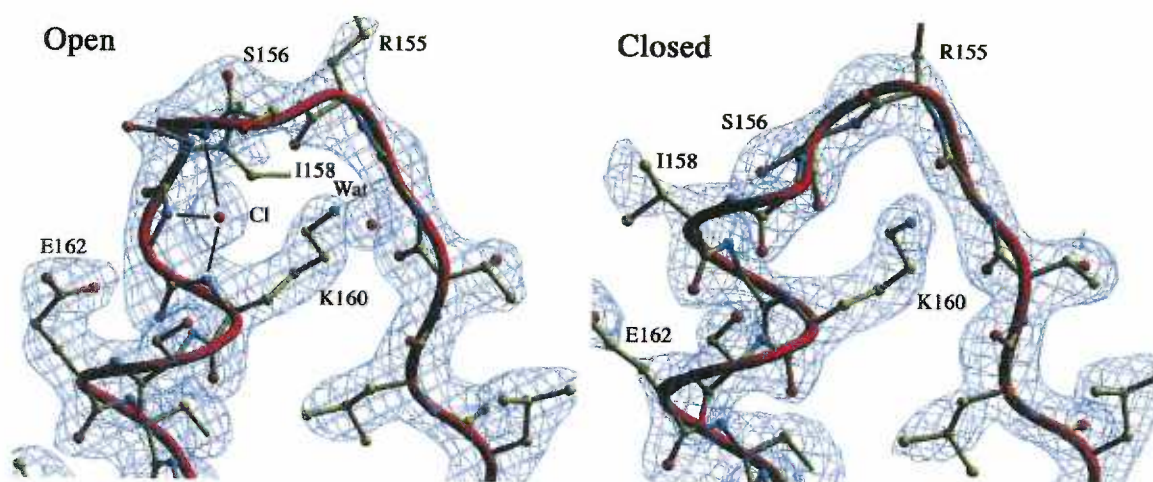


**Figure 4-3:** View of the CTD from the E subunit. Beta sheets are colored green and helices are colored red. Mutated side chains are colored blue and shown in balls and sticks.  $\eta$  denotes  $3_{10}$  helices. The ‘capping motif’ includes  $\eta 1$ ,  $\beta G$ ,  $\beta H$ ,  $\alpha 2$ , and  $\eta 2$ .

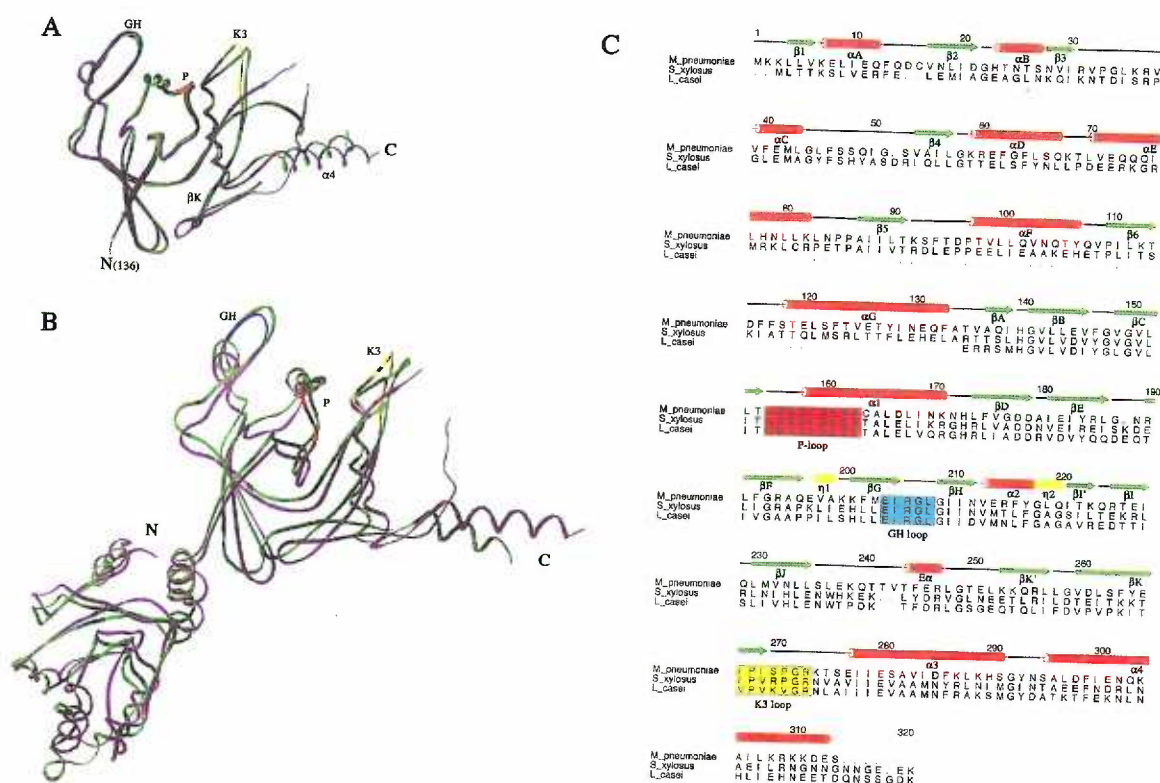


**Figure 4-4:** The *M. pneumoniae* HPrK/P dimer. The E subunit is colored white and the F subunit is colored purple. The N-terminal domain (NTD) and C-terminal domain (CTD) are labelled. The P-loop, (P) (residues 154-160) is shown in red, the GH loop (residues 203-207) in blue and the K3 loop (residues 267-273) in yellow. Primes indicate equivalent secondary structural elements of the NTD dimer partner.

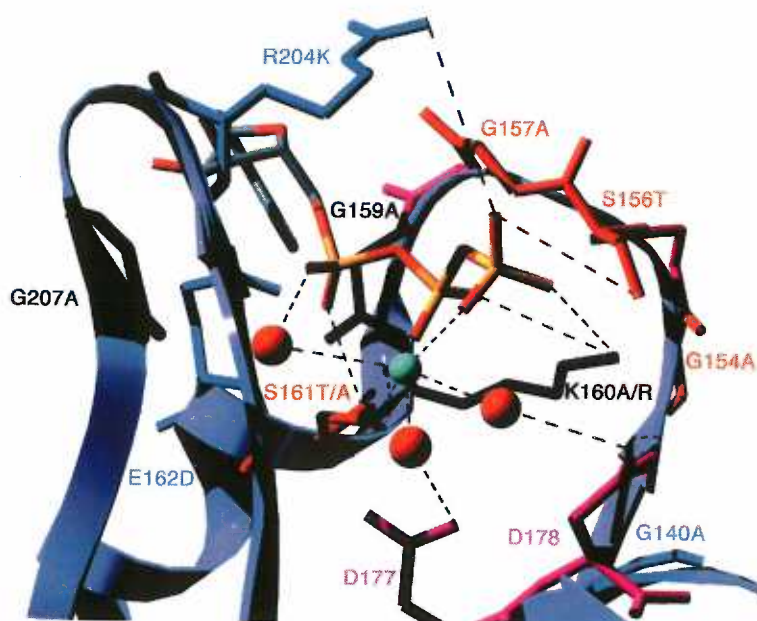




**Figure 4-5:** The open and closed conformations of the P-loop. The electron density clearly shows the different loop conformations of the A and B subunits. **A.** The B subunit: Cl, is the presumed chloride ion described in the text. Note that Ser156 is pointing out of the page while Ile158 is behind the loop and Wat241(Wat) is in front of the loop. **B.** The E subunit: Residues Ile158 and Gly157 display helical dihedral angles. Solvent molecules Wat241 and Wat238 are not present in the closed P-loop conformation. Residue Ser156 is behind the loop and hydrogen bonds with the main chain amide nitrogen of Gly159. The P-loop is shown as a red ribbon, the side chains as appropriately colored balls and sticks, and the solvent molecules as red spheres. The electron density is shown as blue mesh and calculated using sigmaA weighting with data from 30.0-2.5 Å resolution. Note that at a contour level of  $1\sigma$  density completely covers all residues shown but for the sake of clarity the higher contour level of  $1.3\sigma$  is shown here.

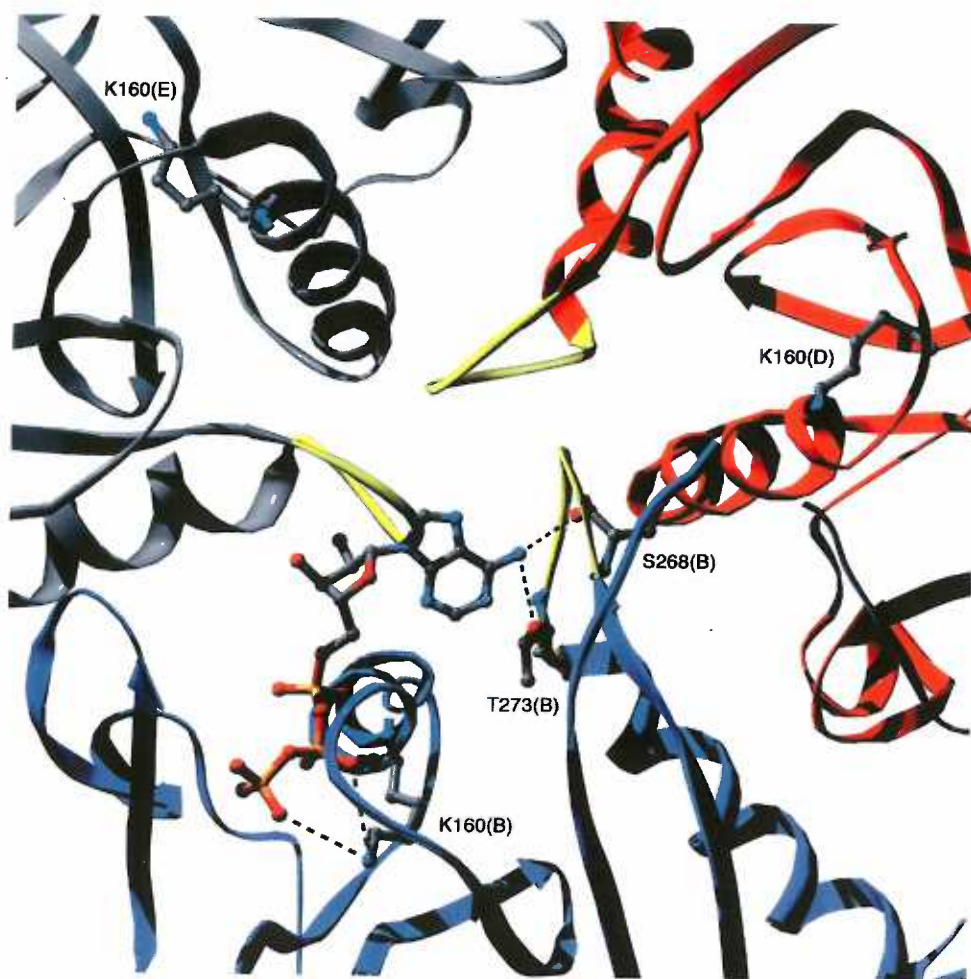


**Figure 4-6:** Superimposition of structural homologues of *M. pneumoniae* HPrK/P. **A.** Overlay of *L. casei* HPrK/P CTD (green) on the CTD of *M. pneumoniae* HPrK/P (purple). In the *M. pneumoniae* structure the P-loop (residues 154-160) is colored red, the GH loop (residues 203-207) is blue and the K3 loop (residues 267-273) is yellow. The different conformations of the two GH loops are not apparent in this view. **B.** Overlay of the entire subunit B of *Staphylococcus xylosus* HPrK/P (green) on the entire subunit B of *M. pneumoniae* HPrK/P (purple). The *M. pneumoniae* HPrK/P P-loop, GH loop, and K3 loop are colored as in A. The N and C termini of the CTDs or subunits are labelled. **C.** A structural alignment of the three extant HPrK/P models; *M. pneumoniae* HPrK/P, 1knx; *S. xylosus* HPrK/P, 1ko7; and *L. casei* HPrK/P CTD, 1jb1.



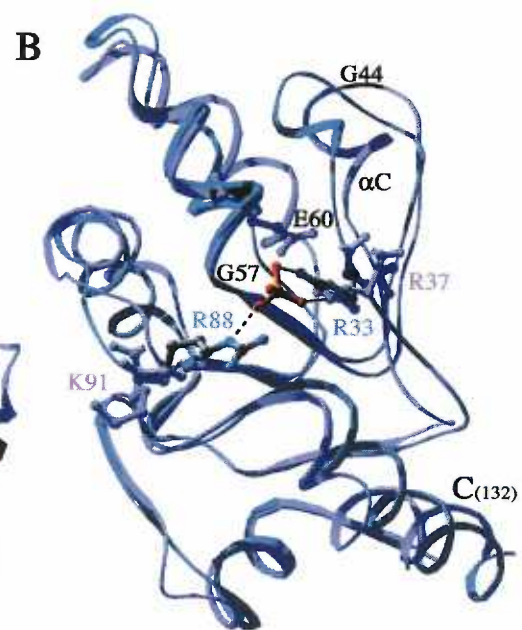
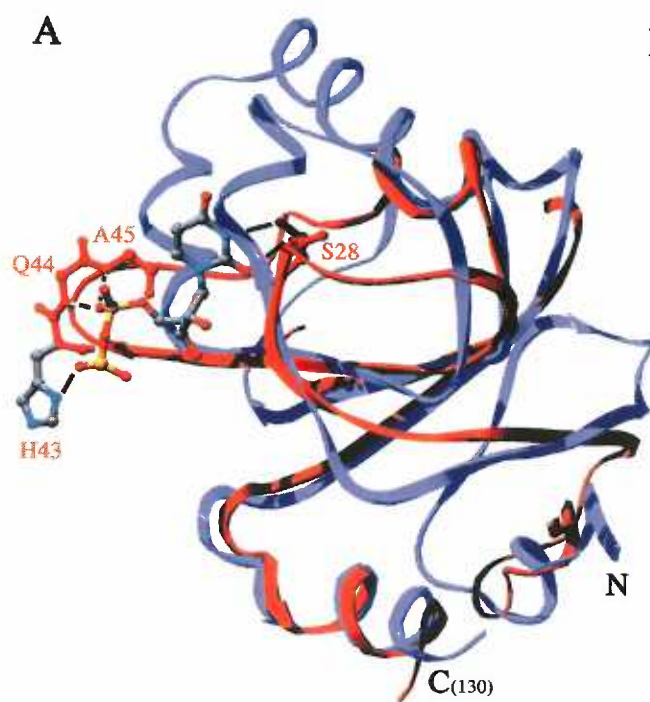
**Figure 4-7:** View of the putative ATP binding pocket of the *M. pneumoniae* HPrK/P. The ATP and  $Mg^{2+}$  (turquoise) from PEPCK are superimposed on the B subunit of the *M. pneumoniae* HPrK/P (purple ribbon); no minimization was performed; solvent molecules (red) were placed approximately where found in PEPCK. The GH loop (residues 203-207) is in the foreground and the P-loop (residues 154-160) is in the background. Dotted lines show the putative hydrogen bonding network of ATP,  $Mg^{2+}$ , solvent, and protein as it might appear in HPrK/P. Key residues are shown as balls and sticks and are colored according to their effect on catalytic activity: Blue residues (E162 and R204), mutations had no effect on kinase activity; Red residues (G154, S156, G157, and S161), mutations reduced the kinase activity; Black residues (G159, K160, and G207), mutations abolished kinase and phosphatase activity; Pink residues have not been mutated in the *M. pneumoniae* HPrK/P.





**Figure 4-8:** View down the molecular 3-fold axis into the putative ATP binding site. The B, D, and E subunits are shown as blue, red, and grey ribbons. The K3 loops of each subunit are colored yellow. ATP from PEPCK, shown as balls and sticks, is superimposed on the B subunit and black dashes denote putative hydrogen bonds. Note the steric clash of the adenine base of ATP with the K3 loop of the adjacent subunit. Two residues, Ser268 and Thr273 in the K3 loop would need to move only  $\sim 2$  Å towards the base to engage in hydrogen bonds to the adenine base. The Lys160 residues are shown to indicate the proximity of the several ATP binding sites.

**Figure 4-9:** Superimposition of the N-terminal domain of the *M. pneumoniae* HPrK/P onto structural homologues. **A.** Overlay of MurE (red) on the F subunit of *M. pneumoniae* HPrK/P (purple). The uridine diphosphate moiety of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-diaminopimelate is shown in balls and sticks along with the MurE residues that contact this ligand, Ser28, His43, Gln44, and Ala45. Hydrogen bonds are shown as black dashes. The last two residues of helix G of HPrK/P have been removed for clarity. **B.** Overlay of the A subunit of *S. xylosus* HPrK/P (blue), including a bound phosphate, and the F subunit of *M. pneumoniae* HPrK/P (purple). This view is rotated ~90 degrees from that shown in A. The side chains of residues R33 and R88 from *S. xylosus* HPrK/P are shown and hydrogen bonds to a bound phosphate are indicated as black dashes. The putative *M. pneumoniae* HPrK/P homologues of R33 and R88 are R37 and K91 respectively, and are shown as balls and sticks. The three strictly conserved residues in all HPrK/Ps, Gly44, Gly57 and Glu60 are labelled in black. Note the significant conformational differences between the two NTDs, whereby residues 39-43 ( $\alpha$ C) in *M. pneumoniae* HPrK/P are shifted by ~5 Å from the corresponding residue of the *S. xylosus* HPrK/P.



## Chapter V

### Conclusion and Perspective

Bacteria consume carbohydrates according to a rigid hierarchy at the top of which is invariably glucose. In most well studied bacteria the choice as to which of the available carbohydrates will be utilized is determined by a complex feedback mechanism that leads to regulation of transcription of genes involved in carbohydrate catabolism. The carbohydrate or some derivative, usually acts as an inducer, which derepresses transcription or inhibits termination of the messenger RNA encoding the enzymes necessary for its catabolism. However, if glucose is present control mechanisms exist, which prevent the induction of transcription of any operon encoding enzymes of secondary carbohydrate catabolism.

The mechanisms of transcriptional control over carbohydrate catabolism vary greatly in detail among the bacterial families even though the outcome, the glucose effect, is the same. In the gram negative *E. coli* the major component of transcriptional control is the complex of cAMP and CRP, which acts primarily as a global transcriptional activator during times of low intracellular energy. Yet the gram positive *B. subtilis* does not

encode a homolog of CRP nor does cAMP appear to have any effect on their carbohydrate metabolism. Instead *B. subtilis* employs a very different mechanism of transcriptional regulation whereby the phosphoprotein p-Ser46-HPr binds to CcpA and the complex acts primarily as a global repressor of secondary carbohydrate catabolism. HPrK/P is the ultimate regulator of this process through its control over the phosphorylation state of residue Ser46 of HPr.

HPrK/P and HPr homologs have been found in 32 organisms to date including *M. pneumoniae*. However, since *M. pneumoniae* do not encode a CcpA homolog it is not clear what role HPrK/P plays in these parasitic organisms. Our crystal structure of the HPrK/P from *M. pneumoniae*, determined to 2.5 Å resolution, reveals that this dual activity enzyme adopts the same overall fold as the CTD of *L. casei* HPrK/P and the *S. xylosus* HPrK/P suggesting that these proteins perform the same function in their respective organisms. Therefore, we expect that our structural studies on *M. pneumoniae* HPrK/P will have general applicability to the study of the HPrK/P enzyme family as a whole, as well as to the regulation of HPr phosphorylation and its role in carbon catabolite repression of gram positive bacteria.



Specifically, we have shown that *M. pneumoniae* HPrK/P is a hexamer, which is composed of a trimer of dimers. Each monomer of the hexamer consists of two domains: an N-terminal domain of 132 residues, a short 3 residue linker, and a 169 residue C-terminal domain. Each monomer maintains extensive contacts, in both domains, with its dimer partner but contacts with other monomers occurs only in the C-terminal domain.

Since there were six monomers in each asymmetric unit we obtained six crystallographically unique copies of the monomer. The abundance of unique copies allowed us to compare the variations of the conformations of the individual P-loops. We found that some of the P-loops adopted what we call open and closed conformations in the absence of ATP, the presumed ligand bound at this loop. The open conformation contained solvent molecules at locations similar to solvent found in the *L. casei* HPrK/P and the *S. xylosus* HPrK/P structures. But we showed that the conserved glycines of the P-loop are required in order to achieve the level of conformational flexibility observed; and this explains the negative effects on activity of alanine mutations of these glycines performed by our collaborators.

A DALI database search using the CTD revealed extensive homologies with PEPCK from *E. coli*, which were also noted for the *L. casei* HPrK/P and the *S. xylosus* HPrK/P structures. The PEPCK structure

contained ATP and  $Mg^{2+}$  bound at its P-loop, which closely overlaid the *M. pneumoniae* HPrK/P P-loop. When the ATP and  $Mg^{2+}$  were modeled onto HPrK/P we were able to explain the effects on activity of additional mutational data obtained by our collaborators. Alanine substitution of Lys160 produces a dead enzyme because the  $\epsilon$  amino group may stabilize the developing charge on the  $\beta$  phosphate of ATP. Substitution of Lys160 with arginine also produces a dead enzyme possibly because the guanidinium group is too bulky to allow productive ATP binding. Furthermore, the ATP base would certainly clash with the K3 loop of the neighboring subunit suggesting that movement of this loop either accompanies, or precedes binding of ATP.

Recent work on HPrK/Ps from other organisms has provided some insight on the catalytic mechanism presumed common to this family. Interestingly, the *B. subtilis* enzyme was shown to dephosphorylate p-Ser46-HPr only in the presence of inorganic phosphate giving the product pyrophosphate [Mijakovic, 2002 #430]. This suggests that the HPrK/P phosphatase reaction occurs through phosphorolysis of the phosphoserine rather than hydrolysis. A companion paper describes the 2.8 Å resolution crystal structure of complexes of HPr and HPrK/P for *L. casei* and proposes catalytic mechanisms for the phosphatase and kinase reactions [Fieulaine,

2002 #431]. These researchers were also able to demonstrate serine phosphorylation of HPr using pyrophosphate or tripolyphosphate but this would not substantially change our mechanistic interpretations as depicted in Figure 4-7 of this thesis.

We propose that the phosphatase mechanism for *M. pneumoniae* HPrK/P also involves phosphorolysis of the phosphoserine. Lys160, Arg204, Mg<sup>2+</sup>, and His139 stabilize the developing negative charge on the phosphate intermediate bound to Ser46-HPr. The phosphate intermediate nucleophilically attacks the inorganic phosphate bound at the P-loop while Asp178 provides a proton for Ser46-HPr. The kinase reaction is approximately the reverse of the phosphatase reaction. However, it is not clear that Asp178 acts as a general base for the kinase reaction. It may act to orient Ser46 for optimal geometry as we noted for protein kinase A.

Finally, a database search using the NTD discovered significant homology with the N-terminal uridyl binding domain of MurE. An overlay of the NTD of *S. xylosus* HPrK/P on the NTD of *M. pneumoniae* HPrK/P is virtually identical, with the exception of the phosphate-binding region of *S. xylosus* HPrK/P; this phosphate-binding region is conserved among HPrK/Ps and overlays the uridyl-binding region of MurE. We propose that this is also the site of binding of inorganic phosphate and FBP in *M. pneumoniae*

HPrK/P; and allosteric regulation follows this binding, which causes global conformational changes of the hexamer that favor kinase activity.

In sum the work presented in this thesis has provided a structural framework for the explanation of a good portion of the functional data available, the means by which a mechanistic description of the activity may be postulated, and the basis for future structural work on *M. pneumoniae* HPrK/P. Furthermore, this work demonstrates that the author has fulfilled the requirements of the degree; he has cultivated the skills of a competent crystallographer and he has made significant contributions to the scientific literature.

## Appendix

### Crystallization and preliminary crystallographic analysis of a complex of CcpA, P~Serine46-HPr, and Cre DNA from

*Bacillus megaterium*.

Gregory S. Allen,<sup>1</sup> Andrea Wagner,<sup>2</sup>

Marco Diehl,<sup>2</sup> Wolfgang Hillen,<sup>2</sup>

and Richard G. Brennan<sup>1\*</sup>

<sup>1</sup> Oregon Health & Science University, Department of Biochemistry and Molecular  
Biology, L224

3181 SW Sam Jackson Park Road, Portland, OR 97201, USA

<sup>2</sup> Lehrstuhl für Mikrobiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg,  
Staudtstr. 5, D-91058 Erlangen, Germany

## Introduction

*B. subtilis*, like most bacteria, maintains a hierarchy of preferences for carbon sources with glucose at the top; this phenomenon, known as carbon catabolite repression (CCR), is the end result of a complex feedback mechanism. In gram positive bacteria, of which *B. subtilis* is the most well known, carbon catabolite protein A (CcpA) is a critical component of CCR. CcpA, in response to a high-energy state or the presence of glucose, represses or activates the transcription of genes encoding a plethora of extracellular enzymes, intracellular carbohydrate metabolic enzymes, some glycolytic enzymes and sporulation enzymes (Hueck *et al.*, 1994; Tobisch *et al.*, 1999; Moreno *et al.*, 2001; Yoshida *et al.*, 2001).

Catabolite control protein A, found in many gram positive bacteria, is a member of the LacI/GalR (Weickert *et al.*, 1992) family of DNA binding proteins (Fig. 1-2). X-ray and solution structures of members of this family reveal that the N-terminal ~60 amino acid residues form an helix-turn-helix DNA binding domain (Schumacher *et al.*, 1994; Lewis *et al.*, 1996; Spronk *et al.*, 1999). The C-terminal inducer binding portions of these proteins resemble periplasmic sugar binding proteins (Schumacher *et al.*, 1993).

In bacilli, CcpA binds to a 14 base-pair pseudo-palindromic consensus sequence, the catabolite response element (CRE) (Kim *et al.*, 1995), found in

or near the promoters of genes known to be glucose repressed (Weickert *et al.*, 1990; Hueck *et al.*, 1994). The binding of CcpA to CRE sites is significantly enhanced in the presence of phospho-S46-HPr (Deutscher *et al.*, 1995) or phospho-S46-Crh, a non-PTS protein with 45% identity to HPr (Galinier *et al.*, 1999). In response to the ratio of  $[P_i]:[ATP/GTP]$ , HPr kinase/phosphatase controls the phosphorylation state of HPr and Crh at Ser46; relatively high concentrations of ATP/GTP favor phosphorylation, while dephosphorylation occurs at relatively high concentrations of  $P_i$  (Jault *et al.*, 2000).

The utilization of the phosphoproteins p-Ser46-HPr, and p-Ser46-Crh in bacilli, as corepressors instead of the usual small molecule inducers, places CcpA in a distinct subfamily of the LacI/GalR family. Kraus *et al.* found that the various CcpA homologues in the low G+C Gram positive bacteria contained additional regions of conservation with respect to other LacI/GalR family members (Kraus *et al.*, 1998). Five mutations, R47S, Y89E, Y295R, A299E, and R303D, of conserved residues *Bacillus megaterium* CcpA (Fig. 1-7) abrogated both catabolite repression *in vivo* and p-Ser46-HPr binding *in vitro* (Kraus *et al.*, 1998). Presumably these conserved regions of CcpA define the binding site of p-Ser46-HPr and p-Ser46-Crh.

In order to more fully characterize the protein-protein interaction between p-Ser46-HPr and CcpA and to gain insight into the mechanism of catabolite repression we have crystallized a CcpA - p-Ser46-HPr - CRE DNA complex. We present here the crystallization and data collection of the complex of CcpA:p-Ser46-HPr:CRE from *Bacillus megaterium* and discuss our preliminary attempts to produce an interpretable electron density map.

## Results & Discussion

The crystal structure analysis of the complex has been very frustrating, and until recently, unsuccessful even though well diffracting crystals were obtained. Three phasing methodologies were attempted in the course of this work; MIR using iodinated DNA, MAD using brominated DNA, and MAD using selenomethionine substituted CcpA. Each of these methods has failed to produce good experimental phases for reasons which are unclear. However, some success was realized in the more recent MAD data collection on selenomethionyl CcpA complex crystals. But we will first discuss the failed attempts.

Crystallization screening experiments were begun with a wide variety of oligonucleotides, which contained the 14 base pair optimized consensus sequence and some additional bases in blunt and overhanging ends (Kraus *et al.*, 1994). Eventually crystals were obtained with a blunt ended 16 mer



(Table A-1). Initial crystallization trials discovered that ~2 M ammonium sulfate, 10 mM magnesium sulfate, and 50 mM morpho-ethane sulfonic acid (MES) buffer at pH 6.0 would produce thin plates of crystals that would break upon mounting in glass capillaries. The thin plates were induced to grow to a usable thickness by the addition of 6 to 10% dimethyl sulfoxide (DMSO) to the crystallization reservoir and protein drops. DMSO apparently increased the solubility of the complex since the required ammonium sulfate concentration increased to ~2.5 M.

Crystals of the complex grown in DMSO were still liable to breakage in glass capillaries and preliminary data suggested they were radiation sensitive, so we sought appropriate cryocooling conditions. Experiments showed that a cryoprotectant solution of 50 mM MES pH 6.0, ~2 M ammonium sulfate, and 20% ethylene glycol would preclude the formation of ice at 100 K. Unfortunately, indexing of the crystals revealed that they took the P1, or triclinic, space group with parameters given in Table A-1. But since these crystals diffracted to a resolution greater than 3.0 Å we decided to proceed with an MIR experimental plan using oligonucleotides iodinated as indicated in Table A-1.

After a reasonable period of time spent attempting to find a solution to the heavy atom Patterson map for the MIR data it was abandoned in favor of

a MAD experiment. Oligonucleotides were obtained that were brominated at the same sites where iodine was located in the MIR experiment and these crystals were grown to adequate size. A four wavelength MAD experiment was performed at the 0.92 Å edge of Bromine (Table A-2). But these data also failed to yield a good solution to the Patterson function using automated or manual methods.

Recently it was suggested that the selenomethionyl CcpA might prove more successful in a MAD experiment because, with 9 selenium atoms the signal would be much greater than in our previous MAD (Bromine) experiment. Using the method of inhibition of methionine biosynthesis in the *E. coli* (Doublié, 1997) overexpression of CcpA we obtained complete methionine substitution with selenomethionine as determined by mass spectroscopy. Crystals of the complex with selenomethionyl CcpA were grown according to the same protocol as described above. However, it was discovered that these crystals were unstable in the ethylene glycol cryoprotectant. It was determined that the replacement of ethylene glycol with glycerol at ~30% resulted in a cryoprotectant in which the complex was stable and also precluded the formation of ice.

The selenomethionyl CcpA complex crystals, cryopreserved in glycerol and suspended in liquid nitrogen, were shipped to Howard

Robinson at the Brookhaven National Light Source. Dr. Robinson found that these crystals displayed orthorhombic symmetry and suggested that he could refine the data, with the DENZO data processing suite, assuming they took the  $P2_12_12_1$  space group. Due to technical difficulties encountered during data collection only the data from the first wavelength was usable. We processed the data from the first 180 degrees of the first wavelength with d\*trek v7.2 (Table A-3).

The average intensities for these data were used in a molecular replacement experiment with the program EPMR (Kissinger *et al.*, 1999). We used the the default parameters of EPMR and the wild type PurR dimer (1PNR) as the search model. EPMR appeared to find a solution with a patterson correlation coefficient of 0.477, where greater than 0.45 is usually found to be correct. However, we have not found convincing evidence for the correctness of these molecular replacement solutions.

## **Conclusion**

We believe that the failure of our early attempts to determine the phases of the CcpA complex are due to the difficulties inherent in working with the space group P1. It is possible the cryoprotectant used in those experiments changed the crystal parameters from  $P2_12_12_1$  to P1.

It is not clear at this time if the  $P2_12_12_1$  symmetry of the selenomethionine complex crystals is due to the selenoprotein or the new cryoprotectant. In future we will attempt to obtain data for a full 3 wavelength MAD experiment with the selenomethionine substituted CcpA complex.

### **Materials & Methods**

Expression of HPr was carried out in *B. megaterium* containing plasmid pWH1520. These cells were grown at 37 °C in superbroth with tetracycline (10 µg/ml) to  $A_{600} = 1$  OD and induced by bringing the media to 0.5% D-xylose. The culture was incubated 5 hours before harvest. The cells were harvested by centrifugation at 4000g for 30 minutes. Cell paste was stored at -80 °C.

Expression of CcpA for the MIR and MAD (Bromouridine) experiments was carried out in *B. megaterium*; the protocol was identical to expression of HPr. For preparation of the MAD (selenomethionine substituted) CcpA protein was overproduced in *E. coli* strain RB791 with plasmid pWH1950OECM. The cells were grown in LB ampicillin (100 µg/ml) to an  $A_{600}=0.6$  OD and then induced by addition of IPTG to 1 mM final concentration. Cells were incubated an additional 2 hours, harvested and

resuspended in 50 mM Tris pH 7.5, 3 mM DTT and 100 mM NaCl then stored at -80 °C.

Expression of *B. subtilis* HPr kinase/phosphatase DH5 $\alpha$  strain

pGP205(hisTag): The cells were grown in LB ampicillin (100  $\mu$ g/ml) to an  $A_{600}$ =0.6 OD and then induced by addition of IPTG to 1 mM final concentration. Cells were incubated for an additional 3 hours, harvested and resuspended in 50 mM Tris pH 7.5, 3 mM DTT and 100 mM NaCl then sonicated on ice. Clarified lysate was purified on Talon (Invitrogen) resin, eluted with 250 mM imidazole, dialyzed overnight against 50 mM Tris pH 7.5, 3 mM DTT and 100 mM NaCl, and concentrated to  $A_{280}$  = 3 OD.

Concentrated enzyme was brought to 10% glycerol and stored at -80 °C.

HPr purification: 24 grams of cell paste was resuspended in 48 mL of 50 mM Tris pH 8.0, 1 mM EDTA, and 1 mM PMSF. Cell suspension was subjected to 3 passes through a French press and clarified by centrifugation at 4 °C and 40,000g for 30 minutes. Lysate was heated to 70 °C for 20 minutes, cooled to room temperature, and centrifuged at 4 °C and 40,000g for 30 minutes. The soluble portion was then passed over a column of Q sepharose fast flow. The flow through from the Q sepharose column was concentrated (Amicon 3kD or 10kD cutoff) and passed over a gel filtration column (Sephacryl HR200) equilibrated with 25mM Tris pH7.6 and 200

mM NaCl. The fractions containing HPr were concentrated (Amicon 3kD or 10kD cutoff) to ~1.0 mg/ml (based on  $\epsilon_{220} = 0.091$ ) and stored at -20 °C.

Phosphorylation of HPr: HPr at ~1 mg/mL was made 10 mM fructose-bis-phosphate, 5 mM ATP, 1 mM DTT, and 10 mM  $\text{MgCl}_2$ . To this reaction mix was added a dilution of purified *B. subtilis* (his6)HPrK/P. The reaction was incubated at 37 °C for 15 min. The reaction was killed at 95 °C for 2 min. Buffer exchange (1:20) and concentration were achieved simultaneously with Centricon 3 filters. The concentrated p-Ser46-HPr at ~20 mg/ml was stored at -20 °C.

*B. megaterium* CcpA purification: For the MIR and MAD (Bromouridine) experiments 24 grams of cell paste was resuspended in 48 mL of 30 mM potassium phosphate pH 7.4, 5% glycerol, and 1% EDTA; for the selenomethionine MAD experiment one liter of induced cells were harvested and resuspended in 25 mL of 30 mM potassium phosphate pH 7.4, 5% glycerol, and 1% EDTA. The cell suspension was subjected to 3 passes through a French press and clarified by centrifugation at 4 °C and 40,000g for 30 minutes. The soluble portion was then passed over a 10 mL column of Q sepharose fast flow. After washing with 10 column volumes of 100 mM potassium phosphate pH 7.4, 5% glycerol, and 1% EDTA a gradient elution was begun. The gradient varied the potassium phosphate only from

100 to 500 mM in 20 column volumes. Fractions containing CcpA were concentrated to  $A_{280} = 30$  OD using Centricon 30 filters and further purified and concentrated to  $A_{280} = 50$  OD by precipitation. Precipitation was effected by either addition of saturated ammonium sulfate or cooling on ice; the precipitate was harvested by centrifugation and resuspended in 100 mM potassium phosphate pH 7.4, 5% glycerol, and 1% EDTA.

**Table A-1** MIR Data For CcpA Complex. VS15 is composed of the top strand iodinated at the second position and a native bottom strand. VS24 is composed of a bottom strand iodinated at the 11<sup>th</sup> position and a native top strand. These data were collected at BL9-1 at SSRL.



	Native	vs24	vs15
A	105.72	104.56	104.62
B	120.41	119.86	120.31
C	131.74	132.05	131.78
ALPHA	102.24	102.02	102.07
BETA	100.43	100.35	100.32
GAMMA	90.43	90.55	90.24
# REFLECTIONS	197,918	97,073	148,006
SHELL OF HIGHEST RES.	3.11 - 3.00	3.27- 3.12	3.14 – 3.03
R <sub>M</sub> OF SHELL	27.4%	38.7%	31.5%
R <sub>M</sub> OVERALL	5.3%		25.3%
COMPLETENESS OF SHELL	82.5%	85.9%	86.4%
COMPLETENESS OVERALL	93.4%		83.9%
I/SIG SHELL	2.3	2.2	2.5

$$R_M = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$$

**VS5**  
 5-CTG TTA GCGC TTT CAG-3  
 3-GAC AAT CGCG AAA GTC-5  
**VS4**

**Table A-2** MAD (Bromouridine) Data For CcpA Complex. The oligonucleotide used in this experiment was composed of the top strand brominated at the second position and the bottom strand brominated at the 11<sup>th</sup> position. These data were collected at BL1-5 at SSRL.

	1	2	3	4
WAVELENGTH	0.920174	0.919935	0.898437	0.953725
A	104.77	104.72	104.88	104.79
B	119.80	119.77	119.94	119.88
C	131.53	131.03	131.25	131.35
ALPHA	102.06	101.70	101.94	102.04
BETA	100.62	100.34	100.49	100.51
GAMMA	89.99	90.36	90.20	90.16
MOSAICITY	0.729	0.778	0.662	0.752
# REFLECTIONS	772,134	760,417	760,616	759,402
OUTER SHELL	3.21 – 3.09	3.22- 3.10	3.22 – 3.10	3.22 – 3.10
R <sub>M</sub> OF SHELL	30.1%	41.5%	50.3%	54.2%
COMPLETENESS OF SHELL	94.2%	93.1%	93.3%	92.5%
I/SIG(I) SHELL	3.2	2.7	2.5	2.9
R <sub>M</sub> OVERALL	14.1%	18.5%	20.3%	17.6%
COMPLETENESS	97.2%	96.5%	96.5%	96.1%

$$R_M = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$$

5-CTG TTA GCGC TTT CAG-3  
3-GAC AAT CGCG AAA GTC-5

A	104.57
B	110.19
C	120.17
# REFLECTIONS	242,722
OUTER SHELL	3.12 - 3.00
I/SIG SHELL	1.0
COMPLETENESS SHELL	99.0%
COMPLETENESS OVERALL	98.2%
R <sub>M</sub> OVERALL	28.1%

**Table A-3** Data from MAD (Selenomethionine) experiment. The crystals take the space group  $P2_12_12_1$ . These statistics are for data between 30.0 – 3.0 Å.

## Bibliography

- Aboud, M. and Burger, M. (1970). "The effect of catabolite repression and of cyclic 3',5' adenosine monophosphate on the translation of the lactose messenger RNA in *Escherichia coli*." Biochem Biophys Res Commun **38**(6): 1023-1032.
- Amster-Choder, O. and Wright, A. (1992). "Modulation of the dimerization of a transcriptional antiterminator protein by phosphorylation." Science **257**(5075): 1395-1398.
- Aurora, R. and Rose, G. D. (1998). "Helix capping." Protein Sci **7**(1): 21-38.
- Barton, G. J. (1993). "ALSCRIPT: a tool to format multiple sequence alignments." Protein Eng **6**(1): 37-40.
- Beyer, W. H. (1987). Chapter XII: Probability and Statistics. Standard Mathematical Tables. Beyer, W. H. Boca Raton, Florida, CRC Press, Inc. **one**: 674.
- Brunger, A. T., Adams, P. D., *et al.* (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-921.
- Brunger, A. T., Kuriyan, J., *et al.* (1987). "Crystallographic R factor refinement by molecular dynamics." Science **235**(4786): 458-461.

- Burmeister, W. P. (2000). "Structural changes in a cryo-cooled protein crystal owing to radiation damage." Acta Crystallogr D Biol Crystallogr **56**(Pt 3): 328-341.
- Dandekar, T., Huynen, M., *et al.* (2000). "Re-annotating the *Mycoplasma pneumoniae* genome sequence: adding value, function and reading frames." Nucleic Acids Res **28**(17): 3278-3288.
- Darbon, E., Servant, P., *et al.* (2002). "Antitermination by GlpP, catabolite repression via CcpA and inducer exclusion triggered by P~GlpK dephosphorylation control *Bacillus subtilis* glpFK expression." Mol Microbiol **43**(4): 1039-1052.
- Das, A. K., Helps, N. R., *et al.* (1996). "Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution." Embo J **15**(24): 6798-6809.
- Deuschle, U., Gentz, R., *et al.* (1986). "lac Repressor blocks transcribing RNA polymerase and terminates transcription." Proc Natl Acad Sci U S A **83**(12): 4134-4137.
- Deutscher, J., Galinier, A., *et al.* (2002). Carbohydrate uptake and metabolism. *Bacillus subtilis* and its closest relatives: from genes to cells. Sonenshein, A. L., Hoch, J. A. and Losick, R. Washington, D. C., ASM Press: 129-150.

- Deutscher, J., Kessler, U., *et al.* (1985). "Streptococcal phosphoenolpyruvate: sugar phosphotransferase system: purification and characterization of a phosphoprotein phosphatase which hydrolyzes the phosphoryl bond in seryl-phosphorylated histidine-containing protein." J Bacteriol **163**(3): 1203-1209.
- Deutscher, J., Kuster, E., *et al.* (1995). "Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in gram-positive bacteria." Mol Microbiol **15**(6): 1049-1053.
- Deutscher, J., Reizer, J., *et al.* (1994). "Loss of protein kinase-catalyzed phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system, by mutation of the ptsH gene confers catabolite repression resistance to several catabolic genes of *Bacillus subtilis*." J Bacteriol **176**(11): 3336-3344.
- Djordjevic, G. M., Tchieu, J. H., *et al.* (2001). "Genes involved in control of galactose uptake in *Lactobacillus brevis* and reconstitution of the regulatory system in *Bacillus subtilis*." J Bacteriol **183**(10): 3224-3236.
- Dossonnet, V., Monedero, V., *et al.* (2000). "Phosphorylation of HPr by the bifunctional HPr Kinase/P-ser-HPr phosphatase from *Lactobacillus*

- Fieulaine, S., Morera, S., *et al.* (2002). "X-ray structure of a bifunctional protein kinase in complex with its protein substrate HPr." Proc Natl Acad Sci U S A **99**(21): 13437-13441.
- Fieulaine, S., Morera, S., *et al.* (2001). "X-ray structure of HPr kinase: a bacterial protein kinase with a P-loop nucleotide-binding domain." Embo J **20**(15): 3917-3927.
- Fraser, C. M., Gocayne, J. D., *et al.* (1995). "The minimal gene complement of *Mycoplasma genitalium*." Science **270**(5235): 397-403.
- Fujita, Y. and Freese, E. (1979). "Purification and properties of fructose-1,6-bisphosphatase of *Bacillus subtilis*." J Biol Chem **254**(12): 5340-5349.
- Galinier, A., Deutscher, J., *et al.* (1999). "Phosphorylation of either crh or HPr mediates binding of CcpA to the bacillus subtilis xyn cre and catabolite repression of the xyn operon." J Mol Biol **286**(2): 307-314.
- Galinier, A., Haiech, J., *et al.* (1997). "The *Bacillus subtilis* crh gene encodes a HPr-like protein involved in carbon catabolite repression." Proc Natl Acad Sci U S A **94**(16): 8439-8444.
- Galinier, A., Kravanja, M., *et al.* (1998). "New protein kinase and protein phosphatase families mediate signal transduction in bacterial catabolite repression." Proc Natl Acad Sci U S A **95**(4): 1823-1828.



- Galinier, A., Lavergne, J. P., *et al.* (2002). "A new family of phosphotransferases with a P-loop motif." J Biol Chem **277**(13): 11362-11367.
- Goldberg, J., Huang, H. B., *et al.* (1995). "Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1." Nature **376**(6543): 745-753.
- Gonzalez, A., Pedelacq, J., *et al.* (1999). "Two-wavelength MAD phasing: in search of the optimal choice of wavelengths." Acta Crystallogr D Biol Crystallogr **55**(Pt 8): 1449-1458.
- Gordon, E., Flouret, B., *et al.* (2001). "Crystal structure of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso- diaminopimelate ligase from *Escherichia coli*." J Biol Chem **276**(14): 10999-11006.
- Gorke, B. and Rak, B. (1999). "Catabolite control of *Escherichia coli* regulatory protein BglG activity by antagonistically acting phosphorylations." Embo J **18**(12): 3370-3379.
- Hanks, S. K. and Quinn, A. M. (1991). "Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members." Methods Enzymol **200**: 38-62.

- Hanson, K. G., Steinhauer, K., *et al.* (2002). "HPr kinase/phosphatase of *Bacillus subtilis*: expression of the gene and effects of mutations on enzyme activity, growth and carbon catabolite repression." Microbiology **148**(Pt 6): 1805-1811.
- Hendrickson, W. A. (1991). "Determination of macromolecular structures from anomalous diffraction of synchrotron radiation." Science **254**(5028): 51-58.
- Hendrickson, W. A. and Ogata, C. M. (1997). "Phase determination from multiwavelength anomalous diffraction measurements." Methods Enzymol **276**: 494-523.
- Henkin, T. M., Grundy, F. J., *et al.* (1991). "Catabolite repression of alpha-amylase gene expression in *Bacillus subtilis* involves a trans-acting gene product homologous to the *Escherichia coli* *lacI* and *galR* repressors." Mol Microbiol **5**(3): 575-584.
- Himmelreich, R., Hilbert, H., *et al.* (1996). "Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*." Nucleic Acids Res **24**(22): 4420-4449.
- Hogema, B. M., Arents, J. C., *et al.* (1998). "Inducer exclusion in *Escherichia coli* by non-PTS substrates: the role of the PEP to

- pyruvate ratio in determining the phosphorylation state of enzyme IIAGlc." Mol Microbiol **30**(3): 487-498.
- Holm, L. and Sander, C. (1995). "Dali: a network tool for protein structure comparison." Trends Biochem Sci **20**(11): 478-480.
- Hueck, C. J., Hillen, W., *et al.* (1994). "Analysis of a cis-active sequence mediating catabolite repression in gram-positive bacteria." Res Microbiol **145**(7): 503-518.
- Hurley, J. H., Faber, H. R., *et al.* (1993). "Structure of the regulatory complex of *Escherichia coli* IIIGlc with glycerol kinase." Science **259**(5095): 673-677.
- Hutchison, C. A., Peterson, S. N., *et al.* (1999). "Global transposon mutagenesis and a minimal *Mycoplasma* genome." Science **286**(5447): 2165-2169.
- Hutter, M. C. and Helms, V. (1999). "Influence of key residues on the reaction mechanism of the cAMP-dependent protein kinase." Protein Sci **8**(12): 2728-2733.
- Jackson, M. D. and Denu, J. M. (2001). "Molecular reactions of protein phosphatases-insights from structure and chemistry." Chem Rev **101**(8): 2313-2340.

- James, R. W. (1962). The optical principles of the diffraction of x-rays.  
Woodbridge, Connecticut, Ox Bow Press.
- Jault, J. M., Fieulaine, S., *et al.* (2000). "The HPr kinase from *Bacillus subtilis* is a homo-oligomeric enzyme which exhibits strong positive cooperativity for nucleotide and fructose 1,6- biphosphate binding." J Biol Chem **275**(3): 1773-1780.
- Jones, T. A., Zou, J. Y., *et al.* (1991). "Improved methods for binding protein models in electron density maps and the location of errors in these models." Acta Crystallogr A **47**(Pt 2): 110-119.
- Kennelly, P. J. (2002). "Protein kinases and protein phosphatases in prokaryotes: a genomic perspective." FEMS Microbiol Lett **206**(1): 1-8.
- Kim, J. H., Guvener, Z. T., *et al.* (1995). "Specificity of DNA binding activity of the *Bacillus subtilis* catabolite control protein CcpA." J Bacteriol **177**(17): 5129-5134.
- Kim, J. H., Shin, D. Y., *et al.* (2001). "Mutational and kinetic evaluation of conserved His-123 in dual specificity protein-tyrosine phosphatase vaccinia H1-related phosphatase: participation of Tyr-78 and Thr-73 residues in tuning the orientation of His-123." J Biol Chem **276**(29): 27568-27574.

- Kim, S. K., Kimura, S., *et al.* (2000). "Dual transcriptional regulation of the *Escherichia coli* phosphate- starvation-inducible *psiE* gene of the phosphate regulon by PhoB and the cyclic AMP (cAMP)-cAMP receptor protein complex." J Bacteriol **182**(19): 5596-5599.
- Kimata, K., Takahashi, H., *et al.* (1997). "cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose transporter gene in *Escherichia coli*." Proc Natl Acad Sci U S A **94**(24): 12914-12919.
- Kissinger, C. R., Gehlhaar, D. K., *et al.* (1999). "Rapid automated molecular replacement by evolutionary search." Acta Crystallogr D Biol Crystallogr **55** ( Pt 2): 484-491.
- Kissinger, C. R., Parge, H. E., *et al.* (1995). "Crystal structures of human calcineurin and the human FKBP12-FK506- calcineurin complex." Nature **378**(6557): 641-644.
- Kleywegt, G. J. (1999). "Experimental assessment of differences between related protein crystal structures." Acta Crystallogr D Biol Crystallogr **55**(11): 1878-1884.
- Knighton, D. R., Zheng, J. H., *et al.* (1991). "Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase." Science **253**(5018): 407-414.

- Kolb, A., Busby, S., *et al.* (1993). "Transcriptional regulation by cAMP and its receptor protein." Annu Rev Biochem **62**: 749-795.
- Kraus, A., Hueck, C., *et al.* (1994). "Catabolite repression of the *Bacillus subtilis* xyl operon involves a cis element functional in the context of an unrelated sequence, and glucose exerts additional xylR-dependent repression." J Bacteriol **176**(6): 1738-1745.
- Kraus, A., Kuster, E., *et al.* (1998). "Identification of a co-repressor binding site in catabolite control protein CcpA." Mol Microbiol **30**(5): 955-963.
- Krause, D. C. (1996). "*Mycoplasma pneumoniae* cytoadherence: unravelling the tie that binds." Mol Microbiol **20**(2): 247-253.
- Kravanja, M., Engelmann, R., *et al.* (1999). "The hprK gene of *Enterococcus faecalis* encodes a novel bifunctional enzyme: the HPr kinase/phosphatase." Mol Microbiol **31**(1): 59-66.
- Kunst, F., Ogasawara, N., *et al.* (1997). "The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*." Nature **390**(6657): 249-256.
- LaPorte, D. C. (1993). "The isocitrate dehydrogenase phosphorylation cycle: regulation and enzymology." J Cell Biochem **51**(1): 14-18.

- Laskowski, R. A., Rullmannn, J. A., *et al.* (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-486.
- Lee, J. and Goldfarb, A. (1991). "lac repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter." Cell **66**(4): 793-798.
- Lee, S. Y., Cho, H. S., *et al.* (2001). "Crystal structure of an activated response regulator bound to its target." Nat Struct Biol **8**(1): 52-56.
- Lewis, M., Chang, G., *et al.* (1996). "Crystal structure of the lactose operon repressor and its complexes with DNA and inducer." Science **271**(5253): 1247-1254.
- Lindwall, G., Chau, M., *et al.* (2000). "A sparse matrix approach to the solubilization of overexpressed proteins." Protein Eng **13**(1): 67-71.
- Madhusudan, Akamine, P., *et al.* (2002). "Crystal structure of a transition state mimic of the catalytic subunit of cAMP-dependent protein kinase." Nat Struct Biol **9**(4): 273-277.
- Manch, K., Notley-McRobb, L., *et al.* (1999). "Mutational adaptation of *Escherichia coli* to glucose limitation involves distinct evolutionary pathways in aerobic and oxygen-limited environments." Genetics **153**(1): 5-12.

- Marquez, J. A., Hasenbein, S., *et al.* (2002). "Structure of the full-length HPr kinase/phosphatase from *Staphylococcus xylosus* at 1.95 Å resolution: Mimicking the product/substrate of the phospho transfer reactions." Proc Natl Acad Sci U S A **99**(6): 3458-3463.
- Matte, A., Tari, L. W., *et al.* (1997). "Structure and mechanism of phosphoenolpyruvate carboxykinase." J Biol Chem **272**(13): 8105-8108.
- Matthews, B. W. (1968). "Solvent content of protein crystals." J Mol Biol **33**(2): 491-497.
- Matthews, K. S. and Nichols, J. C. (1998). "Lactose repressor protein: functional properties and structure." Prog Nucleic Acid Res Mol Biol **58**: 127-164.
- McPherson, A. (1999). Crystallization of Biological Macromolecules. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.
- McRee, D. E. (1999a). Practical Protein Crystallography. San Diego, CA, Academic Press.
- McRee, D. E. (1999b). "XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density." J Struct Biol **125**(2-3): 156-165.



- Mijakovic, I., Poncet, S., *et al.* (2002). "Pyrophosphate-producing protein dephosphorylation by HPr kinase/phosphorylase: a relic of early life?" Proc Natl Acad Sci U S A **99**(21): 13442-13447.
- Milburn, M. V., Tong, L., *et al.* (1990). "Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins." Science **247**(4945): 939-945.
- Miller, S. P., Chen, R., *et al.* (2000). "Locations of the regulatory sites for isocitrate dehydrogenase kinase/phosphatase." J Biol Chem **275**(2): 833-839.
- Miller, S. P., Karschnia, E. J., *et al.* (1996). "Isocitrate dehydrogenase kinase/phosphatase. Kinetic characteristics of the wild-type and two mutant proteins." J Biol Chem **271**(32): 19124-19128.
- Monedero, V., Poncet, S., *et al.* (2001). "Mutations lowering the phosphatase activity of HPr kinase/phosphatase switch off carbon metabolism." Embo J **20**(15): 3928-3937.
- Moreno, M. S., Schneider, B. L., *et al.* (2001). "Catabolite repression mediated by the CcpA protein in *Bacillus subtilis*: novel modes of regulation revealed by whole-genome analyses." Mol Microbiol **39**(5): 1366-1381.

- Mota, L. J., Tavares, P., *et al.* (1999). "Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*." Mol Microbiol **33**(3): 476-489.
- Nick, H. and Gilbert, W. (1985). "Detection in vivo of protein-DNA interactions within the lac operon of *Escherichia coli*." Nature **313**(6005): 795-798.
- Nihashi, J. and Fujita, Y. (1984). "Catabolite repression of inositol dehydrogenase and gluconate kinase syntheses in *Bacillus subtilis*." Biochim Biophys Acta **798**(1): 88-95.
- Ollis, D. and White, S. (1990). "Protein crystallization." Methods Enzymol **182**: 646-659.
- Parkinson, G., Wilson, C., *et al.* (1996). "Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface." J Mol Biol **260**(3): 395-408.
- Pedersen, H., Sogaard-Andersen, L., *et al.* (1991). "Heterologous cooperativity in *Escherichia coli*. The CytR repressor both contacts DNA and the cAMP receptor protein when binding to the deoP2 promoter." J Biol Chem **266**(27): 17804-17808.

- Peitsch, M. C., Wells, T. N., *et al.* (1995). "The Swiss-3DImage collection and PDB-Browser on the World-Wide Web." Trends Biochem Sci **20**(2): 82-84.
- Pettigrew, D. W., Liu, W. Z., *et al.* (1996). "A single amino acid change in *Escherichia coli* glycerol kinase abolishes glucose control of glycerol utilization in vivo." J Bacteriol **178**(10): 2846-2852.
- Pflugrath, J. W. (1999). "The finer things in X-ray diffraction data collection." Acta Crystallogr D Biol Crystallogr **55**(Pt 10): 1718-1725.
- Powell, B. S., Court, D. L., *et al.* (1995). "Novel proteins of the phosphotransferase system encoded within the rpoN operon of *Escherichia coli*. Enzyme IIANtr affects growth on organic nitrogen and the conditional lethality of an *erats* mutant." J Biol Chem **270**(9): 4822-4839.
- Quan, J. A., Schneider, B. L., *et al.* (2002). "Regulation of carbon utilization by sulfur availability in *Escherichia coli* and *Salmonella typhimurium*." Microbiology **148**(Pt 1): 123-131.
- Quentin, Y., Fichant, G., *et al.* (1999). "Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems." J Mol Biol **287**(3): 467-484.

- Razin, S., Yogev, D., *et al.* (1998). "Molecular biology and pathogenicity of mycoplasmas." Microbiol Mol Biol Rev **62**(4): 1094-1156.
- Reizer, J., Bachem, S., *et al.* (1999). "Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of *Bacillus subtilis*." Microbiology **145**(Pt 12): 3419-3429.
- Reizer, J., Bergstedt, U., *et al.* (1996). "Catabolite repression resistance of gnt operon expression in *Bacillus subtilis* conferred by mutation of His-15, the site of phosphoenolpyruvate-dependent phosphorylation of the phosphocarrier protein HPr." J Bacteriol **178**(18): 5480-5486.
- Reizer, J., Hoischen, C., *et al.* (1998). "A novel protein kinase that controls carbon catabolite repression in bacteria." Mol Microbiol **27**(6): 1157-1169.
- Rice, L. M. and Brunger, A. T. (1994). "Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement." Proteins **19**(4): 277-290.
- Riddles, P. W., Blakeley, R. L., *et al.* (1983). "Reassessment of Ellman's reagent." Methods Enzymol **91**: 49-60.

- Rosengarten, R., Citti, C., *et al.* (2000). "Host-pathogen interactions in mycoplasma pathogenesis: virulence and survival strategies of minimalist prokaryotes." Int J Med Microbiol **290**(1): 15-25.
- Rubinson, K. A., Ladner, J. E., *et al.* (2000). "Cryosalts: suppression of ice formation in macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **56**(Pt 8): 996-1001.
- Russell, R. B., Marquez, J. A., *et al.* (2002). "Evolutionary relationship between the bacterial HPr kinase and the ubiquitous PEP-carboxykinase: expanding the P-loop nucleotidyl transferase superfamily." FEBS Lett **517**(1-3): 1-6.
- Rutter, G. A., Tavare, J. M., *et al.* (2000). "Regulation of Mammalian Gene Expression by Glucose." News Physiol Sci **15**: 149-154.
- Ryu, S., Ramseier, T. M., *et al.* (1995). "Effect of the FruR regulator on transcription of the pts operon in *Escherichia coli*." J Biol Chem **270**(6): 2489-2496.
- Saier, M. H., Goldman, S. R., *et al.* (2002). Overall transport capabilities of *Bacillus subtilis*. Bacillus subtilis and its closest relatives: from genes to cells. Sonenshein, A. L., Hoch, J. A. and Losick, R. Washington, D. C., ASM Press: 113-128.

- Saier, M. H., Jr., Feucht, B. U., *et al.* (1975). "Regulation of intracellular adenosine cyclic 3':5'-monophosphate levels in *Escherichia coli* and *Salmonella typhimurium*. Evidence for energy- dependent excretion of the cyclic nucleotide." J Biol Chem **250**(19): 7593-7601.
- Saier, M. H., Jr. and Ramseier, T. M. (1996). "The catabolite repressor/activator (Cra) protein of enteric bacteria." J Bacteriol **178**(12): 3411-3417.
- Saier, M. H., Ramseier, T. M., *et al.* (1995). Regulation of Carbon Utilization. *Escherichia coli* and *Samonella typhymurium*. Washington, D. C., ASM Press: 1325-1343.
- Saraste, M., Sibbald, P. R., *et al.* (1990). "The P-loop--a common motif in ATP- and GTP-binding proteins." Trends Biochem Sci **15**(11): 430-434.
- Savery, N. J., Lloyd, G. S., *et al.* (2002). "Determinants of the C-terminal domain of the *Escherichia coli* RNA polymerase alpha subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters." J Bacteriol **184**(8): 2273-2280.
- Schirmer, F., Ehrt, S., *et al.* (1997). "Expression, inducer spectrum, domain structure, and function of MopR, the regulator of phenol degradation

- in *Acinetobacter calcoaceticus* NCIB8250." J Bacteriol **179**(4): 1329-1336.
- Schlax, P. J., Capp, M. W., *et al.* (1995). "Inhibition of transcription initiation by lac repressor." J Mol Biol **245**(4): 331-350.
- Schumacher, M. A., Choi, K. Y., *et al.* (1995). "Mechanism of corepressor-mediated specific DNA binding by the purine repressor." Cell **83**(1): 147-155.
- Schumacher, M. A., Choi, K. Y., *et al.* (1994). "Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices." Science **266**(5186): 763-770.
- Schumacher, M. A., Macdonald, J. R., *et al.* (1993). "Structural analysis of the purine repressor, an *Escherichia coli* DNA-binding protein." J Biol Chem **268**(17): 12282-12288.
- Shi, L., Potts, M., *et al.* (1998). "The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait." FEMS Microbiol Rev **22**(4): 229-253.
- Shirakihara, Y. and Evans, P. R. (1988). "Crystal structure of the complex of phosphofructokinase from *Escherichia coli* with its reaction products." J Mol Biol **204**(4): 973-994.

- Spronk, C. A., Bonvin, A. M., *et al.* (1999). "The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator." Structure Fold Des 7(12): 1483-1492.
- Steinhauer, K., Allen, G. S., *et al.* (2002). "Crystallization, preliminary X-ray analysis and biophysical characterization of HPr kinase/phosphatase of *Mycoplasma pneumoniae*." Acta Crystallogr D Biol Crystallogr 58(Pt 3): 515-518.
- Steinhauer, K., Jepp, T., Hillen, W., and Stuelke, J. (2002). "A novel mode of control of *M. pneumoniae* HPr kinase/phosphatase activity reflects its parasitic lifestyle." Microbiology 148: 3277-3284.
- Steller, I., Bolotovskiy, R., *et al.* (1997). "An algorithm for automatic indexing of oscillation images using fourier analysis." Journal of Applied Crystallography 30: 1036-1040.
- Stock, A. M., Robinson, V. L., *et al.* (2000). "Two-component signal transduction." Annu Rev Biochem 69: 183-215.
- Stormo, G. D. and Hartzell, G. W., 3rd (1989). "Identifying protein-binding sites from unaligned DNA fragments." Proc Natl Acad Sci U S A 86(4): 1183-1187.
- Stout, G. H. and Jensen, L. H. (1989). X-ray structure determination: a practical guide., John Wiley & Sons.



- Stulke, J. and Hillen, W. (1998). "Coupling physiology and gene regulation in bacteria: the phosphotransferase sugar uptake system delivers the signals." Naturwissenschaften **85**(12): 583-592.
- Stulke, J. and Hillen, W. (2000). "Regulation of carbon catabolism in *Bacillus* species." Annu Rev Microbiol **54**: 849-880.
- Tanaka, T., Saha, S. K., *et al.* (1998). "NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ." Nature **396**(6706): 88-92.
- Terwilliger, T. C. (1997). "Multiwavelength anomalous diffraction phasing of macromolecular structures: analysis of MAD data as single isomorphous replacement with anomalous scattering data using the MADMRG Program." Methods Enzymol **276**: 530-537.
- Terwilliger, T. C. (2000). "Maximum-likelihood density modification." Acta Crystallogr D Biol Crystallogr **56**(Pt 8): 965-972.
- Terwilliger, T. C. and Berendzen, J. (1999). "Automated MAD and MIR structure solution." Acta Crystallogr D Biol Crystallogr **55**(Pt 4): 849-861.
- Titgemeyer, F., Mason, R. E., *et al.* (1994). "Regulation of the raffinose permease of *Escherichia coli* by the glucose- specific enzyme IIA of the phosphoenolpyruvate:sugar phosphotransferase system." J Bacteriol **176**(2): 543-546.

- Tobisch, S., Stulke, J., *et al.* (1999a). "Regulation of the *lic* operon of *Bacillus subtilis* and characterization of potential phosphorylation sites of the LicR regulator protein by site-directed mutagenesis." J Bacteriol **181**(16): 4995-5003.
- Tobisch, S., Zuhlke, D., *et al.* (1999b). "Role of CcpA in regulation of the central pathways of carbon catabolism in *Bacillus subtilis*." J Bacteriol **181**(22): 6996-7004.
- Tong, L. and Rossmann, M. G. (1997). "Rotation function calculations with GLRF program." Methods Enzymol **276**: 594-611.
- Tortosa, P., Aymerich, S., *et al.* (1997). "Multiple phosphorylation of SacY, a *Bacillus subtilis* transcriptional antiterminator negatively controlled by the phosphotransferase system." J Biol Chem **272**(27): 17230-17237.
- Traut, T. W. (1994). "The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites." Eur J Biochem **222**(1): 9-19.
- Tronrud, D. E. (1997). "TNT refinement package." Methods Enzymol **277**: 306-319.

- van Aalten, D. M., DiRusso, C. C., *et al.* (2000). "Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold." Embo J **19**(19): 5167-5177.
- van Tilbeurgh, H. and Declerck, N. (2001a). "Structural insights into the regulation of bacterial signalling proteins containing PRDs." Curr Opin Struct Biol **11**(6): 685-693.
- van Tilbeurgh, H., Le Coq, D., *et al.* (2001b). "Crystal structure of an activated form of the PTS regulation domain from the LicT transcriptional antiterminator." Embo J **20**(14): 3789-3799.
- Versele, M., Lemaire, K., *et al.* (2001). "Sex and sugar in yeast: two distinct GPCR systems." EMBO Rep **2**(7): 574-579.
- Walker, J. E., Saraste, M., *et al.* (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." Embo J **1**(8): 945-951.
- Wang, B. C. (1985). "Resolution of phase ambiguity in macromolecular crystallography." Methods Enzymol **115**: 90-112.
- Weickert, M. J. and Adhya, S. (1992). "A family of bacterial regulators homologous to Gal and Lac repressors." J Biol Chem **267**(22): 15869-15874.

- Weickert, M. J. and Chambliss, G. H. (1990). "Site-directed mutagenesis of a catabolite repression operator sequence in *Bacillus subtilis*." Proc Natl Acad Sci U S A **87**(16): 6238-6242.
- Weik, M., Ravelli, R. B., *et al.* (2000). "Specific chemical and structural damage to proteins produced by synchrotron radiation." Proc Natl Acad Sci U S A **97**(2): 623-628.
- Weiner, J., 3rd, Herrmann, R., *et al.* (2000). "Transcription in *Mycoplasma pneumoniae*." Nucleic Acids Res **28**(22): 4488-4496.
- West, A. H. and Stock, A. M. (2001). "Histidine kinases and response regulator proteins in two-component signaling systems." Trends Biochem Sci **26**(6): 369-376.
- Wylie, C. R. J. (1966). Advanced Engineering Mathematics, McGraw-Hill Book Company.
- Yoshida, K., Kobayashi, K., *et al.* (2001). "Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis*." Nucleic Acids Res **29**(3): 683-692.
- Yoshida, K., Ohmori, H., *et al.* (1995). "*Bacillus subtilis* gnt repressor mutants that diminish gluconate-binding ability." J Bacteriol **177**(16): 4813-4816.

- Yuvaniyama, J., Denu, J. M., *et al.* (1996). "Crystal structure of the dual specificity protein phosphatase VHR." Science **272**(5266): 1328-1331.
- Zhang, Z. Y., Wang, Y., *et al.* (1994). "The Cys(X)5Arg catalytic motif in phosphoester hydrolysis." Biochemistry **33**(51): 15266-15270.