INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UNIVERSITY OF OKLAHOMA

GRADUATE SCHOOL

THE STRUCTURE AND FUNCTION OF YPD1, A HISTIDINE-CONTAINING PHOSPHOTRANSFER PROTEIN FROM YEAST

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

QINGPING XU
Norman, Oklahoma
2001

UMI Number: 3025982

UMI®

UMI Microform 3025982
Copyright 2001 by Bell & Howell Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

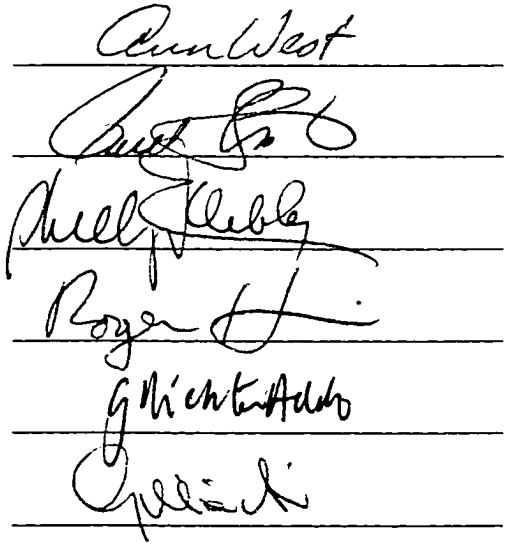
Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© Copyright by QINGPING XU 2001
All Rights Reserved

THE STRUCTURE AND FUNCTION OF YPD1, A HISTIDINE-CONTAINING PHOSPHOTRANSFER PROTEIN FROM YEAST

A Dissertation APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY



To my parents

Acknowledgements

First of all, my gratitude goes to Dr. Ann West who supervised me for the past five years. Dr. West has always supported me in the projects that I was involved in and has been an invaluable resource. Without her guidance and help, none of the results presented here would have existed. I am also thankful to her for giving me the freedom and opportunities for learning, for her great ideas and enthusiasm, for her encouragement, and for the trust that she placed in me to work at my own pace in a joyful, pressure free environment.

I am grateful to my current and past Advisory Committee members, Dr. Paul Cook, Dr. Phillip Klebba, Dr. Gillian Air, Dr. Dick van der Helm, Dr. Roger Harrison and Dr. George Richter-Addo, for their guidance. I have benefitted greatly from their advice. Dr. van der Helm answered many of my questions in crystallography. Dr. Cook helped me with many of my questions in biochemistry.

It has been a pleasure working with my colleagues, in particular, Dr. Fabiola Janiak-Spens, Stace Porter, and the Sparling brothers Jeff and David. I am deeply grateful to Fabiola and Stace for investing time and energy answering my questions and helping me with experiments. Viet Nguyen started the project by cloning the yeast YPD1 gene into a bacterial expression vector. Dr. West worked out purification conditions for YPD1. Dr. Fabiola Janiak-Spens generated and crystallized the YPD1 mutants used in the study described in Chapter 3 and 4. Eric Enwall has been extremely helpful in setting up and managing our Silicon Graphics workstations. Jean Keil helped me through a lot of paperwork. Without these contributors, it would have been impossible for me to finish this study in a timely manner.

I enjoyed a great friendly atmosphere between our lab and our neighbors from Dr. Cook's lab. I am grateful that Dr. Cook and his group generously allowed me to use their equipment and taught me how to use them. I am also thankful to Lalitha Venkatramani and Ranjan Chakraborty, members of Dr. van der Helm's lab, for their help and friendship.

Free exchange of ideas/information and tools is essential for science to progress, protein crystallography is no exception. I am fortunate to work in such a field with a great

sharing spirit and tradition. With many brilliant ideas and freely available techniques/tools, and particularly software programs shared within the community, protein crystallography or structural biology has taken great strides in recent years. My work would be impossible without these tools and techniques. Many people, which I never met in person, shared their thoughts and experiences with me and helped me through many difficult situations at various stages of my research.

I also want to thank Drs. Stephen Sprang and Johann Deisenhofer at University of Texas Southwestern Medical Center, Dr. Ann Stock of University of Medicine and Dentistry of New Jersey, and the staff at the National Synchrotron Light Source (station X12B) at Brookhaven National Laboratory for generous use of their X-ray facilities.

I would like to thank Dr. Ann West and Dr. Ann Stock for giving me the opportunity to participate in the structure determination of the chemotaxis protein CheB.

This dissertation would not have been possible without Chunyan, my wife and friend. Her continuing support and encouragement has seen me through good and bad times. I thank her for taking most of the responsibilities at home, for her understanding when I was away with my research work, for her delicious meals, and above all her love. I would also like to thank my parents, my brothers and sister, for their never-ending support. My elder brother, David, is particularly helpful in almost every aspect of my life.

Table of Contents

List of Tables

List of Figures

Molecular Graphics Figures

All the molecular images were produced with MOLSCRIPT (Kraulis, 1991) or BOBSCRIPT (Esnouf, 1997) and rendered with RASTER3D (Merritt and Bacon, 1997) and GL_RENDER (Lothar Esser). Figures of electron density maps were generated with the program XTALVIEW (McRee, 1993) or SETOR (Evans, 1993). Molecular surface images were generated using the program GRASP (Anthony *et al.*, 1991).

List of Abbreviations

$ArcB^c$	C-terminal domain of ArcB
asu	asymmetric unit
ATP	adenosine triphosphate
C-terminus	carboxy-terminus
kDa	kilo Dalton
FOM	figure of merit
HK	histidine kinase
HPt	histidine-containing phosphotransfer (protein)
MAPK	mitogen activated protein kinase
MAPKK	mitogen activated protein kinase kinase
MAPKKK	mitogen activated protein kinase kinase kinase
MIR	multiple isomorphous replacement
MR	molecular replacement
MIRAS	multiple isomorphous replacement with anomalous scattering
N-terminus	amino-terminus
NCS	non-crystallographic symmetry
NMR	nuclear magnetic resonance
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG(mme)	polyethylene glycol (monomethyl ether)
PTS	phosphoenol pyruvate, sugar phosphotransferase system
SLN1-R1	response regulator domain of SLN1
SSK1-R2	response regulator domain of SSK1
SKN7-R3	response regulator domain of SKN7
RCSB	Research Collaboratory for Structural Bioinformatics

r.m.s.d	root-mean-square deviation
RR	response regulator
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
WT	wild-type

Standard three-letter and one-letter amino acid abbreviations are used,
A = Ala = alanine, C = Cys = cysteine, D = Asp = aspartic acid, E = Glu = glutamic acid, F = Phe = phenylalanine, G = Gly = glycine, H = His = histidine, I = Ile = isoleucine, K = Lys = lysine, L = Leu = leucine, M = Met = methionine, N = Asn = asparagine, P =Pro = proline, Q = Gln = glutamine, R = Arg = arginine, S = Ser = serine, T = Thr = threonine, V = Val = valine, W = Trp = tryptophane, and Y = Tyr = tyrosine

Abstract

THE STRUCTURE AND FUNCTION OF YPD1

YPD1 is an important protein involved in yeast cellular responses to hyperosmolarity and oxidative stress. In *Saccharomyces cerevisiae*, the SLN1-YPD1-SSK1 phosphorelay pathway controls the activation of a downstream mitogen-activated protein (MAP) kinase upon hyperosmotic stress, whereas the SLN1-YPD1-SKN7 pathway is involved in responses to oxidative stress and cell wall biosynthesis. At a molecular level, YPD1 is required for phosphoryl group transfer from the sensor kinase SLN1 to the response regulator proteins SSK1 or SKN7. YPD1 appears to be a versatile protein in accepting/donating phosphoryl group from/to various response regulator domains, some from other species.

The X-ray structure of YPD1 was solved at a resolution of 1.75 Å by conventional multiple isomorphous replacement with anomalous scattering. The tertiary structure of YPD1 consists of six α-helices and a short 3_{10}-helix, four of which form a central four-helix bundle core. The histidine residue that is phosphorylated (His64) is located on the surface of the four-helix bundle. Structure-based comparisons of YPD1 to other proteins having a similar function, such as the *Escherichia coli* ArcB histidine-containing phosphotransfer (HPt) domain and the P1 domain of the CheA kinase, as well as Spo0B from *Bacillus subtilis*, revealed that the helical bundle and several structural features around the phosphorylatable histidine residue are conserved between the prokaryotic and eukaryotic kingdoms. The conservation of YPD1 structure correlates well with observations that YPD1 can functionally substitute for other HPt domains despite limited sequence homology. These findings suggest that the four-helix bundle fold, particularly, the αC-αD helical hairpin motif with the centrally located exposed phosphorylation site, represent important conserved features of other HPt domains that interact with response regulators.

In order to understand the His-Asp phosphotransfer mechanism, a series of YPD1 mutants defective in phosphotransfer ability were further characterized through the determination of their three dimensional structures. The G68Q mutant crystallized in a space group ($P3_121$), which is different from that of the wild type YPD1 protein ($P4_32_12$). The

structure of the G68Q mutant was determined by the method of molecular replacement. The G68Q mutant structure provided further insight into the importance of the exposed histidine with respect to YPD1 function. Based on the analysis of the G68Q structure and biochemical evidence, a conserved hydrophobic molecular surface on YPD1 was identified and postulated to be important in binding response regulator domains.

The role of a hydrogen bond between Gln86 and the phosphorylatable histidine (His64) was also investigated. Structures of three YPD1 mutants, Q86A, Q86E and Q86L revealed that the hydrogen bond to the phosphorylatable histidine can be retained through an intervening water molecule or subtle rearrangement of side-chains. These structures, combined with results from side-directed mutagenesis studies, suggest that the hydrogen bond, although it may not be essential to YPD1 function, might be involved in stabilizing the exposed histidine, and as a result may affect the efficiency of phosphotransfer.

Finally, through molecular modeling and structural comparisons, a general model of how YPD1 and response regulator domains might interact was proposed. It seems likely that YPD1 binds to response regulator domains through conserved hydrophobic surfaces on both the response regulator domains and YPD1. This model supports the hypothesis that molecular interactions between HPt domains and response regulators, in general, are likely to be conserved in other systems as well.

1 Introduction

1.1 Signal transduction and protein phosphorylation

Living organisms must continually monitor the external environment and adjust to it accordingly in order to survive. Single cell organisms such as bacteria have a wide range of regulatory responses to adapt to environmental conditions. For example, when nutrients become available, transport and processing machinery are activated to assimilate them. Likewise, complex multicellular organisms must coordinate their cellular responses in order to survive. The process of converting signals at the cell membrane (such as binding of the receptor to ligand) to specific cellular responses is generally known as signal transduction. The study of sensory-response systems will help to elucidate mechanisms by which organisms detect and adapt to their environment.

Molecular studies have revealed that reversible protein phosphorylation is one of the most widely used mechanisms for regulating biological responses, including intracellular signal transduction (Alberts *et al.*, 1994). The addition of a phosphoryl group to a protein could affect its structural, thermodynamic and kinetic properties (Knowles, 1980; Westheimer, 1987; Witters, 1990; Barford, 1991; Johnson and Barford, 1993). In most cases, the eventual outcome is an alteration in cellular activity and changes in the patterns of genes expressed within the responding cells. The phosphorylation events often function as molecular "switches". Phosphorylation can change the conformational equilibrium between different functional states. The electrostatic properties of the phosphoryl group are usually critical in these reactions, since the phosphoryl group adds two negative charges to the protein which may serve to modify or disrupt electrostatic interactions. Protein phosphorylation can cause a structural change, for example, by attracting a cluster of positively charged side chains. Such a change occurring at one site in a protein can, in turn, alter the protein's conformation elsewhere. This structural change may alter substrate binding and catalytic activity of a phosphorylated enzyme. Phosphorylation and dephosphorylation

reactions, usually under the control of kinases and phosphatases, can result in responses that last less than a second or could span several hours. In some cases, a cascade of phosphorylation and dephosphorylation reactions can amplify a signal from outside the cell membrane, such as hormones and growth factors. Due to these properties, reversible protein phosphorylation is an ideal system for regulating cellular processes (Westheimer, 1987).

In eukaryotic cells, reversible protein phosphorylation is used predominantly to control the activity of proteins. It was estimated that >10% of the 10,000 proteins in a typical mammalian cell are subject to phosphorylation (Alberts *et al.*, 1994). In general, the phosphoryl groups are transferred from ATP molecules by protein kinases. The phosphorylated proteins are often under the regulation of protein phosphatases. Eukaryotic cells contain a large variety of these enzymes, many of which play a central role in intracellular signaling. Residues which are the usual sites of phosphorylation in eukaryotic cells are serine, threonine and tyrosine hydroxyl groups. For example, mitogen-activated protein kinases (MAPK) are a large group of related proteins whose primary functions seem to be in regulating mitosis and signaling cell differentiation (Nishida and Gotoh, 1993; Robinson and Cobb, 1997; Chang and Karin, 2001). MAP kinases become activated when they have been doubly phosphorylated at a specific tyrosine and threonine.

In contrast, the activity of proteins in a bacterium is mainly regulated by small molecules in the cell that bind to specific proteins (Alberts *et al.*, 1994). However, many bacterial protein activities are controlled by phosphorylation. In particular, bacteria have employed a different but conserved phosphotransfer signaling mechanism for eliciting a large variety of adaptive responses to their environment (Hoch and Silhavy, 1995; Mizuno, 1998). In these systems, often known as two-component signaling systems, phosphorylation of proteins occurs on histidine or aspartate residues instead.

1.2 Signal transduction involving two-component systems

Ninfa *et al.* (Ninfa and Magasanik, 1986) proposed the first model for two-component signal transduction with the studies of nitrogen regulation in *E. coli* although individual components of parallel systems were found earlier (*e.g.* Mizuno *et al.*, 1982;

Tommassen *et al.*, 1982). Simple two-component systems are characterized by an auto-phosphorylating histidine protein kinase which, in turn, serves as a phosphoryl donor to a response regulator protein (Hoch and Silhavy, 1995). The functional state of the response regulator protein is modulated by phosphorylation and dephosphorylation of a conserved aspartic acid residue. More complex signaling pathways that involve histidine-to-aspartate phosphoryl transfer reactions between two or more modular protein domains, the so-called phosphorelay systems, have been described in both prokaryotic and eukaryotic organisms (Alex and Simon, 1994; Appleby *et al.*, 1996; Brown and Firtel, 1998; Chang and Steward, 1998; Perraud *et al.*, 1999; Wurgler-Murphy and Saito, 1997).

Two-component signaling systems are prevalent in bacteria. *Escherichia coli* alone, for example, contains more than 30 response regulator proteins (Mizuno, 1997). These systems are involved in the control of metabolic pathways (Falke *et al.*, 1997), cell division (Ohta *et al.*, 1992), osmolarity control (Forst and Roberts, 1994; Egger *et al.*, 1997), antibiotic production (Klein *et al.*, 1993), drug resistance (Arthur *et al.*, 1992; Lai and Kirsh, 1996), virulence (Akerley, 1992; DiRita, 1992; Dziejman and Mekalanos, 1997) and many other bacterial processes (Fabret and Hoch,1998; Hoch and Silhavy, 1995; Quon *et al.*, 1996). Variants of the two-component system in lower eukaryotes, such as *Saccharomyces cerevisiae* (Ota and Varshavsky, 1993; Posas, *et al.*, 1996), *Schizosaccharomyces pombe* (Aoyama *et al.*, 2000; Nguyen *et al.*, 2000; Cottarel, 1997), *Arabidopsis thaliana* (Chang *et al.*, 1993), *Candida albicans* (Calera *et al.*, 1998; Alex *et al.*, 1998; Nagahashi *et al.*, 1998), *Dictyostelium discoideum* (Brown and Firtel, 1998 and reference therein) and *Neurospora crassa* (Alex *et al.*, 1996), have also been reported.

Studies of two-component signaling systems have grown over the past fifteen years. As shown in Figure 1-1, there has been a steady increase of the number of published papers on this subject each year. Due to increased research as well as genomic sequencing efforts, a very large number of histidine kinases and response regulators have been identified and deposited in public databases such as SWISS-PROT (Bairoch and Apweiler, 2000). For example, currently there are 874 response regulator domain sequences assembled in the PFam database (Bateman *et al.*, 2000).

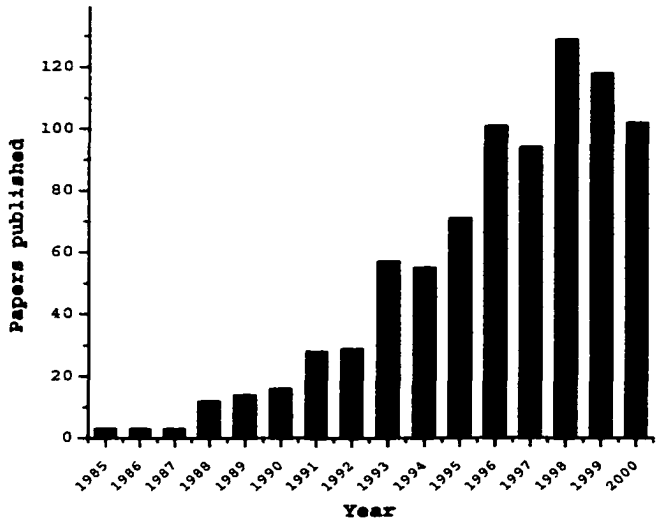
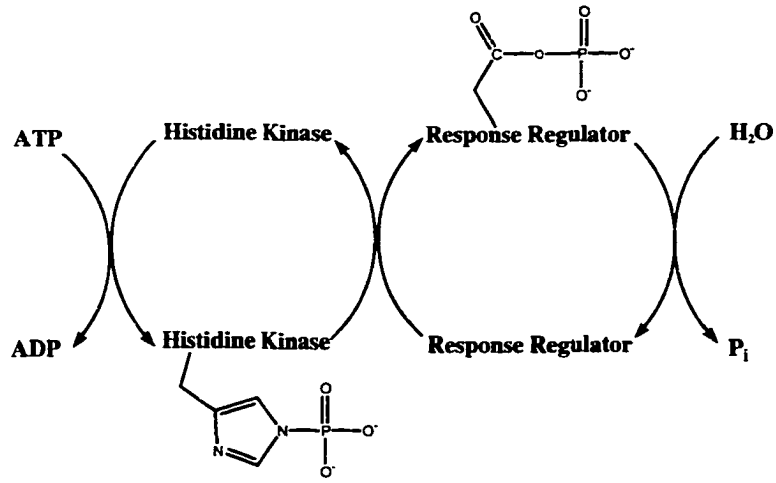


Figure 1-1. Survey of papers published on studies of two-component signal transduction systems. The papers published were estimated via searching PubMed at NCBI (the National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/PubMed/) using keywords such as "two-component signal transduction", "response regulator and histidine kinase", "response regulator CheY", "His-Asp phosphorelay". The result is only an approximation up to the time the survey was conducted (Jan, 2001).

1.2.1 His-Asp phosphotransfer

Although the known two-component signaling systems show variability in terms of domain composition and architecture, and many two-component signaling systems actually consist of more than two components, the basic module of phosphotransfer from histidine to aspartate is conserved. A prototypical two-component signal transduction system consists of a histidine kinase (HK) and a response regulator (RR) (reviewed in Parkinson, 1993; Swanson *et al.*, 1994; Stock *et al.*, 1995; Hoch, 2000; Stock *et al.*, 2000). Histidine kinase activity is modulated through a sensor domain of the kinase. The sensor domain is usually located at the N-terminus of the protein where typically, transmembrane regions serve to position an extracellular domain which is responsive to a specific environmental signal. Through a mechanism which is still not well understood (progress reviewed in Falke and Hazelbauer, 2001), the signal from the outside of the cell is propagated across the inner membrane. The histidine kinase becomes activated and auto-phosphorylates a conserved histidine residue using Mg-ATP as a phosphodonor. The phosphoryl group is then transferred to an aspartate residue on the response regulator, which in turn interacts with a down-



HK-His + ATP <==> HK-His~P +ADP

HK-His~P + RR-Asp <==> HK-His + RR-Asp~P

RR-Asp~P + H_2O <==> RR-Asp + P_i

Figure 1-2. The basic scheme of two-component signal transduction systems.

stream effector (Figure 1-2). In many cases, the response regulator domain is physically attached to a DNA binding domain, whereby the response regulator functions as transcription factor.

A relatively straightforward example of a two-component signaling system is the *E. coli* EnvZ-OmpR osmoregulation system. This system regulates gene expression of two outer membrane porins, OmpF and OmpC, in response to changes in extracellular osmolarity. OmpF and OmpC, each with a different pore size, facilitate the transport of small hydrophilic molecules across the membrane. At low osmolarity, OmpF, the protein with larger pore diameter is predominantly expressed, thus allowing bigger substrates to pass through under dilute solute concentration. Under high solute concentrations, *E. coli* must have a method of detecting this and shutting down the expression of OmpF. This is accomplished with the histidine kinase EnvZ and the response regulator OmpR (Forst and Roberts, 1994).

The osmosensor EnvZ is an archetypical dimeric transmembrane protein which is capable of detecting changes in external solute concentration. Each monomer is composed

of three functional domains: an N-terminal periplasmic sensing domain, a cytoplasmic dimerization domain containing the histidine phosphorylation site and a C-terminal ATP-binding kinase domain. Like many histidine kinases, EnvZ also exhibits phosphatase activity toward its cognate response regulator, OmpR. When EnvZ detects high solute concentration with its extracellular domain, the ratio of kinase to phosphatase activity is modulated such that the level of phosphorylated OmpR in the cell is increased. OmpR is a protein of 239 residues, consisting of two functional domains: an N-terminal response regulator domain and C-terminal domain which is responsible for binding *ompC/ompF* promoters. When OmpR is phosphorylated, its ability to bind DNA is enhanced (Head *et al.*, 1998). Although the details are not very clear, the EnvZ-OmpR two-component regulatory system induces the transcription of OmpC and represses the transcription of OmpF (Pratt and Silhavy, 1995; Pratt *et al.*, 1996).

Many of the current studies on two-component signaling system are focused on structures and functions of individual domains as well as their interactions of the proteins in the two-component systems. Since almost all the histidine kinases and many response regulators are flexible multi-domain proteins, and most histidine kinases are membrane associated, it has been difficult to express and purify these proteins *in vitro*. The "divide-and-conquer" strategy has proven to be very effective in structural determination a large protein. This method aims to understand the structure and function of a multi-domain protein by focusing on the structural and function of one or more sub-domains first. Due to decades of physiological, genetical, biochemical and structural studies, our understanding of the structure and function of individual domains of histidine kinases and response regulators has been greatly improved. Many of these advances have been recently reviewed extensively in the literature (*e.g.* Stock *et al.*, 2000; West and Stock, 2001; Stock and West, 2001; Falke and Hazelbauer, 2001), some of the results are summarized below.

1.2.2 Histidine kinase proteins

The large majority of histidine kinases are architecturally similar to EnvZ. The N-terminal sensor domains, delineated by two membrane spanning hydrophobic sequences and located in the periplasm of the cell, and are generally believed to be involved in the signal sensing. As expected, there is little sequence homology between these sensor

domains. This is consistent with the fact that histidine kinases are responsive to a large number of different environmental signals. The C-terminal cytoplasmic kinase domains, often referred to as "transmitter" domains, are generally more conserved, with an average of 25% sequence homology among different proteins from different species. Members of the histidine kinase superfamily exhibit clusters of highly conserved sequence motifs that are presumed to be involved in important functions such as substrate binding, catalysis, and structure stability. These conserved primary sequence motifs have been designated as the H, N, G1, F, and G2 boxes. Grebe and Stock (1999) systematically analyzed the sequences of the histidine kinase superfamily and categorized more than 300 histidine kinases into different subfamilies based on sequence similarities within the kinase domain. The histidine kinases function as dimers. Although there are no experimental structure models, biochemical data support the hypothesis that the transmembrane portion of the histidine kinase is a four-helix bundle formed by the dimerization of the two transmembrane helices from each histidine kinase monomer. Additionally, biochemical and mutagenesis studies have suggested that the histidine kinase involves a trans-autophosphorylation reaction in which the catalytic domain of one histidine kinase subunit phosphorylates a specific histidine on the other histidine kinase subunit in the dimer.

There are no definitive models for the mechanism of how the signals are detected and transmitted across the membrane. However, recently NMR structures of two cytoplasmic domains of EnvZ, including the kinase domain and dimerization domain, have provided important information about the histidine kinase structure and function. The tertiary structure of the EnvZ catalytic domain (PDB code 1BXD, Tanaka *et al.*, 1998), shown in Figure 1-3, can be described as a two-layered α/β sandwich consisting of a flat 5-stranded β sheet on one side and three α helices on the other. Additionally, a small antiparallel β sheet is located on one end of the larger β sheet. The G1-box, N-box and F-box are all near the ATP binding site. The N-box is located on the second helix. The F and G2 sequence motif are located on a highly mobile central loop. The overall structure of the EnvZ catalytic domain resembles heat-shock protein 90 (HSP90, Prodromou *et al.*, 1997) and DNA gyrase B (Wigley *et al.*, 1991) and MutL (Ban and Yang, 1998), all of which bind ATP. The ATP-binding region reveals similar protein-folds as well.

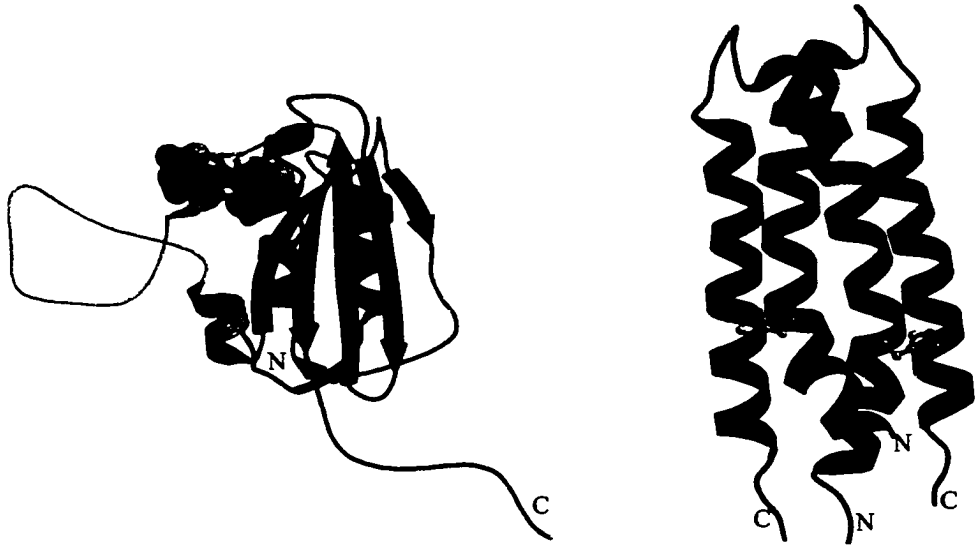


Figure 1-3. NMR solution structures of the catalytic domain of the *E. coli* EnvZ histidine kinase (left) bound to ATP analog AMP-PNP (shown as CPK spheres) and the homodimeric core domain (right). The side chain for the histidine phosphorylation site is shown in ball-and-stick representation.

Solution NMR structures of the dimerization domain of EnvZ (PDB code 1JOY) revealed that the autophosphorylation site of these histidine kinases is located on the surface of an up-down-up-down helix bundle (Tomomori *et al.*, 1999). The bundle is formed by a homodimer of two helix-turn-helix motifs. The turn region is made up of nine residues. The H-box sequence motif is located in the middle of the N-terminal alpha helix as shown in Figure 1-3. The two polypeptide chains are held together via hydrophobic interactions.

1.2.3 Response regulator proteins

Response regulators usually function at the end of a phosphotransfer pathway. They regulate output responses using phosphorylation as a "molecular switch". Many response regulators have more than one domain, a conserved regulatory domain containing the invariant aspartic acid side chain as well as a variable C-terminal effector domain. The regulatory domain itself can function as an output regulator as in CheY. However, adding an effector domain vastly increases the types of responses that the two-component systems can regulate. The majority of the effector domains are DNA binding domains. Some effector domains have enzymatic activities. For example, in the chemotaxis protein CheB, the

esterase activity of the effector domain is regulated by the phosphorylation state of its regulatory domain (reviewed in Djordjevic and Stock, 1998).

The regulatory domains of response regulators, also known as receiver domains, in general share ~20% overall sequence identity among diverse species (Volz, 1993; Volz, 1995). Several key active site residues within response regulator domains are completely conserved across the whole family. These domains can be phosphorylated at an aspartate residue via histidine kinases, histidine-containing phosphotransfer (HPt) domains, or sometimes small molecule phopho-donors. The phosphoaspartic acid is chemically unstable but its hydrolysis rate can vary depending on the response regulator. Typical half-lives of phosphorylated response regulator domains range from seconds to days.

Members of the response regulator superfamily are speculated to have the same tertiary structure as CheY, a key component in bacterial chemotaxis, and the first known crystal structure of a response regulator (Stock *et al.*, 1993). This assumption has been consolidated by several other response regulator structures from different species (reviewed in Robinson *et al.*, 2000), such as Spo0F (Madhusudan *et al.*, 1996), NarL (Baikalov *et al.*, 1996), PhoB (Sola *et al.*, 1999), and CheB (Djordjevic *et al.*, 1998). The regulatory domain, consisting of approximately 125 residues, has a well defined $(\beta/\alpha)_5$ three-layer sandwich fold. This fold consists of a hydrophobic central 5-stranded parallel β sheet, protected on both sides by amphipathic helices. The residues that are highly conserved in sequences in the CheY family are clustered around the active site. The active site is located at the C-terminal ends of the β sheet. This active site location is typical in proteins containing a Rossmann fold (Branden and Tooze, 1991). Three highly conserved aspartate residues at the active site, Asp12, Asp13 and Asp57, are involved in the coordination of a Mg^{2+} ion which is essential for phosphorylation of the invariant Asp57 (Figure 1-4).

It has long been speculated that response regulator domains share a similar activation mechanism upon phosphorylation. A phosphorylation-dependent conformational change is believed to be essential for response regulator activation. Significant effort has been devoted to understanding the details of the conformational changes upon phosphorylaton of response regulator domains. However, the phosphorylated state of response regulators are difficult to study due to autophosphatase activity of regulatory domains. Recent

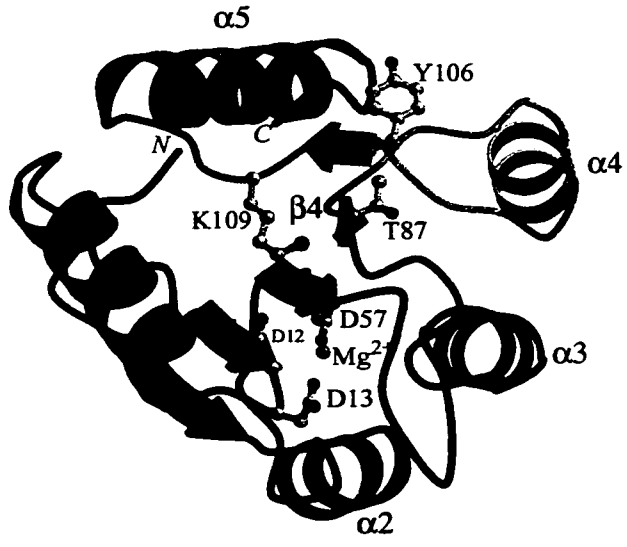


Figure 1-4. Structure of CheY (PDB code 2CHE, Stock *et al.*, 1993). The secondary structural elements are labelled sequentially from β1 to α5. Functional important residues referred in the text are shown in ball-and-stick.

advances, such as the determination of the structure of BeF_3/regulatory domain complexes and NMR studies of phosphorylated state in solution, have provided insights about the nature of these structural changes (Birck *et al.*, 1999; Kern *et al.*, 1999; Lewis *et al.*, 1999; Yan *et al.*, 1999; Cho *et al.*, 2000; Lee *et al.*, 2001). These studies suggest that a common mechanism is involved in the phosphorylation-induced conformational change. A highly conserved residue, Lys109 in CheY, is positioned to interact with the phosphoryl group. Two other nearby conserved residues, Thr87 and Tyr106, were observed to have different conformations between the phosphorylated and unphosphorylated forms. In several response regulator domain structures, the Ser/Thr (Thr87 in CheY) side-chain moves towards the phosphorylated asparate residue, generating a space that is filled by repositioning of the Phe/Tyr residue (Tyr106 in CheY). This creates a "domino" effect that changes a surface known to be important for protein-protein interaction. This surface is formed by residues from α3, β4, α4, β5 and adjacent loops. The magnitude of the structural changes upon phosphorylation appears to be protein specific. In some cases, only small structural changes are found. However, in phosphorylated NtrC, larger changes involving rearrangements of molecular surfaces and thus alteration of molecular properties have been observed (Kern *et al.*, 1999).

These studies suggest that response regulator domains function as generic on/off switches. Phosphorylation shifts the equilibrium between two functionally distinct response regulator conformations. Structural changes of response regulator domains can be utilized or integrated in the subsequent reactions. For example, the conformational changes of the regulatory domain can affect the effector domains. Or the change of surface properties of the phosphorylated response regulator can promote favorable protein-protein interactions which otherwise were not favorable.

1.2.4 Practical implications of studies of two-component regulatory systems

Studies of two-component systems have provided knowledge concerning the underlying phosphotransfer mechanism. It is important to understand why and how such a simple signaling scheme can be tailored and adapted to achieve so many different responses. Since two-component systems are ubiquitous in bacteria and they exist in many pathogenic bacteria, the study of histidine kinases and response regulators has great practical implications because they are good targets for anti-microbial drug development (Desnottes *et al.*, 1996; Barrett *et al.* 1998; Barrett and Hoch, 1998; Fabret and Hoch, 1998; Frosco and Barrett, 1999). Furthermore, certain features are universally shared by all two-component systems. Therefore, it seems possible to identify or design inhibitors to a single target in a single pathogen or multiple targets in multiple pathogens.

In addition to the potential antimicrobial utilization of inhibitors to the two-component systems, the use of such inhibitors to suppress antibiotic resistance is also a distinct possibility (Guenzi *et al.*, 1994; Salyers *et al.* 1995; Lai and Kirsch, 1996; Gunn *et al.*, 1998). One of the better understood two-component systems involved in antibiotic resistance is the VanS:VanR system in *Enterococcus faecium*. The VanS:VanR system controls an operon responsible for the production of the peptidoglycan-modifying enzymes VanA and VanH (Arthur *et al.*, 1992). The VanH dehydrogenase reduces pyruvate to D-lactate (D-Lac) and the VanA ligase catalyzes synthesis of the depsipeptide D-alanyl-D-lactate (D-Ala-D-Lac). The depsipeptide replaces the native dipeptide D-Ala-D-Ala and leads to production of peptidoglycan precursors which bind glycopeptides with reduced affinity. For the glycopeptide vancomycin, it has an affinity 1000-fold less for the altered peptidoglycan

than it does for the native peptidoglycan (Lai and Kirsch, 1996; Walsh *et al.*, 1996). As a result, this subtle change in the peptidoglycan prevents vancomycin from binding to the growing peptidoglycan and inhibiting growth. Down-regulation of the VanS:VanR two-component system will lead to the suppression of the production of the downstream cell wall-modifying enzymes, VanA and VanH, and thus restore vancomycin susceptibility.

Since the theme of two-component signal transduction is conserved, there may be similar opportunities for the pharmaceutical application of two-component systems in yeast and fungi. Since two-component signaling systems have not been found in higher eukaryotes, the inhibitors designed to target bacteria and lower eukaryotes might have few side-effects in humans.

1.3 Multi-component His-Asp-His-Asp phosphorelay systems

The basic two-component His-Asp phosphotransfer mechanism is often embedded in more extensive phosphotransfer networks wherein the phosphoryl group is passed to a second histidine and then to a second aspartate. These types of phosphotransfer systems involving more than two components is often referred to as a "multi-step" or His-Asp-His-Asp phosphorelay system. For simplicity, the notation H1-D1-H2-D2 has been used to describe each module involved in the phosphorelay process. For example, H1 refers to the histidine kinase module. In a generic phosphorelay system, the phosphoryl group is transferred from H1 to D2 employing four sequential phosphorylation events (H1 autophosphorylation using ATP, H1 to D1, D1 to H2, H2 to D2). These phosphotransfers alternate between histidine and aspartate residues. In many cases, the histidine kinase, the first aspartate-containing receiver domain and sometimes the HPt domain are encoded together in a single large peptide, termed a "hybrid histidine kinase". About 20~30% of all known histidine protein kinases belong to this class (Grebe and Stock, 1999). This suggests that the His-Asp-His-Asp signaling strategy is widely utilized. All known eukaryotic two-component signaling pathways employ this strategy.

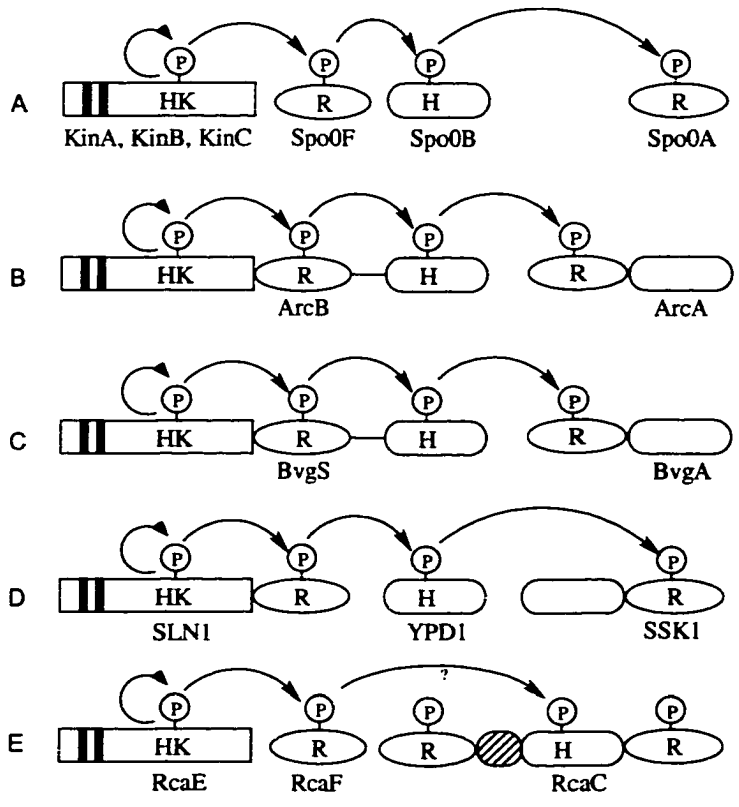


Figure 1-5. Examples of His-Asp-His-Asp phosphorelay pathways (see text for detailed descriptions). Variations in domain linkage are observed for different phosphorelay systems even though the basic His-Asp scheme is conserved.

1.3.1 Phosphorelay systems and architecture

Burbulys *et al.* (1991) were the first to describe a multi-step phosphorelay system (Kin-Spo0) governing the initiation of sporulation in *Bacillus subtilis*. The Kin-Spo0 pathway (Figure 1-5A) involves a four-protein phosphorelay system which ultimately activates the Spo0A transcriptional regulator (Burbulys *et al.*, 1991; Grossman, 1995). The phosphorelay begins with the autophosphorylation of one of three sensor kinases, KinA, KinB or KinC. The phosphoryl group is then transferred to an aspartate residue in Spo0F, a single domain response regulator. Spo0F serves as a phosphodonor for the next component in the series, Spo0B, a histidine-containing phosphotransfer (HPt) protein. Finally, the phosphoryl group completes its journey by transfer to an aspartate in Spo0A (Hoch, 1995).

Another example of multi-component phosphorelay pathway is the ArcB-ArcA system (Lynch and Lin, 1996; Iuchi and Weiner, 1996) that regulates aerobic/anaerobic

response in *E. coli* (Figure 1-5B). In the ArcB-ArcA phosphorelay system, the first three steps of the four-step phosphorelay occur within a single protein, ArcB. ArcB is a typical "hybrid histidine kinase. It resides in the cytoplasmic membrane and contains both a kinase domain (H1), a response regulator domain (D1) and a C-terminal HPt domain (H2). ArcB autophosphorylates a histidine residue (His292) in its histidine kinase domain (H1) triggered by appropriate environmental signal. The phosphoryl group is then sequentially transferred to an aspartate (Asp576) in the ArcB response regulator domain (D1), then to a histidine (His717) in the C-terminal domain, and finally to an aspartate (Asp54) of a physically independent response regulator (D2), ArcA. The BvgS-BvgA system, that modulates the transcriptional regulation of virulence factors in *Bordetella pertussis*, is architecturally similar to the ArcB-ArcA system (Akerley and Miller, 1996; Uhl and Miller, 1996). There are many other hybrid histidine kinases similar to ArcB and BvgS, such as the bacterial proteins BarA, EvgS, LemA, and TorS (Appleby *et al.*, 1996; Inouye, 1996; Brown and Firtel, 1998).

The basic modular domains (H1, D1, H2, D2) of phosphorelay systems often fuse together to form larger proteins. The combination of domains can vary dramatically (Figure 1-5) although the general phosphorelay scheme (from H1 to D2) is conserved. In the Kin-Spo0 pathway, the four phosphorylation sites are located on independent proteins. The SLN1-YPD1-SSK1 yeast phosphorelay system (Posas *et al.*, 1996) offers another variation involving a three-protein implementation of the His-Asp-His-Asp phosphorelay scheme (Figure 1-5D). In this system, the hybrid kinase SLN1 contains the histidine kinase and first response regulator domain (H1-D1), while YPD1 and SSK1 contain the second histidine phosphorylation site (H2) and response regulator domain (D2), respectively. This pathway will be discussed in detail in the next section.

To realize the His-Asp-His-Asp phosphorelay, there exist other possible architectural arrangements. For example, in the pathway of RcaE-RcaF-RcaC which governs chromatic adaptation in cyanobacteria, RcaE and RcaF act as the sensor kinase (H1) and response regulator (D1, Kehoe and Grossman, 1997) respectively. The protein RcaC, however, contains two intact regulator domains, one at each terminus (Chiang *et al.*, 1992). This represents another interesting scheme of phosphorelay (Figure 1-5E).

Although the signal is in general transmitted from H1 to D2, it is important to realize that phosphotransfer between two components in the phosphorelay system may not be exclusively linear or one directional. For example, it has been noted that H1 to D2 (in addition to H2-D2) phosphotransfer is possible *in vitro* (Tsuzuki *et al.*, 1995; Georgellis *et al.*, 1998). For the SLN1-YPD1-SSK1 system, it has been shown that SLN1-R1 and YPD1 phosphotransfer is bidirectional *in vitro*, whereas phosphotransfer from YPD1 to SSK1-R2 is predominately one directional favoring phosphorylation of SSK1 (Janiak-Spens *et al.*, 1999).

Interestingly, the His-Asp-His-Asp relay pathways found in lower eukaroytic organisms such as yeast and plant, are integrated with exclusively eukaryotic MAP kinase pathways as found in SLN1/SSK1 and ETR1/CTR1 from yeast and *Arabidopsis thaliana* respectively (Wurgler-Murphy *et al.*, 1997; Posas *et al.*, 1998; Woeste and Kieber, 1998), and also in *Schizosaccharomyces pombe* (Shiozaki *et al.*, 1997).

1.3.2 Why multi-component phosphorelay systems?

It is not completely understood why these multi-component phosphorelays are needed. Unlike the eukaryotic phosphorylation cascade, the multi-step relay scheme generally offers no signal amplification beyond the initial autophosphorylation site (*i.e.* first histidine site). Since many of the reported phosphorelay systems govern major developmental commitments or other critical responses, these pathways must be carefully regulated to prevent premature triggering of the potentially dangerous physiological processes. The multi-step relay might provide more sites for multiple regulatory checkpoints (Appleby *et al.*, 1996). Phosphatases are often found to act on specific phosphorylation sites and thus provide an additional mechanism for regulating signal flow in phosphorelay systems. For example, a number of phosphatases have been identified that regulate different components of the phosphorelay in the Kin-Spo0 phosphorelay system. Spo0E is a phosphatase that dephosphorylates the Spo0A~P protein. The response regulator Spo0F is regulated by three phosphatases RapA, RapB and RapE (Perego and Hoch, 1996).

Alternatively, the multiple phosphorylation sites of the phosphorelay could provide junction points for communicating with other signaling pathways. It has also been sug-

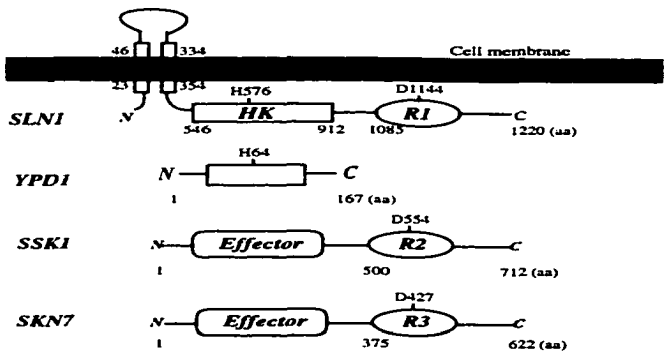


Figure 1-6. Components of the yeast phosphorelay system and nomenclature. The response regulator domains associated with SLN1, SSK1 and SKN7 are designated as R1, R2 and R3 respectively.

gested that the presence of multiple relay steps in one protein might increase signaling efficiency and reduce non-specific crosstalk from other pathways (Appleby *et al.*, 1996).

It is obvious that higher organisms favor phosphorylation on Ser/Thr/Tyr over the His/Asp phosphorylation. In higher organisms, the simpler His-Asp signaling transduction scheme gives way to more elaborate and more tightly regulated signaling pathways which often involve a cascade of phosphorylation events.

1.3.3 Phosphorelay system in yeast

In the yeast osmoregulation system (Mager and Varela, 1993), the two-component signaling pathway consists of a sensor kinase SLN1, a phosphotransfer protein YPD1 and a response regulator SSK1 (Posas *et al.*, 1996). The domain organization of these proteins is shown Figure 1-6. The sensor kinase SLN1 (Ota and Varshavsky, 1993) consists of an extra-cellular domain delineated by two predicted transmembrane α-helices at its N-terminus, a histidine kinase (HK) and a response regulator domain (R1) at its C-terminus. SLN1 exists as a dimer in the membrane. Although the extracellular domain of SLN1 is found functionally necessary and is implicated in SLN1 dimerization, the actual signal SLN1 senses remains unclear (Tao *et al.*, 1999). It is hypothesized that SLN1 is sensitive to structural changes within the membrane induced by osmotic stress, instead of one or more particular ligands. The SLN1 protein autophosphorylates at a histidine residue using ATP within the histidine kinase domain and then the phosphoryl group is transferred to an aspartate residue in the response regulator domain (R1). SSK1 has a response regulator domain (R2) at its C-terminus and an effector domain at its N-terminus whose function is not cur-

rently known. YPD1 is a phosphorelay protein which transfers the phosphoryl group from the aspartate residue of R1 to an aspartate residue of R2 via a histidine-phosphorylated YPD1 intermediate (Figure 1-7).

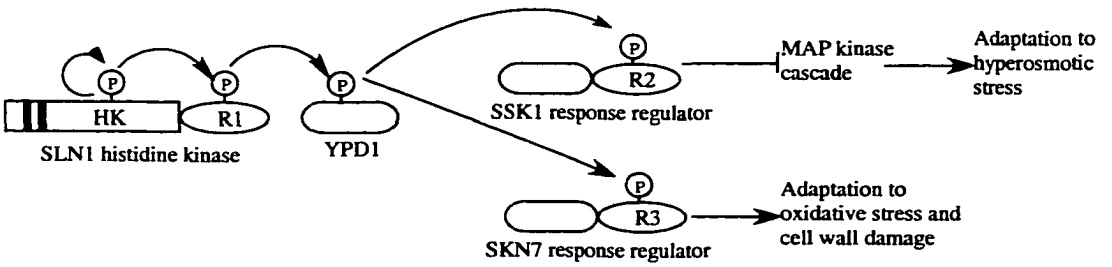


Figure 1-7. Schematic diagram of His-Asp-His-Asp phosphorelay signaling pathways in *S. cerevisiae*. Arrows indicate His-Asp phosphoryl transfer reactions that occur within pathways. The dark bars in SLN1 denote transmembrane segments.

Under normal conditions, SLN1-R1, YPD1 and SSK1-R2 remain constitutively phosphorylated at the expense of ATP. Under hyperosmotic shock, the SLN1 histidine kinase function is repressed through an unknown mechanism, which eventually leads to shutdown of the phosphorelay. The unphosphorylated form of SSK1 can interact with and activate (through its unphosphorylated R2 domain) the downstream MAP kinase cascade which eventually will result in increased glycerol production (Posas *et al.*, 1996; Posas and Saito, 1998).

The MAP kinase cascade consists of three proteins: a MAP kinase (MAPK) HOG1, a MAP kinase kinase (MAPKK) PBS2 which is proposed to be a scaffold protein which holds HOG1 and SSK2 together, and two redundant MAP kinase kinase kinases (MAPKKK) SSK2 and SSK22 (reviewed in Posas *et al.*, 1998). The unphosphorylated SSK1 interacts with SSK2 and activates the kinase activity of SSK2; phospho-SSK2 in turn will activate PBS2, which will activate the MAPK HOG1 protein (Posas and Saito, 1998). The HOG1 protein, upon phosphorylation, translocates into the nucleus where it phosphorylates corresponding transcription factors regulating the gene expression of GPD1 (glycerol-3-phosphate dehydrogenase) and other genes (Blomberg and Alder, 1989; Brewster *et al.*, 1993; Albertyn *et al.*, 1994a; Albertyn *et al.*, 1994b).

In yeast, there exists another independent transmembrane sensor protein SHO1 for hyperosmolarity sensing. The SHO1 protein, which consists of four predicted transmem-

brane regions and a C-terminal cytoplasmic domain containing a Src-homology 3 (SH3) domain, is structurally unrelated to SLN1. At high osmolarity, SHO1 can activate the MAPKK PBS2 independent of the SLN1-YPD1-SSK1 phosphorelay (Posas *et al.*, 1996).

Additionally, the SLN1-YPD1 pathway also donates a phosphoryl group to SKN7 (Brown *et al.*, 1993; Brown *et al.*, 1994; Yu *et al.*, 1995; Krems *et al.*, 1996; Fassler *et al.*, 1997; Morgan *et al.*, 1997; Ketela *et al.*, 1998; Li *et al.*, 1998), the only other known response regulator (R3) in yeast besides SSK1 and SLN1. The SKN7 protein is shown to be multi-functional and is responsible for responses to cell wall damage and oxidative stress although its detailed functional role in these processes is not clearly known (Singh, 2000). It is interesting to note that SKN7, mainly functioning as an transcription factor, is localized predominately in the nucleus. YPD1, however, is presumably located in the cytoplasm because it interacts with the histidine kinase SLN1, which is in turn anchored to the cell membrane. It is currently unclear how the SLN1-YPD1-SSK1 pathway and SLN1-YPD1-SKN7 pathway are integrated and regulated.

Although phosphatases have been discovered in other phosphorelay systems such as in Kin-Spo0 and ArcB-ArcA, no phosphatases have been identified in the yeast phosphorelay system.

The framework of the yeast osmoregulatory signaling system was mainly established by Haruo Saito's group at Harvard University (Posas *et al.*, 1996; Posas and Saito, 1998; Wurgler-Murphy and Saito, 1997; Maeda *et al.*, 1994; Maeda *et al.*, 1995). The yeast phosphorelay provides a useful model system for investigating, at a biochemical and structural level, the role of phosphorylation and dephosphorylation in regulating protein function. The novelty of this system is that it combines the bacterial two-component paradigm of histidine to aspartate phosphoryl transfer (the SLN1-YPD1-SSK1/SKN7 phosphorelay) with serine/threonine/tyrosine phosphorylation of the proteins of the MAP kinase cascade which is more commonly observed in eukaryotic organisms. The downstream MAP kinase cascade is in an "OFF" state when SSK1 is phosphorylated. This is in contrast with most two-component signaling systems in which phosphorylated response regulators will activate downstream pathways.

The rapid development of genomic sequencing allows a more rapid identification of two-component systems. Homology searches yielded no other response regulators in yeast besides the three protein discussed above (Cherry *et al.*, 1997; Cherry *et al.*, 1998). It is not surprising that a similar two-component pathway was identified in *Schizosaccharomyces pombe*: Mprl (YPD1) / MCS4 (SSK1). The response regulator domain of MCS4 is highly homologous to the SSK1 response regulator domain (>60% sequence similarity, Shieh *et al.*, 1997; Shiozaki *et al.*, 1997; Cottarel, 1997). However, the N-terminal sequences of these proteins are quite different. Mprl is a protein highly homologous to YPD1 with comparable sequence similarity as MCS4 *vs* SSK1 response regulator domains (Aoyama *et al.*, 2000; Nguyen *et al.*, 2000). Interestingly, Mprl has an N-terminal domain which is absent in YPD1. Considering the highly conserved sequences and architecture of the Mprl-MCS4 pathway and the budding yeast YPD1-SSK1 pathway, it is intriguing that Mprl and MCS4 have evolved for transmitting oxidative stress rather than osmostress signals.

1.4 YPD1, a yeast HPt protein

1.4.1 HPt domains

The multi-step phosphorelay systems often utilize a domain of 100-300 residues, termed an HPt (Histidine-containing phosphotransfer) or H2 domain. Unlike histidine kinases, the HPt domains generally do not exhibit any catalytic function. The HPt domains appear to serve solely as a histidine-phosphorylated intermediate molecule (or substrate) in His-Asp phosphorelay systems by acquiring/transferring a phosphoryl group from/to another regulatory domain. Considering the H1-D1-H2-D2 phosphorelay in which the phosphoryl group is transferred linearly from H1 to D2, it is obvious that H1 (histidine kinase) and H2 (HPt) need to interact with the same regulatory domain D1. It is not surprising, therefore, that there might be similar interactions with which HPt and histidine kinases recognize response regulator domains. For one HPt domain H2, it must recognize and interact with both response regulator domains D1 and D2 to fulfill its role as a phosphotransfer protein.

The HPt proteins vary in size and domain architecture. HPt domains have been found as independent monomeric proteins, and in at least one case as a dimer. They are also found fused to upstream hybrid histidine kinases forming D1-H2 or H1-D1-H2 polypeptides. For example, YPD1 is a cytoplasmic, single domain protein of 167 amino acid residues. The ArcB HPt domain ($ArcB^C$), of about 120 residues in length, is located at the C-terminus of the sensor kinase ArcB. In contrast, Spo0B from *Bacillus subtilis* functions as a dimer with each monomer consisting of 195 residues.

The HPt domains are difficult to recognize solely by sequence similarity. The sequence homology is very low between HPt domains or between HPt domains and histidine kinases. However, structural comparisons reveal strong conservation amongst HPt proteins. Several structures of HPt domains have revealed a similar 4-helix bundle architecture, in which the phosphorylatable histidine residue is located in the middle of an α helix with its side chain almost completely exposed to solvent (Zhou *et al.*, 1995; Kato *et al.*, 1997; Varughese *et al.*, 1998; Xu and West, 1999).

1.4.2 Properties of YPD1

YPD1 is a member of the newly defined family of HPt domains. YPD1 is a versatile molecule that can interact with more than one response regulator. *In vivo*, YPD1 is able to interact with SLN1-R1, SSK1-R2 and SKN7-R3. CheY, a bacterial response-regulator protein, can also serve as a phosphoryl donor to YPD1 *in vitro* (Janiak-Spens *et al.*, 1999). More surprisingly, it has also been shown that YPD1 can substitute for other related HPt proteins in different species. For example, RdeA can complement deficiency in the slime mold *Dictyostelium* (Chang *et al.*, 1998) as well as ATHP1, ATHP2 and ATHP3 deficiencies in the plant *Arabidopsis thaliana* (Miyata *et al.*, 1998; Suzuki *et al.*, 1998). Furthermore, $ArcB^C$, the C-terminal domain of a sensor kinase involved in bacterial anaerobic response was shown to be able to partially substitute for YPD1 in the *in vitro* phosphorelay

assay (Janiak-Spens and Xu, unpublished). Table 1-1 lists some of the known HPt domains

Table 1-1. HPt domains that are YPD1 functional homologs

Proteins	Species	Function	Comments	References
RdeA	*Dictyostelium discoideum*	Modulating rate of development	The mutant is complemented by transformation with the Ypd1 gene.	Chang *et al.* (1998)
ArcB^c	*Escherichia coli*	Response to anaerobic conditions	Can replace YPD1 in the yeast *in vitro* phosphorelay, however less efficiently	Janiak-Spens, *et al.* (1999) Janiak-Spens and Xu, unpublished
AHP1	*Arabidopsis thaliana*	Might be involved in ethylene- and cytokinin-dependent signal transduction pathways	AHP1 exhibits *in vivo* ability to complement a mutational lesion of the yeast YPD1 gene.	Miyata *et al.* (1998) Suzuki *et al.* (1998) Urao *et al.* (2000)
Spy1 (or Mpr1)	*Schizosaccharomyces pombe*	Transmit oxidative stress signals	The spy1(+) gene showed an ability to complement a mutational lesion of the *Saccharomyces cerevisiae* YPD1 gene	Aoyama *et al.* (2000) Nguyen *et al.* (2000)
CaYPD1	*Candida albican*	Currently unknown	CaYPD1 but not CaYPD1-H69Q complements the lack of YPD1 in *S. cerevisiae*.	Celera *et al.* (2000)

that can replace YPD1 functionally, or *vice versa*.

The phosphorylation site on YPD1 was determined to be on His64 as demonstrated by the fact that a YPD1 His64 mutant cannot be phosphorylated and disrupts the phosphotransfer from SLN1 to SSK1 in both *in vivo* and *in vitro* assays (Posas *et al.*, 1996; Janiak-Spens *et al.*, 1999). The calculated molecular weight for YPD1 is 19.8kD. The secondary structure prediction of YPD1 based on the PHD program (Rost, 1996) suggested that YPD1 would be all helical. This was later confirmed by the crystal structure of YPD1 which will be described in detail in the following chapters.

1.5 Summary

Two-component signaling pathways are ubiquitous in bacterial organisms where they regulate adaptive response to a wide variety of extracellular stimuli. Although less prevalent, two-component signaling strategies are also utilized in the eukaryotic and archaebacterial kingdoms. At the time our studies of YPD1 was initiated, it was not known

to what extent the eukaryotic two-component proteins resemble their bacterial homologs with respect to biochemical function and three-dimensional structure.

The determination of macromolecular structures is a crucial step towards understanding structure and function. Elucidation of the structures of the interacting proteins in the phosphorelay system will lead to a better understanding of the mechanism of molecular recognition and phosphotransfer in these systems.

In this dissertation, the focus of study was on the structure and function of YPD1 in order to obtain a detailed picture of the environment surrounding the site of phosphorylation (His64), to identify molecular surfaces that are used for recognition and interaction with response-regulator domains, and to combine this knowledge with information from biochemical studies in order to gain insights regarding the phosphorelay mechanism. Using X-ray crystallography as a tool, the crystal structure of YPD1 from *Saccharomyces cerevisiae* was determined. Structures of several YPD1 mutants G68Q, Q86A, Q86L and Q86E were subsequently analyzed in order to gain a better understanding of the roles of these conserved residues in the function of YPD1. The crystal structure of YPD1 was the first structure determined of a eukaryotic homologue of a two-component phosphorelay protein which revealed evolutionary conservation of the tertiary fold, consisting of a four-helical bundle core. Structural interpretations of the YPD1 and other HPt domain structures have led to hypotheses concerning the roles of the important conserved residues. These hypotheses can then be tested by biochemical methods such as site-directed mutagenesis. Structural characterizations of these mutants will help to decipher experimental results from biochemical studies. Structural analysis and amino acid sequence comparisons of YPD1 to other HPt domains have helped to define common characteristics of the HPt family.

2 Structure Determination of YPD1

2.1 Introduction

As discussed in Chapter 1, YPD1 functions as a HPt protein in a multistep phosphorelay system involved in yeast responses to hyperosmolarity and oxidative stress. At a molecular level, YPD1 transfers phosphoryl groups from the sensor kinase SLN1 to the response regulator proteins SSK1 or SKN7 (Ketela *et al.*, 1998; Li *et al.*, 1998; Posas *et al.*, 1996). As a result, YPD1 can interact with three homologous regulatory domains R1, R2 and R3 associated with SLN1, SSK1 and SKN7, respectively. YPD1 is an essential phosphoprotein intermediate during phosphoryl transfer from SLN1 to SSK1 since the deletion of the *ypd1* gene is lethal (Posas *et al.*, 1996). The site of phosphorylation on YPD1 is His64 (Posas *et al.*, 1996; Janiak-Spens *et al.*, 1999).

YPD1 functions similarly to other members of the HPt family despite less than significant sequence homology. Structures of several HPt domains, such as the P1 domain of CheA, the HPt domain of ArcB, as well as Spo0B, have been determined by X-ray crystallography or NMR (Zhou *et al.*, 1995; Kato *et al.*, 1997; Varughese *et al.*, 1998). Structural comparisons of these bacterial HPt domains revealed a common four-helix bundle motif. It was not known, however, to what extent the emerging class of eukaryotic two-component regulatory proteins resembled their bacterial counterparts.

This chapter describes the X-ray structure of YPD1, which was initially determined at a resolution of 2.7 Å (Xu and West, 1999). High resolution data were subsequently obtained at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL) and a 1.75 Å structure of YPD1 was then obtained. Section 2.2 summarizes the materials and methods used in this study, the description and analysis of the YPD1 structure are presented in Section 2.3.

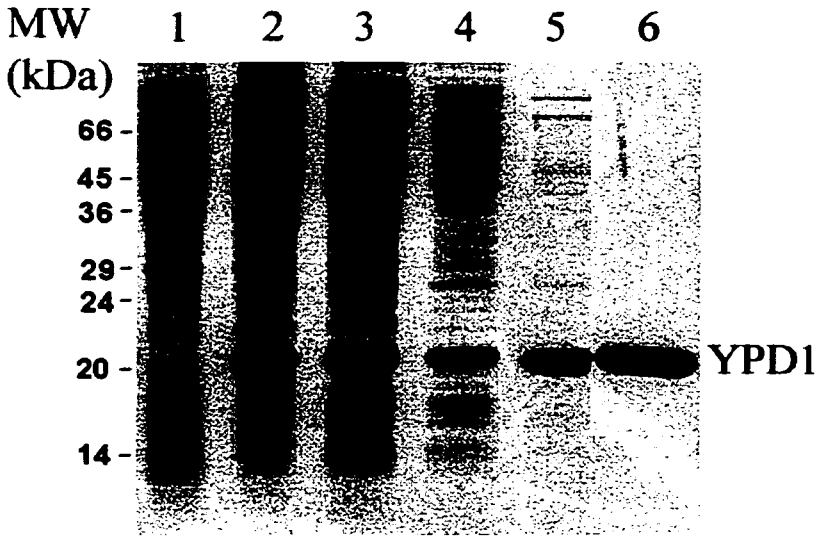


Figure 2-1. Purification of YPD1. Samples were removed at various stages of purification and analyzed by 15% SDS-PAGE. Lane 1 and 2, whole cell lysate of DH5α cells with and without the YPD1 expression plasmid pVN1; lane 3, sonication extract; lane 4, sample after ammonium sulfate step; lane 5, combined sample after Q Sepharose anion exchange chromatography; lane 6, pooled sample after Sephadex G50 column. Low molecular weight standard markers are indicated on the left.

2.2 Materials and Methods

2.2.1 Expression and purification

The yeast YPD1 protein was overexpressed and purified from *Escherichia coli*. A DNA fragment containing the 501 base-pair gene encoding YPD1 was amplified by PCR using yeast genomic DNA as the template. The resulting PCR product and the parent vector pME43 (Simms *et al.*, 1987) was digested with NdeI and PstI. The resulting 2.7 kb pUC-vector from pME43 was gel purified and then ligated to the YPD1 gene fragment overnight at 288 K using T4 DNA ligase (New England Biolabs). DH5α cells were then transformed and the desired recombinant plasmid pVN1 was obtained (Xu *et al.*, 1999).

YPD1 can be purified using the following steps, ammonium sulfate precipitation, anion-exchange column chromatography followed by gel filtration chromatography, all carried out at 277 K (Figure 2-1). A 1 L culture of pVN1-containing DH5α cells was grown overnight at 310 K in Luria broth supplemented with 50 μg/ml ampicillin. Cells were harvested by centrifugation and resuspended in 0.1 M sodium phosphate pH 7.0, 1 mM EDTA

and 1.4 mM β-mercaptoethanol (βME). Cells were lysed by sonication. Saturated ammonium sulfate was added slowly to the sonicated supernatant to a final concentration of 55% saturated ammonium sulfate. The protein precipitate was then collected by centrifugation and resuspended in approximately 5 ml of 50 mM Bis-Tris pH 6.5, 1.4 mM βME. The sample was dialyzed against the same buffer and was then loaded onto a 5 ml HiTrap Q Sepharose Fast Flow (Pharmacia) anion-exchange column pre-equilibrated in 50 mM Bis-Tris, pH 6.5 buffer. Protein was eluted from the column using a linear salt gradient of 0-500 mM NaCl in the starting buffer. YPDl-containing fractions, identified by SDS-PAGE, were combined and then concentrated to approximately 8 ml using an Amicon ultrafiltration device with a YM10 membrane. The sample was loaded onto a Sephadex G50 (Sigma) gel-filtration column equilibrated in 50 mM sodium phosphate pH 7.0, 1 mM EDTA, 1.4 mM βME. Fractions containing pure YPD1 protein were pooled and concentrated. Typical yields for purified YPD1 were ~40 mg per 1L culture.

2.2.2 Crystallization

The purified YPD1 protein sample was dialyzed against 50 mM sodium acetate, pH 6.5, 1 mM EDTA, 1.4 mM βME, and then concentrated to ~12-15 mg/ml. The sample was filter-sterilized before crystallization trials. The Hampton Research Crystal Screen I reagents, designed using the sparse-matrix screening technique (Jancarik and Kim, 1991), were used to obtain the first crystals (only #10[1] resulted in crystals). The first crystals were grown at room temperature by the hanging-drop vapor-diffusion method (McPherson, 1998). Further optimization of the crystallization conditions yielded reproducible diffraction quality crystals (Figure 2-2) using polyethylene glycol 4000 (Fluka) as a precipitant. In the vapor diffusion setup, the reservoir was composed of 30-33% (w/v) PEG 4000, 0.1 mM sodium acetate pH 5.0 and 0.2 M ammonium acetate. The hanging drop consisted of 1.5 µl of 10-12 mg/ml protein solution and 1.5 µl of reservoir solution.

2.2.3 Preparation of heavy atom derivatives

For preparation of heavy atom derivatives, concentrated stock solutions of $Hg(CH_3COO)_2$, (Sigma) and K_2PtCl_4 (Aldrich) were prepared in artificial mother liquor.

1. **Salt:** 0.2M Ammonium acetate; **Buffer:** 0.1 Sodium acetate trihydrate, pH 4.6; **Precipitant:** 30% w/v polyethylene glycol 4000.

YPD1 crystals were soaked with mercuric acetate at an approximate final concentration of 4 mM for three hours or 1 mM overnight at room temperature. Crystals that had been pre-equilibrated in artificial mother liquor at a higher pH (0.1 M Tris-HCl pH 7.5) were soaked with K_2PtCl_4, at a final concentration of about 2 mM for two days at room temperature.

2.2.4 X-ray data collection and processing

The YPD1 crystals belong to space group $P4_32_12$ (unit cell dimensions, a=b= 52.71 Å, c=244.02 Å at room temperature and a=b=52.16 Å, c=241.50 Å at 103 K) with two independent molecules in the asymmetric unit. The crystals were grown in the presence of 32% (w /v) polyethylene glycol 4000 which served as a cryoprotectant allowing for direct freezing of crystals in liquid propane and storage in liquid nitrogen prior to data collection at 103 K. All data except the high resolution native data were collected at the University of Texas Southwestern Medical School and at Rutgers University/University of Medicine and Dentistry of New Jersey using an RAXIS IIc image plate system and mirror-focused CuKα radiation from a Rigaku RU-200 rotating anode X-ray generator operated at 50 kV and 100 mA. Diffraction data were indexed and integrated using DENZO (Otwinowski and Minor, 1997) and then processed using SCALA (Evans, 1993). The heavy-atom derivative and

Figure 2-2. Crystal of YPD1.

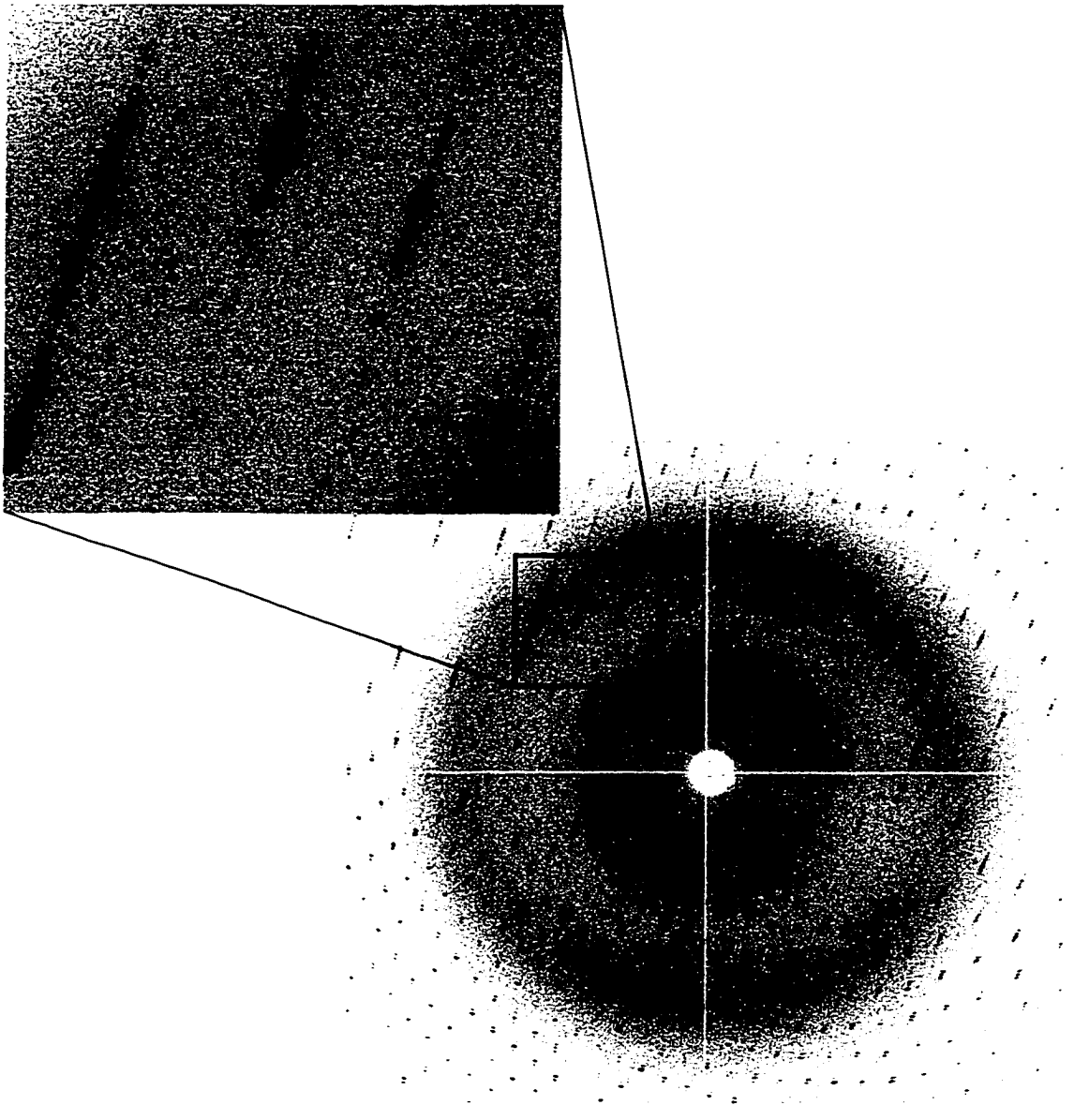


Figure 2-3. A typical diffraction pattern of YPD1 native crystal collected at NSLS station X12B at Brookhaven National Laboratory.

native datasets were scaled together using FHSCAL and SCALEIT in the CCP4 suite of programs (CCP4, 1994).

The 1.75 Å resolution X-ray diffraction data (Figure 2-3) were collected using 1.0Å wavelength at 100 K at the NSLS station X12B at Brookhaven National Laboratory. The data were recorded with a Quantum 4 CCD detector using oscillations of 0.5 degree from a single crystal. The raw data were indexed, integrated, merged and scaled using D*TREK

software (Pflugrath, 1999). The data collection and processing statistics for crystals of the native and heavy atom derivatives are summarized in Table 2-1.

Table 2-1. Data collection and phasing statistics

Data collection						
Crystal	Native I	HgAc I	K_2PtCl_4	Native II	HgAc II	Native III
Temperature (K)	298	298	298	103	103	103
Resolutions)	2.7	3.0	3.0	2.7	3.0	1.75
Mean I/σ(I)	9.7(11.2)[c]	9.1(6.3)	12.0(26.3)	6.9(6.9)	7.1(11.3)	27.1(9.9)
Observations	52,152	22,438	35,563	38,739	20,276	207,329
Unique reflections	9674	6660	6711	9451	5199	31681
Completeness(%)	91.6(82)[c]	90.6(90.1)	89.9(88.4)	96.3(90.5)	71.7(77.7)	92.1(87.5)
R_{merge} (%)[a]	6.6(7.8)[c]	6.3(9.4)	5.1(7.5)	7.2(8.6)	6.2(7.3)	4.9(13.7)
R_{deriv} (%)[b]		29.1	20.2		33.5	

Phasing				
Resolution used (Å)	3.0	3.0	3.0	
Number of sites per molecule	1	0.5	1	
R_{cullis} (centric)[c]	0.559	0.889	0.675	
Phasing power (centric/anom)[d]	2.61	0.766	1.57/2.44	
Figure of merit (combined)	0.596			

[a] $R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma|I|$, where I is the intensity for a give reflection, and $\langle I\rangle$ is the average intensity for multiple measurements of this reflection.

[b] $R_{deriv}=\Sigma|F_{PH}-F_P|/\Sigma|F_P|$, F_{PH} and F_P are structure factor amplitudes for derivative and native.

[c] $R_{cullis}=\Sigma||F_{PH}\pm F_P|-|F_H||/\Sigma|F_{PH}\pm F_P|$, where F_H is the calculated heavy atom structure factor amplitudes.

[d] Phasing power=root mean square of (F_H/E), where E is the residual lack of closure error.

[e] Values given in parenthesis correspond to the highest resolution shell.

2.2.5 MIRAS phasing and electron density map calculations

Two mercury binding sites (at the single cysteine residue, Cys82, in each monomer) were located by difference Patterson map analysis (Figure 2-4) and by the use of direct methods implemented in the program SHELXS (Sheldrick *et al.*, 1993). One platinum binding site was identified in a cross difference Fourier map calculated using phase infor-

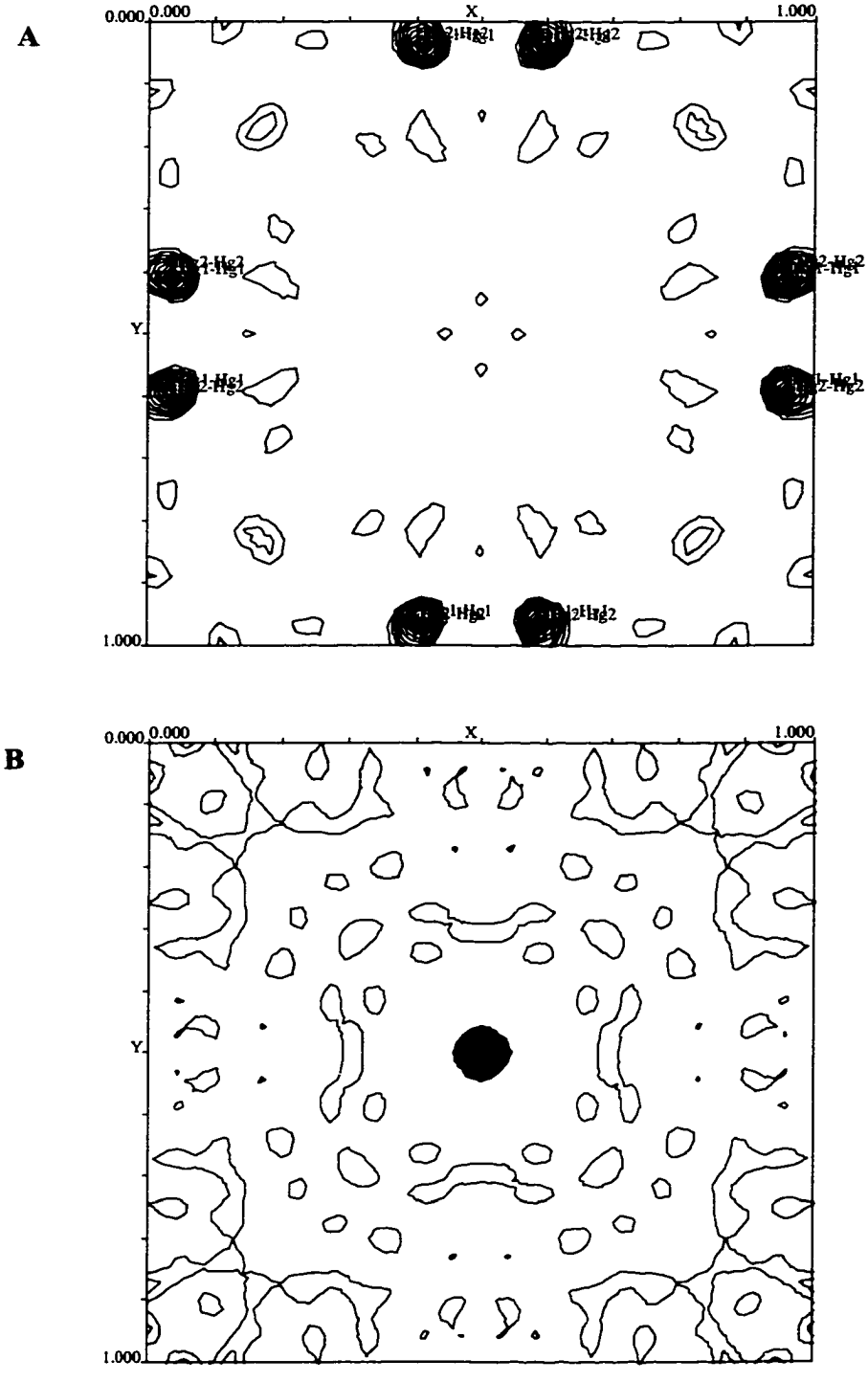


Figure 2-4. A. Difference Patterson map of YPD1 mercury derivative at section z=0.25; **B**. Native Patterson map of YPD1 (calculated with $|F_{obs}|^2$) at section z=-0.009.

mation obtained from the mercury derivative. Since the data was collected at different temperatures, the heavy-atom positions and phases to 3.0 Å were initially refined separately by the maximum likelihood method as implemented in SHARP (*de la* Fortelle and Bricogne, 1997) and then refined together based on the temperature at which the data sets were collected. The heavy atom phasing statistics are listed in Table 2-1. Although the $Hg(CH_3COO)_2$, dataset collected at 103 K was only ~50% complete with respect to Friedel pairs, the final phases obtained from joint refinement included some phase contribution based on anomalous diffraction data from this crystal. The initial electron density map calculated using MIRAS phases was improved using density modification methods in SOLOMON (Abrahams and Leslie, 1996) without molecular averaging. The modified electron density map showed improvement (Figure 2-5A) particularly in the definition of molecular boundaries and identification of three helical segments. The space group enantiomer was assigned by checking the handedness of the helices.

A peak with the height at (0.5, 0.5, -0.08) on a native Patterson map implies that there exists a pseudo-translation between the two molecules in the asymmetric unit (Figure 2-4B). An accurate transformation matrix was found by using NCS6D and IMP (Kleywegt and Jones, 1999), a six-dimensional brute force and refinement method. Two spheres (radius 15 Å) of electron density, each centered at one of the two mercury heavy atom sites, were then carved out from of MIRAS map. The transformation matrix was then derived and refined by superimposing the two spheres of electron densities.

The MIRAS electron density map was significantly improved (Figure 2-5B) by incorporating two-fold molecular averaging, along with solvent flattening and histogram matching implemented in the program DM (Cowtan, 1994).

2.2.6 Model building and refinement

A polyalanine atomic model was manually built into the DM improved electron density map using XtalView (McRee, 1993) with the help of ESSENS/SOLEX (Kleywegt and Jones, 1997) for polypeptide chain tracing. The starting model consisted of backbone atoms for 165 of the 167 residues, the exception being two residues at the N-terminus. Atoms for the side-chains were then added. Two loop regions (amino acid residues 19-24

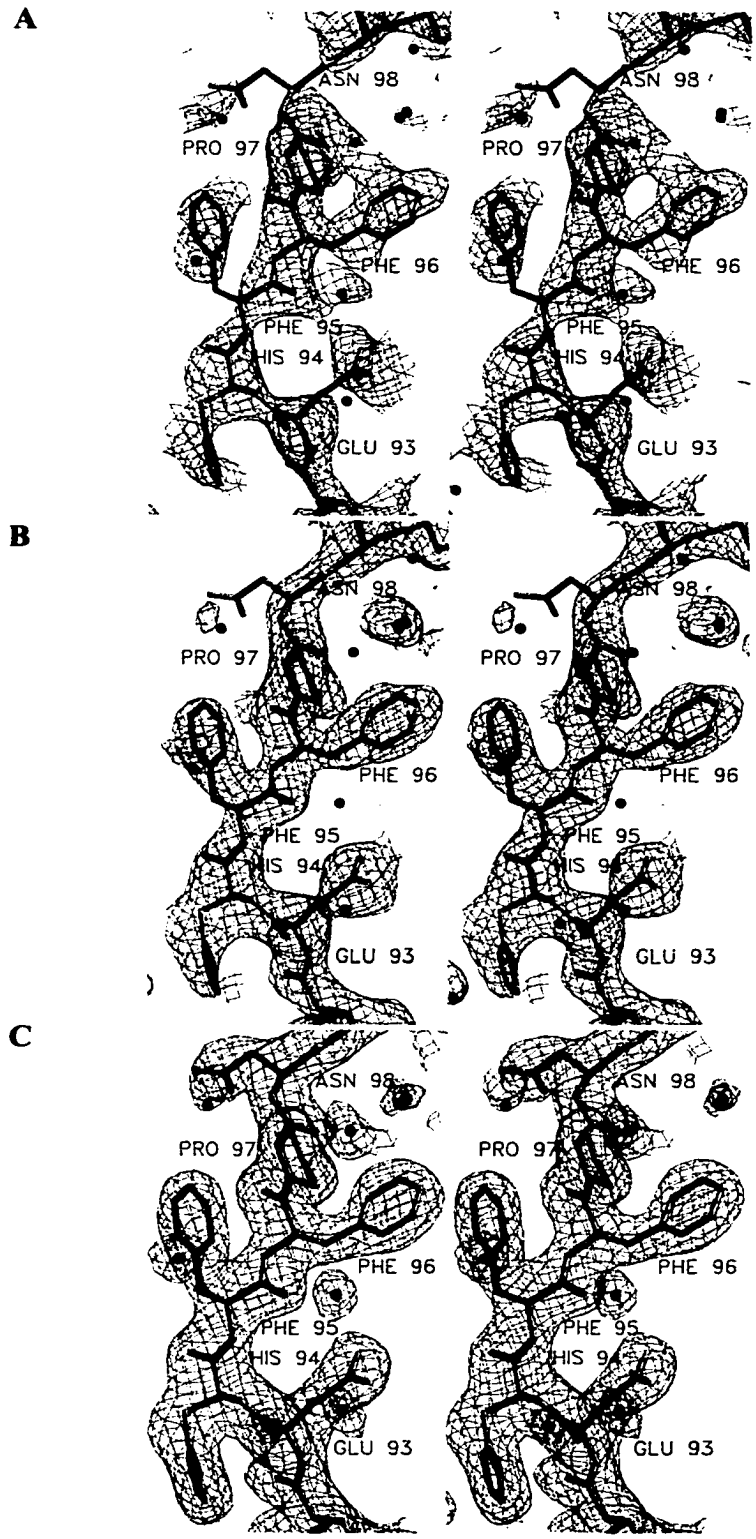


Figure 2-5. Density maps at various stages of structure determination overlayed with the final refined model at 2.7 Å resolution. **A.** original map with MIRAS phases; **B.** improved density map obtained by density modification; **C.** Final 2.7Å $2F_o$-F_c map.

and 125-134) showed weak electron density initially, so side-chain atoms for these residues were added later during refinement.

The overall position and orientation of the starting model was first refined with rigid body refinement. Due to the limitation of the resolution, the non-crystallographic symmetry (NCS) was strictly applied by assuming that the two monomers in the asymmetric unit (*asu*) are identical. The starting model was subjected to torsion angle molecular dynamics refinement in XPLOR (Brünger *et al.*, 1987; Rice and Brünger, 1994) starting at 5000 K using data from 8.0-2.7 Å (8% of the data were excluded from refinement and used to calculate R_{free}). The R_{factor} and R_{free} values for the initial model after rigid body refinement were 0.505 and 0.527, respectively. After one round of torsion angle simulated annealing, the R_{factor} dropped to 0.313 (R_{free} = 0.383). After this stage, all low-resolution data were used in subsequent refinement steps and a bulk solvent correction was applied. Following several rounds of model rebuilding, torsion angle refinement with NCS constraints, and group B-factor refinement, the R_{factor} was 0.237 (R_{free} = 0.290). Model rebuilding, based on a SIGMAA-weighted electron density map (Read, 1986), was performed using the program O (Jones *et al.*, 1991) and OOPS (Kleywegt and Jones, 1996). The model was further refined using REFMAC (Murshudov *et al.*, 1997) with NCS constraints replaced by tight NCS restraints (σ=0.05 Å) with the exception of the loop region (residues 125-134) for which medium NCS restraints (σ=0.5 Å) were applied. This loop region exhibited weak density and was refined to a high temperature factor and, as such, is not as well defined as the rest of the structure. A total of 104 water molecules were added gradually during the final stages of refinement. The two independent molecules in the asymmetric unit are nearly identical (r.m.s.d = 0.099 Å). The final model at 2.7 Å, which includes residues 3-167, was examined using the programs PROMOTIF (Hutchinson and Thornton, 1996), PROCHECK (Laskowski *et al.*, 1993), and WHAT IF (Vriend, 1990). Backbone dihedral angles for all the residues including the loop region (residues 125-134) fall into allowable regions of a Ramachandran plot with 93% of the residues in the most favored area as defined in PROCHECK (Table 2-2).

2.2.7 Refinement of the 1.75 Å model

The 2.7 Å structure of YPD1 was then used as the starting model for further high resolution refinement. The starting molecules were edited to omit all the water molecules and the temperature factors were set to 17.0 $Å^2$ according to Wilson plot of experimental observations ($|F_{obs}|^2$). All the data were used in the refinement except 5% of the total reflections, which were used as the R_{free} set to monitor the refinement process. The models were first subjected to rigid body refinement using X-PLOR in increasing steps of resolution until all data were incorporated. After he rigid body refinement, the R_{factor}/R_{free} dropped from 59.3/56.5 to 40.6/41.5. Positional parameters and isotropic B-factors were initially refined using simulated annealing implemented in program CNS, with inclusion of the bulk solvent correction. The model was rebuilt based on a SIGMAA weighted ($2mF_o$-DF_c) map. The XtalView suite of programs was used for all model rebuilding and map calculations. At the later stages of refinement, ARP/wARP and REFMAC were used to refine and locate water molecules. The final model was checked in the composite omit map, rebuilt using XtalView and refined using CNS (Brünger *et al.*, 1998). The final refinement statistics are summarized in Table 2-2.

2.3 Results and Discussion

2.3.1 Structure determination of YPD1 and model quality

The X-ray structure of the yeast phosphorelay protein YPD1 was determined and reported at a resolution of 2.7 Å (Xu and West, 1999) and later improved to 1.75 Å. Initial phase information was derived from mercury and platinum isomorphous heavy atom derivatives and anomalous diffraction data collected from a mercury acetate-soaked crystal.

Each asymmetric unit contains two independent YPD1 monomers with each of the 2.7 Å YPD1 models containing 165 of 167 total amino acid residues. The crystallographic R-factor of the final 2.7 Å model was 0.201 (R_{free}=0.254). The 1.75 Å model was refined to a final R_{factor} is 19.5 (R_{free}=22.5). The estimated mean coordinate error was 0.19 Å on basis of the dependence of R_{cryst} values on resolution, and 0.23 Å if R_{free} was used in the analysis (Luzzati, 1952). Refinement statistics for both the 2.7 Å model and 1.75 Å model are summarized in Table 2-2.

Table 2-2. Model and refinement statistics

Crystals	Native I	Native III
Space group	$P4_32_12$	$P4_32_12$
Unit cell dimensions (Å)	a=52.71, c=244.02	a=51.9, c=240.19
Resolution (Å)	30-2.7	30-1.75
R_{cryst} (R_{free})[a]	20.1/25.4	19.5/22.5
Number of protein atoms	2666	2552
Number of solvents	104	333
Average B factor (Å^2)	27.68	17.0
R.m.s.d bond distance (Å)	0.007	0.0065
R.m.s.d bond angle (deg.)	1.664	1.28
Ramachandran plot		
Residues in most favorable region (%)	93.2	95.9
Residues in additional favorable region(%)	6.8	4.1

[a]$R_{cryst}=\Sigma|F_o-F_c|/\Sigma|F_o|$, where F_o *and* F_c are the observed and calculated structural factors respectively. R_{free} was calculated using 8% of randomly selected reflections, which were not used in the refinement of the coordinates. R_{cryst} was calculated using the rest of reflections.

Compared to the 2.7 Å model, the 1.75 Å YPD1 model does not include a flexible loop region (residue 126-131) which was originally reported. The higher resolution structure also contains more ordered water molecules. A total of 2552 protein atoms and 333 water molecules were included in the final 1.75 Å structural model. The 1.75 Å model shows improved stereochemistry over the 2.7 Å model. The r.m.s.d for bond length, bond angle, dihedral angle are 0.0065, 1.28 and 19.5 respectively. In the Ramachandran plot (Figure 2-6), 95.9% of the main-chain dihedral angles for non-glycine residues of the 1.75 Å model fell into the most favored region and the other 4.1% fell into the additionally allowed region as defined in the program PROCHECK. The electron density for the main-chain was continuous and clearly defined except for the N-terminal Thr2 (Met1 is not present in the purified protein used for crystallization as indicated by mass spectroscopy, data not shown) and three loop regions (22-24 and 126-131 of molecule A, 127-131 of molecule B). No atoms were built in for the last two flexible loops due to lack of defined density.

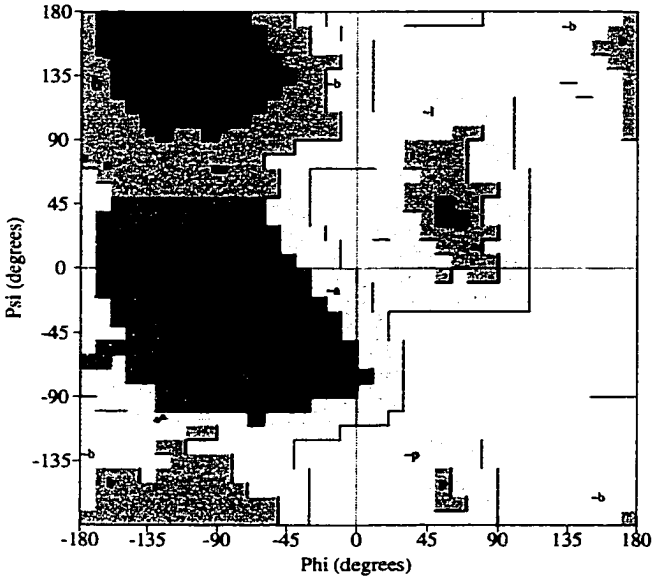


Figure 2-6. (Left) Ramachandran plot of the 1.75Å model.

Figure 2-7. (Right) Backbone differences between various YPD1 models. The r.m.s.d is proportional to the radius of the sausage at that position.

The structure of YPD1 was also independently determined to a similar resolution by Song *et al.* (1999). When comparing molecule A of the 1.75 Å model to other models of equivalent or higher resolution, the Cα r.m.s.d was generally between 0.35-0.42 Å, while the r.m.s.d for the entire model is around 0.9 Å. A large part of the four-helix bundle (30-90), including helix B, short linker between B and C, helix C and helix D, helix G, and helix E appear to be more invariant by having small r.m.s.d. Larger r.m.s.d were observed at helix A, helix F, and the N-terminus of helix B. Two areas showing the largest r.m.s.d are the C-terminus of YPD1 (162-167) and the flexible loop (124-133). Figure 2-7 shows the backbone difference between the following YPD1 models, the 2.7 Å YPD1 model (293K, PDB code 1QSP), 1.8 Å YPD1 structure by Song *et al.* (293 K, PDB code 1C02), and our current 1.75 Å YPD1 model (100 K). While these structural differences in part might be introduced by slight differences in the refinement procedures, they more likely reflect structural flexibility of difference parts of the YPD1 molecule. The four-helix bundle core is likely to be more rigid than the other portions of YPD1 in solution. It is worth pointing out that the crystallization conditions for obtaining crystals in the $P4_32_12$ space group obtained by Song *et*

al. are different from ours. There also exist some subtle differences in inter-molecular packing within the different crystals.

The first molecule (molecule-A) of the 1.75 Å structure will be referred to in the future discussion of the YPD1 structure unless noted otherwise.

2.3.2 Overall tertiary structure

YPD1 is classified as an all alpha protein with an up-down, left-handed twisted four-helix bundle fold core according to the SCOP database (Murzin *et al.*, 1995). Other proteins in the family include apolipoprotein (PDB code 1BZ4, 1OR4 and 1NFN *etc*), aspartate receptor ligand-binding domain (PDB code 2ASR and 1VLS *etc*), Cytochrome b562 (PDB code 256B *etc*), cytochrome c' (PDB code 1A7V *etc)*, hemerythrin (PDB code 2HMZ *etc*), TMV-like viral coat proteins (PDB code 1EI7 *etc*), acyl-CoA dehydrogenase (flavoprotein) C-terminal domain (PDB code 1BUC *etc*), FKBP12-rapamycin-binding domain of FKBP-rapamycin-associated protein (FRAP, PDB code 2FAP *etc*), proteasome activator reg (alpha, PDB code 1AVO), histidine-containing phosphotransfer domain $ArcB^c$ (PDB code 1A0B and 2A0B). Although many of these proteins do not share the same function as YPD1, it has been commonly observed that the four-helix bundle motif can support a wide variety of biological functions. The Cα trace and a ribbon representation of the YPD1 monomer is shown in Figure 2-8A. The overall tertiary fold of YPD1 is composed of six α-helices (labeled as A, B, C, D, E and G) and one 3_{10}-helix (labeled as F). These helices represent about 71% of the total 167 amino acid sequence. The last turn of the first helix (helix A) is a 3_{10} helix. The seven helices and two γ-turns of YPD1 are listed in Table 2-3. The inter-helical interactions (distance, angle, and interaction strength) of YPD1 are summarized in Table 2-4.

The molecule has an elongated shape with overall dimensions of approximately 60 Å x 40 Å x 30 Å. In the orientation shown in Figure 2-8, the two long helices B and G, together with the two short helices C and D, form an up-down-up-down four-helix bundle. The longest helix in the structure, helix G, is about 45 Å long and is slightly bent and as a result exhibits a relatively large deviation from ideal helix geometry.

The phosphorylatable residue (His64) is positioned in the middle of helix C and is completely exposed to solvent. Helices C and D, each about 25 Å long, are tightly associ-

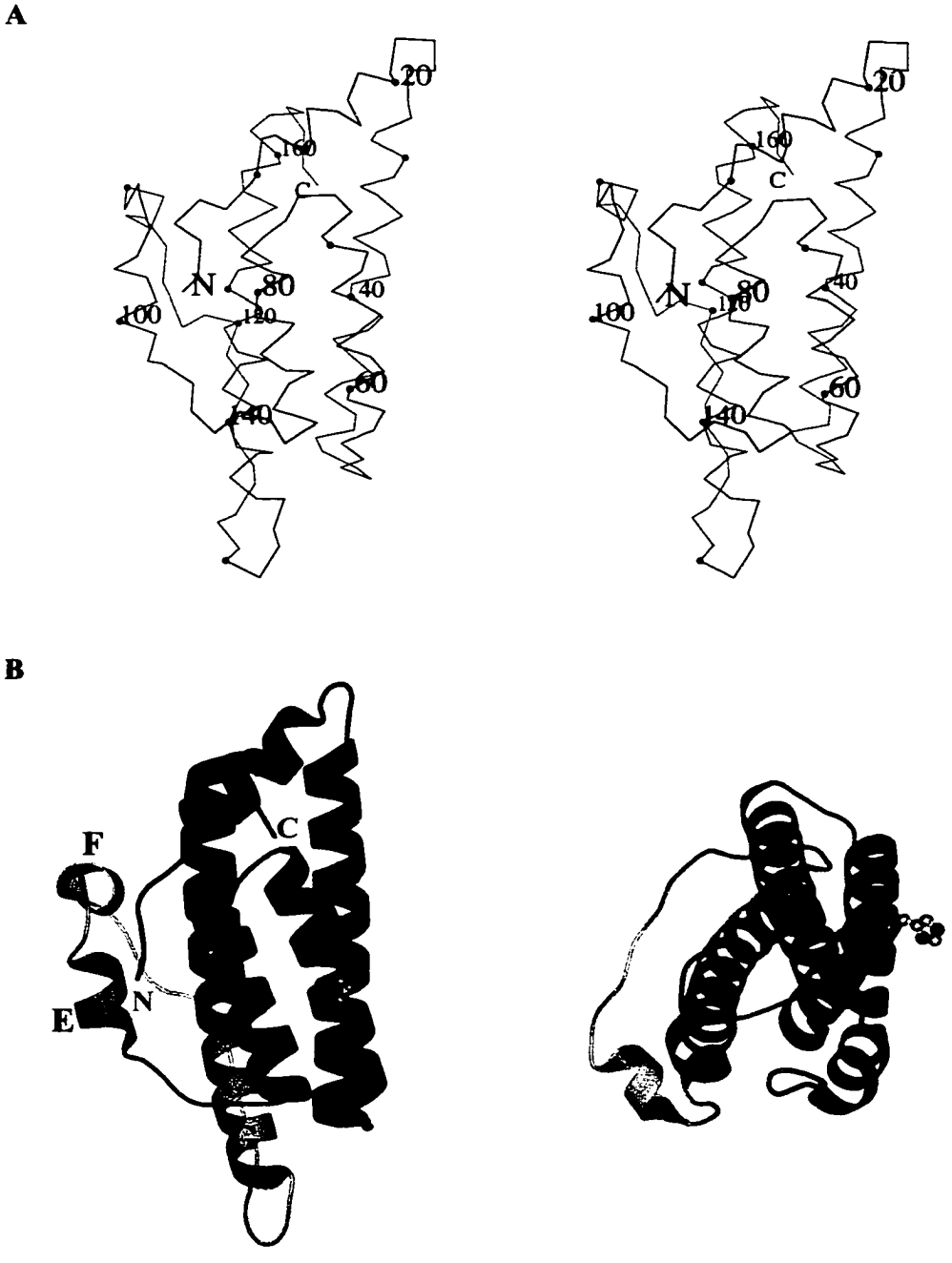


Figure 2-8. Structure of YPD1. **A.** Stereoview of a Cα trace of YPD1 with every twentieth residue is numbered. **B.** Ribbon representation of YPD1 in two orientations. The phosphorylation site, His64, is shown in ball and stick model. The helices are sequentially labelled from A to G (black). The N- and C-terminus are labelled in red.

Table 2-3. Secondary structure of YPD1

Identifier	Seq #	SS type	Length (Å)	Rise (Å)	Residue per turn	Deviation from ideal (°)
T1	9-11	γ-turn	5.6[¥]	-	-	-
A	10-19	α-helix	15.31	1.51	3.86	14.1
A'	20-22	3_{10}-helix	-	-	-	-
B	26-52	α-helix	40.41	1.48	3.68	15.6
C	55-73	α-helix	28.60	1.52	3.65	12.6
T2	74-76	γ-turn	5.2[¥]	-	-	-
D	75-90	α-helix	23.83	1.49	3.66	10.6
E	98-104	α-helix	10.34	1.51	3.72	11.7
F	108-113	3_{10}-helix	11.38	1.85	3.32	20.5
G	134-164	α-helix	44.55	1.44	3.83	27.4

[¥]This value corresponds to Cα(i) to Cα(i+2) distance.

Table 2-4. Inter-helical interactions within YPD1

Helix-1	Helix-2	Distance (Å)	angle (°)	No. of interacting residues		
				Total	Helix 1	Helix 2
A	B	7.6	130.6	4	2	3
A	C	7.0	63.7	5	3	3
A	D	7.3	114.3	1	1	1
A	G	14.1	-51.0	5	3	4
B	C	9.7	-165.3	23	11	10
B	D	13.2	28.9	3	2	3
B	G	10.7	-156.7	21	11	10
C	D	7.0	-157.6	20	10	9
C	G	11.7	25.6	6	2	4
D	E	12.5	147.4	3	2	2
D	G	8.3	-165.2	19	7	11
E	F	8.0	120.2	3	2	2
E	G	11.0	-33.2	6	4	3
F	G	10.5	87.7	4	2	3

ated. Besides closely knit hydrophobic interactions buried inside the four-helix bundle, there are four parallel hydrogen bond(s) or salt bridges linking helix C and D, Asp60 and Arg90, His64 and Gln86, and Lys67 and Glu83. Helices C and D are connected by a very tight γ-turn (74-76). The $C_\alpha(i)$ to $C_\alpha(i+2)$ distance is 5.2 Å. This turn is both structurally and functionally important as indicated by the observation that a G75A mutation rendered YPD1 unstable (Janiak-Spens and West, 2000). It is likely that this mutation disrupts the formation of a stable four-helix bundle core. The interhelical distance between Helix C and D is 7.3 Å, which is less than what is commonly observed in other four-helix bundle structures. Helices C and D presumably contribute to the binding surface for interaction of YPD1 with response regulator domains.

Helices D and G are connected by a 45 residue long loop that contains two short helices (E and F), a feature that has not been observed in other classic four-helix bundle structures. This long connecting loop is tethered to helix G predominantly through hydrophobic interactions. This loop is located away from the phosphorylation site. Therefore, it might not be directly involved in phosphotransfer but rather structurally critical. The importance of this loop to the structural integrity of YPD1 is confirmed by recent experiments (Nadine Keller, unpublished results): a deletion of the loop or a single/double mutation on the loop rendered YPD1 completely or partially insoluble.

Located on the same surface of the four-helix bundle as the phosphorylatable histidine, helix A appears to be both structurally and functionally important. Helix A is oriented at an angle of about 50 degrees to the four-helix bundle. It makes hydrophobic contacts with helices B and G, as well as the hydrophobic reverse turn connecting helices C and D. The distance between the His64 phosphorylation site and the middle of helix A is about 15 Å. Helix A is geometrically plausible to make contacts to response regulator domains. Thus, helix A can extend the potential binding surface on YPD1 for response regulator domains and may itself contribute to the binding interface. A detailed analysis of potential roles for helix A in response regulator binding and recognition is presented in Chapter 5.

2.3.3 A conserved HPt domain fold

2.3.3.1 Conservation of structure

All the known HPt domain-containing proteins vary significantly in sequence and domain arrangements. A large majority of prokaryotic HPt domains are part of multi-domain sensor kinases, whereas the eukaryotic HPt domains described thus far appear to have evolved to function as separate independent proteins. Structural information is currently available for several prokaryotic HPt domains, a low-resolution NMR structure of the histidine-containing P1 domain of the *Escherichia coli* histidine kinase CheA (Zhou *et al.*, 1995), both a 1.6 Å X-ray and NMR structure of the HPt domain ($ArcB^c$) of the anaerobic sensor kinase ArcB from *Escherichia coli* (Kato *et al.*, 1997; Kato *et al.*, 1999; Ikegami *et al.*, 2001), and an X-ray structure at 2.6 Å resolution of the dimeric Spo0B phosphorelay protein from *Bacillus subtilis* (Varughese *et al.*, 1998). Additionally, a structure of the dimerization domain of EnvZ was also determined by NMR (Tomomori *et al.*, 1999). As shown in Figure 2-9A, the YPD1, $ArcB^c$, CheA P1 domain, Spo0B, and EnvZ dimerization domain structures all have a four-helix bundle motif in common that contains the conserved histidine residue that is phosphorylated. Spo0B differs from the other HPt domains in that it functions as a dimer, however, it is evident from the structure that each monomer contains two long antiparallel α-helices that, in the dimer, come together to form the central four-helix bundle. Although the overall fold of the CheA P1 domain and YPD1 is similar, these structures are not directly superimposable. Superimposition of the helices that hold the phosphorylatable histidine residue showed that there are large differences in the position of helix D in YPD1 and the corresponding helix in CheA. However, the differences observed in the overall fold and active-site region of CheA, as compared to YPD1 and $ArcB^c$, could be due in part to the low resolution of the NMR structure. Spo0B and YPD1 are not directly superimposable either, no obvious similarities were found between the residue arrangements around the active-site histidine side-chain (His30) in Spo0B (Varughese *et al.*, 1998) and the phosphorylation site of either YPD1 or $ArcB^c$. Despite the lack of the sequence similarity (~12% with structural based sequence alignment), the structure of the yeast YPD1 phosphorelay protein is remarkably similar to the bacterial ArcB

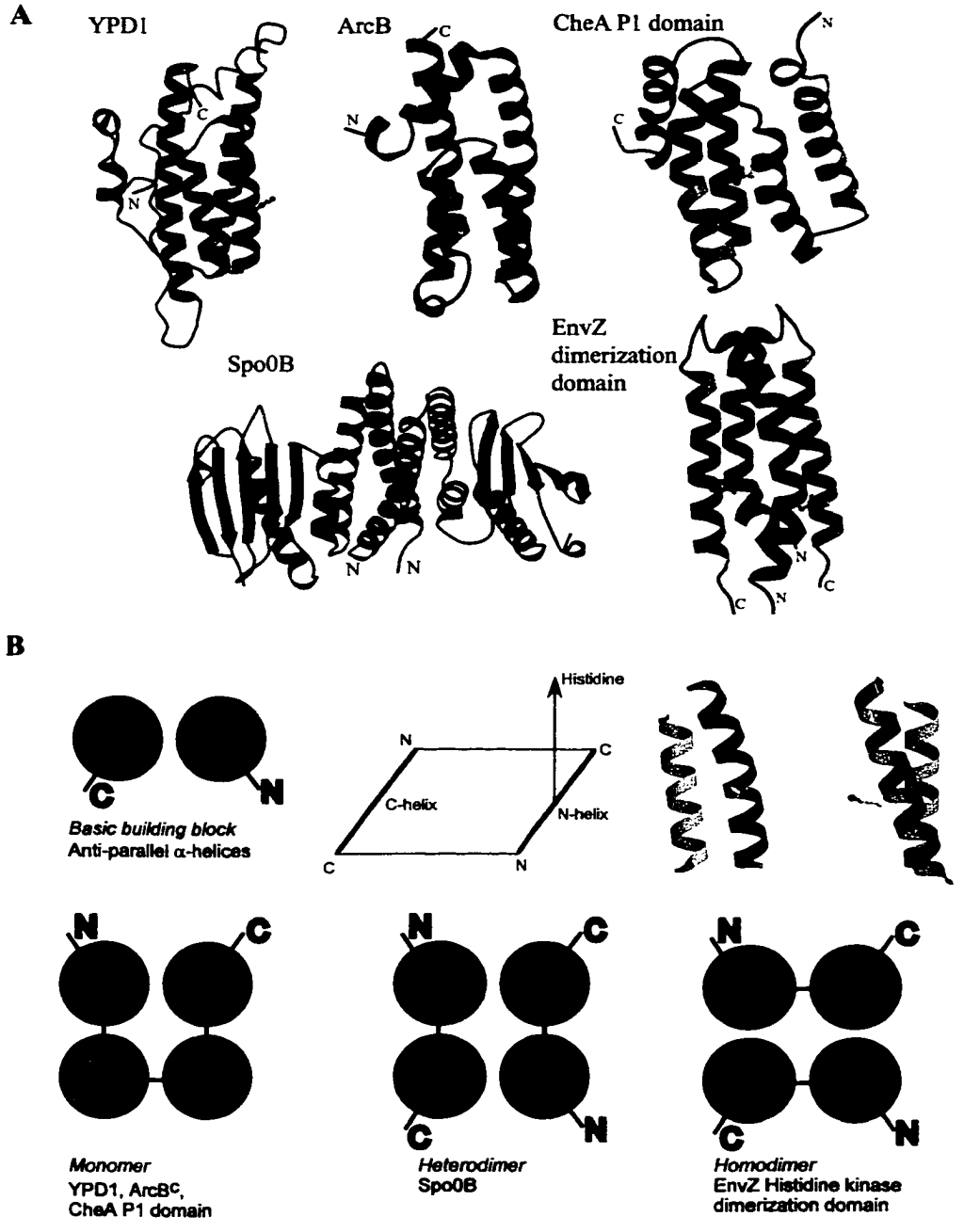


Figure 2-9. A. Known structures of HPt domains and histidine kinases domain with a phospho-histidine site. All these structures share a similar anti-parallel four-helix bundle, highlighted in pink and cyan here. The portions of structures that are not similar are colored in blue. The histidine residues are shown in ball and stick model. **B.** Topology diagrams of the anti-parallel four-helix bundle containing histidine phosphorylation sites in two-component signaling systems.

HPt domain, a detailed comparison between YPD1 and ArcB^c will be presented in the next section.

Topological analysis of the four-helix bundle formation of various known HPt domains and the EnvZ HK dimerization domain provides further insights about HPts and/ or possibly HKs. Since all these proteins/domains have to interact with structurally conserved response regulator domains by receiving or donating a phosphoryl group, it is reasonable to expect that these HPts and HKs might have a structurally conserved core as well. All the currently known HPt/HK structures support this notion except that the HPt family, in general, might be structurally more diversified than the response regulator superfamily.

The topology diagrams of the four-helix bundles within HPts and HK, are shown in Figure 2-9B. Aside from the linkers between the helices, it is obvious that these four-helix bundles are topologically equivalent. Each four-helix bundle can be formed by two pairs of anti-parallel α-helices in a head-to-head, shoulder-to-shoulder fashion. We define an anti-parallel α-helix unit (called an H-unit) as composed of two α-helices perpendicular to the phosphorylatable histidine residue (side-chain orientation). In an H-unit, the phosphorylatable histidine is always located in the middle of the first helix (called an N-helix, color cyan). The second α-helix (called a C-helix, color pink) is sequentially greater (not necessarily continuous from the first helix) than the first helix. The geometrical and spacial arrangement of an H-unit is illustrated in the top row of Figure 2-9B. In YPD1, the H-unit contains helices C and D. There is only one H-unit in the monomeric YPD1, ArcB^c and CheA P1 domains. An H-unit in Spo0B is formed by two helices each contributed by a different monomer, the four-helix bundle, as a result, can be thought of as a hetero-dimer of two H-units. For the EnvZ dimerization domain, since an H-unit is formed by two helices in the same monomer, the four-helix bundle is a homo-dimer of the two H-units. In the case of dimers, as we have seen in Spo0B and EnvZ HK, packing H-units in a head-to-head and back-to-back manner always produces four-helix bundles such that two phosphorylatable histidines are in *trans* positions.

The use of the H-unit allows us to unify all currently known HPt and HK structures under one framework. Besides the significance of it in categorizing these four-helix bundles, H-units are likely functionally important as well. As we will see in following sections,

the H-unit region contains the most significant sequence similarity (Figure 2-13) within the HPt family.

2.3.3.2 YPD1 and ArcB^c structural comparisons

The high resolution structures of YPD1 and ArcB^c allow for the comparison of these two structures in greater detail. Both of them share the same fold with respect to the four-helix bundle and helix A, with the exception that YPD1 has an extended loop region that joins the third (αD) and fourth (αG) helices of the bundle. The r.m.s.d for 111 aligned Cα atoms is 2.4 Å, as calculated by DALI (Holm and Sander, 1993). A stereoview superimposition of the YPD1 and ArcB HPt domain structures (Cα backbone) is shown in Figure 2-10A.

The molecular surface around His64 in YPD1, specifically helices A, B, C and D, is well conserved as compared to ArcB^c (Figure 2-10B). The portion of the YPD1 structure is mapped onto the CPK model (Figure 2-10C). It is likely that these helices contribute to the response regulator binding (Chapter 3 and 5). Helix A of YPD1 is parallel to the second helix of ArcB^c, however the corresponding helix in ArcB^c is displaced outward (~1.5 Å) due to a kinked conformation of the two helices in the back. The backbone conformation of helices C and D in YPD1 is essentially the same in the ArcB HPt domain. A superimposition of YPD1 (residues 56-86) and ArcB^c (residues 709-739) gives r.m.s.d values of 0.359 Å for main-chain atoms. In both the YPD1 and ArcB^c structures, these two helices are joined by a conserved three-residue reverse turn and are packed close together through hydrophobic interactions and interhelical hydrogen bonds. The local environment around the phosphorylation site in YPD1 and the ArcB HPt domain is well conserved. The phosphorylatable histidine residues in both structures are completely exposed to solvent and in the same side-chain conformation with the imidazole ring slanted slightly towards the C-terminal end of the helix.

The similarity of YPD1 and ArcB^c is also reflected in structural features such as the buried waters and charges. In the 1.75 Å YPD1 model, each monomer contains one pair of buried charge residues (Arg77 and Glu151) and one pair of completely buried waters (WAT500/501; WAT503/506) are involved in an extensive hydrogen bond network involving N-terminal of helix D (Arg77), N-terminus of helix A (Pro5, Ile8), helix E

(Leu102, Thr105), and helix G (Gln147, Glu151). The hydrogen bond interactions between the buried waters and relevant residues are shown in Figure 2-11. These buried waters are highly ordered. The B factors for the buried water molecules are around 10.0 $Å^2$, compared to the average B factor for buried atoms in YPD1 (12.7 $Å^2$). WAT501 forms hydrogen bonds to Ile8 O, Arg77 N, Ile78 N and WAT500; WAT500 is hydrogen bonded to

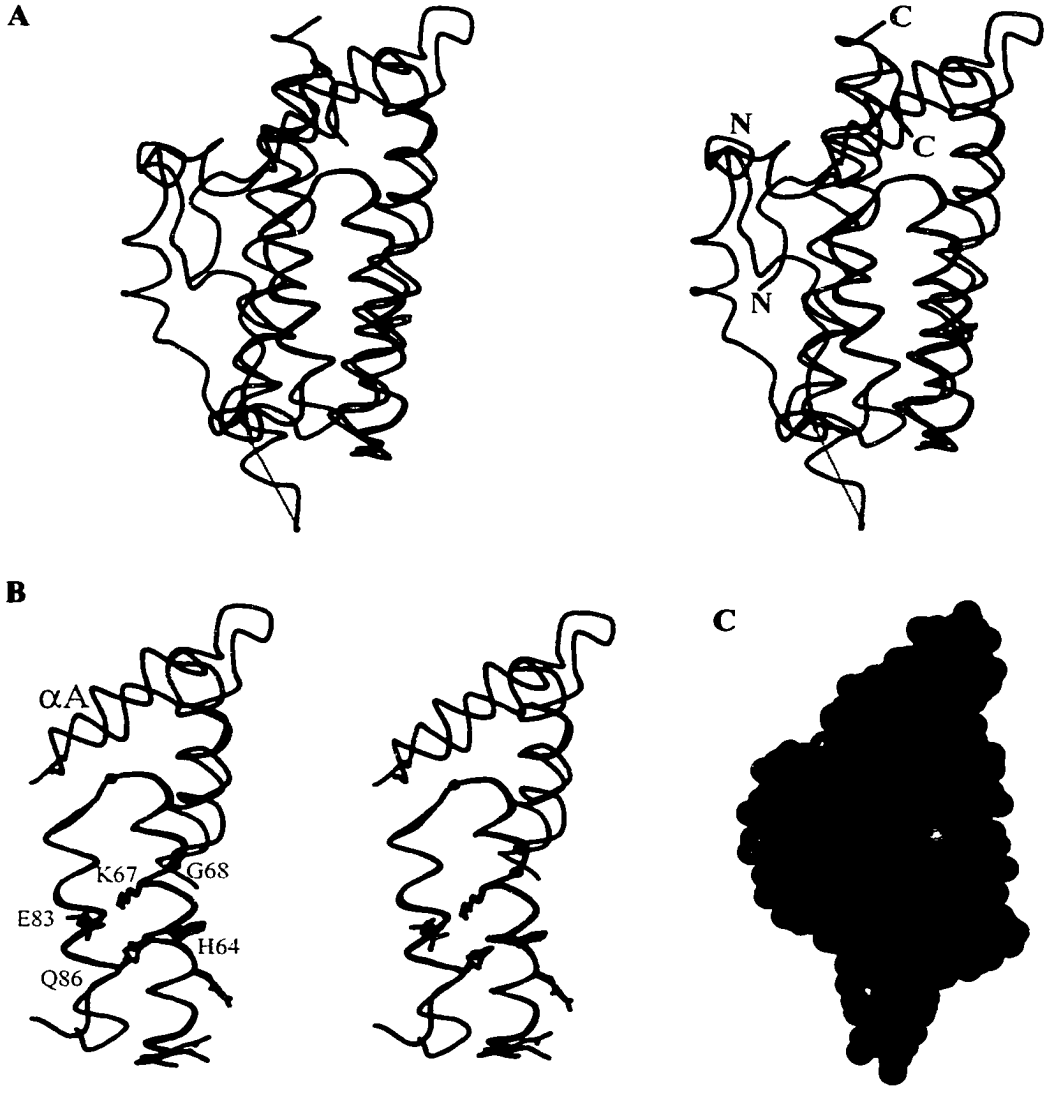


Figure 2-10. A. Overlay of Cα backbones of YPD1 (cyan) and the ArcB HPt domain (red). **B.** Structural comparison of the YPD1 and ArcB H-unit. Regions corresponding to helix A and the N-terminal portion of helix B of YPD1 is also shown above. YPD1 is shown in cyan and the conserved YPD1 residues are labelled. $ArcB^c$ is shown in red. **C.** CPK representation of YPD1 with the His64 residue highlighted in yellow and blue. The portion of YPD1 shown in B is colored in red.

WAT501, Glu151 Oε1, Glu151 O. The charged residues Arg77 and Glu151 are neutralized by each other.

Such an arrangement of buried waters and charged groups in YPD1 show very strong homology to a similar arrangement in $ArcB^c$. In fact, WAT501 is equivalent of WAT801 in $ArcB^c$. When $ArcB^c$ is superimposed onto YPD1, the distance between YPD1-WAT501 and $ArcB^c$-WAT801 is 0.5 Å. YPD1-WAT501 is hydrogen bonded to Ile8 O, Arg77 N, Ile78 N and WAT500 while in $ArcB^c$ WAT801 is hydrogen bonded to Leu662 O, His728 N, Leu729 N, Asp762 Oδ2. In YPD1, the salt bridge between Arg77 and Glu151 links helix D and helix G, while in $ArcB^c$, a similar interaction between equivalent helices via a zinc ion was observed at a similar location in the crystal structure (Figure 2-12). Thus, it is likely these conserved structural features have similar functional/structural roles. These clusters of buried charges or water molecules appear to act as molecular 'glue' and very likely contributes to the integrity of the overall structure.

2.3.4 Conservation of phosphorylation sites

The known members of the HPt protein family show much greater sequence variability compared to the sequence similarity within the response regulator domain family, thus it is difficult to align full length protein sequences. However, there are some significant sequence similarities in a segment of sequence about 35 residues around the phosphorylatable histidine. This region consists of the so-called H-unit defined above. In YPD1, the sequence consists of helix C and helix D. The structure of helix C and D in YPD1 is well defined. The anti-parallel helical hairpin is stabilized by an extensive network of buried hydrophobic interactions. Other stabilizing forces are contributed by, most noticeably, the four parallel inter-helical hydrogen bonds (see Figure 2-13B) as well as the "N-cap" of helix C. Manual sequence alignment of this segment of YPD1 with other proteins that possess HPt-like activity (Figure 2-13), using the common histidine as an anchor, indicates that conserved features of YPD1 (and $ArcB^c$) are also present in many other HPt domain family members for which structures are not yet available.

First, the repetitive pattern of hydrophobic residues that appear at locations i, $i + 3$, $i + 4$, and $i + 7$ (on helices C and D of YPD1) are well conserved. This provides a strong

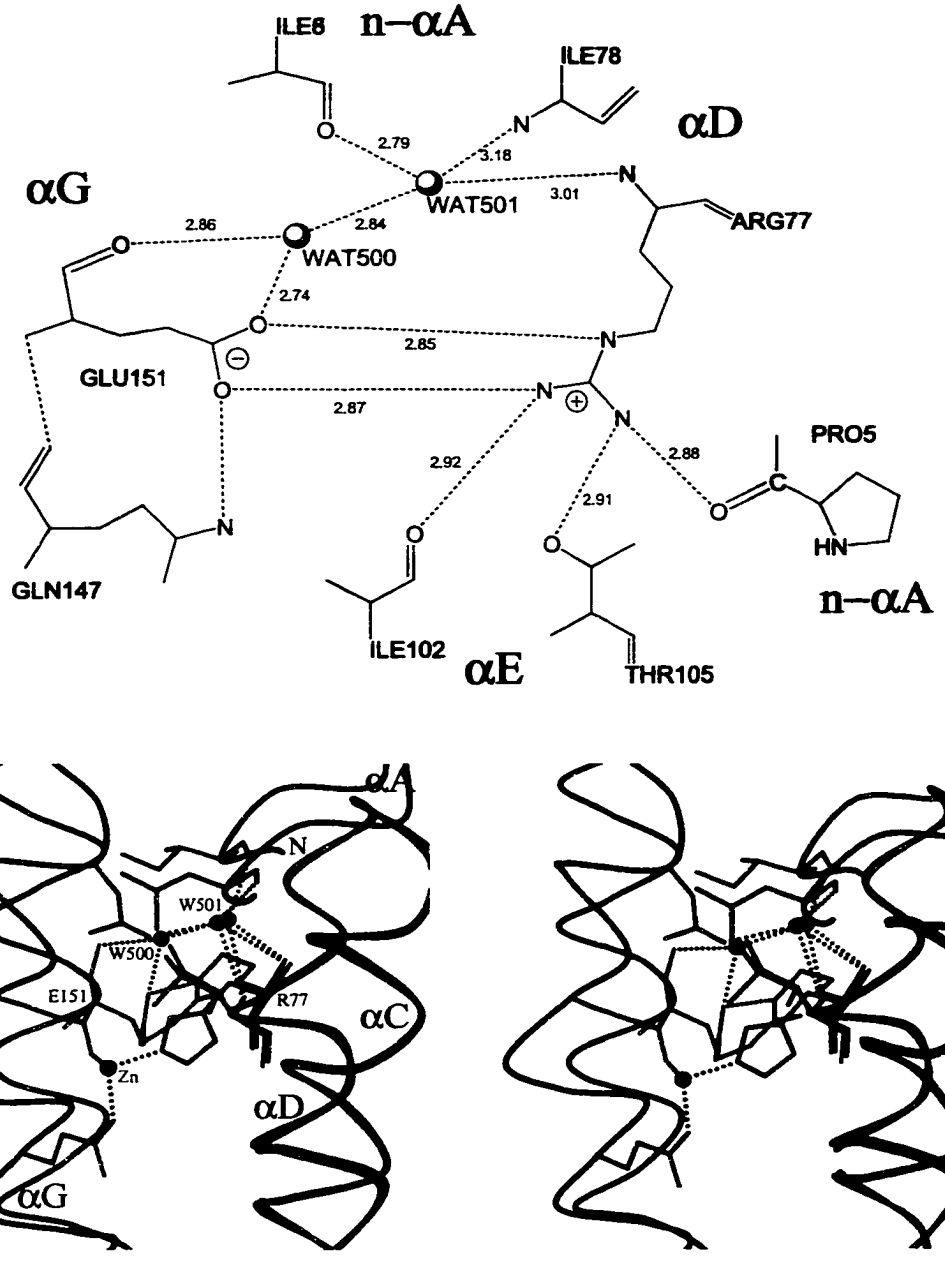


Figure 2-11. (Top) Interactions between buried charge residues, water molecules, as well as various YPD1 secondary structures. Dashed lines indicate hydrogen bonds. The numbers near the dashed lines are distances in Angstroms. Residues within a particular secondary structure are shaded together.

Figure 2-12. (Bottom) Comparison of buried waters and charges. YPD1 (cyan) structure is overlayed onto $ArcB^{c}$ (red). The buried waters are located in similar position in both structures with similar hydrogen bond patterns.

indication of a common helical conformation in this region within the HPt family. The hydrophobic reverse turn between helix C and D (positions +9, +10, +11 in Figure 2-13, relative to the phosphorylatable histidine residue) is also highly conserved. In the YPD1 and $ArcB^c$ structures, this turn is buried and covered by an additional helix (helix A in YPD1). In particular, a highly conserved glycine at position +10 is structurally important since a point mutation from glycine to alanine at this position rendered YPD1 insoluble (Janiak-Spens and West, 2000). A possible explanation for this result is that the mutation prevents the formation of a reverse turn since it is geometrically impossible (due to steric hindrance), and thus disrupts the YPD1 four-helix core.

Furthermore, residues at position -9 (Figure 2-13), located at the N-terminus of helix C in YPD1, are largely conserved as residues having side-chains that can serve as hydrogen bond acceptors. In YPD1, a helix "N-cap" is formed by hydrogen bonds between the side-chain carboxylate group of Asn55 and the main-chain amide groups of Thr57 and Glu58. In the ArcB HPt domain, a similar arrangement is found whereby the side-chain CO group of Asp708 forms a hydrogen bond to the main-chain Nη group of Gly711 which helps to stabilize the helix dipole. The conserved hydrophobic reverse turn between helices C and D and the "N-cap" stabilization of helix C in YPD1 is presumably important for the structural integrity of the antiparallel helical motif. Additionally, in order for the two helices to pack closely together, the size and arrangement of the interacting residues between the two helices are critically important.

The phosphorylatable histidine residue and surrounding residues that are part of helix C and D in YPD1 are illustrated in Figure 2-13B by well-defined electron density (Figure 2-14). The histidine side-chain adopts one of the common rotamer conformations for histidine with χ_1=178.5° and χ_2=72.8°. The most striking feature of the region around His64 is that the surrounding residues seem to be arranged in such a way as to maximize the accessibility of His64. Two strategies appear to be employed to achieve such a goal. First, residues with small or no side-chains are located on the outward facing surface of the C-terminal portion of helix C. The glycines at positions +4 (Gly68 of YPD1) and small residues at position +5, +6 and +7 also well conserved. Biochemical studies of the G68Q mutant confirmed the importance of Gly68 at this position in YPD1/response regulator

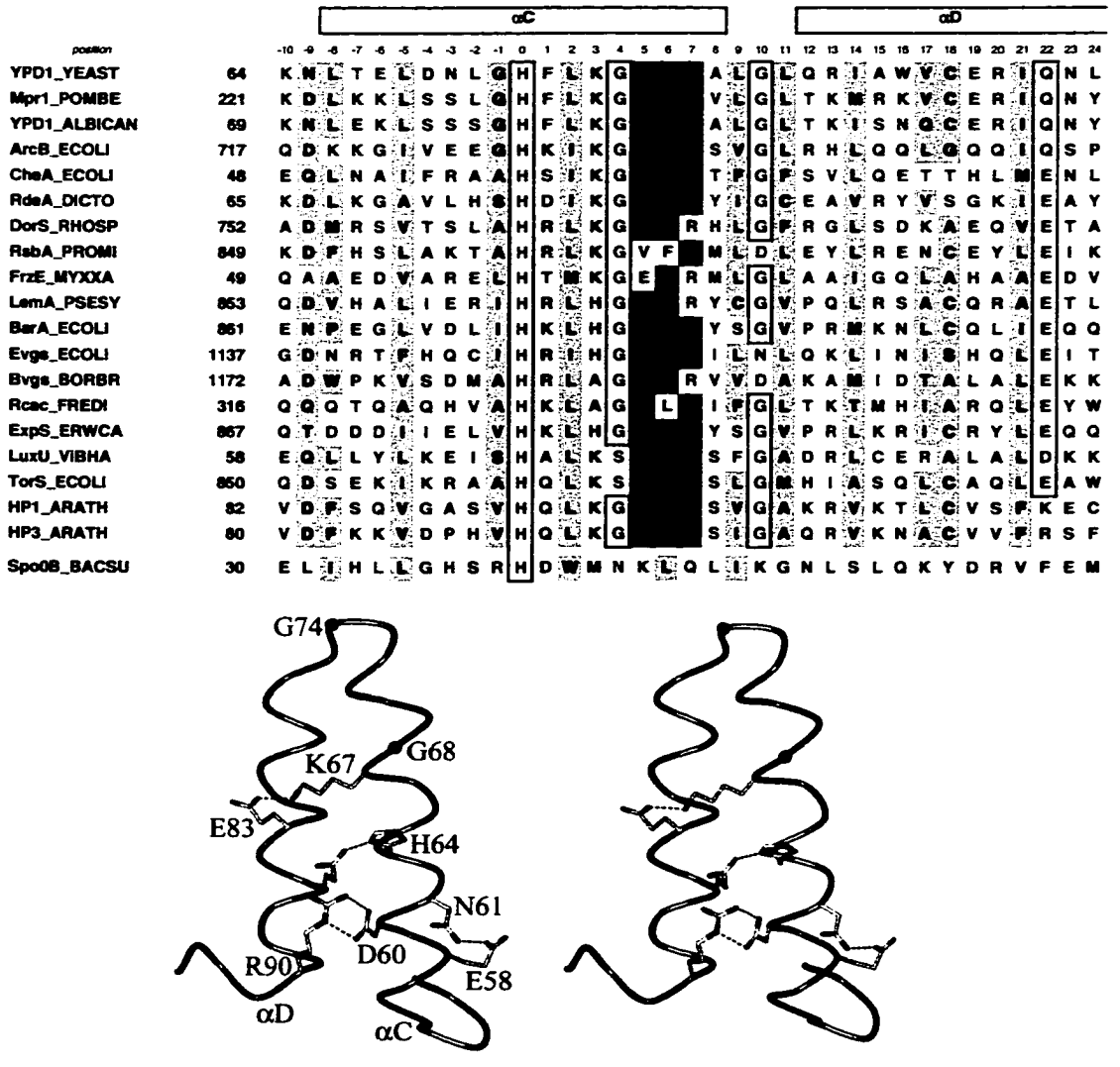


Figure 2-13. Conservation in the αC-αD region of YPD1 among HPt domains. (**Top figure**) sequence alignment of this region of various HPt domains is shown. Shown above the sequence alignment, is the secondary structure of the αC and αD helices of YPD1. The newly discovered HPts such as Mpr1 from *S. pombe* and YPD1 from *Candida albicans* shows very high degree of sequence identity in this region. The numbering scheme across the top of the sequence alignment refers to positions of residues relative to the histidine phosphorylation site. The numbers in the second column refer to the position of the histidine phosphorylation site with respect to the primary sequence. Boxed residues are highly conserved amongst HPt domains. Light gray shading indicates hydrophobic residues with a periodicity of (i, i + 3, i + 4) typically found in amphipathic antiparallel helices. Dark gray shading highlights residues (position 5, 6 and 7) that are conserved in size/shape. Residues shaded in blue are positively charged side-chains near the phosphorylatable histidine residue that may be functionally important (see text). Residues at position -9, highlighted in pink, have side-chains that can serve as hydrogen bond acceptors in forming a helix "N-cap". Spo0B, a well-characterized HPt domain from *B. subtilis*, does not fit the alignment well (Tzeng *et al.*, 1998). (**Bottom figure**) Corresponding helices C and D of YPD1 are shown with potentially important residues in ball-and-stick (stereoview). The inter-helical hydrogen bonds are shown by dashed lines. Abbreviations, POMBE, *Schizosaccharomyces pombe*; ALBICAN, *Candida albicans*; ECOLI, *Escherichia coli*; DICTO, *Dictyostelium discoideum*; RHOSP, *Rhodobacter sphaeroides*; PROMI, *Proteus mirabilis*; MYXXA, *Myxococcus xanthus*; PSESY, *Pseudomonas syringae*; BORBR, *Bordetella bronchiseptica*; FREDI, *Fremyella diplosiphon*; ERWCA, *Erwinia carotovora*; VIBHA, *Vibrio harveyi*; ARATH, *Arabidopsis thaliana*; BACSU, *Bacillus subtilis*.

interactions (Janiak-Spens and West, 2000). Secondly, large side-chains around His64 are pulled away from the histidine side-chain. For example, Asn61 and Lys67, that are on the same surface of helix C as His64, are both pulled away from the histidine side-chain through hydrogen bonds to other residues. The side-chain for Asn61 (position -3) is hydrogen bonded to Glu58 (position -6) which essentially "pulls" the side-chain away from the imidazole ring of His64. A similar effect is observed in $ArcB^c$ in which Glu712 forms a salt bridge with nearby Lys710 and thus adopts a similar conformation as Asn61 in YPD1. On the opposite side of the phosphorylatable histidine residue in YPD1, Lys67 (position +3) is also positioned away from the imidazole ring by forming a hydrogen bond with Glu83 (position +19) on the adjacent helix D. Again, an interesting parallel is seen in $ArcB^c$ in which Lys720 forms a hydrogen bond to Gln736, thus positioning the lysine side-chain away from the phosphorylatable histidine residue.

The combined effect is that the Nε atom of His64, the site of phosphorylation, is fully exposed. Such an arrangement will likely make the imidazole ring readily accessible to serve as a donor/acceptor to response regulator domains. The solvent accessible area (SA) for the His64 side-chain in native conformation is 105 $Å^2$. In order to assess how the histidine His64 is exposed and to explain why another surface exposed histidine of YPD1, His94, is unable to serve as a phosphorylation site, we calculated the solvent accessible area for the His94 side-chain (YPD1-His94) and the maximum solvent accessible area for His64

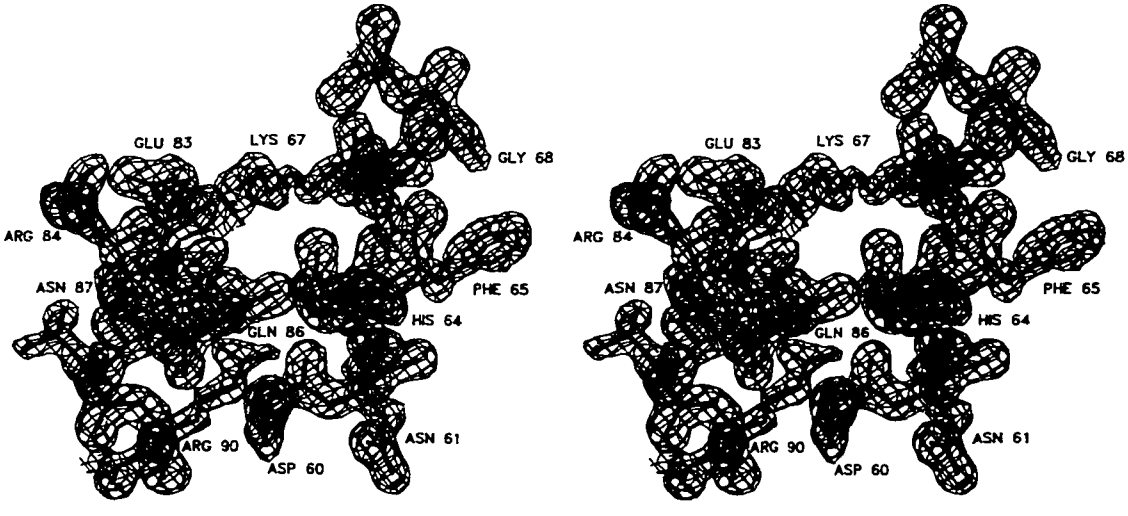


Figure 2-14. 2Fo-Fc electron density map of the phosphorylation site of 1.75Å YPD1 model in stereo view. The contour level is 1.5σ.

(YPD1-His64$_{max}$). The results are summarized in Table 2-5. The maximum solvent acces-

Table 2-5. Histidine side-chain solvent accessible area

Side-chain	YPD1-His64	YPD1-His64$_{max}$	ArcB^c-His717	YPD1-His94
Cβ	2.1	11.1	4.5	27.4
Cγ	1.9	3.0	2.9	3.4
Cδ	26.7	26.7	28.9	18.7
Nδ	1.8	21.9	3.5	13.5
Cε	41.4	43.7	42.8	22.9
Nε	30.6	30.6	29.7	12.4
Total	104.5	137.0	112.3	99.1

sible area for His64 side-chain (YPD1-His64$_{max}$) is calculated in the same manner as the His64 side-chain, but with all the side-chains surrounding the His64 ignored. First, it is obvious that both Nδ and Nε of His94 have limited solvent exposure while in contrast the Nε of His64 has SA of 30.6 Å^2. Second, by comparing the solvent accessible area for each atom of the side-chain between YPD1-His64 and YPD1-His64$_{max}$, it is evident that the difference is mostly contributed by changes in Nδ atom (20 Å^2), which was originally buried (SA~2 Å^2) by a hydrogen bond to Gln86 Oε atom and C$_β$ atom (9 Å^2). The solvent accessible area for the phosphorylation site Nε stays the same (30.6 Å^2). This calculation suggests that such an arrangement of the phosphorylation site at His64 in YPD1 achieves two purposes: 1) makes the histidine readily accessible to the active sites of response regulators by exposing the Nε phosphorylation site as much as possible; 2) at the same time provides a means of stabilization for the histidine side-chain (via a hydrogen bond) to achieve optimal reactivity. Due to the similarity in their phosphorylation sites, the distribution of solvent accessible area for ArcB^c is, as expected, almost identical to YPD1.

The Nδ atom of His64 forms a hydrogen bond to the half-buried Gln86 carbonyl oxygen atom on the neighboring helix D. This hydrogen bond interaction, which requires a precise arrangement of the two helices relative to one another is also present in ArcB^c. Similar hydrogen bond patterns have been observed in other proteins that also function via a phospho-histidine intermediate. In fact, the phosphorylatable histidines sites often have a

similar hydrogen bond in other proteins such as succinyl-CoA synthetase (PDB code 1SCU, Wolodko *et al.*, 1994), nucleoside diphosphate (NDP) kinase (PDB code 1NSQ, Moréra *et al.*, 1995) and enzyme IIA of the phosphoenol pyruvate-dependent phospho-transferase (PTS) system (Nunn *et al.*, 1996). It seems likely that the hydrogen bonds in these histidine phosphorylated proteins serve similar purposes, *i.e.* to optimally orient the histidine imidazole ring for phosphoryl transfer and possibly to affect the reactivity of the histidine.

The overall electrostatic surface of YPD1 (Figure 2-15) is mostly negative. Two clusters of negative charges are located at two flexible loops 21-24 and 124-133 (opposite end of four-helix bundle). The negative clusters around 21-24 is conserved in close YPD1 homologs such as Mpr1 from *S. pombe* and YPD1 from *Candida albicans*. The roles of these negatively charged residues are not known currently. One possible role might be to direct YPD1 in an appropriate orientation toward reacting partners. The molecular surface above the His64 is mostly neutral, several charged residues are scattered to the left and below His64. Electrostatic interactions within the conserved anti-parallel helix motif may be both structurally and functionally important for HPt domains. First, the two interhelical salt bridges (Arg90-Asp60, Lys67 and Glu83) contributes to the overall stability of helix C

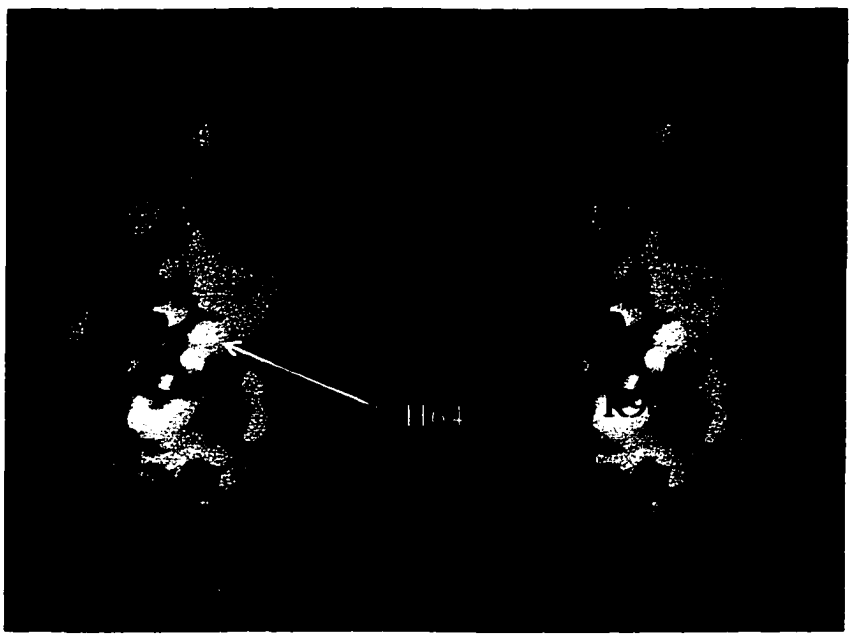


Figure 2-15. Stereoview of the electrostatic potential surface of YPD1 (red, negative, blue, positive, white, neutral).

and D. There is a positive patch near His64 due to the presence of Lys67 and Arg90. The lysine residue at position +3 is quite conserved in terms of charge within the HPt protein family. The orientation of this lysine residue is completely superimposable in the $ArcB^c$ structure. There are several possible roles for positively charged residues at this location. First, these positively charged residues on the HPt domains may enhance the binding affinity between response regulators and HPt domains. Experiments based on a reverse yeast two-hybrid assay (Bartel and Fields, 1997) seem to suggest that Lys67 might be involved in binding SSK1-R2 (Stace Porter, personal communication). Secondly, there is a possibility that the positively charged residues could provide a stabilizing force by hydrogen bonding to the oxygen atoms of the phosphorylated histidine residue and/or may provide electrostatic charge neutralization. Based on the crystal structure, both Lys67 and Arg90 are not in favorable positions to form hydrogen bonds with phosphate oxygen atoms when His64 is phosphorylated without undergoing a significant side-chain movement. Site-directed mutagenesis (Janiak-Spens and West, 2000) showed that mutations of both K67A and R90A decrease the half-life of the phospho-histidine. As a result, the positive charges contribute to the stability of the phospho-histidine on YPD1, possibly through charge neutralization. Biochemical experiments also indicate that mutation of these residues (Lys67 and Arg90) does not critically affect phosphotransfer between YPD1 and various response regulator domains (Janiak-Spens and West, 2000).

2.4 Conclusions

In this chapter, the crystal structure of YPD1 at 1.75Å resolution was described. The structure of YPD1 consists of a four-helix bundle core which is conserved in all currently known HPt domain structures. Despite the limited degree of sequence homology, the conservation of structure and function between the yeast YPDl and bacterial ArcB HPt domain is quite striking. The overall fold, key residues around the phosphorylation site, and potential response regulator binding interface are well conserved. Conservation of the YPD1 structure is supported by the observation that YPD1 can interact with multiple response regulator domains.

The phosphorylation site of YPD1 is located on the surface of the four-helix bundle. and is highly solvent exposed. The surrounding residues contribute to the exposure and sta-

bility of the phosphorylation site histidine. As such, the arrangement of the phosphorylation site creates an optimal configuration for phosphotransfer.

Base on the analysis of the YPD1 structure, it is possible that the four-helix bundle fold, particularly the αC-αD helical hairpin motif with the centrally located exposed histidine residue, will be an important conserved feature of other HPt domains that interact with CheY-like response regulators.

3 Structure of the G68Q Mutant, Implications for YPD1 and Response Regulator Interaction

3.1 Introduction

The previous chapter described the crystal structure of YPD1. When compared with other atomic structures of HPt proteins (or domains), such as the C-terminal domain of ArcB ($ArcB^C$), P1 domain of CheA and Spo0B, it was evident that they all share a four-helix bundle core despite the fact that the overall sequence homology between HPt proteins/domains from different species is very weak. The histidine phosphorylation sites are located in the middle of an α-helix that forms part of the four-helix bundle. Analysis of the sequence alignment of the two anti-parallel helices (αC and αD) near the site of phosphorylation revealed stronger sequence homology. Close inspection of this region of YPD1 revealed that the side-chain of the phosphorylatable histidine is completely exposed to solvent. It has been hypothesized that such an arrangement helps to make His64 readily accessible to the active site of a response regulator domains (Xu and West, 1999).

The exposure of the histidine side-chain is achieved by the arrangement of other side-chains around His64. First, the bulkier side-chains around the His64 are oriented away from the histidine. Second, there are residues with small or no side-chains in close proximity of the histidine. More specifically, Gly68 is highly conserved among many HPt domains (see Figure 2-13 of Chapter 2). Since the position of Gly68 is in close proximity to the active site residue His64, we speculated that Gly68 is essential for the function of YPD1, the space created by having no or small side-chain being critical for the accessibility of His64 by response regulator proteins. To investigate this hypothesis, we used site-directed mutagenesis to mutate the Gly68 to a residue with a bulkier side-chain. In order to minimize

the effects of electrostatic and hydrophobic interactions, we chose to mutate Gly68 to Gln68, hereafter denoted as G68Q. *In vitro* assays using a reconstituted phosphorelay system showed that the G68Q mutant is severely impaired in its ability to transfer phosphoryl groups from SLN1-R1 to SSK1-R2 (Janiak-Spens and West, 2000).

The 2.2 Å structure of the G68Q mutant is presented in this chapter. By comparison and analysis of this structure, a rationale for the loss of phosphotransfer activity of the G68Q mutant as well as other experimental findings will be presented from a structural point of view. The results of a gel mobility shift assay for this mutant is also presented, which is consistent with our interpretation of the effect of the G68Q mutation.

3.2 Materials and Methods

3.2.1 Construction of the G68Q mutant, purification and crystallization

Site-directed mutagenesis, purification and crystallization of the G68Q mutant were carried out by Dr. Janiak-Spens as described previously (Janiak-Spens and West, 2000). The purification procedure for G68Q was similar to that of native YPD1 (Xu *et al.*, 1999). The hexagonal rod-like form of the G68Q mutant crystals were obtained by seeding in 0.1 M sodium acetate pH 5.0 buffer with 0.2 M ammonium acetate and 28-30% PEG 4000 using a hanging-drop vapor diffusion setup and resembled the Q86E mutant crystals (shown in Figure 4-1) in morphology. The starting protein concentration was 13 mg/ml. The original microseeds were obtained in 0.1 M sodium acetate pH 5.0 buffer with 0.2 M ammonium acetate and 35% PEG 4000, and a protein concentration of 19.0 mg/ml. The crystal used for data collection had dimensions of 0.1 mm x 0.1 mm x 0.35 mm, and was harvested after two weeks of growth.

3.2.2 Data collection and processing

Data were collected using a wavelength of 1.0 Å at NLSL station X12B at Brookhaven National Laboratory (Figure 3-1). The crystal used for data collection were flash frozen beforehand in liquid propane in the original crystallization mother liquid containing ~30% PEG 4000. The crystal was kept at 100 K in a N_2 gas stream during data collection. The data were recorded with a Quantum 4 CCD detector using oscillations of 1.0°.

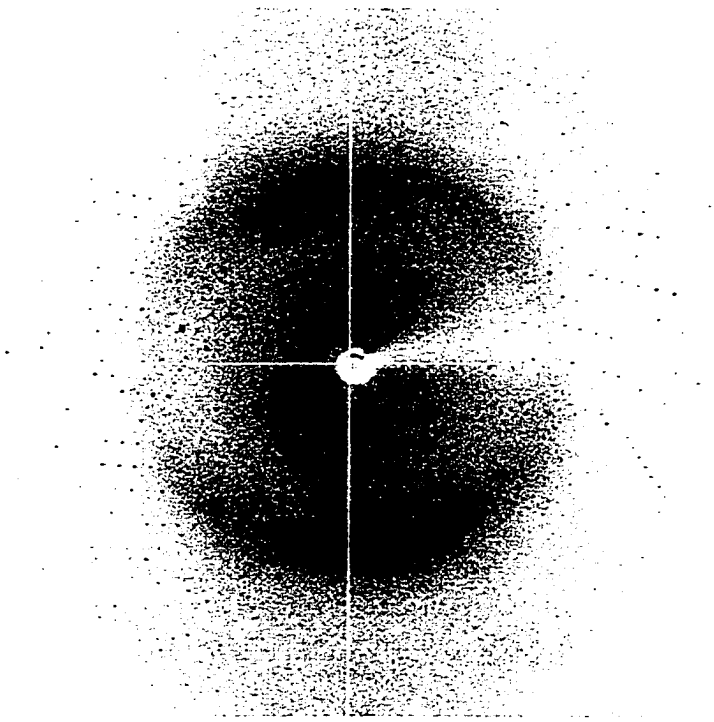


Figure 3-1. Diffraction pattern of the crystal of G68Q mutant.

The data were indexed, integrated, merged and scaled using the DENZO/SCALEPACK software (Otwinowski and Minor, 1997). The G68Q mutant crystallized in space group $P3_121/P3_221$ with cell dimensions of $a=b=75.994$ Å, $c=65.878$ Å, $\alpha=\beta=90°$, $\gamma=120°$. There was one molecule per asymmetric unit, which corresponds to a *Matthews' constant* of 2.9 $Å^3$ or ~60% solvent content (Matthews, 1968). The hexagonal G68Q mutant crystal diffracted to 2.2 Å. The data collection and processing statistics for the G68Q crystal are summarized in Table 3-1.

3.2.3 Molecular replacement and refinement of the G68Q structure

The G68Q mutant structure was determined using the molecular replacement method using the refined structure of wild type YPD1 as the starting model with B factors set to 25.0 based on Wilson statistics. The orientation and position of the monomer in the asymmetric unit was determined using the program REPLACE (Tong and Rossman, 1990). The space group was determined to be $P3_121$ by inspection of the electron density maps calculated in space group $P3_121$ and $P3_221$.

Table 3-1. Data collection and refinement statistics for G68Q mutant

Data Collection	
Space group	$P3_121$
Unit cell dimensions (Å)	a=75.994 c=65.878
Resolution (Å)	40-2.2
Total observations	78761
Unique observations	11498
Completeness (%)	99.8(99.9)[c]
Redundancy	6.85(6.1)
I/σ(I)	35.1(16.8)
R_{merge}[a]	4.9(13.0)
Refinement	
R_{cryst} (R_{free})[b]	19.1(25.6)
Number of protein atoms	1342
Number of solvents	98
Average B factor (Å^2)	
protein	28.60
solvent	34.89
overall	29.03
R.m.s.d bond distance (Å)	0.0064
R.m.s.d bond angle (deg.)	1.14
Ramachandran plot	
Residues in most favorable region (%)	94.9
Residues in additional favorable region(%)	5.1

[a] $R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma|I|$, where I is the intensity for a give reflection, and <I> is the average intensity for multiple measurements of this reflection.
[b] $R_{cryst}=\Sigma|F_o-F_c|/\Sigma|F_o|$, where F_o *and* F_c are the observed and calculated structural factors respectively. R_{free} was calculated using 8% of randomly selected reflections, which were not used in the refinement of the coordinates. R_{cryst} was calculated using the rest of reflections.
[c] Values given in parenthesis correspond to the highest resolution shell (2.28-2.2Å).

The refinement procedure for G68Q was similar to that of WT YPD1 as described in Chapter 2. Ten percent of the reflections were used as the R_{free} set. Rigid body refinement using XPLOR improved the R_{factor} from 49.5 to 47.3 (R_{free} from 53.4 to 47.0). One round of torsion angle simulated annealing and isotropic B refinement dropped the R_{factor} to 31.2 (R_{free}=39.1). The model was rebuilt using a SIGMAA-weighted map and displayed in XtalView (McRee, 1993). All the main-chain electron density was continuous and clearly visible including the loop 125-132, which was not visible in the 1.75 Å WT electron density maps. The Gln68 side-chain was built into the map at a later stage of refinement when the electron density for all the side-chain atoms were clearly defined. The R_{factor} for the final model was 19.1 (R_{free}=25.6). Refinement statistics for the 2.2 Å G68Q mutant structure are summarized in Table 3-1.

3.3 Results and Discussion

3.3.1 Structure of the G68Q mutant

The G68Q mutant was crystallized in space group $P3_121$. The structure of the G68Q mutant was determined at a resolution of 2.2 Å by using the molecular replacement method. The model contains all the residues in YPD1 (2-167). The average B factor for the monomer is 28.60 $Å^2$, which is higher than that of the 1.75 Å WT model (~17 $Å^2$). The value is consistent with the Wilson plot analysis that gave an estimation of the overall B value as being 27.8$Å^2$. The side-chain of Q68 showed well defined electron density (Figure 3-2). All the non-glycine residues of the G68Q model fell into the allowed region of a Ramachandran plot in which 94.9% belong to the most favored region as defined in the program PROCHECK (Laskowski *et al.*, 1993).

There were no overall structural changes between G68Q and YPD1 WT. The Cα r.m.s.d between the two structures was 0.6 Å. The arrangement of the four-helix bundle is intact with r.m.s.d of 0.25 Å between Cα atoms. However, the largest structural conformational difference was observed in a region containing residues 10-26, which includes helix A. The C-terminus of helix A underwent a larger shift in position than its N-terminus. The largest shift for side-chain positions was about 7 Å. The last turn of helix A 3_{10}-helix was unwound in G68Q. As a result of the movement of helix A, the N-terminal position of helix

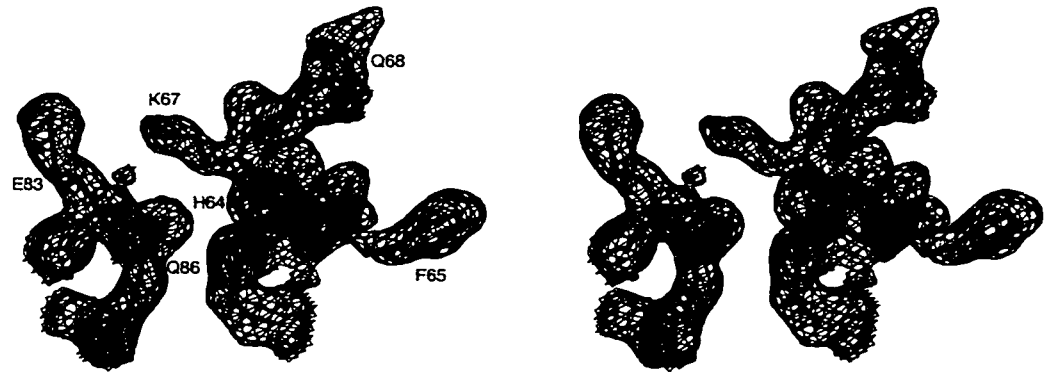


Figure 3-2. Stereo view of a portion of the final composite omit map and the 2.2 Å G68Q mutant structure, showing part of helix C and helix D. The contour level is 1.20σ.

B and the C-terminal of the G68Q structure adjusted positions accordingly. Among these, the side-chain orientation for Tyr162 is changed. Additionally, some minor differences were observed in the linker region between helix D and G (Figure 3-3A).

The rigidity of the four-helix bundle core and flexibility of the loops and helix A observed in the G68Q and WT crystal forms may reflect the dynamics of YPD1 in solution. Maintaining the rigidity of the four-helix bundle is likely to be important for structure integrity of the molecule, molecular recognition, and also possibly for the phosphotransfer reaction.

The αC-αD region of G68Q remains similar to that of WT (Figure 3-3B). The side-chain of Gln68 packs along the axis of helix C but faces away from the His64. As a result, the conformation of His64 and its immediate neighbors are not disturbed by this mutation. In the structure of G68Q, the salt bridge observed in the WT structure between Lys67 and Glu83 is no longer well defined because of the poor electron density of the Lys67 side-chain Nζ atom. This suggests that there could be some flexibility for the Lys67 side-chain in solution.

Using a probe sphere of radius of 1.4 Å, the solvent accessibility of the His64 side-chain for both the WT and G68Q mutant was calculated in order to estimate the effects of the G68Q mutation. The values for the WT are the average of the two molecules in the asymmetric unit. The G68Q mutation decreases the solvent accessible area for the His64

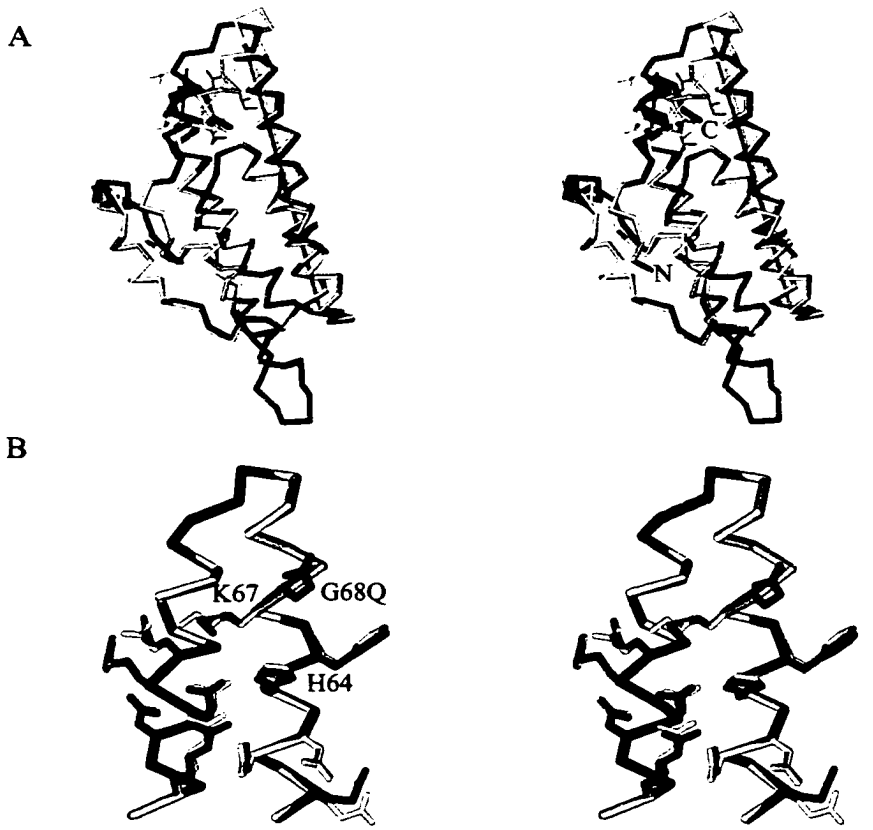


Figure 3-3. Superimposition of the G68Q mutant (green) with the wild type structure (gray) in stereoview. **A.** Overlay of Cα backbone with Ile13, Leu14, Ile17, Ile18 and His64 side-chains. Cylinders along the axes of helix A and helix B indicate the magnitude of the movement. **B.** overlay of the environment around the phosphorylation site, including the side-chain conformation of the glutamine substitution.

side-chain by about 10%. The solvent accessible area for His64 in WT is 100.1 $Å^2$, while the solvent accessible area for the His64 in G68Q is 91.5 $Å^2$. The main portion of the His64 which is affected in the G68Q mutant is the main-chain O (25%) and side-chain Cδ2 (55%), contributed by the bulkier side-chain of glutamine as well as a small side-chain movement of Phe65. The mutation does not seem to have a significant effect on the solvent accessible area of the phosphorylation site Nε (Figure 3-4).

3.3.2 Movement of helix A

The most significant conformational change in G68Q is the large movement of helix A (Figure 3-5). The Cα r.m.s.d between G68Q (residues 11-25) and WT (residues 11-25) is 3.0 Å. The movement of helix A can be decomposed into two components. First, the

Figure 3-4. The glutamine substitution at position G68 (yellow) in comparison to the wide type surface of YPD1 (colored by hydrophobicity), shown in stereoview. Hydrophobic residues are shown in blue, hydrophilic residues are shown in red.

central axis of helix A rotates about 30° upwards from its WT position centered on the N-terminal region of the helix (Leu9). Second, helix A twists around its axis for about 1/4 turn. As a result of the twisting movement, the originally exposed Ile13 and Ile17 side-chains become buried while Leu14 and Ile18 become exposed. The combined effects of the alternate conformation of helix A are that similar hydrophobic interactions as observed in the WT structure, are maintained. In the unit cell, helix A is located between helix B (in contact with 37-58) and helix G (in contact with residues 130-140), contributed by two

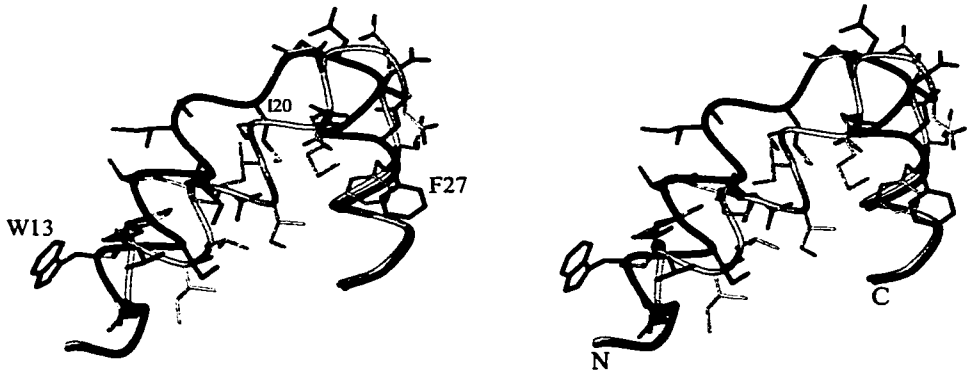


Figure 3-5. Superimposition of helix A and the N-terminus of helix B (residues 7-32) between YPD1 WT (gray) and the G68Q mutant (green) in stereoview. Side-chains are shown in stick model.

symmetrically related molecules respectively. The otherwise exposed hydrophobic residues on helix A (Trp11, Leu14, Ile18) are protected by the hydrophobic patch consisting of residues Val128, Ile136, Leu140 of the symmetry related molecule.

The movement of helix A affects the molecular surface properties of YPD1. In the WT YPD1 structure, there is a hydrophobic patch consisting of Ile13, Ile17, Met20, Phe27, Leu31, Ala72, and Leu73. However, because of the helix A movement in the G68Q mutant described above, hydrophobic residues on helix A which contribute to this hydrophobic patch is buried. For example, Ile17 becomes completely buried in the structure of the G68Q mutant.

In the crystal lattice of the G68Q mutant, Trp11, located at the N-terminus of helix A, is packed against a cluster of glutamates/aspartates from the loop between helix F and helix G of a symmetry-related monomer. Although there is well defined electron density for Trp11, the side-chain for Trp11 cannot fit into the electron density without violating geometrical constraints (data not shown). It is possible that unfavorable interactions between Trp11 and negatively charges residues contribute to the displacement of Trp11 and helix A from its original position (Figure 3-6). Although the movement of helix A is likely due to the crystal packing, it is possible that the flexibility of helix A is intrinsic. It is not known whether such conformational flexibility of helix A is functionally related, for example, in moderating interactions between YPD1 and response regulators.

3.3.3 Crystal packing

Since the phosphorylation site of YPD1 (αC-αD surface) is fully exposed to solvent and partially involved in crystal packing, it is important to evaluate the effects of crystal packing on the phosphorylation site configuration as well as the overall YPD1 structure. Based on the similarity of the phosphorylation sites in both WT YPD1 and the G68Q mutant, determined from different crystal forms, we concluded that the crystal packing artifacts are small, and the configuration of the phosphorylation site in both the crystal structures are likely representative of their configurations in solution.

In the tetragonal crystal form of WT YPD1, the position of Gly68 is close to three hydrophobic residues Phe27 (3.85 Å) and Leu31 (4.61 Å) and Ile17 (6.83 Å) of a symmetry-related molecule. The introduction of a glutamine residue at position 68 may cause

steric hindrance in terms of crystal packing. This may explain why G68Q molecules did not favor adopting the molecular packing as was observed in the WT tetragonal form. The molecules are packed more loosely in the unit cell in the hexagonal G68Q crystal form (~60% solvent content) as compared to the tetragonal WT crystal form (~40% solvent content). There is a large solvent channel of radius ~20 Å along the *c* axis in the crystal form of G68Q. Such molecular packing may allow G68Q molecules more freedom for movement (consistent with the overall B factor value of 28.60 $Å^2$).

In both crystal forms, the His64 residues are close to aspartates or glutamates due to molecular packing, this suggests that the phosphorylation site has a certain degree of affinity for or at least tolerance of negatively charged residues. In the WT crystal, Lys67 and Arg90 form ionic interactions with an aspartate residue (Asp24) of a symmetry-related molecule, while His64 is not directly involved in the crystal packing. The Nε of His64, the site of phosphorylation, is within hydrogen bond distance to a water molecule, which in turn is in contact with the main chain carbonyl group of Phe27 of a symmetry-related molecule. In the hexagonal form of the G68Q mutant, Lys67 is not involved in crystal packing. The Nε of His64 is hydrogen bonded (distance 3.05 Å) to Glu122 Oε1 of a symmetry-related molecule (Figure 3-6).

3.3.4 Effects of G68Q mutation

Biochemical characterization of the G68Q mutant (Janiak-Spens and West, 2000), via an *in vitro* phosphorelay system, indicated that the G68Q mutant is adversely affected in its phosphotransfer ability. The steady-state level of phosphorylation of the G68Q mutant via phosphotransfer from SLN1-R1 is ~30% as compared to the level of phosphorylation for wild type YPD1. Phosphotransfer to SSK1-R2 from the phospho-G68Q mutant is also much less efficient.

Since the conformation of His64 and surrounding residues of the G68Q mutant remain largely intact in the G68Q mutant, the loss of activity in the G68Q mutant is not likely to be caused by conformational changes in or around the phosphorylation site. It seems reasonable to conclude that the loss of activity in the G68Q mutant is mainly caused by steric hindrance, *i.e.*, the bulkier side-chain of a glutamine poses steric hindrance for binding of response regulator domains. As a result, the level of phosphorylation of YPD1

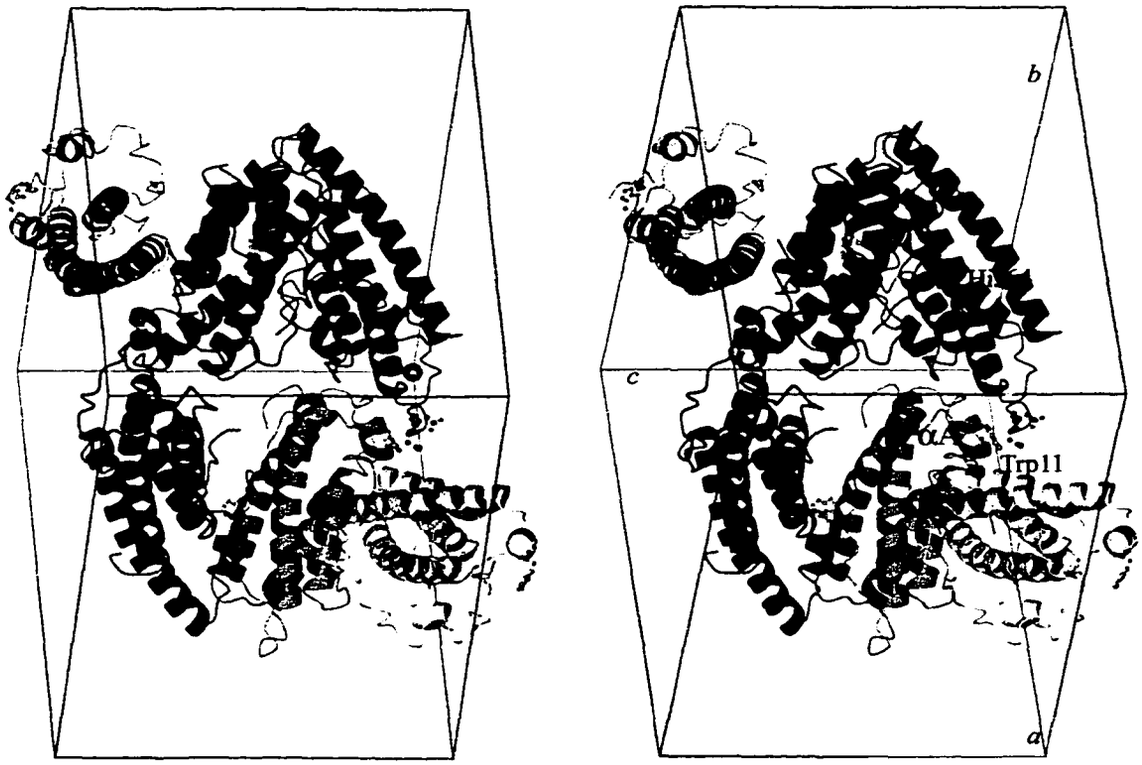


Figure 3-6. Crystal packing diagram of the G68Q mutant in space group $P3_121$ in stereoview. Each molecule in the unit cell is shown in a different color.

reaction will be affected. The idea that the bulkier side-chain prevents binding between YPD1 and response regulator domains is supported by results of a native gel shift assay (Figure 3-7). It had previously been demonstrated that YPD1 is able to protect phospho-SSK1-R2 from hydrolysis and as a result extends the half-life of phospho-SSK1-R2 (Janiak-Spens *et al.*, 2000). Further characterization indicated that SSK1-R2 forms a complex with YPD1 only when the response regulator domain is phosphorylated (Figure 3-7B, lane 7). Interestingly, the G68Q mutant disrupts the formation of this complex (Figure 3-7B, lane 8).

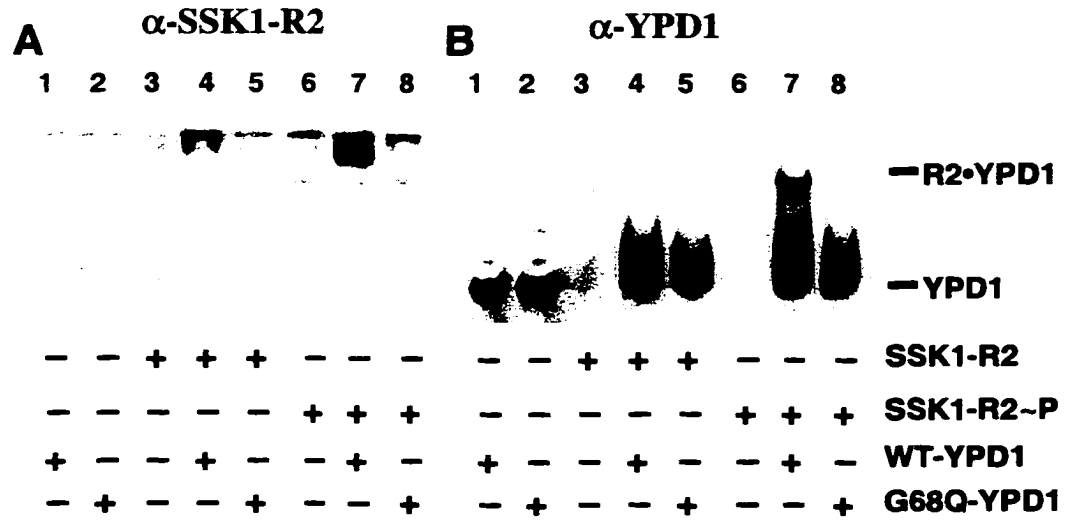


Figure 3-7. Native gel shift assay demonstrating a complex formed between YPD1 and SSK1-R2 (Courtesy of Dr. Fabiola Janiak-Spens).

3.3.5 Binding interface on YPD1

Based on biochemical and structural studies of the G68Q mutant, it is evident that the Gly68 site is involved in close contact with response regulator domains upon binding. The overall surface of YPD1 is mostly hydrophilic. There are large patches of neutral surface area above His64, delineated by a hydrophobic perimeter shown in Figure 3-8A). The surface area for this patch is ~690 $Å^2$. This area is mainly contributed by residues from helix A, helix C and the N-terminal portion of helix B, including Ile13, I17, Met20, Phe27, Leu31, Gly68, Ser69, Ala72, Leu73 and Gly74. Interestingly, a similar hydrophobic surface area is also present in $ArcB^c$ (Figure 3-8B). Sequence alignment of these residues between YPD1, $ArcB^c$ and other YPD1 homologs reveals similar conservation patterns (Figure 3-9). It seems reasonable to suggest that this patch may contribute to the interacting interface of YPD1.

Evidence supporting the viewpoint that this region may be involved in binding response regulator come from mutational studies of the YPD1 homologue $ArcB^c$ (Matsushika and Mizuno, 1998; Kato *et al.*, 1999). In $ArcB^c$, random mutagenesis and site-directed mutagenesis identified the following residues on the helix containing the phosphorylatable histidine His715 (equivalent to helix C of YPD1) as being important for molecular recog-

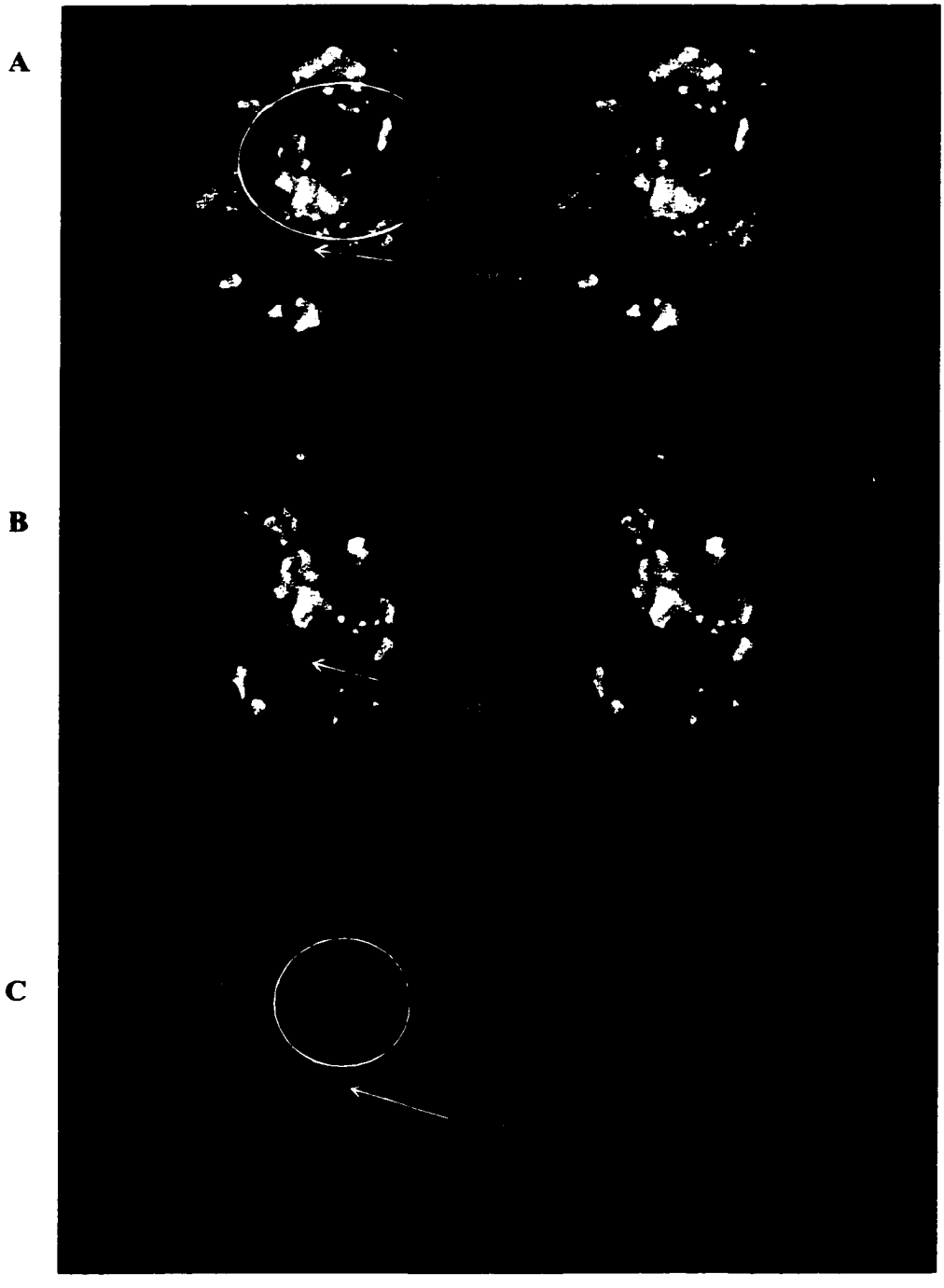


Figure 3-8. Comparison of the molecular surfaces of YPD1 and the ArcB HPt domain in stereoview. Hydrophobic residues are shown in blue, hydrophilic residues are shown in red. **A.** molecular surface of YPD1 colored according to hydrophobicity; **B.** molecular surface of $ArcB^{c}$ colored according to hydrophobicity; **C.** Superimposition of YPD1 (yellow, transparent) and $ArcB^{c}$ (red) molecular surfaces.

	αA			αB		αC				
	13	17	20	27	31	68	69	72	73	74
YPD1	I	I	M	F	L	G	S	A	L	G
MPR1	V	L	M	F	I	G	S	A	V	G
CaYPD1p	V	I	M	F	L	G	S	V	L	G
ArcB	M	Y	V	L	G	G	A	S	V	G
	667	671	675	679	683	719	720	723	724	725

Figure 3-9. Sequence alignment of residues contributing to the hydrophobic patch defined in the text between YPD1 and close YPD1 homologs. Both Mpr1 and CaYPD1p have global sequence similarity to YPD1 and thus reliable sequence alignment can be obtained. The sequence alignment with ArcB was obtained by structural alignment using crystal structures of YPD1 (PDB code 1QSP) and ArcB^{c} (PDB code 2A0B). The number on the top corresponds to YPD1 sequence. The number at the bottom corresponds to the C-terminal domain of ArcB sequence. Corresponding secondary structure of YPD1 is shown at the top.

nition and/or phosphotransfer activity: K716E, G719D, and G722E, which are equivalent to YPD1 Phe65, Gly68, Ala71. Two other mutations in ArcB^{c} of Leu679 and G683 (equivalent to Phe27 and Leu31 on helix B of YPD1), result in loss of function. Since both YPD1 and ArcB HPt interact with CheY, it is not surprising that the molecular surface, especially around the hydrophobic region discussed above, of these proteins are similar to each other (Figure 3-8C). In the ArcB HPt domain, the overall molecular surface surrounding the active-site histidine residue is concave mainly because of kinks in the two long helices of the four-helix bundle due to a proline residue in the middle of one helix. The YPD1 molecular surface encompassing the phosphorylatable histidine residue (formed mainly by helices A, C, and D) is more or less flat. The concave shape is therefore not a general feature of all HPt domains.

It seems likely that HPt domains share a common recognition mechanism involving surface complementary and hydrophobic interactions. Affinity between HPts and certain response regulator domains, such as YPD1 toward different response regulator domains in yeast, might come from localized favorable interactions.

3.4 Conclusions

In summary, the mutation of Gly68 to Gln68 does not affect the reactivity of His64 by altering the histidine conformation itself, but rather the bulkier volume of a glutamine side-chain generates steric hindrance and as a result blocks the binding of YPD1 to response regulator domains. In fact, combined with results from studies of other mutations

located near His64, the accessibility of His64 appears to be the most important factor in the phosphotransfer mechanism. Additionally, since YPD1 needs to interact with more than one response regulator, the rigidity of four-helix bundle as well as critical side-chain arrangements, ensures universal accessibility of His64 to all interacting partners. Our conclusions correlate with similar studies on $ArcB^c$, a close functional and structural homologue of YPD1.

4 Crystal Structures of Gln86 Mutants, Possible Roles for the Hydrogen Bond Between Gln86 and His64

4.1 Introduction

The phosphorylatable histidines in both YPD1 and ArcB^c are hydrogen bonded to a glutamine residue from the nearby helix. Sequence comparisons suggest that this hydrogen bond is likely to be conserved in other HPt domains as well. The functional importance of the hydrogen bond between His64 and Gln86 is not well understood. Based on structural interpretation of YPD1 and ArcB^c, there are at least three possible roles for this hydrogen bond. First, it may provide structural stability by contributing to the inter-helical interactions between helices C and D. Another possible role for this hydrogen bond is to immobilize (*i.e.* limit the freedom of) the His64 side-chain for phosphorylation since without the hydrogen bond the imidazole ring of the histidine may be free to rotate. Additionally, this hydrogen bond might affect other properties of His64, for example the *pKa* of His64, and thus may be important in the kinetic process (McEvoy and Dahlquist, 1997). In contrast, a mutational study of ArcB^c has suggested that the conserved hydrogen bond (His717 and Gln739) may not be important functionally (Matsushika and Muzuno, 1998). Furthermore, solution NMR studies of ArcB^c also suggest the possibility that this hydrogen bond might not exist in ArcB^c (Ikegami *et al.*, 2001).

Biochemical characterization of several Gln86 mutants of YPD1 indicate that certain mutations can affect the His-Asp phosphoryl transfer reaction. However, these *in vitro* studies also suggests that the Gln86-His64 hydrogen bond may not be essential for the phosphotransfer activity of YPD1 (Janiak-Spens and West, 2000). Further studies on the

kinetic properties and structures of these mutants need to be carried out in order to better understand the role(s) of the hydrogen bond between His64 and Gln86 in YPD1.

To this end, we studied high resolution crystal structures of a series of Gln86 mutants, Q86A, Q86L and Q86E, which have been partially characterized biochemically (Janiak-Spens and West, 2000). The phosphorylation sites in these mutants maintain a similar hydrogen bond network as in the wild type protein. The Q86E mutant achieves the same purpose with a slight rearrangement of the side-chain of Glu86. In both the Q86A and Q86L structures, water molecules were found to form hydrogen bonds with His64 Nδ.

4.2 Materials and Methods

4.2.1 Cloning, purification and crystallization

The cloning, expression, and biochemical characterization of the YPD1 Q86A, Q86L and Q86E mutants were carried out by Dr. Janiak-Spens and reported in detail previously (Janiak-Spens and West, 2000). The purification procedure for these mutants was the same as that of WT YPD1. The Q86A and Q86L mutants crystallized in a tetragonal space group using the same conditions as WT YPD1 (summarized in Chapter 2). The Q86E crystals (Figure 4-1) on the other hand crystallized in a trigonal space group and were obtained using a similar procedure as described for the G68Q mutant (see Chapter 3).

Figure 4-1. Q86E crystals in a different crystal form from the wild type crystals.

4.2.2 Data collection and processing

All X-ray data were collected at 100 K using a 1 degree oscillation angle on an Raxis IV^{++} image plate detector and Rigaku RUH3 rotating anode X-ray generator operated at 50 kV and 100 mA. Artificial mother liquid containing 30% or higher PEG 4000 was used as a cryoprotectant. The Q86A and Q86L data were indexed, integrated and scaled using the software D*TREK (Pflugrath, 1999), whereas Q86E data were processed with DENZO/SCALEPACK (Otwinowski and Minor, 1997). The Q86E mutant crystallized in space group $P3_121$ with cell dimensions of $a=b=75.869$ Å, $c=65.577$ Å, $\alpha=\beta=90°$, $\gamma=120°$. There is one molecule per asymmetric unit corresponding to a Matthews' constant of 2.9 Å^3 (Matthews, 1968). The Q86A and Q86L mutants have the same space group and similar cell dimensions as WT YPD1. The trigonal Q86E mutant crystal diffracted to 1.9 Å. The tetragonal crystals of Q86A and Q86L diffracted to 2.0 Å. However, due to the long cell dimension along the *c* axis, a sacrifice of higher resolution data had to be made in order to obtain a better I/σ (signal to noise) ratio. As a result, 2.0 Å and 2.09 Å data have been obtained for Q86L and Q86A, respectively. The data collection statistics for the Q86E, Q86L and Q86A structures are summarized in Table 4-1.

4.2.3 Molecular replacement and refinement of the Q86E mutant structure

In order to avoid model bias, the Q86E mutant structure was determined using the molecular replacement method using the 1.8 Å YPD1 structure (Song *et al.*, 1999) as the starting model with the N-terminal 23 residues omitted. The orientation and position of the monomer in the asymmetric unit was determined using the program CNS (Brünger *et al.*, 1998). In space group $P3_121$, the cross rotation function produces the highest peak ($\theta_1=126.74°$, $\theta_2=90.51°$, $\theta_3=73.08°$) with peak height of 0.1091 (the second highest peak height was 0.0630). The translational search with the orientation obtained from the cross-rotation search yielded a #1 peak with height 0.582 (#2 peak height =0.235) with good inter-molecular packing. Rigid body refinement of this solution gave R_{factor}/R_{free} values of 0.415/0.412 (starting from 0.417/0.415). In space group $P3_221$, the cross rotational search produced the same peaks as expected, while the translational search gave a highest peak of height 0.322 (second highest peak is 0.215). Rigid body refinement of this solution gave R_{factor}/R_{free} values of 0.5242/0.5234 (starting value 0.5225/0.5244). Thus, the space group

Table 4-1. Data collection and refinement statistics

Data collection			
Crystal	Q86L	Q86A	Q86E
Temperature (K)	103	103	103
Cell dimensions (Å)	a=51.79 c=239.80	a=52.22 c=240.43	a=75.87 c=65.58
Space group	$P4_32_12$	$P4_32_12$	$P3_121$
Resolution (Å)	2.0	2.09	1.9
Mean I/σ(I)	21.8(6.0)	27.5(15.6)	24.7(6.1)
Observations	78826	105149	101581
Unique reflections	20962	19896	17449
Completeness(%)	90.3(74.1)	91.0(73.6)	99.4(98.7)
R_{merge} (%)[a]	6.2(12.3)	4.4(16.4)	4.4(30.7)
Refinement			
R_{cryst} (R_{free})[b]	20.5(25.6)	21.0(25.4)	20.6(23.1)
Number of protein atoms	2555	2581	1336
Number of solvents	254	312	75
Average B factor (Å^2)			
protein	15.67	17.78	30.49
solvent	22.68	27.44	30.42
overall	16.30	18.73	30.53
R.m.s.d bond distance (Å)	0.0067	0.0053	0.0093
R.m.s.d bond angle (deg.)	1.25	1.07	1.21
Ramachandran plot			
Residues in most favorable region (%)	93.9	94.6	97.4
Residues in additional favorable region(%)	6.1	5.4	2.6

[a] $R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma|I|$, where I is the intensity for a give reflection, and <I> is the average intensity for multiple measurements of this reflection.
[b] $R_{cryst}=\Sigma|F_o-F_c|/\Sigma|F_o|$, where F_o *and* F_c are the observed and calculated structural factors respectively. R_{free} was calculated using 8% of randomly selected reflections, which were not used in the refinement of the coordinates. R_{cryst} was calculated using the rest of the reflections.
[c] Values given in parenthesis correspond to the highest resolution shells.

for Q86E was determined to be $P3_121$. The solution generated by this procedure was equivalent to that of the G68Q structure described in Chapter 3.

The refinement procedure for Q86E was similar to that of WT YPD1 as described previously. Ten percent of the reflections were used as the R_{free} set. One round of torsion angle simulated annealing and isotropic B refinement dropped the R_{factor} to 0.319 (R_{free}=0.340). The model, including the N-terminal 22 amino acids, was rebuilt using XtalView based on a $2F_o$-F_c omit map, and then refined using the least squares method. The R_{factor} after this step was 0.2547 (R_{free}=0.2743). Water molecules were added in subsequent steps and were checked against electron density maps. Finally, two alternate conformations for Lys67 were built into the omit map. The R_{factor} for the final model was 0.206 (R_{free}=0.231). Refinement statistics for the 1.90Å structure of the Q86E mutant are also summarized in Table 4-1.

4.2.4 Refinement of Q86A and Q86L mutant structures

The refinement procedure for the Q86A and the Q86L structures was similar to that of WT YPD1 as described previously. Again the 1.8 Å YPD1 model determined by Song *et al.* (1999) was used as a starting model to avoid possible model bias. The starting model was stripped of water molecules and the B factors were uniformly set to 15.0 $Å^2$ according to Wilson statistics. The crystallographic programs suite CNS was used for the refinement of both mutants. All the data were used in the refinement and a bulk solvent correction implemented in CNS (Brünger *et al.*, 1998) was used to model the disordered solvents. the program XFIT (McRee, 1993) was used to display and rebuild the models throughout the refinement process.

For the mutant Q86A, the starting R_{factor}/R_{free} after rigid body refinement was 0.4012/0.3943. The non-crystallographic system operators were calculated from the rigid body model. Subsequent simulated annealing refinement imposed the non-crystallographic symmetry strictly. The R_{factor}/R_{free} after this round of refinement was 0.3400/0.3598, after which, B-factor refinement reduced the R_{factor}/R_{free} to 0.3085/0.3251. A $2F_o$-F_c omit electron density map was calculated and the model was rebuilt according to the omit map. The non-crystallographic symmetry was relieved in subsequent rounds of refinement and rebuilding. Water molecules were located in batches in the $2F_o$-F_c and F_o-F_c electron den-

sity maps during final stages of refinement. The final model contained all residues from Ser2 to Leu167 and 312 solvent molecules. The flexible loop region (residues 124-133) in the WT YPD1 model was not visible in the final electron density map for the Q86A mutant. As a result, these residues were not included in the final model. The final crystallographic R_{factor}/R_{free} was 0.2104/0.2539.

Similar steps were used to refine the structure of the Q86L mutant. The starting R_{factor}/R_{free} after rigid body refinement was 0.3850/0.3947, and after one round of stimulated annealing refinement of the coordinates followed by B-factor refinement using strict NCS constraints dropped to 0.3078/0.3364. The model was then rebuilt and refined again using least squares minimization. Water molecules were located in batches using the CNS program. The final crystallographic R_{factor}/R_{free} was 0.2050/0.2562. Refinement and model statistics for the Q86A and Q86L structures are summarized in Table 4-1.

4.3 Results and Discussion

4.3.1 Structures of the Q86 mutants

Three structures of YPD1 Gln86 mutants, including Q86A, Q86L and Q86E, have been determined by the molecular replacement method (Rossmann and Blow, 1962). Overall, the structures of the Q86L and Q86A mutants are very similar to the WT YPD1 structure. The r.m.s.d for the Cα positions between Q86L (or Q86A) and WT YPD1 is only ~0.18 Å. The Q86E structure overall is very similar to the G68Q mutant structure described in Chapter 3. The hydrogen bond network around the phosphorylation sites in the three mutants are similar to what was observed in the WT crystal structure. The histidines in all three mutants maintain the same conformation as in WT and show well defined electron density (Figure 4-2A-C).

In the Q86L crystal structure, a water molecule, located nearby Leu86, is within hydrogen bond distance (2.62 Å) to the His64 Nδ. Thus, this hydrogen bond ostensibly replaces the original hydrogen bond that existed between Gln86 and His64 in the WT structure. This water molecule is further secured in place by additional hydrogen bond interactions to nearby residues, specifically Arg90 and Asp24. The distance between the O atom of the water molecule and Arg90 Nη1 is 3.18 Å. The hydrogen bond to Asp24 is due to

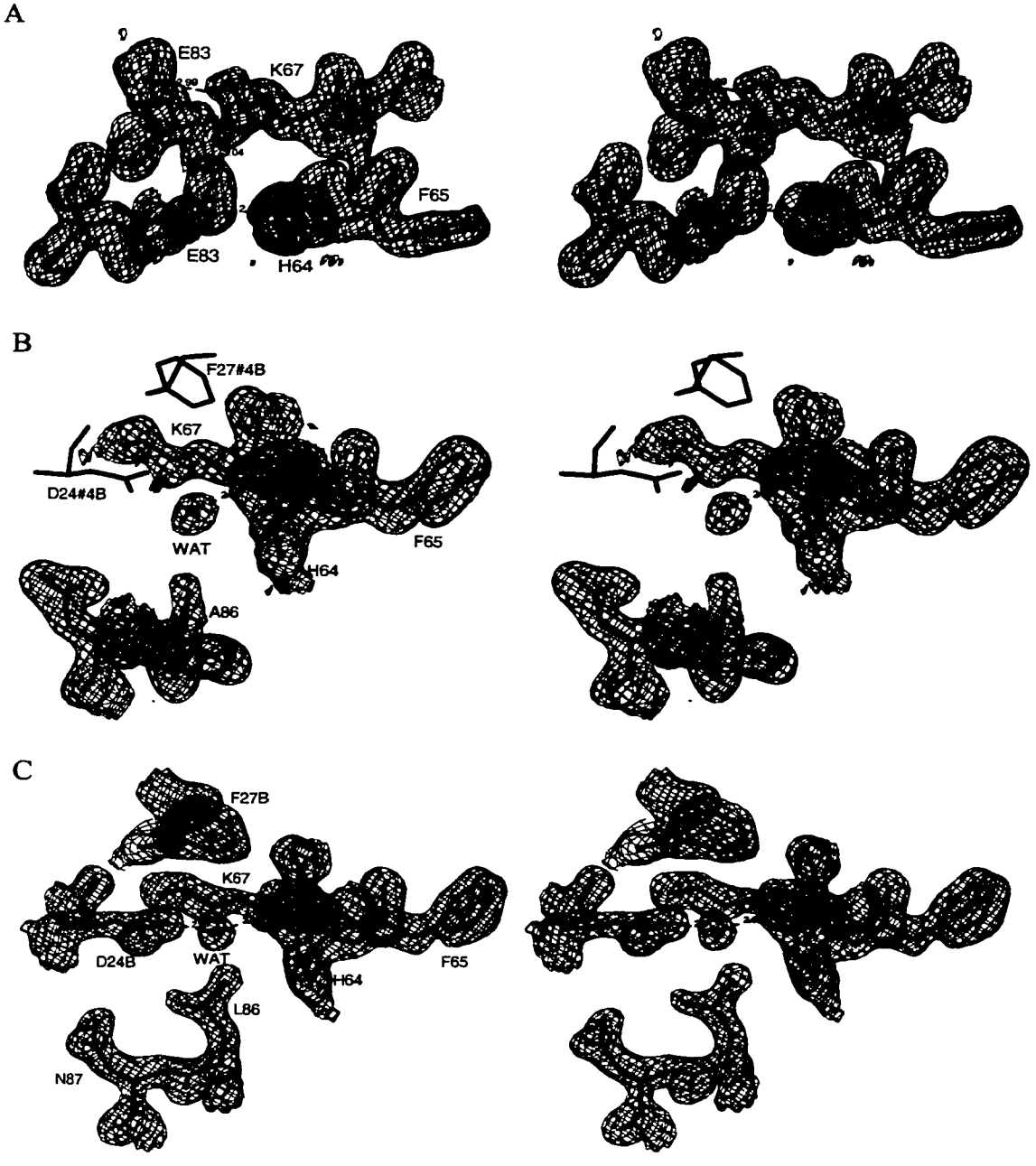


Figure 4-2. Electron density maps ($2F_o$-F_c) around the phosphorylation site of YPD1 mutants. **A.** Q86E contoured at 1σ; **B.** Q86A contoured at 1σ. The residues from a neighboring molecule in the crystal lattice are shown in sticks without electron density; **C.** Q86L contoured at 1.5σ

crystal packing and thus is not likely to be present in solution. In the other Q86L monomer in the asymmetric unit, however, no water molecule was observed near the His64 Nδ atom. Instead, Nδ is in direct contact with Asp24 of another monomer (Figure 4-3A).

The phosphorylation site in the Q86A mutant is very similar to that of the Q86L mutant. A water molecule is located at a similar location as the Oε atom of Gln86 in the WT YPD1 structure. When the two structure are overlaid onto each other (Figure 4-3B), the distance between this water molecule (in the Q86A mutant structure) and Gln86 Oε is 0.96 Å (in the WT structure).

The glutamate substitution in the Q86E mutant adopts a different conformation from the glutamine residue in the WT structure (Figure 4-3C). Compared to the glutamine side chain, the terminal group of Glu86 is rotated about 90 degrees, possibly through a favorable electrostatic interaction with a nearby positively charged residue Lys67. As a result, two conformations were observed for Lys67, one of the conformers maintains the interaction with Glu83 as seen in the WT structure, while the second conformer interacts with Glu86. Due to the new side-chain arrangement, Glu83 is also within hydrogen bond distance to Nδ of His64 as well as Nη1 of Arg90. Interestingly, the side-chain conformations of Lys67 and Glu86 are very similar to the WT structure of YPD1 (PDB code 1C02) determined independently by Song *et al.* (1999) using crystals obtained with different crystallization conditions (Figure 4-3D).

4.3.2 Structural roles for the Gln86-His64 hydrogen bond

4.3.2.1 Effects of Gln86 mutations on phosphorylation of YPD1

The Gln86 mutants maintain their ability to mediate phosphoryl group transfer to and from response regulator domains *in vitro* (Janiak-Spens & West, 2000). These observations indicated that the hydrogen bond between Gln86 and His64 might not be essential for YPD1-mediated phosphotransfer. It was assumed that the mutations Q86A and Q86L would result in disruption of the hydrogen bond since both alanine and leucine lack the ability to be a hydrogen bond donor or acceptor. However, as demonstrated by the crystal structures of Q86A and Q86L, the environment and exposed nature of the phosphorylatable histidine provide ample opportunity for a hydrogen bond to be formed with a water mole-

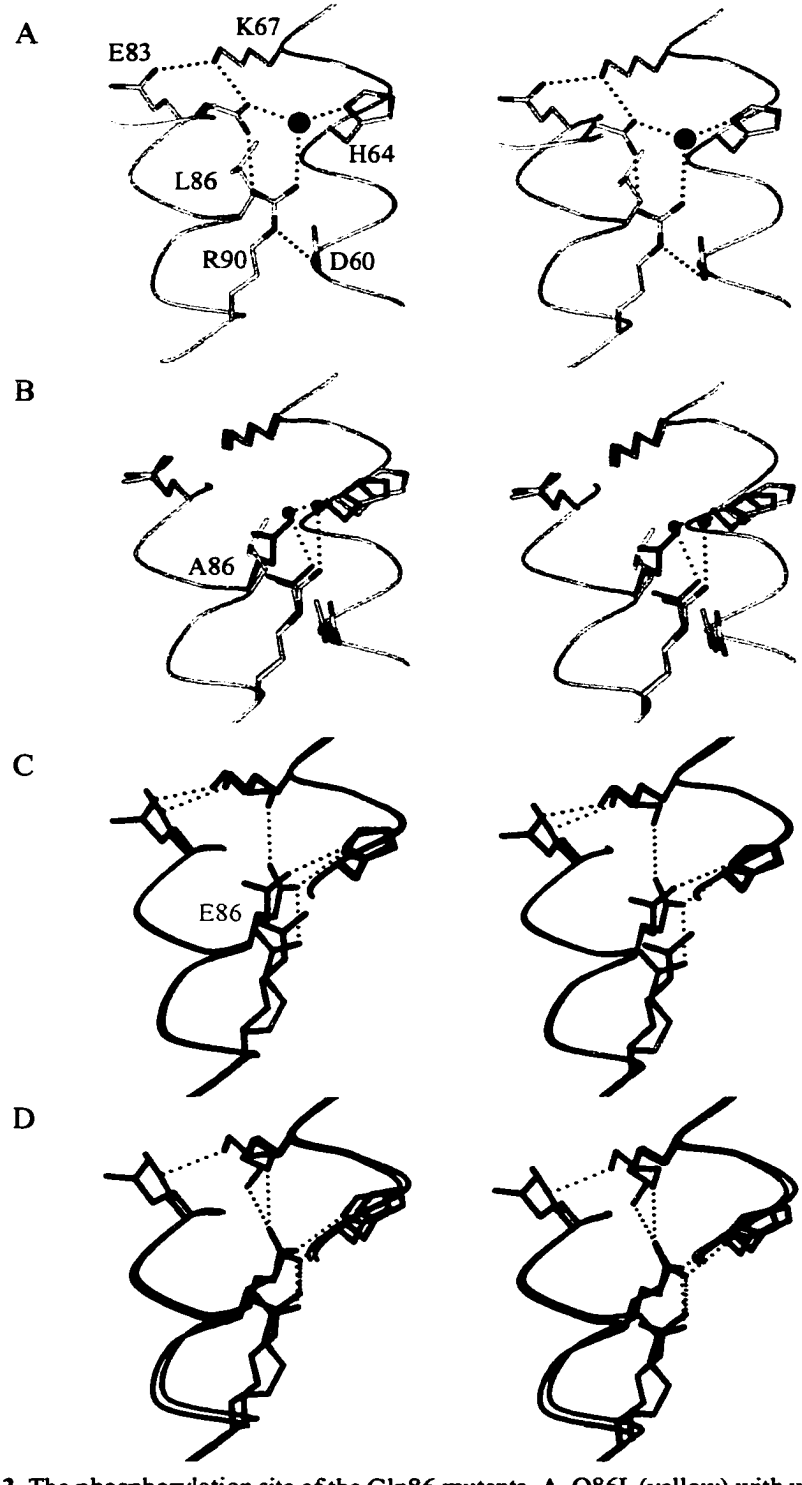


Figure 4-3. The phosphorylation site of the Gln86 mutants. **A.** Q86L (yellow) with water molecule (red sphere) and hydrogen bond interactions (dashed lines); **B.** Side-chains for Ala86, His64, Arg90 and the water molecule (blue) in Q86A are overlaid onto Q86L (yellow). The side-chain for Gln86 of the WT structure is shown in green; **C.** Q86E (cyan) is superimposed onto WT (dark red, 1QSP); **D.** Q86E (cyan) is superimposed onto the Song *et al.* YPD1 structure (1C02-A, green).

cule. Thus, it appears that a hydrogen bond involving the histidine Nδ1 atom is indeed an important structure feature required for YPD1 function.

4.3.2.2 Orientation and mobility of the phosphorylatable histidine

It has been suggested that the hydrogen bond to the solvent exposed histidine residue helps to orient the histidine Nε atom in an optimal conformation for phosphotransfer. Several crystal structures of HPt domains, such as Spo0B, $ArcB^c$ and YPD1, support the role of such a hydrogen bond in restricting the mobility of the solvent exposed histidine. In both YPD1 and $ArcB^c$, in which this hydrogen bond was observed, the histidine electron density is clearly defined. In Spo0B, however, the equivalent hydrogen bond is not observed because the electron density for the histidine side-chain in the Spo0B crystal structure is not well defined, possibly indicating more flexibility for this histidine residue (Varughese *et al.*, 1998).

It is important to note that an optimal conformation of the histidine can be induced by its interacting partners, as exemplified by Spo0B. As indicated by the crystal structure of the Spo0B/Spo0F complex, the histidine was spacially restricted to the well defined observed conformation (similar to the YPD1 histidine conformation, see Chapter 5) by the side-chain of Ala83 within the β4–α4 loop of Spo0F (Zapf *et al.*, 2000).

4.3.3 (Possible) multiple conformations for Gln86

When the phosphorylation site of our 1.75Å model is compared with that of the 1.8 Å YPD1 model determined by Song *et al.* (1999) at room temperature (PDB code 1C02), we notice some side-chain differences concerning several important residues around the phosphorylation site, in particular, His64, Lys67 and most noticeably Gln86. First, the Lys67-Glu83 salt bridge is not observed in the Song *et al.* structure (1C02) due to terminal side-chain movement of Lys67, however the interaction between Lys67 and Gln83 is still maintained through a mediating water molecule. Second, the side-chain of Gln86 in the Song *et al.* structure adopts a different conformation from our model. As a result, the His64 side-chain was pushed slightly away. The Arg90 side-chain position was shifted upward as well in order to accommodate this side-chain conformation. As a result, Gln86 Oε is within hydrogen bond distance of Arg90, which is not possible with the side-chain conformations observed in our models. The distance between His64 Nδ and Gln86 Oε is ~3.5Å, while the

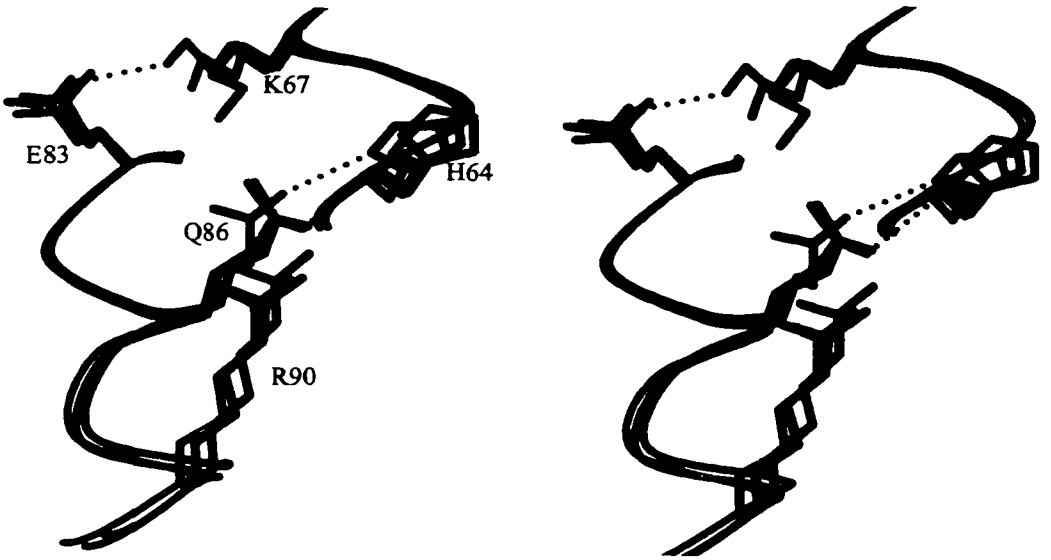


Figure 4-4. Multiple side-chain conformations observed in the phosphorylation site of YPD1. Three WT structures of YPD1 are shown above: the 1.8 Å structure determined by Song *et al.* (1999) in dark gray; the 1.75 Å structure described in Chapter 2, shown in dark red; and the 2.3 Å structure of YPD1 determined in space group P1 (Song *et al.*, 1999), shown in cyan.

distance in our current model is 2.7-3.0 Å. Thus, the hydrogen bond we observed (in 1QSP) might be stronger (Figure 4-4).

This observation suggests that there exists two side-chain conformations for Gln86 in the WT structure of YPD1. It may be possible that the two conformers can switch from one to the other in solution. As indicated by the structure of the Q86E mutant, the addition of a single negative charge at the Gln86 position can induce Glu86 to favor one of the two conformers. It is currently unclear what the functional significance is of multiple conformations for Gln86. An energy minimization of the modeled complex between YPD1 and CheY (see Chapter 5) indicated that Gln86 may favor the conformation observed in the Song *et al.* structure (1C02) and the Q86E mutant since this allows packing of the two molecules with less steric hindrance. Additionally, the strength of the hydrogen bond might correlate with the mobility of His64 side-chain. Control of the mobility of His64 via hydrogen bond strength, achieved through different side-chain conformations, might have functional implications. Some degree of flexibility for the histidine side-chain might be needed to accommodate slight differences in the active sites of the response regulator domains (SLN1-R1, SSK1-R2, SKN7-R3).

4.3.4 Tautomeric states of the phosphorylatable histidine (His64)

For YPD1 to be phosphorylated, the exposed His64 Nε must act as the attacking nucleophile. A hydrogen bond to the NδH could cause redistribution of electrons in the imidazole ring of the phosphorylatable histidine His64, and as a result, increasing the nucleophilicity of Nε and allowing it to attack the phosphoryl group as the nucleophile in the nucleophilic substitution (Sn2) reaction (Figure 4-5).

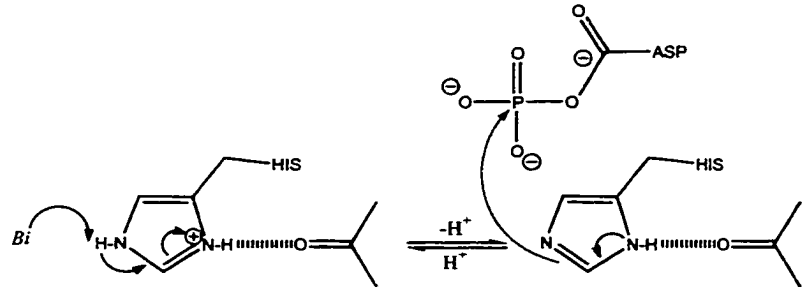


Figure 4-5. A hydrogen bond to the NδH of the imidazole ring can increase the nucleophilicity of the Nε phosphorylation site.

Current experimental results are controversial over the role of the hydrogen bond involving the phosphorylatable histidines. A solution NMR study of the YPD1 homolog $ArcB^c$ (Ikegami *et al.*, 2001) indicates that the imidazole ring of the phosphorylation site, His717, has a similar *pKa* value (6.76) to that of a solvent-exposed histidine imidazole ring (~6.9). Since the YPD1 phophosphorylation site is similar to the corresponding $ArcB^c$ site, it is expected that the *pKa* value for YPD1 is also similar. The NMR study of the tautomeric states of His717 on $ArcB^c$ indicated that the Nε and Nδ protonated states are equally possible. This result is different from results obtained for His48 of CheA P1 domain. The pH titration experiment of CheA indicated that the Nε-deprotonated tautomer predominates at high pH (Zhou & Dahlquist, 1997). The authors suggested that NδH is protected by a possible hydrogen bond and NδH is the hydrogen donor (Figure 4-5). Although in both the structures of YPD1 and $ArcB^c$, such a hydrogen bond to the Nδ atom of the phosphorylable histidine was also present (Gln86 and His64 in YPD1, Gln739 and His717 in $ArcB^c$), one feasible reason for the difference between CheA and $ArcB^c$ (or YPD1) is that the hydrogen bond partner to His48 of CheA is glutamate or aspartate instead of glutamine[1]. The negative charges on glutamate or aspartate attracts the proton to the Nδ and thus changes the

nucleophilicity of Nε. Glutamine, however, forms a weaker hydrogen bond and may not significantly affect the tautomeric distributions of the histidine residue. As a result, the influence of the *pKa* value by the hydrogen bond between Gln86 and His64 might be small.

The *pKa* values of His64 for Q86L and Q86A mutants are currently unknown. It is expected that they are similar to the WT YPD1 since its side-chain environment remains very similar to that of free histidines in water. On the other hand, a mutation from a neutral glutamine to a negatively charged glutamate will likely affect the *pKa* of the phosphorylatable histidine by increasing the nucleophilicity of Nε of His64 while maintaining its hydrogen bond to His64. However, preliminary experiments indicate that the *pKa* of His64 in the Q86E mutant remains more or less the same as WT (Dr. Janiak-Spens, unpublished). There is no apparent explanation for this observation. It is possible that the negative charge on Glu86 is evenly distributed due to its interaction with Arg90 and Lys67 and thus reduces its ability to affect the *pKa* of His64.

4.3.5 His-Asp phosphotransfer mechanism

Unlike histidine kinases, YPD1 lacks the ability to autophosphorylate using ATP. YPD1 cannot be phosphorylated using small molecule phospho-donors either. Instead, YPD1 appears to be a relatively passive but essential phosphotransfer intermediate, which can interact with multiple response regulator domains. In contrast to many phosphorylation sites which are located on the N-terminus of a helix and the phosphoryl group can then be stabilized by helix dipole, the phosphorylation site of YPD1 (or other HPt domains and histidine kinases for which structures are known) are located in the middle of a helix and as a result the contribution by the helix dipole is not significant. Instead, the stabilization could come from nearby positively charge residues. Phosphoryl groups on YPD1 can be removed by hydrolysis or phospho-accepting response regulator domains. The structure of YPD1 correlates well with its function as a phosphotransfer intermediate. The residues around the phosphorylation site are spread out on a flat surface made of two antiparallel helices and the histidine residue is solvent exposed. As a result, YPD1 lacks a traditional "active site"

1. The hydrogen bond partner of His48 is Glu70 according to a recently published high resolution crystal structure of the CheA P1 domain (Mourey *et al.*, 2001).

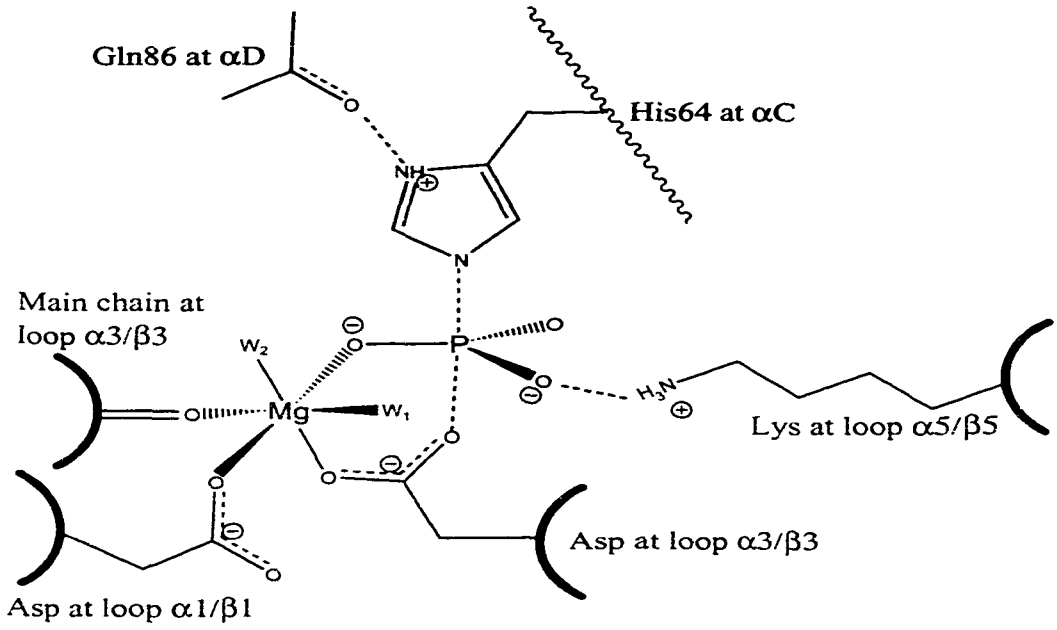


Figure 4-6. A transition state model for His-Asp phosphotransfer between YPD1 (black) and a generic response regulator domain (red). The essential magnesium ion is hexa-coordinated by several conserved Asp or Glu side-chains and ordered water molecules resulting in an overall trigonal bi-pyramidal geometry. The penta-coordinated phosphoryl group is stabilized and properly aligned with His64 of YPD1 for the phosphotransfer reaction by the magnesium ion and a conserved lysine residue. The location of important residues are labelled according to CheY secondary structure described in Chapter 1.

for binding small molecules. The exposed histidine is readily accessible to the active sites of response regulator domains.

The phosphorylation mechanism of response regulator domains such as CheY has been proposed based on structural and biochemical studies as well as similarity with other phosphoryl transfer reactions (Stock *et al.*, 1993). The penta-coordinated phosphoryl group is stabilized and properly oriented by an essential magnesium ion and a conserved lysine residue. The magnesium ion is octahedrally coordinated by several conserved Asp or Glu residues, including the aspartate that is phosphorylated (Asp57 in CheY), a main chain carbonyl and water molecules. The His-Asp phosphotransfer model can be derived from the above basic model. In fact, the structure of the Spo0B/Spo0F complex supports the transition state model above (Zapf *et al.*, 2000; Stock and West, 2001). A transition state model for YPD1 and a generic response regulator domain can be built similarly (Figure 4-6). It is obvious from this proposed model, that only the phosphorylatable histidine is required for

phosphotransfer to occur while all other components are contributed by the response regulator domain. It is not surprising that response regulator domains are enzymes in this case since the response regulator proteins are the ones which initiate the necessary downstream response. The observations that response regulator domains, such as CheY, can be phosphorylated by small molecule phosphodonors also indicates that conserved response regulator domains are catalytic enzymes in phosphorylation reactions. As a result, HPt domains such as YPD1 most likely do not possess enzymatic function. The overall structure of YPD1 provides a framework for the phosphorylatable histidine. An important aspect of the YPD1 scaffold is to provide a means for molecular recognition and specificity as discussed in the next chapter.

Thus, it seems likely that residues in YPD1 except the phosphorylatable histidine are not directly involved in forming the phosphotransfer transition state. The only other residue that might affect the phosphotransfer reaction is Gln86 which could form a hydrogen bond to His64. Like in most other proteins involving phosphorylatable histidines, its role might be to stabilize and properly orient the histidine. Other functionally important residues, may not be directly involved in the transition state, but could be important for binding or stabilization of the phosphoryl group once YPD1 is phosphorylated.

4.4 Conclusions

As revealed by the three structures of YPD1 mutants (Q86A, Q86L and Q86E), the hydrogen bond network near the phosphorylation site of YPD1 can be preserved by addition of a water molecule (as in the Q86A and Q86L mutants) or a slight arrangement of the side-chain (as in the Q86E mutant) close to the Gln86 location. These rearranged hydrogen bonds could provide an explanation for the retention of phosphotransfer ability in YPD1 mutants.

Although it cannot be determined conclusively whether the hydrogen bond is essential, it appears to be functionally important. One of its more likely functions may be to orient and stabilize the phosphorylatable histidine. Another possible role is to affect the *pKa* and thus the reactivity of His64. Further biochemical and kinetic characterization is needed in order to provide a more definite answer.

5 Modeling His-Asp Phosphotransfer in Yeast, Molecular Basis for Recognition and Specificity

5.1 Introduction

YPD1 appears to be a promiscuous HPt domain that can interact with multiple response regulators. In addition to its ability to interact with three response regulator domains (SLN1-R1, SSK1-R2 and SKN7-R3) in yeast, YPD1 is able to receive phosphoryl groups from bacterial CheY *in vitro* (Janiak-Spens *et al.*, 1999). It has also been demonstrated that YPD1 can function in other organisms by substituting for their corresponding HPt domains (Chang *et al.*, 1998). However, the response regulator domains that YPD1 can interact with share only limited sequence homology. Furthermore, *in vitro* experiments indicate that YPD1 has a higher affinity for phospho-SSK1-R2 as compared to other response regulator domains (Janiak-Spens *et al.*, 2000). These interesting properties of YPD1 raise the following two questions. First, how is molecular recognition between YPD1 and the response regulator domains achieved? Moreover, what factors contribute to the affinity that has for different response regulator domains?

All currently known structures of response regulator domains have a similar tertiary fold, represented by the bacterial chemotaxis protein CheY (Stock *et al.*, 1993; Volz, 1995) and described in Chapter 1. Sequence alignment of the CheY superfamily indicated that residues at the active site, the hydrophobic β-sheet core, as well as other functionally critical residues are highly conserved (Volz, 1995). Despite the fact that the overall sequence identity within the family is often as low as ~20%, the highly conserved landmark sequences in the CheY superfamily allows for accurate sequence alignment of the response regulator domains, including those from yeast.

A HPt domain in a typical multistep phosphorelay system interacts with more than one response regulator domain, which likely have similar three-dimensional structures. Insights about molecular interactions can be obtained by comparing the molecular surfaces of these response regulator domains. However, it is usually difficult and time consuming to determine the structures for every individual response regulator domains experimentally. Given the highly conserved nature of the tertiary fold of the response regulator family, they offer an excellent opportunity for predicting the three-dimensional structure of new response regulator domains using homology-based modeling approaches.

Comparative (or homology) protein structure modeling uses experimentally determined structures (templates) to predict the conformation of another protein (target). This is possible because proteins with similar sequences have similar structures. First, a known structure is identified as a homolog of the target. The target sequences are aligned with the template sequences. Accurate sequence alignment is most critical since the final model quality is highly correlated to the accuracy of the sequence alignment. The target model is built by copying backbone elements from this template. Side chain atoms can be placed by a wide variety of methods such as backbone-dependent rotamer libraries (Bower *et al.*, 1997), segment matching followed by minimization (Levitt, 1992) and self-consistent mean field optimization (Koehl and Delarue, 1994; Koehl and Delarue, 1996). Loops are in general more difficult to model. Although the models generated by homology modeling are not as accurate as experimentally determined structures, reliable theoretical models can be obtained with r.m.s.d as low as ~2.0 Å for large (especially buried) portions of structures dependent on the sequence similarity and quality of the sequence alignment (Koehl and Levitt, 1999). For residues on the surface, it is more difficult to model the side-chain conformations accurately because these residues are often more flexible due to less spacial constraints. However, these models are usually sufficient for mapping the surface residues close to their correct positions.

To better understand the molecular interactions and phosphotransfer mechanism between YPD1 and response regulator domains, structures of complexes between YPD1 and response regulator domains would be highly desirable and would provide us more definite answers. However, complexes between YPD1 and response regulators are difficult to

obtain since the interactions between YPD1 and response regulator domains are likely to be transient (Kern *et al.*, 1999; Zapf *et al.*, 2000) and dependent on the phosphorylation state of the interacting partners. The Spo0B/Spo0F crystal structure is the only biological complex structure currently available between a HPt protein and a response regulator (Zapf *et al.*, 2000). Despite sharing a central four-helix bundle, Spo0B shows no local and global sequence homology to other HPt domains. In addition, unlike many other HPt proteins, Spo0B functions as a dimer. Thus, it is not known whether the interaction between Spo0B and Spo0F can be generalized to other functionally similar systems.

Since the interaction between macromolecules is affected mainly by molecular surfaces between interacting partners, it seems likely that the range of affinities displayed by YPD1 toward different response regulators can be attributed to the molecular surface properties of the response regulators and YPD1. In this chapter, I attempt to address the molecular interactions between YPD1 and response regulator domains through the use of computer modeling. The homology models of the three response regulator domains SLN1-R1, SSK1-R2 and SKN7-R3 in yeast were built by homology modeling based on the crystal structure of CheY. Analysis of the homology models provided information about molecular surface properties of the yeast response regulator domains. Two different approaches were used to construct the complex between YPD1 and CheY. First, YPD1 was docked to CheY based on geometrical constraints. Second, we modeled the known structures of CheY and YPD1 into Spo0B/Spo0F complex. Based on the analysis of these models, a molecular basis for recognition and specificity was hypothesized.

5.2 Materials and Methods

5.2.1 Computational equipment and programs

All modeling and graphic visualization was carried out on SGI graphics workstations (Silicon Graphics Incorporated, USA). Amino acid sequence alignments were produced using ClustalW (Thompson *et al.*, 1994). The homology models were produced using the program WHAT IF (Vriend, 1990). The alignment of structures was carried out using the least square minimization method implemented in the program LSQMAN (Kley-

wegt and Jones, 1999). The program GRASP (Anthony *et al.*, 1991) was used to visualize and analyze molecular surface properties.

5.2.2 Sequence alignment of regulatory domains

The multiple sequence alignment of the regulatory domains were performed in a two step fashion. First, the initial seed sequence alignment based on a 3D structural alignment of several response regulator domains (PDB code 1NTR, 1RNL, 1SRR-A, 1TMY, 3CHY) was obtained from the COMPASS database (Sowdhamini *et al.*, 1998). Second, the sequences of selected response regulator domains were aligned to the above aligned sequences using the program Clustal W. Sequences that were of interest to us were regulatory domains known to or likely to interact with YPD1 or ArcB, including the three response regulator domains from yeast, the SSK1 homolog MCS4 from *Schizosaccharomyces pombe*, the ArcB response regulator domain from *E. coli*, as well as the template bacterial CheY. The sequence alignment for these sequences is shown in Figure 5-1 (initial seed sequences are not shown here).

5.2.3 Homology modeling

Using CheY (PDB code 2CHE) as the template, the side chains of different regulatory domains were modeled onto the backbone of CheY using the WHAT IF program. The loop inserts for R1, R2 and R3 were not modeled since they are located on the surface opposite from the active site. The resulting models show excellent geometry, indicating that the side-chains of our selected response regulator domains can be accommodated by the backbone of CheY.

5.2.4 Building YPD1-response regulator complexes

The complex of YPD1 and RR was first built with computer docking algorithm implemented in the program FTDOCK (Gabb *et al.*, 1997). This algorithm uses the Fast Fourier Transform (FFT) to enumerate all possible combinations between two macromolecules based on shape complementary (Katchalski-Katzir *et al.*, 1992). In general, a large list of possible solutions are produced. Geometrical constraints based on experimental knowledge are then used to filter out the results. In our case, two distance constraints based on the proposed reaction mechanism and conserved interactions were used (see Table 5-1).

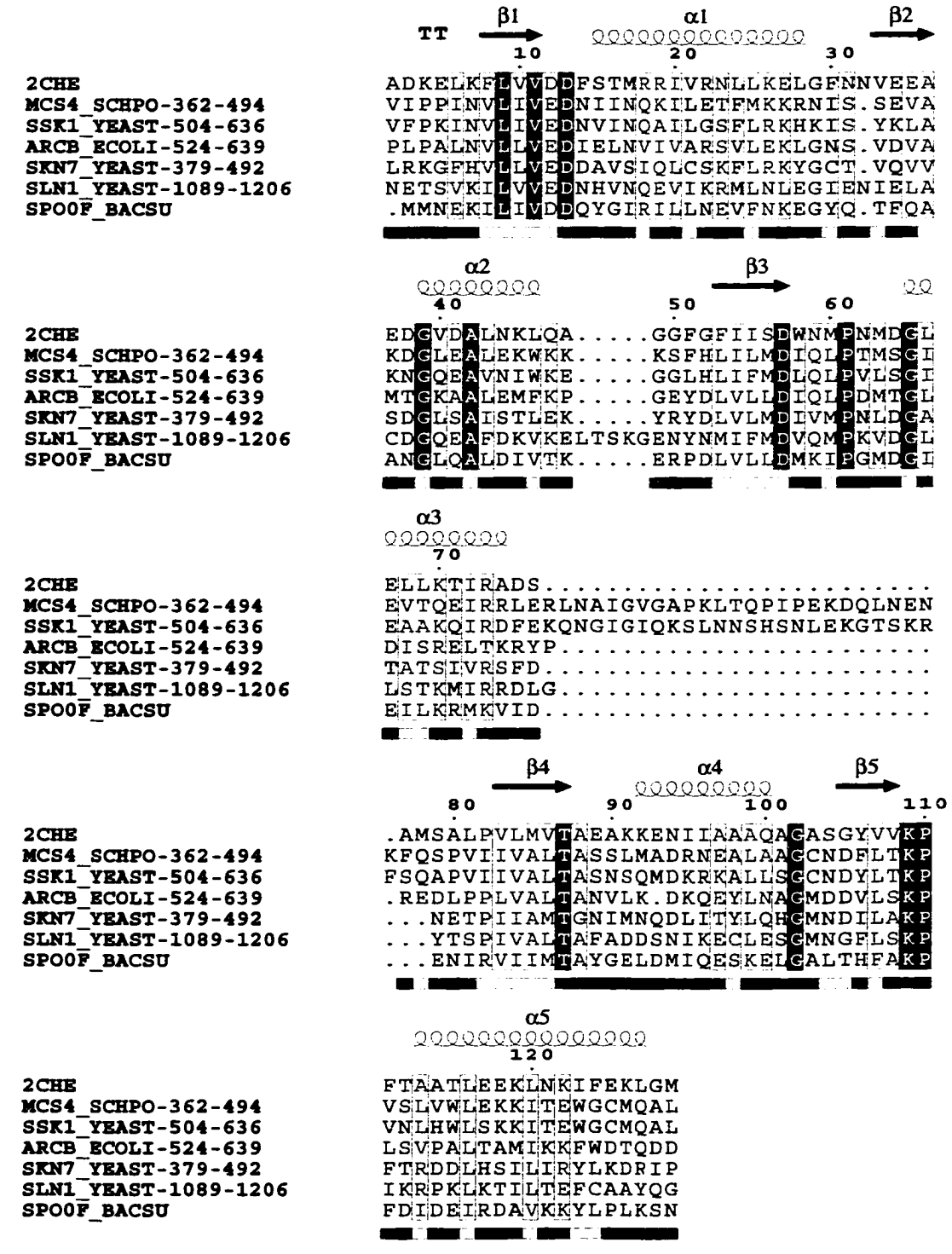


Figure 5-1. Sequence alignment of the bacterial chemotaxis protein CheY (template, PDB code 2CHE) with its homologs (targets) in yeast SLN1(R1), SSK1(R2), and SKN7(R3), *S. pombe* MCS4, *E. coli* ArcB regulatory domains and Spo0F from *Bacillus subtilis*. The secondary structure of CheY is shown on top whereas the solvent accessibility (white, buried; blue, solvent exposed) for each residue is shown at the bottom.

ArcB^c, a structural homolog of YPD1 (see Chapter 2), was initially used in the docking calculation for the following reasons: 1) ArcB^c, like YPD1, can interact with CheY as a phosphodonor (Yaku *et al.*, 1997; Matsushika and Mizuno, 1998; Janiak-Spens *et al.*, 1999); 2) its concave binding surface, because of a protruding helix A, provides more geometrical constraints; and 3) biochemical information about the interaction between ArcB^c and response regulator domains, on which distance constraints can be derived, are available (Kato *et al.*, 1999). First, for phosphotransfer to occur, the histidine and asparatate must be in close contact, most of the solutions by FTDOCK can be eliminated due to this constraint. Secondly, random mutagenesis of surface exposed residues of ArcB^c identified mutations which disrupt interactions between ArcB^c and response regulator domains. Specifically, two residues, Leu679 and Gly683 were implicated in binding the ArcA response regulator (Kato *et al.*, 1999). To summarize, the following distance constraints were used to exclude solutions that did not satisfy the above requirements (Table 5-1):

Table 5-1. Distance constraints

	CheY	**ArcB^c**	**Distance (Å)**
1	D57-CheY	H715-ArcB^c	5.0
2	*-CheY	L679-ArcB^c	5.0
3	*-CheY	G683-ArcB^c	5.0

Each row mean residue x from molecule A (x-A) is within 5 Å from residue y of molecule B (y-B), *-A mean any residue(s) from molecule A.

Upon applying the first filter, all but twelve solutions were eliminated. Only one solution was left after applying the second or third contraints (only one of condition 2 and 3 is needed).

The docking model between ArcB^c and CheY generated using the procedure above was found to be very similar to the biological complex of Spo0B/Spo0F, in the alignment of the phosphorylatable histidine to the active site aspartate residue as well as the four-helix bundle orientation relative to conserved response regulator domains. As a result, a model complex between YPD1 and CheY (or a response regulator domain) was generated based on the Spo0B/Spo0F complex by matching YPD1 to Spo0B and CheY to Spo0F. YPD1 was aligned with the Spo0B dimer in the Spo0B/Spo0F complex (PDB code 1F51) such that the phosphorylatable histidines from both proteins have similar Cα positions. A por-

tion of helix C of YPD1 (residues 60-72) was superimposed with a segment of the first helix (α1) of Spo0B (residues 26-38) using the phosphorylatable histidine positions as a reference point (Figure 5-4). The positioning of CheY or another response regulator domain was achieved by superimposing it with Spo0F in the Spo0B/Spo0F complex (PDB code 1F51).

5.3 Results and Discussion

5.3.1 Conservation of response regulator binding surfaces

Functionally important residues located in and around the active site as well as residues that form the hydrophobic core of response regulator domains are highly conserved. Based on the sequence alignment shown in Figure 5-1, critical residues of the selected response regulator domains, including Asp13, Asp57, Thr87, Lys109, were identified as being highly conserved (Figure 5-2A). The Asp12 position of CheY are glutamates in many other response regulator domains. The hydrophobic core, consisting of the central β sheet, is conserved as well. Tyr106 in CheY has been implicated as being involved in the phosphorylation-dependent activation mechanism of response regulator domains. This position in the sequence alignment (Figure 5-1) is not as well conserved as expected but remains hydrophobic in character. Only SSK1 has a tyrosine residue at this position, implying that SSK1 may follow a similar activation mechanism as CheY. In MCS4 and SLN1, a conserved phenylalanine occupies this position, while in ArcB and SKN7 there is a valine and isoleucine, respectively. It is worth noting that some of these regulatory domains, such as SLN1 and ArcB, are intermediates of multi-step phosphorelay systems, they are not directly involved in phosphorylation-dependent activation of downstream effectors.

SSK1-R2 as well as the MCS4 response regulator domain have a ~20 residues long insertion between α3 and β4. SLN1-R1 has a short loop insertion between α2 and β3. However, all of the insertions are located at the opposite surface from the response regulator domain active sites. Since apparently there are no long loop additions around the active sites, the binding interfaces of these response regulator domains are likely to be conserved in surface shape. This is not surprising since all of them can bind YPD1, presumably in a similar fashion.

5.3.2 Surface properties of yeast response regulator domains

The molecular surface of the globular yeast response regulator domain involving in binding YPD1 are likely to be centered around the active site, consisting of β_i/α_i (i=1,5) loops as well as the N-terminus of helix α_i (i=1,5). The electrostatic and hydrophobic molecular surfaces around the active sites of the yeast response regulator domains SLN1-R1, SSK1-R2, SKN7-R3, as well as bacterial CheY, are shown in Figure 5-2(B-E). All of the response regulator domains have acidic active sites due to three highly conserved aspartate/glutamate residues. The electrostatic surfaces surrounding the active sites are mostly neutral. There are no other large charged clusters around this area. In addition to these general shared features, there are significant differences among the electrostatic molecular surfaces of the yeast response regulator domains. For example, the molecular surface of SSK1-R2 is largely neutral and positive.

The distribution of hydrophobic patches around the active sites, however, is well conserved among yeast response regulator domains. The largest hydrophobic patch, centered around the conserved residues, Ile20 and P110, which are located at helix α1 and the loop between β5 and α5, respectively. Another smaller conserved hydrophobic patch is contributed by the loop that joins β3 to α3. Although there are no assigned roles for these conserved residues for SLN1-R1, SSK1-R2 and SKN7-R3, the corresponding residues in other homologous response regulator, such as Spo0F, are involved in inter-molecular interactions as assessed by biochemical as well as structural studies (Tzeng and Hoch, 1997; Zapf *et al.*, 2000).

5.3.3 Interactions between YPD1 and response regulators

In the modeled complex between YPD1 and CheY, the direction of the four-helix bundle of YPD1 is approximately perpendicular to the β-sheet of CheY (Figure 5-3). The phosphorylatable histidine (His64) and phosphorylatable (Asp57) were well aligned for the phosphotransfer reaction. The distance between His64 Nε of YPD1 and Asp57 Oδ1 of CheY is 5.0 Å. As expected, the β/α loops around the CheY active site (including loops 1, 3, 4 and 5) are in contact with YPD1. The surface on YPD1 which contributes to the binding of CheY are helix A, C, and D. Helix C of YPD1 sits on top of the CheY active site with

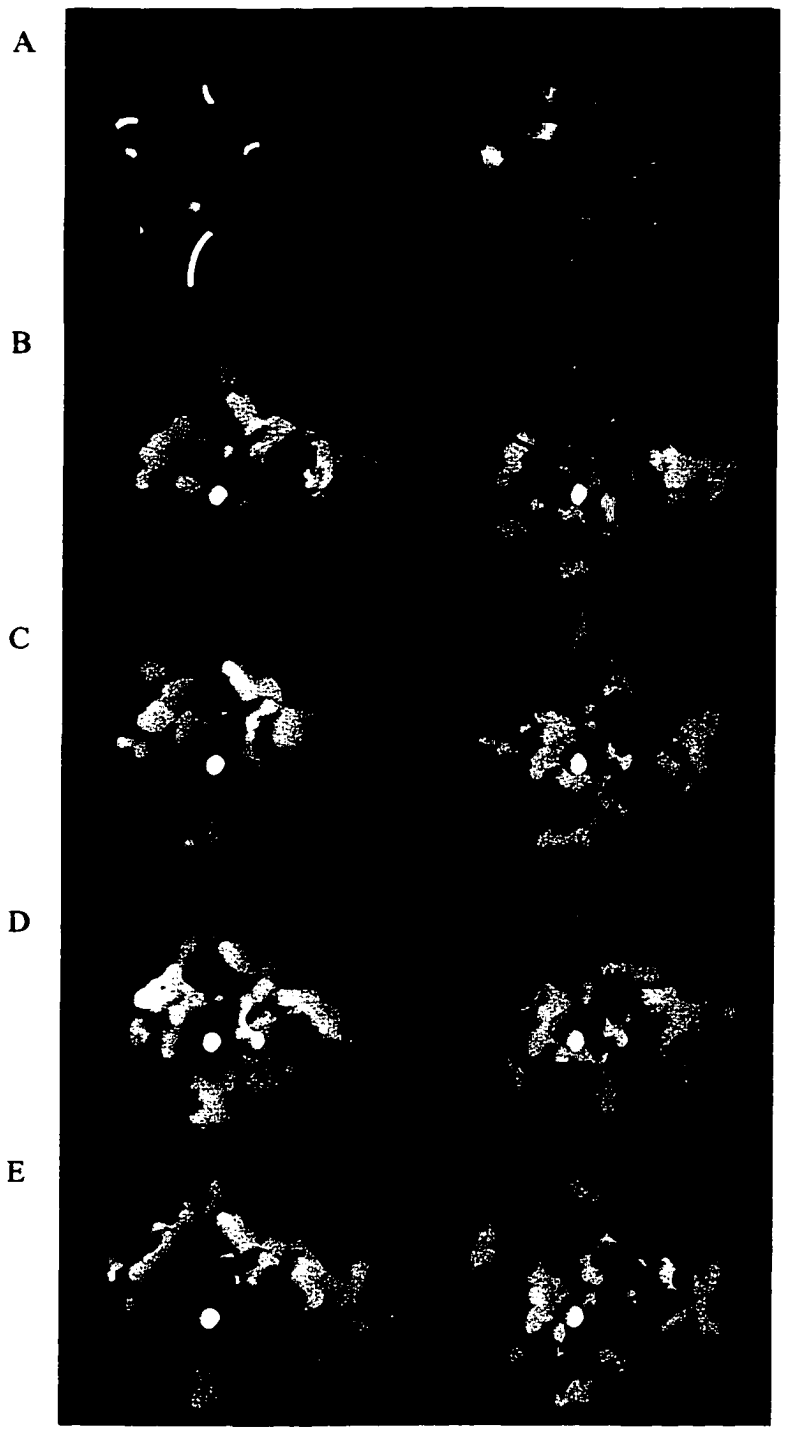


Figure 5-2. Electrostatic potential (left) and hydrophobic surfaces (right) of CheY, SLN1-R1, SSK1-R2 and SKN7-R3 models. Negatively charged surfaces are shown in red and positively charged surfaces are shown in blue. On figures shown to the right, the hydrophobic residues are shown in grey. **A.** (left) ribbon representation of CheY with some critical residues labelled, the molecular surfaces shown in B,C, D and E were generated using the same orientation; (right) mapping of sequence conservation among response regulator domains used in this study onto CheY secondary structure, with highly conserved positions shown in red; **B.** CheY; **C.** SLN1-R1; **D.** SSK1-R2; **E.** SKN7-R3.

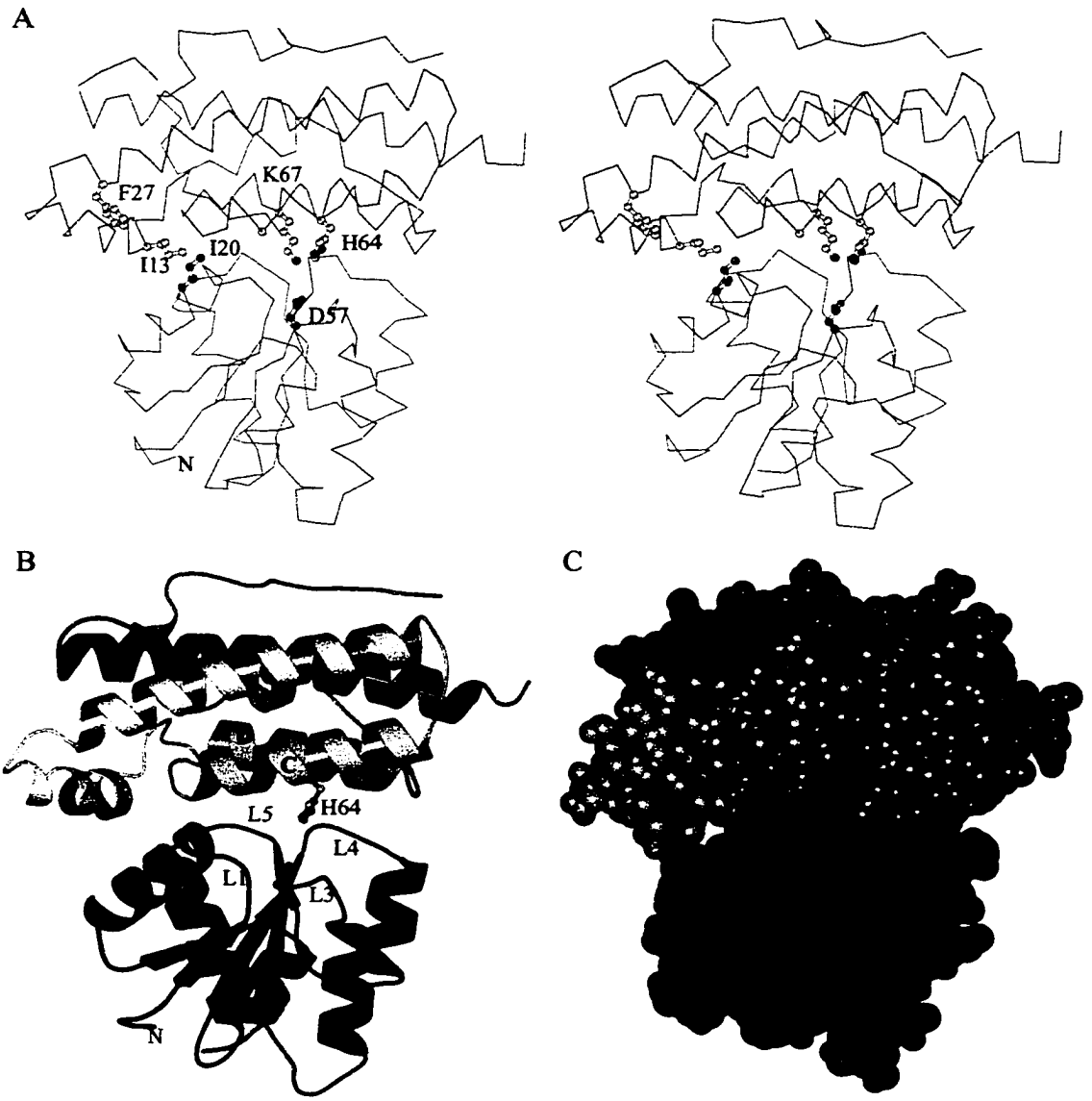


Figure 5-3. Hypothetical model of a complex between YPD1 and CheY. **A.** Stereoview of the Cα backbone of the complex (YPD1, red, CheY, cyan); **B.** Ribbon rendering of the complex, with His64 and Asp57 shown in ball and stick model (CheY, bottom, shown in blue and cyan; YPD1, top, shown in yellow and red); **C.** CPK representation of the model. The coloring scheme is the same as **B.**

loop 4 and 5 of CheY on one side and loops 1 and 3 on the other side. Helix D of YPD1 makes contact with loops 4 and 5 of CheY. Interestingly, the first helix (α1) of CheY is in close proximity to helix A of YPD1. The angle between these two helices is approximately 50 degrees. Hydrophobic interactions between these two helices is the most significant as compared to other areas of contact. The conserved hydrophobic patch centered at Ile20 of CheY on helix α1 and Pro110 on loop 5 (see previous section) are aligned with the hydro-

phobic surface of helix A consisting of Ile13, Ile17 and Met20. It is interesting that both of these hydrophobic surfaces are conserved (see Chapter 3 for hydrophobic surface conservation on HPt domains). Additionally, Gly68 and Ala71 of YPD1 lie close to the conserved Pro110 within the loop 5 of CheY.

Steric hindrance is observed between loop 4 of CheY and helix D of YPD1 within the docked complex constructed using the procedure described above. Compared to Spo0B, the distance between helix C and D of YPD1 is shorter than the distance between the antiparallel helices of Spo0B (α1, α2'; Figure 5-5B). As a result, the conformation of the loops on the response regulator which helix D from YPD1 are in contact with (especially loop 4) may need to be adjusted in order to accommodate the positioning of helix D in YPD1.

Shape complementary between YPD1 and a response regulator is thought to be important for their interaction (Xu and West, 1999). However, the docking procedure using YPD1 in place of $ArcB^c$ failed to generate a similar solution. Since the main difference between YPD1 and $ArcB^c$ is that $ArcB^c$ has a more protruding helix at the equivalent position of helix A of YPD1 and thus a more concave surface, it seems likely that positioning of helix A is important for achieving a complementary surface. The protruding helix can readily make contact with the CheY surface in $ArcB^c$. However, helix A of YPD1 is within 5-7 Å of the surface of CheY which is not close enough for interactions. The positioning of helix A in YPD1 generates a flat surface and thus it might be less restrictive in term of interaction with response regulator domains. Manual refinement of the docked complex with both YPD1 and CheY as rigid bodies suggested that the C-terminus of helix A and N-terminus of helix B can not make direct contact with CheY without encountering steric hindrance in other areas of the protein-protein interface. These discrepancies can be accommodated by conformational changes of the flexible loops associated with the response regulator domain binding interface; or a slightly different binding orientation between YPD1 and the response regulator domain; or the movement of helix A toward the conserved hydrophobic patch on the response regulator domain or a combination of any of the above.

It appears surprising that YPD1 can bind to response regulator domains in a similar fashion as Spo0B because Spo0B and YPD1 lack sequence homology at both a global and

local sequence level. However, it is important to note that the binding surface on Spo0F is well conserved in the response regulator domains SLN1-R1, SSK1-R2 and SKN7-R3 and CheY. The Cα r.m.s.d between Spo0F (PDB code 1F51) and CheY (PDB code 2CHE) is ~1.7 Å. Large deviations are located at the loops as well as amphipathic helices, for example, the N-terminus of α5 and loop 4. The active site aspartate residues in the two response regulators have very similar conformations (Figure 5-4). The main hydrophobic patch on Spo0F, which is involved in Spo0B binding, consists of hydrophobic residues (Ile15, Leu18 from helix α1; Pro105, Phe106 located at the loop between β5 and α5) which are highly conserved among other response regulator domains, including SLN1-R1, SSK1-R2 and SKN7-R3 and CheY (see previous section). It is therefore reasonable to expect a similar molecular recognition mechanism in other HPt-response regulator systems, as first proposed by Zapf *et al.* (2000).

Given the similarity of the hydrophobic recognition surfaces among response regulator domains, it was interesting to see whether a similar recognition mechanism is utilized by different HPt domains, such as YPD1 and Spo0B. Structural comparisons of YPD1 and $ArcB^c$ with Spo0B reveals similarity beyond a sequence level. As shown in Figure 5-5(B-C), the potential binding surfaces of YPD1 and $ArcB^c$ are superimposed onto helices α1 and α2' of Spo0B. The corresponding secondary structure elements from YPD1 include

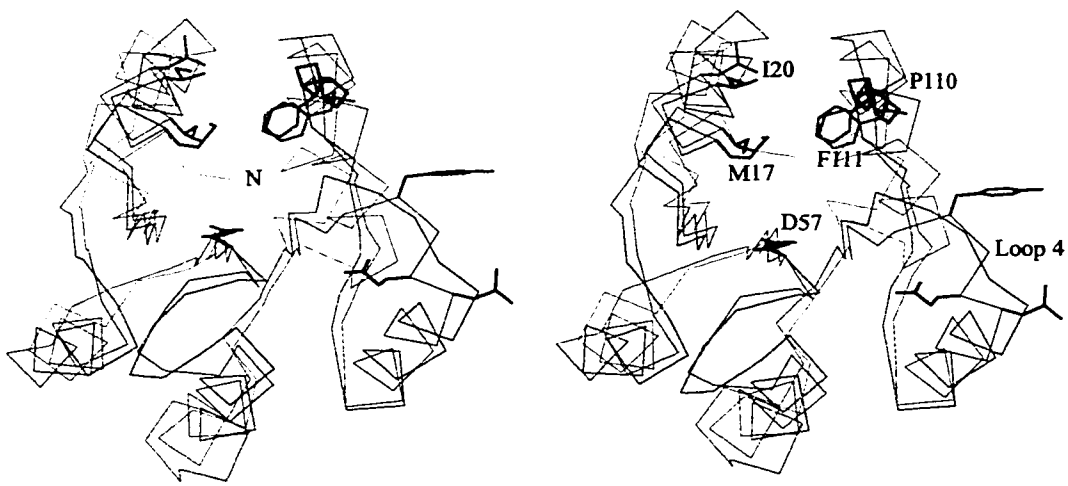
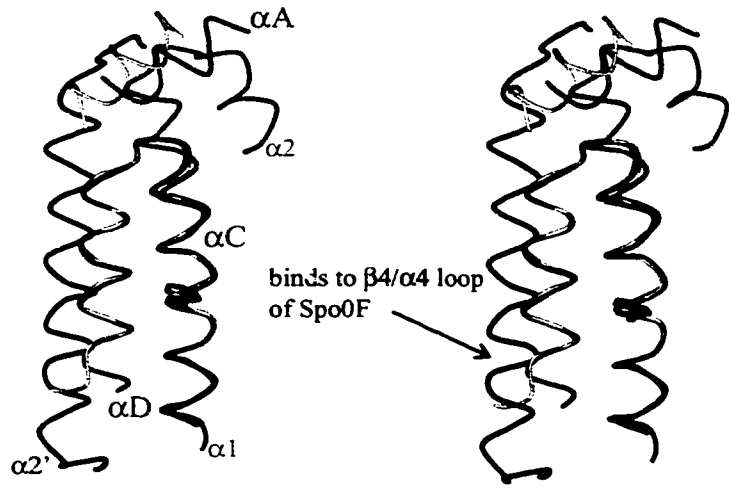


Figure 5-4. Superimposition of the Cα backbone of CheY (PDB code 2CHE, colored red) with Spo0F (PDB code 1F51, colored cyan). The conserved hydrophobic residues on helix α1 and loop 5 (β5/α5) are shown in stick representation. The residues from CheY are labelled.

A

	αA			αB		αC				
	13	17	20	27	31	68	69	72	73	74
YPD1	I	I	M	F	L	G	S	A	L	G
MPR1	V	L	M	F	I	G	S	A	V	G
CaYPD1p	V	I	M	F	L	G	S	V	L	G
ArcB	M	Y	V	L	G	G	A	S	V	G
	667	671	675	679	683	719	720	723	724	725

B



C

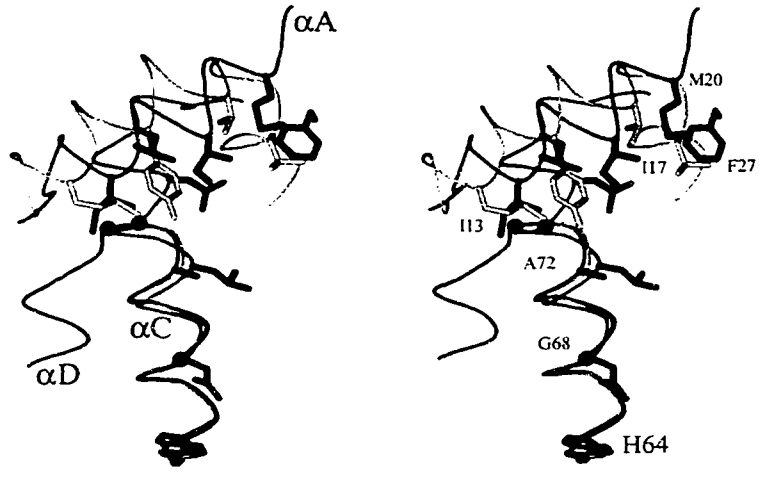


Figure 5-5. A. The potential binding interface on YPD1 is likely to consist of conserved hydrophobic residues (see Chapter 3 for details); **B.** Superimposition of the Spo0B binding interface with YPD1(red) and ArcB^{c} (yellow), shown in stereoview. The structural alignment was generated by superimposing a portion of the helices (Cα only) containing the phosphorylatable histidines, with the histidine positions as a reference. The histidines were in a similar location and side-chain conformation; helix D of YPD1 is parallel to the corresponding helix in Spo0B (shown in green). The portion of helix α2' in Spo0B which binds to loop 4 of Spo0F is highlighted in blue; helix A of YPD1 occupies a similar position as the C-terminal portion of α1 helix and the turn leading to α2 (cyan) of Spo0B. **C.** A close-up stereoview of the structural alignment around helices A and C of YPD1. The side-chains are shown in stick representation with the same color scheme as in **B**, YPD1, red; ArcB^{c}, yellow; Spo0B, cyan.

helix A, helices C and D, as well as the N-terminus of helix B, are all located on the same side as His64. The structural alignment was generated by aligning the helix with the phosphorylatable histidines as a reference point. First, it is apparent that both histidines adopt a very similar side-chain conformation. Moreover, helix D of YPD1 is parallel to the corresponding helix in Spo0B ($\alpha 2'$, shown in green).

In the structure of Spo0B, the C-terminus of the longer helix $\alpha 1$ provides a hydrophobic recognition site, in contact with a conserved hydrophobic patch on Spo0F (Zapf *et al.*, 2000). YPD1 and $ArcB^c$, however, have much shorter helices. It is possible than the extra support can come from helix A, which occupies a similar location as the C-terminus of helix $\alpha 1$ in Spo0B. In the Spo0B/Spo0F complex, a conserved hydrophobic residue Leu18 of Spo0F (equivalent to Ile20 of CheY) interacts with Gly41, Asn42 and Leu45 of Spo0B. Similar hydrophobic interactions are observed in the modeled complexes, in which conserved hydrophobic residues Ile13 from YPD1 interacts with Ile20 of CheY (Figure 5-5C). As a result of these arrangements, the overall hydrophobic recognition surfaces of Spo0B and YPD1 (or $ArcB^c$) are quite similar.

5.3.4 HPt domain specificity

In order for YPD1 to interact with more than one response regulator domain, each with distinctive surface features, the YPD1 binding surface needs to have two properties. First, the surface properties of YPD1 must be general enough to bind all three response regulator domains. Based on analysis of these modeled complexes and the known structure of the Spo0B/Spo0F complex, we hypothesize that the conserved hydrophobic patches on both HPts and response regulator domains are essential for molecular recognition. Additionally, the arrangement of side-chains around the histidine phosphorylation site creates a flat binding surface with a solvent exposed histidine. These properties of the YPD1 binding surface may be sufficient to allow YPD1 to bind multiple response regulators and provide accessibility of His64 on YPD1 to the active sites of response regulator domains. Second, YPD1 may need to have higher specificity for one response regulator over the other. The question that remains is: which residues contribute to specificity and binding affinity?

In the yeast phosphorelay system, YPD1 functions at a branch point between two phosphorelay signalling pathways (SLN1-YPD1-SSK1 or SLN1-YPD1-SKN7) and medi-

ates phosphoryl group transfer to either SSK1 or SKN7. Since there are no other response regulator proteins in yeast, YPD1 is unlikely to affect other pathways due to cross-talk. On the other hand, a promiscuous HPt domain like YPD1 or $ArcB^c$ may be deleterious in organisms with many two-component pathways since cross-talk between pathways may lead to fatal consequences. One possible way of avoiding cross-talk may be to restrict the freedom and accessibility of non-specific HPt domains, for instance, by incorporating them into a multidomain sensor kinase or having them exist at low copy number inside the cell. More importantly, the specificity between a HPt domain and its cognate response regulator domain is likely due to their unique structural properties. Chimeric studies with the EvgS and BvgS hybrid sensor kinases, which are architecturally similar to ArcB, suggest that there is specificity between the HPt domain and its cognate response regulator domain (Perraud *et al.*, 1998; Perraud *et al.*, 1999). Sequence alignment between the putative binding interface for YPD1 and BvgS/EvgS reveals significant differences (see Chapter 2). For example, the highly conserved lysine residue (Lys67 of YPD1) near the phosphorylation site is an alanine residue in BvgS. Additionally, the hydrophobic reverse turn conserved in other HPt domains is not as well conserved in BvgS and EvgS. Overall, it seems reasonable to conclude that limited sequence similarity amongst HPt domains can generate differences between surface properties and thus binding properties of HPt domains.

Based on the structure of the Spo0B/Spo0F complex, Zapf *et al.* (2000) suggested that the specificity of interactions is due to nonconserved residues especially those from the β4–α4 loop (loop 4). This loop of Spo0F interacts with a segment of the second helix (residues 263-270) of the other Spo0B monomer (α2') in the Spo0B/Spo0F complex. Interestingly, this segment of α2' maps to the C-terminus (residue 86-92) of helix D of YPD1 (shown in Figure 5-5B). Several charged residues, such as Lys67, Glu83 and Arg90, are located in this area. In the docked complex, Lys67 can form a hydrogen bond with a main-chain oxygen atom in loop 4 of CheY. On the other hand, the sequence variability within β4–α4 loop amongst selected response regulator domains is not very significant (Figure 5-1). As a result, this may not be the only region which contributes to specificity within the system of YPD1 and its associated response regulator domains. Other localized specific interactions might be important as well.

5.4 Conclusions

The modeling studies, based on sequence analysis, docking and structural comparisons, strongly suggest the notion that molecular recognition between HPts and response regulator domains is based on a common mechanism. Protein-protein recognition is likely to be based on the interactions between conserved hydrophobic patches, as well as shape complementary. Several hydrophobic surfaces on response regulator domains and HPts are conserved, even when there is no apparent sequence homology. Although members of the HPt family show much greater sequence variability than the response regulator domain superfamily, it appears that patterns and locations of interacting hydrophobic patches might be very important in the HPt family. As a result, the HPts may bind response regulator domains in a similar manner which in turn suggests a common phosphotransfer mechanism.

The conclusions here are consistent with currently available experimental data, they can be verified by further biochemical experiments such site-directed mutagenesis, and ideally, a structure of a complex between YPD1 and a cognate response regulator. The computer generated models also provide a structural basis for designing further experiments. Additionaliy, these modeled complexes are useful as starting models for molecular replacement once crystals of complexes between YPD1 and regulatory domains are obtained.

6 Conclusions and General Discussion

The crystal structure of YPD1 from *Saccharomyces cerevisiae* was the first structure determined of a eukaryotic homologue of a two-component phosphorelay protein. The structure revealed evolutionary conservation of the tertiary fold which consists of a four-helical bundle core. The phosphorylation site of YPD1, His64, is located on the surface of the four-helix bundle and is highly solvent exposed. The surrounding residues contribute to the exposure and stability of the phosphorylation site histidine. As a result, the arrangement of the phosphorylation site creates an optimal configuration for phosphotransfer. Structures of several YPD1 mutants, G68Q, Q86A, Q86L and Q86E, were subsequently analyzed in order to gain a better understanding of the roles of these conserved residues in the function of YPD1. Structural interpretations of YPD1 and other HPt domain structures have led to hypotheses concerning roles of important conserved residues.

Structural analysis and amino acid sequence comparisons of YPD1 to other HPt domains have helped to define common characteristics of the HPt family. Based on the analysis of the YPD1 structure and other HPt domains, it is possible that the four-helix bundle fold, particularly the "H-unit" structural motif with the centrally located exposed histidine residue, will be an important conserved feature of other HPt domains that interact with CheY-like response regulators. Furthermore, a conserved hydrophobic patch on the YPD1 surface was identified as a potential recognition site for YPD1 to bind response regulator domains.

The 2.1 Å crystal structure of the CheA P1 domain was determined recently (Mourey *et al.*, 2001). The crystal structure of the CheA P1 domain corroborates well with the analysis and conclusions made in this dissertation. As expected, the four-helix bundle core is well conserved between CheA P1 domain and YPD1 (as well as $ArcB^{C}$) with r.m.s.d of 1.2 Å for 73 common Cα atoms (Mourey *et al.*, 2001; 1.5 Å for $ArcB^{C}$). More strikingly,

the anti-parallel α-hairpin (H-unit as defined in Chapter 2) is highly conserved between the CheA P1 domain and YPD1 (as well as $ArcB^c$) with Cα r.m.s.d between YPD1 (residues 60-88) and CheA (residues 44-72) of 0.77 Å (0.74 Å for main chain atoms). The phosphorylatable histidine of CheA P1 domain (His48) has the same conformation as that of YPD1 and $ArcB^c$. The hydrogen bond between His48 and Glu70, as correctly predicted by sequence alignment in Figure 2-13, is equivalent to the hydrogen bond between His64 and Gln86 in YPD1. Interestingly, if we mutate Gln86 of YPD1 to a glutamate (the corresponding residue in CheA is Glu70), the resulting conformations of Lys67 and Glu86, as described in the crystal structure of Q86E in Chapter 4, are similar to what were observed in the crystal structure of the CheA P1 domain (Lys51 and Glu70, respectively).

Based on the high resolution structures of several HPt proteins such as $ArcB^c$, CheA P1 domain and Spo0B (Kato *et al.*, 1997; Mourey *et al.*, 2001; Song *et al.*, 1999; Xu and West, 1999; Varughese *et al.*, 1998; Zapf *et al.*, 2000), the H-unit appears to be highly conserved. As shown in Figure 6-1, the YPD1 H-unit is superimposed onto that of $ArcB^c$, CheA P1 domain and Spo0B. First, the structural arrangement of the two anti-parallel helices is conserved. In the case of YPD1 and $ArcB^c$, or the CheA P1 domain in which there is sequence homology within the H-unit (Figure 2-13), the H-units are almost identical.

Figure 6-1. Structural conservation of H-unit (defined in Chapter 2). YPD1 H-unit (red) is superimposed onto $ArcB^c$ (green), CheA P1 domain (yellow) as well as Spo0B (cyan and green), respectively. The phosphorylatable histidines and other conserved residues are shown as stick model and CPK spheres in case of glycines.

Despite lack of obvious sequence homology between YPD1 and Spo0B, one main difference between their respective H-units is the distance between the two anti-parallel helices. Second, it is obvious that conformations of all these phosphorylatable histidines are very similar. These observations suggest that a common mechanism is utilized for the His-Asp phosphotransfer. In other words, the transition state geometry and the conformations of those important residues associated with the transition state, especially the histidine and aspartate residues that are directly involved in the phosphorylation reaction, are conserved. As a result, the complex between a HPt domain and a response regulator domain is likely to be structurally conserved. In Chapter 5, we have shown that this is possible with the modeling study of the YPD1 and CheY complex.

If interactions between HPt domains and response regulators are conserved, then the knowledge gained in one system could be applied in other homologous systems. For example, one immediate application of the above analysis is that the complex between the CheA P1 domain, P2 domain (CheY binding domain; Welch *et al.*, 1998; McEvoy *et al.*, 1998) and CheY can readily be modeled. Specifically, the computational complex for the CheA P1 domain and P2 domain and CheY can be generated in two steps. First, the CheA P1 domain and CheY complex (P1/CheY) can be built upon the hypothetical complex of YPD1/CheY (described in Chapter 5) by replacing YPD1 with CheA P1 domain which can be done by superimposing the H-unit of YPD1 and CheA P1 domain. Second, the P2/P1/CheY complex can be built by superimposing the common part (CheY), which is present in the P1/CheY computer complex as well as the experimental P2/CheY complex (PDB code 1EAY). The resulting ternary complex is shown in Figure 6-2A. The linker region (residues 132-158) between the P1 and P2 domain is missing. The distance between the C-terminal residue of the P1 domain (residue 131) and the N-terminal residue of the P2 domain (residue 159) is about 17 Å. This distance could be easily accomondated by 27 residues (residues 132-158), for example, via a loop containing a short helix. The model of the ternary complex suggests that the binding of CheY to CheA could be a two-step process (hinge-lock model). The initial binding of the CheA P2 domain to CheY helps to restrict the motion of the P1 domain and put it in a more or less correct orientation towards CheY (*analogy*: control the movement of a door with a "hinge"). The second binding site, con-

A

B

YPD1	I13	I17	M20	F27	L31	G68	S69	A72	L73	G74
ArcB	M667	Y671	V675	L679	G683	G719	A720	S723	V724	G725
CheA_P1	—	I5		M3	F8	G52	G53	T56	F57	G58

Figure 6-2. A. Computer model for the complex between P1 (yellow) and P2 domains (cyan) of CheA with CheY (red) rendered in ribbon (left) and cpk (right) representations. The missing link (residues 132-158) between the P1 and the P2 domain is shown as red dotted line. The distance between the C-terminal of the P1 domain (residue 131) and the N-terminal residue of the P2 domain (residue 159) is about 17 Å according the modeled complex. **B.** Sequence alignment (based on structural comparison) of the residues in the conserved hydrophobic patch (see Chapter 5 for details) between the CheA P1 domain and YPD1 and ArcB^c.

sisting of the complementary conserved hydrophobic patches on the P1 domain as well as CheY (located at the diagonal end of the P1-P2 linker as shown in Figure 6-2A; also see Zapf *et al.* (2000) and Chapter 5), might be involved in orienting the P1 domain and CheY optimally and sealing their interfaces (*analogy*: close the door with a "lock"). The corresponding conserved hydrophobic patch on the CheA P1 domain is located at the N-terminal region of the first helix (Met3, Ile5, Phe8, Phe12) as well as the C-terminal portion of the second helix (Gly52, Gly53, Thr56, Phe57, Gly58). The sequence alignment of these residues with corresponding residues of YPD1 and ArcB^c is shown in Figure 6-2B. Since the phosphotransfer between CheY and the P1 domain of CheA can still occur without the presence of the P2 domain, the role of the P2 domain is thus likely to increase the efficiency and fidelity of the phosphotransfer between CheA and CheY (Stewart *et al.*, 2000). There-

fore, the CheA/CheY system might represent an example of how affinity could be achieved for a promiscuous HPt domain.

Although the conservation of structure and function might be the underlying principle of the two-component signaling systems, it is clear that each system has its own unique properties. It is because of these unique properties that two-component systems can be utilized for adaptation to a wide variety of responses. Therefore, it is important to study each individual system in order to find out what the differences are between two-component pathways and to understand how the unique properties of each pathway affect its function.

References

[1] Abrahams, J.P. & Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. sect. D,* **52**, 30-42.

[2] Akerley, B.J., Monack, D., Falkow, S. & Miller, J.F. (1992). The bvgAS locus negatively controls motility and synthesis of flagella in *Bordetella bronchiseptica*. *J Bacteriol.* **174**, 980-90.

[3] Akerley, B.J. & Miller, J.F. (1996). Understanding signal transduction during bacterial infection. *Trends Microbiol.* **4**(4), 141-6.

[4] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J.D. (1994) *Molecular Biology of the Cell*. Garland, New York.

[5] Albertyn, J., Hohmann, S. & Prior, B.A. (1994a). Characterization of the osmotic-stress response in *Saccharomyces cerevisiae*: Osmotic stress and glucose repression regulate glycerol-3-phosphate dehydrogenase independently. *Curr. Genet.* **25**, 12-18.

[6] Albertyn, J., Hohmann, S., Thevelein, J.M. & Prior, B.A. (1994b). *GPD1*, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in *Saccharomyces cerevisiae*, and its expression is regulated by the high-osmolarity glycerol response pathway. *Mol. Cell. Biol.* **14**, 4135-4144.

[7] Alex, L., Borkovich, K. & Simon, M.I. (1996). Hyphal development in *Neurospora crassa*, Involvement of a two-component histidine kinase. *Proc Natl Acad Sci USA*, **93**, 9416-21.

[8] Alex, L.A. & Simon, M.I. (1994). Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. *Trends Genet.* **10**(4), 133-8.

[9] Alex, L.A., Korch, C., Selitrennikoff, C.P. & Simon, M.I. (1998). COS1, a two-component histidine kinase that is involved in hyphal development in the opportunistic pathogen *Candida albicans*. *Proc Natl Acad Sci USA*. **95**(12), 7069-73.

[10] Anthony, N., Sharp, K.A. & Honig, B. (1991). Protein folding and association, insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins. Struct. Funct. Genet.* **11**, 281-296.

[11] Aoyama, K., Mitsubayashi, Y., Aiba, H. & Mizuno, T. (2000) Spy1, a histidine-containing phosphotransfer signaling protein, regulates the fission yeast cell cycle through the Mcs4 response regulator. *J Bacteriol.* **182**(17), 4868-74.

[12] Appleby, J.L., Parkinson, J.S. & Bourret, R.B. (1996). Signal transduction via the multi-step phosphorelay: not necessarily a road less traveled. *Cell* **86**, 845-848.

[13] Arthur, M., Molinas, C. & Courvalin, P. (1992). The VanS-VanR two-component regulatory system controls synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. *J Bacteriol.* **174**, 2582-91.

[14] Bairoch A. & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45-48.

[15] Ban, C. & Yang, W. (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* **95**(4), 541-52.

[16] Barford D. (1991). Molecular mechanisms for the control of enzymic activity by protein phosphorylation. *Biochim Biophys Acta.* 1133(1), 55-62.

[17] Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R.P. & Dickerson, R.E. (1996). Structure of the *Escherichia coli* response regulator NarL. *Biochemistry* **35**(34), 11053-61.

[18] Barrett, J.F. & Hoch J.A. (1998). Two-component signal transduction as a target for microbial anti-infective therapy. *Antimicrob Agents Chemother*, **42**, 1529-36.

[19] Barrett, J.F., Goldschmidt, R.M., Lawrence, L.E., Foleno, B., Chen, R., Demers, J.P., Johnson, S., Kanojia, R., Fernandez, J., Bernstein, J., Licata, L., Donetz, A., Huang, S., Hlasta, D.J., Macielag, M.J., Ohemeng, K., Frechette, R., Frosco, M.B., Klaubert, D.H., Whiteley, J.M., Wang, L. & Hoch, J.A. (1998). Antibacterial agents that inhibit two-component signal transduction systems. *Proc Natl Acad Sci USA.* **95**(9), 5317-22.

[20] Bartel, P.L. & Fields, S. (1997). *The Yeast Two-Hybrid System*. Oxford University Press.

[21] Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. & Sonnhammer, E.L.L. (2000). The Pfam protein families database. *Nucleic Acids Research.* **28**, 263-266.

[22] Birck, C., Mourey, L., Gouet, P., Fabry, B., Schumacher, J., Rousseau, P., Kahn, D. & Samama, J.P. (1999). Conformational changes induced by phosphorylation of the FixJ receiver domain. *Structure Fold Des.* 7(12), 1505-15.

[23] Blomberg, A. & Adler, L. (1989). Roles of glycerol and glycerol-3-phosphate dehydrogenase (NAD^+) in acquired osmotolerance of *Saccharomyces cerevisiae. J. Bacteriol.* **171**, 1087-1092.

[24] Bower, M.J., Cohen, F.E., & Dunbrack, Jr. R.L. (1997). Prediction of protein sidechain conformations from a backbone-dependent rotamer library, A new homology modeling tool for Proteins, Application to Side-chain prediction. *J. Mol. Biol.* **267**, 1268-1282.

[25] Branden, C. & Tooze, J. (1991). *Introduction to protein structure.* New York, Garland Publishing.

[26] Brewster, J.L., de Valoir, T., Dwyer, N.D., Winter, E. & Gustin, M.C. (1993). An osmosensing signal transduction pathway in yeast. *Science* **259**, 1760-1763.

[27] Brown, J.L., Bussey, H. & Stewart, R.C. (1994). Yeast Skn7p functions in a eukaryotic two-component regulatory pathway. *EMBO J.* **13**, 5186-5194.

[28] Brown, J.L., North, S. & Bussey, H. (1993). *SKN7*, a yeast multicopy suppressor of a mutation affecting cell wall β-glucan assembly, encodes a product with domains homologous to prokaryotic two-component regulators and to heat shock transcription factors. *J. Bacteriol.* **175**, 6908-6915.

[29] Brown, J.M. & Firtel, R.A. (1998). Phosphorelay signalling: New tricks for an ancient pathway. *Curr. Biol.* **8**, R662-R665.

[30] Brünger, A.T., Adams, P.D., Clore, G.M., Delano,W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998). Crystallography & NMR System, A New Software Suite for Macromolecular Structure Determination. *Acta Cryst. D***54**, 905-921

[31] Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-460.

[32] Burbulys, D., Trach, K.A. & Hoch, J.A. (1991). Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell* **64**(3), 545-52.

[33] Calera, J.A., Choi, G.H. & Calderone, R.A. (1998). Identification of a putative histidine kinase two-component phosphorelay gene (CaHK1) in *Candida albicans*. *Yeast* **14**, 665-74.

[34] Calera, J.A., Herman, D., Calderone, R. (2000) Identification of YPD1, a gene of *Candida albicans* which encodes a two-component phosphohistidine intermediate protein. *Yeast* **16**(11), 1053-9.

[35] Chang, L. & Karin, M. (2001). Mammalian MAP kinase signalling cascades. *Nature* **410**, 37-40.

[36] Chang, C. & Stewart, R.C. (1998). The two-component system, regulation of diverse signaling pathways in prokaryotes and eukaryotes. *Plant Physiol.* **117**,723-731.

[37] Chang, C., Kwok, S.F., Bleecker, A.B. & Meyerowitz, E.M. (1993). *Arabidopsis* ethylene-response gene ETR1, similarity of product to two-component regulators. *Science* **262**(5133), 539-44.

[38] Chang, W.-T., Thomason, P.A., Gross, J.D. & Newell, P.C. (1998). Evidence that the RdeA protein is a component of a multistep phosphorelay modulating rate of development in *Dictyostelium*. *EMBO J.* **17**, 2809-2816.

[39] Cherry, J.M., Adler, C., Ball, C., Chervitz, S.A., Dwight, S.S., Hester, E.T., Jia, Y., Juvik, B., Roe, T., Schroeder, M., Weng, S. & Botstein, D. (1998). SGD, *Saccharomyces* Genome Database. *Nuc. Acids Res.* **26**, 73-80.

[40] Cherry, J.M., Ball, C., Weng, S., Juvik, G, Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R.K. & Botstein, D. (1997). Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* **387** (supplement), 67-73.

[41] Chiang, G.G, Schaefer, M.R. & Grossman, A.R. (1992). Complementation of a red-light-indifferent cyanobacterial mutant. *Proc Natl Acad Sci USA.* **89**(20), 9415-9.

[42] Cho, H.S., Lee, S.Y., Yan, D., Pan, X., Parkinson, J.S., Kustu, S., Wemmer, D.E., & Pelton, J.G. (2000). NMR structure of activated CheY. *J Mol Biol.* **297**(3), 543-51.

[43] Cottarel, G. (1997). Mcs4, a two-component system response regulator homologue, regulates the Schizosaccharomyces pombe cell cycle control. *Genetics* **147**(3), 1043-51.

[44] Cowtan, K.D. (1994). "DM", an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM newsletter on protein crystallography* **31**, 34-38.

[45] de la Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 472-494.

[46] Desnottes, J.F. (1996). New targets and strategies for the development of antibacterial agents. *Trends Biotechnol.* **14**, 134-40.

[47] DiRita, V.J. (1992). Co-ordinate expression of virulence genes by toxR in *Vibrio cholera. Mol Microbiol.* **6**, 451-8.

[48] Djordjevic, S., Goudreau, P.N., Xu, Q., Stock, A.M. & West, A.H. (1998) Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. *Proc Natl Acad Sci USA.* **95**(4), 1381-6.

[49] Djordjevic, S. & Stock, A.M. (1998). Structural analysis of bacterial chemotaxis proteins, components of a dynamic signaling system. *J Struct Biol.* **124**(2-3), 189-200.

[50] Dziejman, M. & Mekalanos, J.J. (1995). Two-component signal transduction and its role in the expression of bacterial virulence factors. In *Two-Component Signal Transduction*. J.A. Hoch and T.J. Silhavy (Eds.), ASM Press, Washington, D.C., 305-7.

[51] Egger, L.A., Park, H. & Inouye, M. (1997). Signal transduction via the histidyl-aspartyl phosphorelay. *Genes Cells* **2**, 167-84.

[52] Esnouf, R.M. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 133-138.

[53] Evans, P.R. (1993). Data reduction. In *Proceedings of the CCP4 Study Weekend, Data Collection and Processing.* (Sawyer, L., Isaacs, N. & Bailey, S., eds), SERC Daresbury Laboratory, England.

[54] Evans, S.V. (1993). SETOR, hardware lighted three-dimensional solid model representation of macromolecules. *J. Mol. Graph.* **11**, 134-138.

[55] Fabret, C. & Hoch, J.A. (1998). A two-component signal transduction system essential for growth of *Bacillus subtilis*, Implications for anti-infective therapy. *J Bacteriol.* **180**, 6375-83.

[56] Falke, J.J., Bass, R.B., Butler, S.L., Chervitz, S.A. & Danielson MA. (1997). The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu Rev Cell Dev Biol.* **13**, 457-512.

[57] Falke, J.J. & Hazelbauer, G.L. (2001) Transmembrane signaling in bacterial chemoreceptors. *Trends Biochem Sci.* **26**(4), 257-265.

[58] Fassler, J.S., Gray, W.M., Malone, C.L., Tao, W., Lin, H. & Deschenes, R.J. (1997). Activated alleles of yeast *SLN1* increase Mcm1-dependent reporter gene expression and diminish signaling through the Hog1 osmosensing pathway. *J. Biol. Chem.* **272**, 13365-13371.

[59] Forst, S.A. & Roberts, D.L. (1994). Signal transduction by the EnvZ-OmpR phosphotransfer system in bacteria. *Res Microbiol.* **145**, 363-73.

[60] Frosco M. & Barrett, J.F. (1999). Bacterial two-component systems as antimicrobial drug discovery targets. *Drug News & Perspectives.* **12**(5), 293-299.

[61] Gabb, H.A., Jackson, R.M. & Sternberg, M.J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol.* **272**(1), 106-20.

[62] Georgellis, D., Kwon, O., De Wulf P. & Lin, E.C. (1998). Signal decay through a reverse phosphorelay in the Arc two-component signal transduction system. *J Biol Chem.* **273**(49), 32864-9.

[63] Grebe, T. & Stock, J.B. (1999) The histidine kinase superfamily. *Adv. Microbial Phys.* **41**, 139-227.

[64] Grossman, A.D. (1995). Genetic Networks Controlling the Initiation of Sporulation and the Development of Competence in *Bacillus subtilis*. *Annu. Rev. Genetics.* **29**, 477-508.

[65] Guenzi, E., Gasc, A.M., Sicard, M.A. & Hakenbeck, R. (1994). A two-component signal-transducing system is involved in competence and penicillin susceptibility in laboratory mutants of *Streptococcus pneumoniae*. *Mol Microbiol.* **12**(3), 505-15.

[66] Gunn, J.S., Lim, K.B., Krueger, J., Kim, K., Guo, L., Hackett, M. & Miller, S.I. (1998). PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Mol Microbiol.* **27**(6), 1171-82.

[67] Head, C.G., Tardy, A. & Kenney, L.J. (1998). Relative binding affinities of OmpR and OmpR-phosphate at the ompF and ompC regulatory sites. *J Mol Biol.* **281**(5), 857-70.

[68] Hoch, J.A. (2000). Two-component and phosphorelay signal transduction. *Curr Opin Microbiol.* **3**(2), 165-70.

[69] Hoch, J.A. & Silhavy, T.J., Eds. (1995). *Two-component signal transduction*. American Society for Microbiology Press, Washington, D.C.

[70] Hoch, J.A. (1995). Control of cellular development in sporulating bacteria by the phosphorelay two-component signal transduction system. In *Two-Component Signal Transduction*. J.A. Hoch and T.J. Silhavy (Eds.), ASM Press, Washington, D.C., 129-44.

[71] Holm, L. & Sander, C. (1997). Dali/FSSP classification of three-dimensional protein folds. *Nucl. Acids Res.* **25**, 231-234.

[72] Hubbard, S.T. & Thornton, J.M. (1993). 'NACCESS', Computer program, Dept. of Biochem. and Mol. Biol. University College London.

[73] Hutchinson, E.G. & Thornton, J.M. (1996). PROMOTIF-A program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212-220.

[74] Ikegami, T., Okada, T., Ohki, I., Hirayama, J., Mizuno, T. & Shirakawa, M. (2001). Solution structure and dynamic character of the histidine-containing phosphotransfer domain of anaerobic sensor kinase ArcB from *Escherichia coli*. *Biochemistry* **40**(2), 375-86.

[75] Inouye, M. (1996). His-Asp phosphorelay. Two-components or more? *Cell* **85**, 13-4.

[76] Iuchi, S. & Weiner, L. (1996). Cellular and molecular physiology of *Escherichia coli* in the adaptation to aerobic environments. *J Biochem (Tokyo)*. **120**(6), 1055-63.

[77] Jancarik, J. & Kim, S.-H. (1991). Sparse matrix sampling, a screening method for crystallization of proteins. *J. Appl. Cryst.*, **24**, 409-411.

[78] Janiak-Spens, F., Sparling, J.M., Gurfinkel, M. & West, A.H. (1999). Differential stabilities of phosphorylated response regulator domains reflect functional roles of the yeast osmoregulatory SLN1 and SSK1 proteins. *J. Bacteriol.* **181**, 411-417.

[79] Janiak-Spens, F., Sparling, D.P. & West, A.H. (2000). Novel role for an HPt domain in stabilizing the phosphorylated state of a response regulator domain. *J Bacteriol.* **182**(23), 6673-8.

[80] Janiak-Spens, F. & West, A.H. (2000). Functional roles of conserved amino acid residues surrounding the phosphorylatable histidine of the yeast phosphorelay protein YPD1. *Mol Microbiol.* **37**(1), 136-44.

[81] Johnson, L.N., and Barford, D. (1993). The effects of phosphorylation on structure and function of proteins. *Ann. Rev. Biophys. Biomol. Struct.* **22**, 199-232.

[82] Jones, S. & Thornton, J.M. (1996). Principles of protein-protein interactions. *Proc. Natl. Acad. Sci., USA.* **93**, 13-20.

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

[84] Katchalski-Katzir, E, Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C. & Vakser, I.A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci USA.* **89**(6), 2195-9.

[85] Kato, M., Mizuno, T., Shimizu, T. & Hakoshima, T. (1999). Refined structure of the histidine-containing phosphotransfer (HPt) domain of the anaerobic sensor kinase ArcB from *Escherichia coli* at 1.57Å resolution. *Acta Crystallogr D Biol Crystallogr.* **D55**(11), 1842-9.

[86] Kato, M., Mizuno, T., Shimizu, T. & Hakoshima, T. (1997). Insights into multistep phosphorelay from the crystal structure of the C-terminal HPt domain of ArcB. *Cell* **88**, 717-723.

[87] Ketela, T., Brown, J.L., Stewart, R.C. & Bussey, H. (1998). Yeast Skn7p activity is modulated by the Sln1p-Ypd1p osmosensor and contributes to regulation of the HOG1 pathway. *Mol. Gen. Genet.* **259**, 372-378.

[88] Kehoe, D.M., Grossman, A.R. (1997). New classes of mutants in complementary chromatic adaptation provide evidence for a novel four-step phosphorelay system. *J Bacteriol.* **179**(12), 3914-21.

[89] Kern, D., Volkman, B.F, Luginbuhl, P., Nohaile, M.J., Kustu, S. & Wemmer, D.E. (1999). Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature* **402**(6764), 894-8.

[90] Klein, C., Kaletta, C. & Entian, K.-D. (1993). Biosynthesis of the lantibiotic subtilin is regulated by a histidine kinase/response regulator System. *Appl Environ Micro.* **59**, 296-303.

[91] Kleywegt, G.J. & Jones, A.T. (1996). Efficient rebuilding of protein structures. *Acta Crystallogr. D***52**, 829-832.

[92] Kleywegt, G.J. & Jones, T.A. (1997). Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallogr. D***53**, 179-185.

[93] Kleywegt, G.J. & Jones, T.A. (1999). O and associated programs. In *International Tables for Crystallography, Vol. F,* (Arnold, E. and Rossmann, M.G., *eds*), Chapter 25.2.6., International Union of Crystallography and Kluwer Academic Publishers, Dordrecht/Boston/London, *In press.*

[94] Koehl, P. & Delarue, M. (1994). Application of a Self-consistent Mean Field Theory to Predict Protein Side-chains Conformation and Estimate Their Conformational Entropy. *J. Mol. Biol.* **239**, 249-275.

[95] Koehl, P. & Levitt, M. (1999). A brighter future for protein structure prediction. *Nat Struct Biol.* **6**(2), 108-11.

[96] Koehl, P. & Delarue, M. (1996). Mean-field minimization methods for biological macromolecules. *Curr Opin Struct Biol.* **6**(2), 222-6.

[97] Knowles, J.R. (1980). Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem.* **49**, 879-919.

[98] Kraulis, P.J. (1991). MOLSCRIPT, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

[99] Krems, B., Charizanis, C. & Entian, K.-D. (1996). The response regulator-like protein Pos9/Skn7 of *Saccharomyces cerevisiae* is involved in oxidative stress resistance. *Curr. Genet.* **29**, 327-334.

[100] Lai, M.H. and Kirsch, D.R. (1996). Induction signals for vancomycin resistance encoded by the Van A gene cluster in Enterococcus faecium. *Antimicrob Agents Chemother*, **40**, 1645-8.

[101] Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK, a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

[102] Lee, S.Y., Cho, H.S., Pelton, J.G., Yan, D., Henderson, R.K., King, D.S., Huang, L., Kustu, S., Berry, E.A. & Wemmer, D.E. (2001). Crystal structure of an activated response regulator bound to its target. *Nat Struct Biol.* **8**(1), 52-6.

[103] Levitt, M. (1992). Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol.* **226**(2), 507-33.

[104] Lewis, R.J., Brannigan, J.A., Muchova, K., Barak, I. & Wilkinson, A.J. (1999). Phosphorylated aspartate in the structure of a response regulator protein. *J Mol Biol.* **294**(1), 9-15.

[105] Li, S., Ault, A., Malone, C.L., Raitt, D., Dean, S., Johnston, L.H., Deschenes, R.J. & Fassler, J.S. (1998). The yeast histidine protein kinase, Sln1p, mediates phosphotransfer to two response regulators, Ssk1p and Skn7p. *EMBO J.* **17**, 6952-6962.

[106] Liao, D.-I., Silverton, E., Seok, Y.-J., Lee, B.R., Peterkofsky, A. & Davies, D.R. (1996). The first step in sugar transport, crystal structure of the amino terminal domain of enzyme I of the *E. coli* PEP:sugar phosphotransferase system and a model of the phosphotransfer complex with HPr. *Structure* **4**, 861-872.

[107] Luzzati, P.V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Cryst.* **5**, 802-810.

[108] Lynch, A.S. & Lin, E.C.C. (1996). Transcriptional control mediated by the ArcA two-component response regulator protein of *Escherichia coli*, Characterization of DNA binding at target promoter. *J Bacteriol.* **178**, 6238-49.

[109] Madhusudan, Zapf J., Whiteley, J.M., Hoch, J.A., Xuong, N.H. & Varughese, K.I. (1996). Crystal structure of a phosphatase-resistant mutant of sporulation response regulator Spo0F from *Bacillus subtilis*. *Structure* **4**(6), 679-90.

[110] Maeda, T., Takekawa, M. & Saito, H. (1995). Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor. *Science* **269**, 554-558.

[111] Maeda, T., Wurgler-Murphy, S.M. & Saito, H. (1994). A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**, 242-245.

[112] Mager, W. H. & Varela, J.C.S. (1993). Osmostress response of the yeast *Saccharomyces*. *Mol. Microbiol.* **10**, 253-258.

[113] Matsushika, A. & Mizuno, T. (1998). The structure and function of the histidine-containing phosphotransfer (HPt) signaling domain of the *Escherichia coli* ArcB sensor. *J Biochem (Tokyo)*. **124**(2), 440-5.

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

[115] McEvoy, M.M. & Dahlquist, F.W. (1997). Phosphohistidines in bacterial signaling. *Curr. Opin. Struct. Biol.* **7**, 793-797.

[116] McEvoy, M.M., Hausrath, A.C., Randolph, G.B., Remington, S.J. & Dahlquist, F.W. (1998). Two binding modes reveal flexibility in kinase/response regulator interactions in the bacterial chemotaxis pathway. *Proc Natl Acad Sci USA*. **95**(13), 7333-8.

[117] McRee, D.E. (1993). *Practical protein crystallography*, Academic Press, San Diego, CA.

[118] McPherson, A. (1998). *Crystallization of Biological Macromolecules*, Cold Spring Harbor Laboratory Press.

[119] Merritt, E.A. & Bacon, D.J. (1997). Raster3D: Photorealistic Molecular Graphics. Methods in Enzymology **277**, 505-524.

[120] Miyata, S., Urao, T., Yamaguchi-Shinozaki, K. & Shinozaki, K. (1998). Characterization of genes for two-component phosphorelay mediators with a single HPt domain in Arabidopsis thaliana. *FEBS Lett.* **437**(1-2), 11-4.

[121] Mizuno, T. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res.* **4**, 161-168.

[122] Mizuno, T. (1998). His-Asp phosphotransfer signal transduction. *J. Biochem* (Tokyo). **123**, 555-63.

[123] Mizuno, T., Wurtzel, E.T. & Inouye, M. (1982). Osmoregulation of gene expression. II. DNA sequence of the envZ gene of the ompB operon of *Escherichia coli* and characterization of its gene product. *J Biol Chem.* **257**, 13692- 8.

[124] Moréra, S., Chiadmi, M., LeBras, G., Lascu, I. & Janin, J. (1995). Mechanism of phosphate transfer by nucleoside diphosphate kinase, X-ray structures of the phosphohistidine intermediate of the enzymes from *Drosophila* and *Dictyostelium*. *Biochemistry* **34**, 11062-11070.

[125] Morgan, B.A., Banks, G.R., Toone, W.M., Raitt, D., Kuge, S. & Johnston, L.H. (1997). The Skn7 response regulator controls gene expression in the oxidative stress response of the budding yeast *Saccharomyces cerevisiae*. *EMBO J.* **16**, 1035-1044.

[126] Mourey, L., Da Re S., Pedelacq, J.D., Tolstykh, T., Faurie, C., Guillet, V., Stock, J.B. & Samama, J.P. (2001). Crystal structure of the CheA histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis. *J Biol Chem.* **276**(33), 31074-82.

[127] Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D***53**, 240-255.

[128] Murzin A.G., Brenner S.E., Hubbard T. & Chothia C. (1995). SCOP, a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.

[129] Nagahashi, S., Mio, T., Ono, N., Yamada-Okabe, T., Arisawa, M., Bussey, H. & Yamada-Okabe, H. (1998). Isolation of CaSLN1 and CaNIK1, the genes for osmosensing histidine kinase homologues, from the pathogenic fungus *Candida albicans*. *Microbiology* **144** (2), 425-32.

[130] Nguyen, A.N., Lee, A., Place, W. & Shiozaki, K. (2000) Multistep phosphorelay proteins transmit oxidative stress signals to the fission yeast stress-activated protein kinase. *Mol Biol Cell.* **11**(4), 1169-81.

[131] Ninfa, A.J. & Magasanik, B. (1986). Covalent modification of the glnG product, NRI, by the glnL product, NRII, regulates the transcription of the glnALG operon in *Escherichia coli*. *Proc Natl Acad Sci USA*. **83**, 5909-13.

[132] Nishida, E. & Gotoh, Y. (1993). The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends Biochem Sci.* **18**(4), 128-31.

[133] Nunn, R.S., Markovic-Housley, Z., Génovésio-Taverne, J.-C., Flükiger, K., Rizkallah, P.J., Jansonius, J.N., Schirmer, T. & Erni, B. (1996). Structure of the IIA domain of the mannose transporter from *Escherichia coli* at 1.7 Å resolution. *J. Mol. Biol.* **259**, 502-511.

[134] Ohta, N., Lane, T., Ninfa, E.G., Sommer, J.M. & Newton, A. (1992). A histidine protein kinase homologue required for regulation of bacterial cell division and differentiation. *Proc Natl Acad Sci USA.* **89**(21), 10297-301.

[135] Ota, I.M. & Varshavsky, A. (1993). A yeast protein similar to bacterial two-component regulators. *Science* **262**, 566-569.

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

[137] Parkinson, J.S. (1993). Signal transduction schemes of bacteria. *Cell*, **73**, 857-71.

[138] Perego, M. & Hoch, J.A. (1996). Protein aspartate phosphatases control the output of two- component signal transduction systems. *Trends Genet.* **12**, 97-101.

[139] Perraud, A.-L., Kimmel, B., Weiss, V. & Gross, R. (1998). Specificity of the BvgAS and EvgAS phosphorelay is mediated by the C-terminal HPt domains of the sensor proteins. *Mol. Microbiol.* **27**, 875-887.

[140] Perraud, A.-L., Weiss, V. & Gross, R. (1999). Signalling pathways in two-component phosphorelay systems. *Trends Microbiol.* **7**, 115-120.

[141] Pflugrath, J.W. (1999). The finer things in X-ray diffraction data collection. *Acta Cryst.* **D55**, 1718-1725.

[142] Posas, F. & Saito, H. (1998). Activation of the yeast SSK2 MAP kinase kinase kinase by the SSK1 two-component response regulator. *EMBO J.* **17**, 1385-1394.

[143] Posas, F., Wurgler-Murphy, S.M., Maeda, T., Witten, E.A., Thai, T.C. & Saito, H. (1996). Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SNL1-YPD1-SSK1 "two-component" osmosensor. *Cell* **86**, 865-875.

[144] Posas, F., Takekawa, M. & Saito, H. (1998). Signal transduction by MAP kinase cascades in budding yeast. *Curr Opin Microbiol.* **1**(2), 175-82.

[145] Pratt, L.A., Hsing, W., Gibson, K.E. & Silhavy, T.J. (1996). From acids to osmZ, multiple factors influence synthesis of the OmpF and OmpC porins in *Escherichia coli. Mol Microbiol.* **20**(5), 911-7.

[146] Pratt, L.A. & Silhavy, T.J. (1995). Porin regulon of *Escherichia coli.* In *Two-Component Signal Transduction.* J.A. Hoch and T.J. Silhavy (Eds.), ASM Press, Washington, D.C., 105-127.

[147] Prodromou, C., Roe, S.M., O'Brien, R., Ladbury, J.E., Piper, P.W. & Pearl, L.H. (1997). Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **90**(1), 65-75.

[148] Quon, K.C., Marczynski, G.T. & Shapiro, L. (1996). Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* **84**, 83-93.

[149] Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A***42**, 140-149.

[150] Rice, L.M. & Brünger, A.T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic refinement. *Proteins, Str., Funct., Genet.* **19**, 277-290.

[151] Robinson, M.J., Cobb, M.H. (1997). Mitogen-activated protein kinase pathways. *Curr Opin Cell Biol.* **9**(2), 180-6.

[152] Robinson, V.L., Buckler, D.R. & Stock, A.M. (2000). A tale of two components: a novel kinase and a regulatory switch. *Nat Struct Biol.* **7**(8), 626-33.

[153] Rossmann, M.G. & Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-31.

[154] Rost, B. (1996). PHD, predicting one-dimensional protein structure by profile based neural networks. *Meth Enzymol.* **266**, 525-539.

[155] Salyers, A.A., Shoemaker, N.B. and Stevens, A.M. (1995). Tetracycline regulation of conjugal transfer genes. In *Two-Component Signal Transduction*. J.A. Hoch and T.J. Silhavy (Eds.), ASM Press, Washington, D.C., 393-400.

[156] SERC (UK) Collaborative Computational Project, N. (1994). The CCP4 suite, Programs for protein crystallography. *Acta Crystallogr.* *D***50**, 760-763.

[157] Sheldrick, G.M., Dauter, Z., Wilson, K.S., Hope, H. & Sieker, L.C. (1993). The application of direct methods and patterson interpretation to high-resolution native protein data. *Acta Crystallogr.* *D***49**, 18-23.

[158] Shieh, J.C., Wilkinson, M.G., Buck, V., Morgan, B.A., Makino, K. & Millar, J.B. (1997). The Mcs4 response regulator coordinately controls the stress-activated Wak1-Wis1-Sty1 MAP kinase pathway and fission yeast cell cycle. *Genes Dev.* **11**(8), 1008-22.

[159] Shiozaki, K., Shiozaki, M. & Russell P. (1997). Mcs4 mitotic catastrophe suppressor regulates the fission yeast cell cycle through the Wik1-Wis1-Spc1 kinase cascade. *Mol Biol Cell.* **8**(3), 409-19.

[160] Singh, K.K. (2000). The *Saccharomyces cerevisiae* Sln1p-Ssk1p two-component system mediates response to oxidative stress and in an oxidant-specific fashion. *Free Radic Biol Med.* **29**(10), 1043-50.

[161] Sola, M., Gomis-Ruth, F.X., Serrano, L., Gonzalez, A. & Coll, M. (1999). Three-dimensional crystal structure of the transcription factor PhoB receiver domain. *J Mol Biol.* **285**(2), 675-87.

[162] Song, H.K., Lee, J.Y., Lee, M.G., Moon, J., Min, K., Yang, J.K. & Suh, S.W. (1999). Insights into eukaryotic multistep phosphorelay signal transduction revealed by the crystal structure of Ypd1p from *Saccharomyces cerevisiae*. *J Mol Biol.* **293**(4), 753-61.

[163] Sowdhamini, R., Burke, D.F, Huang, J-F, Mizuguchi, K. Nagarajaram, H.A., Srinivasan, N., Steward, R.E. & Blundell T.L. (1998). CAMPASS: A database of structurally aligned protein superfamilies. *Structure* **6**(9), 1087-94.

[164] Stewart, R.C., Jahreis, K. & Parkinson, J.S. (2000). Rapid phosphotransfer to CheY from a CheA protein lacking the CheY-binding domain. *Biochemistry.* **39**(43), 13157-65.

[165] Stock, A.M., Martinez-Hackert, E., Rasmussen, B.F., West, A.H., Stock, J.B., Ringe, D. & Petsko, G.A. (1993). Structure of the Mg^{2+}-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry* **32**, 13375-13380.

[166] Stock, A.M., Robinson, V.L., & Goudreau, P.N. (2000). Two-component signal transduction. *Ann. Rev. Biochem.* **69**, 183-215.

[167] Stock, A.M. & West, A.H. (2001). Response regulator proteins and their interactions with histidine protein kinases, in "*Histidine kinases in signal transduction*", (M. Inouye and R. Dutta, eds.), Manuscript accepted for publication.

[168] Stock, J.B., Surette, M.G., Levit, M. & Park, P. (1995). Two-component signal transduction systems, Structure-function relationships and mechanisms of catalysis. In *Two-component signal transduction* (Hoch, J.A. & Silhavy, T.J., eds.), American Society for Microbiology, Washington, D.C., 25- 51.

[169] Suzuki, T., Imamura, A., Ueguchi, C. & Mizuno, T. (1998) Histidine-containing phosphotransfer (HPt) signal transducers implicated in His-to-Asp phosphorelay in *Arabidopsis. Plant Cell Physiol.* **39**(12), 1258-68.

[170] Swanson, R.V., Alex, L.A. & Simon, M.I. (1994). Histidine and aspartate phosphorylation: Two-component systems and the limits of homology. *Trends Biol Sci.* **19**, 485-90.

[171] Tanaka, T., Saha, S.K., Tomomori, C., Ishima, R., Liu, D., Tong, K.I., Park, H., Dutta, R., Qin, L., Swindells, M.B., Yamazaki, T., Ono, A.M., Kainosho, M., Inouye, M. & Ikura, M. (1998). Structure NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ. *Nature* **396**(6706), 88-92.

[172] Tao, W., Deschenes, R.J. & Fassler, J.S. (1999). Intracellular glycerol levels modulate the activity of Sln1p, a *Saccharomyces cerevisiae* two-component regulator. *J Biol Chem.* **274**(1), 360-7.

[173] Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W, improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

[174] Tommassen, J., de Geus, P., Lugtenberg, B., Hackett, J. & Reeves, P. (1982). Regulation of the pho regulon of *Escherichia coli* K- 12. Cloning of the regulatory genes phoB and phoR and identification of their gene products. *J Mol Biol.* **157**, 265- 74.

[175] Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S.K., Zhu, Y., Ishima, R., Liu, D., Tong, K.I., Kurokawa, H., Qian, H., Inouye, M. & Ikura, M. (1999). Solution structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ. *Nature Struct Biol.* **6**(8), 729-34.

[176] Tong, L. & Rossmann, M.G. (1990). The locked rotation function. *Acta Cryst.* **A46**, 783-792.

[177] Tsuzuki, M., Ishege, K. & Mizuno, T. (1995). Phosphotransfer circuitry of the putative multi-signal transducer, ArcB, of *Escherichia coli*: *in vitro* studies with mutants. *Mol. Microbiol.* **18**, 953–962.

[178] Tzeng, Y.-L. & Hoch, J.A. (1997). Molecular recognition in signal transduction: The interaction surfaces of Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis. *J. Mol. Biol.* **272**, 200-212.

[179] Uhl, M.A. & Miller, J.F. (1996). Integration of multiple domains in a two-component sensor protein, the *Bordetella pertussis* BvgAS phosphorelay. *EMBO J.* **15**(5), 1028-36.

[180] Urao, T., Miyata, S., Yamaguchi-Shinozaki, K. & Shinozaki, K. (2000) Possible His to Asp phosphorelay signaling in an Arabidopsis two-component system. *FEBS Lett.* **478**(3), 227-32.

[181] van Montfort, R.L.M., Pijning, T., Kalk, K.H., Hangyi, I., Kouwijzer, M.L.C.E., Robillard, G.T. & Dijkstra, B.W. (1998). The structure of the *Escherichia coli* phosphotransferase $IIA^{mannitol}$ reveals a novel fold with two conformations of the active site. *Structure* **6**, 377-388.

[182] Varughese, K.I., Madusudan, J.Z., Zhou, X.Z., Whiteley, J.M. & Hoch, J.A. (1998). Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. *Mol. Cell* **2**, 485-493.

[183] Volz, K. (1995). Structural and functional conservation in response regulators. In *Two-Component Signal Transduction*. J.A. Hoch and T.J. Silhavy (Eds.), ASM Press, Washington, D.C., 53-64.

[184] Volz, K. (1993). Structural conservation in the CheY superfamily. *Biochemistry* **32**, 11741-53.

[185] Vriend, G. (1990) WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52-56.

[186] Walsh, C.T., Fisher, S.L., Park, I.S., Prahalad, M. & Wu, Z. (1996). Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chem Biol.* **3**(1), 21-8.

[187] Welch, M., Chinardet, N., Mourey, L., Birck, C. & Samama, J.P. (1998). Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. *Nat Struct Biol.* **5**(1), 25-9.

[188] West, A.H., Martinez-Hackert, E. & Stock, A.M. (1995). Crystal structure of the catalytic domain of the chemotaxis receptor methylesterase, CheB. *J Mol Biol.* **250**(2), 276-90.

[189] West, A.H. & Stock, A.M. (2001). Histidine kinases and response regulator proteins in two-component signalling systems. *Trends in Biochem. Sci.* **26**(6), 301-8.

[190] Westheimer, F.H. (1987). Why nature chose phosphates? *Science* **235**, 1173-1178.

[191] Wigley, D.B., Davies, G.J., Dodson, E.J., Maxwell, A. & Dodson, G. (1991). Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature* **351**(6328), 624-9.

[192] Witters, L.A. (1990). Protein phosphorylation and dephosphorylation. *Cur. Cell Bio.* **2**, 212-220.

[193] Woeste, K. & Kieber, J.J. (1998). The molecular basis of ethylene signalling in *Arabidopsis*. *Philos Trans R Soc Lond B Biol Sci.* **353**(1374), 1431-8.

[194] Wolodko, W.T., Fraser, M.E., James, M.N.G. & Bridger, W.A. (1994). The crystal structure of succinyl-CoA synthetase from *Escherichia coli* at 2.5 Å resolution. *J. Biol. Chem.* **269**, 10883-10890.

[195] Wurgler-Murphy, S.M. & Saito, H. (1997). Two-component signal transducers and MAPK cascades. *Trends Biochem. Sci.* **22**, 172-176.

[196] Xu, Q. & West, A.H. (1999). Conservation of Structure and Function Among Histidine-Containing Phosphotransfer (HPt) Domains as Revealed by the Crystal Structure of YPD1. *J. Mol. Biol.* **292**, 1039-1050.

[197] Xu, Q., Nguyen, V. & West, A.H. (1999). Purification, Crystallization, and Preliminary X-Ray Diffraction Analysis of the Yeast Phosphorelay Protein YPD1. *Acta Cryst.* D**55**, 291-293.

[198] Yaku, H., Kato, M., Hakoshima, T., Tsuzuki, M. & Mizuno, T. (1997). Interaction between the CheY response regulator and the histidine-containing phosphotransfer (HPt) domain of the ArcB sensory kinase in *Escherichia coli*. *FEBS Lett.* **408**(3), 337-40.

[199] Yan, D., Cho, H.S., Hastings, C.A., Igo, M.M., Lee, S.Y., Pelton, J.G., Stewart, V., Wemmer, D.E. & Kustu, S. (1999). Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. *Proc Natl Acad Sci USA.* **96**(26), 14789-94.

[200] Yu, G., Deschenes, R.J. & Fassler, J.S. (1995). The essential transcription factor, Mcm1, is a downstream target of Sln1, a yeast "two-component" regulator. *J. Biol. Chem.* **270**, 8739-8743.

[201] Zapf, J., Sen, U., Madhusudan, M., Hoch, J.A. & Varughese, K.I. (2000). A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure Fold Des.* **8**(8), 851-62.

[202] Zhou, H. & Dahlquist, F.W. (1997). Phosphotransfer site of the chemotaxis-specific protein kinase CheA as revealed by NMR. *Biochemistry* **36**, 699-710.

[203] Zhou, H., Lowry, D.F., Swanson, R.V., Simon, M.I. & Dahlquist, F.W. (1995). NMR studies of the phosphotransfer domain of the histidine kinase CheA from Escherichia coli, assignments, secondary structure, general fold, and backbone dynamics. *Biochemistry* **34**, 13858-13870.