Structural Basis for Control by Phosphorylation

Louise N. Johnson* and Richard J. Lewis

Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU

Received December 6, 2000

Contents

I. Introduction

Posttranslational modification by phosphorylation is a ubiquitous regulatory mechanism in both eukaryotes and prokaryotes. Intracellular phosphorylation by protein kinases, triggered in response to extracellular signals, provides a mechanism for the cell to switch on or switch off many diverse processes. These processes include metabolic pathways, kinase cascade activation, membrane transport, gene transcription, and motor mechanisms. The reverse reaction of dephosphorylation is catalyzed by protein phosphatases that are controlled by response to different stimuli so that phosphorylation and dephosphorylation are separately controlled events. In eukaryotes, the protein kinase domain responsible for phosphorylation on serine, threonine, or tyrosine residues is the first, second, and third most common domain in the genome sequences of yeast (*S. cerevisiae),* the worm (*C. elegans*), and the fly (*D. melanogaster*), respectively, indicating the importance of phospho-signaling in higher organisms. The human genome contains 575 eukaryotic protein kinase domains, representing 2% of the total genome and the third most populous domain.[1] In prokaryotes, signaling by phosphorylation is equally important. In *E. coli* there are 62 genes that encode proteins involved in dual histidine kinase/response regulator systems, representing approximately 1.5% of the entire genome.

In this review we ask what are the structural consequences of phosphorylation and how is this structural response correlated with signal? We limit the examples for the most part to those where the structures of both the phospho- and non-phospho protein are known and coordinates available in the Protein Data Bank. We also limit our examples to phosphorylation of intact proteins and exclude the numerous structural examples of phospho−peptide binding to cognate proteins such as phospho−tyrosine peptide binding to SH2 (e.g., ref 2) and PTB domains (e.g., ref 3) (reviewed in ref 4) and phospho−serine peptide binding to 14−3−3 domains.[5] The phospho−peptide interactions give insights into the recognition by other proteins of the phospho−amino acid, an important result especially for the design of ligands that might interfere with this association, but these structures do not illuminate how the recognition site is engineered in the target protein as a consequence of phosphorylation nor the response of the phospho−protein on recognition. We also exclude results where a phospho−protein is an intermediate in a reaction pathway such as in the phosphatases,[6,7] topoisomerases,[8] and Cre recombinase.[9] In these examples, phosphorylation is part of the catalytic pathway and does not result in a response in the target protein that signals to other proteins. Finally, phenylalanine hydroxylase is an allosteric protein regulated by phenylalanine, tetrahydrobiopterin, and phosphorylation on serine 16. The non-phospho and phospho

* To whom correspondence should be addressed. Phone: 01865 275365. Fax: 01865 510454. E-mail: louise@biop.ox.ac.uk.

10.1021/cr000225s CCC: $36.00
Published on Web 07/11/2001

Louise Johnson obtained her Ph.D. degree in 1965 at the Royal Institution, London, where she worked on the binding of inhibitors to lysozyme, the first enzyme structure to be solved. After a postdoctoral year at Yale University, she has held academic positions in the University of Oxford since 1967. She is currently David Phillips Professor of Molecular Biophysics at the University of Oxford, a position that she has held since 1990. Her main research interests are in structural biology with the aim of seeking an explanation for biological function in terms of three-dimensional structures of macromolecules. She has made contributions in the fields of control by phosphorylation, carbohydrate recognition and catalytic mechanisms, time-resolved crystallography, and regulatory proteins of the cell cycle. She is a fellow of the Royal Society and Member of EMBO.

Richard J. Lewis received his B.A. and D.Phil. degrees in Biochemistry from the University of Oxford in 1991 and 1994, respectively, before commencing postdoctoral studies in the Chemistry Department of the University of York until October, 2000. He is currently a Wellcome Trust Research Career Development Fellow in the Laboratory of Molecular Biophysics, University of Oxford. His research interests include the molecular basis of control in phosphorylation-regulated biological processes, prokaryotic signal transduction, and protein/protein recognition with an emphasis on *Bacillus* species.

structures of phenylalanine hydroxylase showed no differences, and the region containing the phospho−serine was disordered, suggesting that other allosteric effectors need to be present to produce the activation changes.[10] We omit a discussion of "silent" phosphorylation.

Addition of a phosphoryl group to serine, threonine, or tyrosine (in eukaryotes and occasionally in prokaryotes) or to histidine or aspartic acid (in prokaryotes) residues confers properties that can have profound effects on protein conformation and function. The phosphoryl group with p$K_a \sim 6.7$ is likely to be dianionic at physiological pH. The property of a double-negative charge and the capacity for forming extensive hydrogen-bond networks with the four phosphoryl oxygens confers special characteristics. Analysis of protein/phosphate interactions in protein structures has shown that two types of interactions are dominant. In one, the phosphoryl group interacts with main-chain nitrogens at the start of an α-helix and utilizes the positive charge of the helix dipole for charge neutralization. In the other the phosphate group interacts with the side chain of one or more arginine residues. The guanidinium group of an arginine residue is well suited for interactions with a phosphate group by virtue of its planar structure and its ability to form multiple hydrogen bonds. Because of its resonance stabilization, the guanidinium group is a poor protein donor (p$K_a > 12$) and cannot function as a general acid catalyst in the hydrolysis of phosphorylated amino acids. Electrostatic interactions between arginine and phosphoryl groups provide tight binding sites that play a dominant role in recognition and conformational response. In addition, a number of other residues are also involved at phospho-recognition sites, and these include lysine, tyrosine, serine, threonine, asparagine, histidine, and, additionally in prokaryotes, metal ions.

The first phospho−protein structures, namely, rabbit muscle glycogen phosphorylase (reviewed in ref 11) and *E. coli* isocitrate dehydrogenase (reviewed in ref 12), revealed two different mechanisms for the structural response to phosphorylation. Glycogen phosphorylase is a dimer and is allosterically activated by phosphorylation. Phosphorylation on a serine, Ser14, results in a conformational change of the N-terminal residues such that Ser14 shifts 50 Å to change contacts from intrasubunit to intersubunit. The intersubunit contacts of the phosphoryl−serine, pSer14, to two arginines (one in its own subunit and the other in the other subunit) and other contacts of the residues that surround Ser14 are correlated with a conformational response that results in a changed orientation of the two subunits of the dimer and changes in tertiary structure at the catalytic site leading to activation. Isocitrate dehydrogenase is inhibited by phosphorylation. The mechanism for control by phosphorylation is simple. In this enzyme there is no change in conformation on phosphorylation. Instead, inhibition is achieved by phosphorylation of a serine residue, Ser113, at the start of an α-helix, which blocks access to the catalytic site of the negatively charged isocitrate substrate, through both steric and electrostatic effects.

Our understanding of the regulatory properties of phosphorylation has increased most notably with studies on the protein kinases, the serine/threonine kinases (namely, cyclic AMP-dependent protein kinase (cAPK), cyclin-dependent protein kinase 2 (CDK2), and extracellular regulated kinase (ERK2 or MAPK)), the tyrosine kinases (namely, insulin receptor tyrosine kinase (IRK)), and the family of Src kinases. Structural studies have shown the role of the phospho group in the activation segment of the kinase as an organizing center that localizes and organizes the activation segment for substrate recognition. Additional structural information about inhibitory phosphorylation includes inhibitory phos-

phorylation on Tyr15 in CDK2 and recognition of the phospho−tyrosine located in the C-terminal region by the SH2 domain of Src. More recent structural examples have shown how phosphorylation may cause disorder (e.g., K^+ channel inactivation domain), how phosphorylation may cause order, structural changes, and protein association (e.g., KIX/pKID from the CBP co-activator protein/CREB transactivation domain system and some prokaryotic response regulator signaling proteins such as CheY), how phosphorylation can promote dimerization (e.g., ERK2, STAT3B, and STAT1 and other prokaryotic response regulators such as FixJ and Spo0A), and how phosphorylation can inhibit protein/protein association (e.g., the prokaryotic HPr/enzyme I association in the sugar phosphotransferase system) and also cause protein/protein dissociation in the anti-anti-σ^F/anti-σ^F (SpoIIAA/SpoIIAB) interaction. Thus, we begin to have a portfolio of structural examples that demonstrates the most frequently observed aspects of phosphorylation, namely, enzyme activation, enzyme inhibition, protein association, protein dissociation, and order/disorder phenomena.

II. Eukaryotic Protein Phosphorylation

A. Eukaryotic Protein Kinases

The 28 eukaryotic protein kinase structures that are known to date have shown that they all have a very similar fold, as anticipated from conservation of certain residues throughout the kinase domain. The fold comprises an N-terminal lobe of mostly β-sheet with one α-helix termed the C-helix and a C-terminal domain that is mostly α-helical with a small amount of sheet. The catalytic site for the ATP moiety of the substrate is located at the lobe interface. Many protein kinases, but not all, require phosphorylation on a threonine (Thr160 in human CDK2) or tyrosine residue (e.g., Tyr416 in Src) that is located in a region termed the activation segment running from the conserved DFG motif to the conserved APE motif (residues 145−172 in CDK2). (This region is sometimes called the 'T loop', but we prefer the term 'activation segment' because the region contains some secondary structure, because it includes both phospho−threonine and phospho−tyrosine residues in different kinases, and because it plays a key role in kinase activation). We use the symbol "−" to denote covalent modification and the symbol "/" to denote noncovalent association. Phospho residues or phospho proteins are prefixed by "p" (e.g., pThr160 or pCDK2).

The first observation of phosphorylation on the activation segment was reported for cyclic AMP-dependent protein kinase (cAPK) in 1979,[13] but it was not until 1990 that mutagenesis studies showed that phosphorylation of Thr197 was important for recognition of the inhibitory regulatory subunit[14] and not until 1993 that it was recognized that threonine autophosphorylation was essential for kinase activity.[15] The crystal structure of cAPK, the first protein kinase structure to be solved in 1991, demonstrated the key contacts made by the phosphoryl threonine (pThr197) and provided ideas for its role as an organizing center critical for activation.[16,17] Structural studies with an inhibitory peptide showed interactions of the peptide with part of the activation segment, but the side chains of the peptide also made key specific interactions with side chains of cAPK in other regions.[17−19] For tyrosine kinases, autophosphorylation of v-Src at position Tyr416 had been shown in the early 1980s and its significance for the activation of the cellular counterpart c-Src kinase established in 1987 (reviewed in ref 20).

As more and more kinases were discovered and the presence of kinase cascades established in which activation of one kinase led to the activation of a following kinase, it became recognized that control by phosphorylation in the activation segment is a property of most, but not all, kinases.[21] All protein kinases contain an aspartic acid residue (Asp166 in cAPK; Asp127 in CDK2) that participates in catalysis. In kinases regulated by phosphorylation in their activation segment, an arginine residue precedes the catalytic aspartate. These are the so-called RD kinases. The arginine contacts the phosphoryl group of the phosphorylated residue in the activation segment. The contact does not change the position of the catalytic aspartate, but it assists in the correct alignment of the activation segment for interaction with the substrate. The structural studies on those kinases that are not regulated by phosphorylation have shown that these kinases have other mechanisms for localization of their activation segments. In phosphorylase kinase domain, a glutamate takes the place of the phosphoryl group.[22] The kinase domain is constitutively active without the requirement for posttranslational modification, and the activation segment is held in the correct position to accept the peptide substrate.[23] In twitchin kinase, a cluster of nonpolar groups is likely to take the place of the ionic interactions.[24] In casein kinase I[25,26] and in the cell cycle checkpoint kinase, Chk1,[27] a dianion (sulfate) can occupy the positively charged pocket that accommodates the phosphoryl group in phospho-regulated kinases, but studies have shown that removal of the dianion does not change the structure. In these examples the activation segment is stabilized by secondary structural elements and by side-chain interactions. Titin kinase represents an unusual example. The kinase is activated by phosphorylation of a tyrosine residue (Tyr170) that is outside the activation segment (two residues beyond the E of the APE motif). It is not an RD kinase. In the inactive non-phosphorylated state, the tyrosine is directed into the catalytic site.[28] It is proposed that on phosphorylation, the pTyr is displaced and that it may make a contact to the basic residue preferred in the P + 1 position for substrates of titin kinase. Evidently there are several different ways of stabilizing the activation segment. Those kinases controlled by activation segment phosphorylation represent some of the most important enzymes in signaling cascades.

1. Serine/Threonine Kinases

CDK2. Cyclin-dependent kinases (CDKs) play a central role in the coordination of the eukaryotic cell

cycle (reviewed in ref 29). Their activation is a two-step process that requires cyclin binding and phosphorylation by the CDK-activating kinase (CAK) on a threonine (Thr160 in CDK2) in the activation segment. pCDK2 in complex with cyclin A becomes active as cells enter the S phase. The complex phosphorylates downstream targets that include the tumor-suppressor proteins, pRb and the related p107, and other proteins that regulate transcription and replication. CDK2 has become the most intensely studied and the best understood structurally of the CDKs.

The crystal structure of the free inactive form of CDK2 was solved in 1993.[30] This revealed that the activation segment was located in a region that partially blocked the ATP binding site (Figure 1a). The two lobes of the kinase structure were in a closed conformation, and the C-helix was in the wrong orientation to promote crucial contacts with the triphosphate moiety of ATP, contacts that were known from the structure of active cAPK. Structural studies have shown that phospho-CDK2 (pCDK2) phosphorylated on Thr160 is similar in structure to free inactive CDK2, except that the part of the glycine-rich loop (residues 10−20) and the activation segment from residues 153−164 have become disordered (Figure 1b).[31] A partial rationalization for the disorder is provided by the observation that the activation segment in free inactive CDK2 places Thr160 close to a glutamate (Glu12), and hence, on phosphorylation there would be electrostatic repulsion. Phospho-CDK2 exhibits only 0.3% of the activity of fully active pCDK2/cyclin A. Evidently phosphorylation and the electrostatic repulsion are not sufficient to trigger the essential conformational changes required to activate the kinase, but some pliability is induced in the enzyme that allows a basal level of activity. The crystal structure of CDK2/cyclin A[32] showed that association of cyclin A with CDK2 led to significant changes in conformation that seemingly allowed a near active conformation, but the complex also exhibits less than 0.3% of the activity exhibited by the fully active phospho-CDK2/cyclin A complex. The major conformational changes involve opening of the bilobal structure, rearrangement of secondary structural elements, and reorganization of the activation segment (Figure 1c). The C helix (which contains the motif PSTAIRE) shifts and rotates by 90°, resulting in tight packing of the C-helix against the N-terminal lobe with extensive nonpolar interactions. The movement brings the side chain of Glu51 into the catalytic site where it hydrogen bonds with Lys33, an important contact residue to the α phosphate of ATP bound at the catalytic site. In free inactive CDK2, residues from the activation segment Gly147-Gly153 form a short helix (L12 helix) that packs against the C helix (Figure 1a). On binding cyclin A, this short helix melts to allow movement of the C helix and the L12 residues form a reverse turn followed by a β sheet (residues 150−152) with the β6 strand (residues 122−124) of the C-terminal lobe. The conformational changes observed on binding of cyclin A to CDK2 result in the displacement of the activation segment from the catalytic site and exposure of Thr160. A nearby glutamate, Glu162, occupies the phospho-recognition site (Figure 1c). There is no change in the conformation of cyclin A in binding CDK2.[33]

In the structure of the fully active pCDK2/cyclin A[34] complex, the pThr160 turns into the CDK2 molecule and the phosphate group contacts three arginines, Arg50 from the PSTAIRE C helix, whose position is adjusted by the movement of the C-helix on binding cyclin A, Arg126 that precedes the catalytic aspartate, and Arg150 from the activation segment whose position is adjusted by the melting of the L12 helix on association with cyclin A[34,35] (Figure 1d). The major change in conformation on forming the active pCDK2/cyclin A complex compared with the partially active (<0.3%) CDK2/cyclin A complex is a reorganization of the activation segment in the region between residues 152−163. The phospho−threonine group acts as an organizing center, reminiscent of the seryl−phosphate in glycogen phosphorylase. Its position appears to be determined by the need to neutralize the positive charge cluster of the three arginine residues (Figure 1e). The geometry of coordination is close to the ideal arrangement seen in model compounds. Arg50 and Arg150 also hydrogen bond to cyclin A main-chain carbonyls, and Arg126 hydrogen bonds to CDK2 atoms. The phosphate group is buried in an environment that is likely to strengthen the phospho−arginine interactions stabilizing the conformation in the vicinity.

The change in conformation engineered by the phospho−threonine in the CDK2/cyclin A complex is crucial for the recognition of the peptide substrate (Figure 1f).[36] CDK2 is specific for a proline residue (P + 1 site) following the phosphorylatable serine (P0 site) and a basic residue 3 residues C-terminal to the serine (P + 3 site). The peptide used in the crystallographic binding studies was derived from an optimal peptide substrate with sequence HHASPRK, using the single amino acid code. The peptide is bound in an extended conformation across the catalytic site on the surface of the kinase contacting only the C-terminal lobe and the activation segment. The pocket to accommodate the proline residue in the position P + 1 is created by a special conformation of the activation segment in which a valine residue (Val164) adopts a left-handed conformation stabilized by contact from its carbonyl oxygen to an arginine (Arg169). Binding of any residue except proline would be unfavorable because of an uncompensated hydrogen bond from the substrate's main-chain nitrogen. In the non-phosphorylated CDK2/cyclin A complex, Val163 adopts a different conformation and blocks the peptide binding site, thus explaining how the conformational changes produced by phosphorylation on the activation segment lead to activation through creating the peptide substrate recognition site. The phospho−threonine also plays a direct role in substrate recognition. The lysine in the P + 3 position contacts pThr160 (Figure 1f).

Recent kinetic results with a peptide substrate have shown an increase in k_{cat} of 843-fold and a decrease in K_m of 137-fold, resulting in an overall increase in catalytic efficiency (k_{cat}/K_m) on phospho-

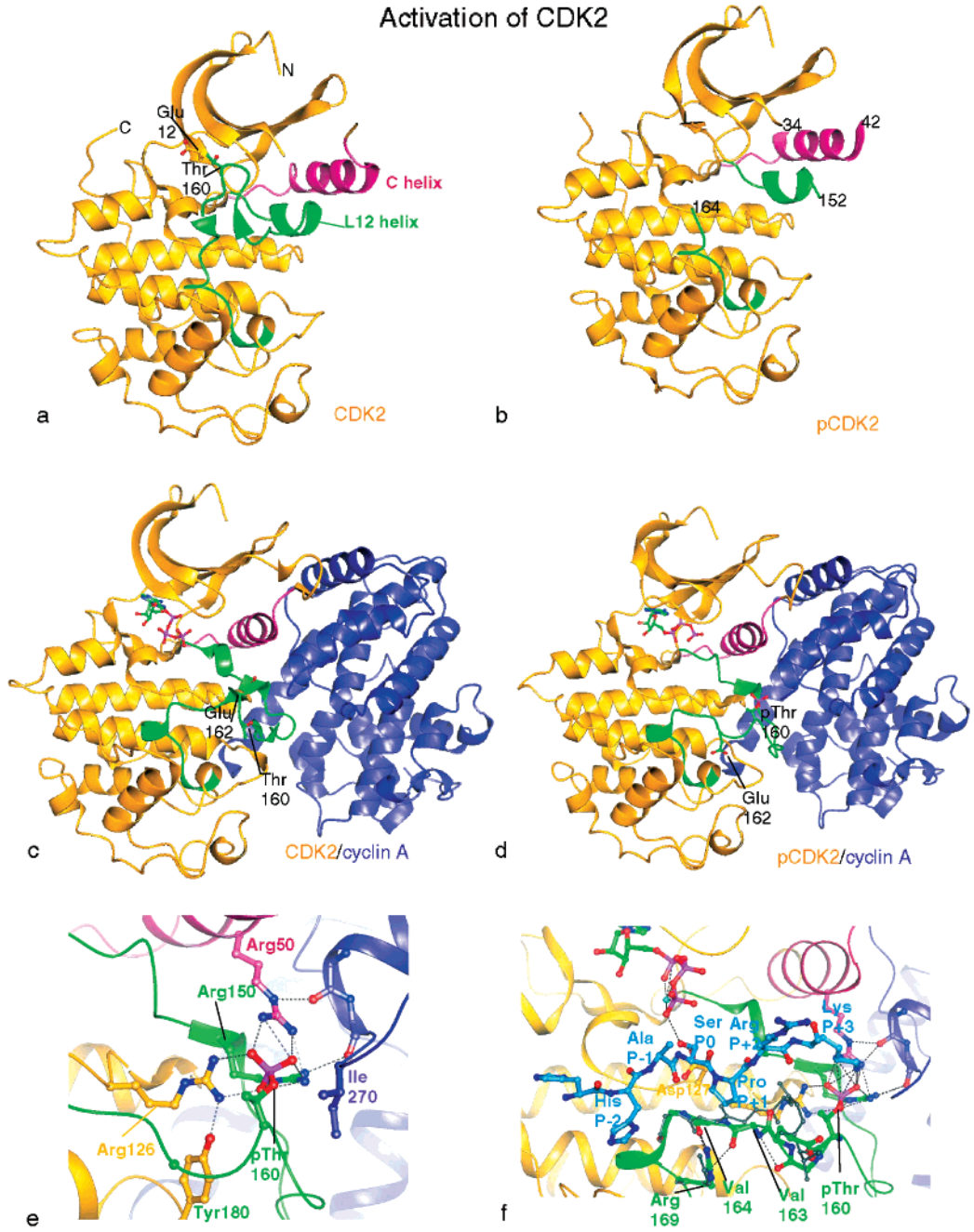


Figure 1. Structural response of CDK2 to activation by cyclin A and phosphorylation. Color scheme: CDK2 (yellow); C-helix (magenta); activation segment from residues 145−172 (green); cyclin A (purple). (a) Inactive apo-CDK2.[30] The positions of Thr160, Glu12, and the L12 helix in the activation segment are marked. There is a break in the electron density for residues 34−42 just before the C helix. (b) pCDK2.[31] Phosphorylation results in disorder of residues 153−164 in the activation segment but no other conformational changes from apo-CDK2. (c) CDK2/cyclin A complex.[32] There is a significant shift in the C-helix from its position in CDK2, creating the ATP triphosphate binding site, and in the activation segment resulting in exposure of Thr160 and burying of Glu162 into the phosphoryl recognition site. (d) pCDK2/cyclin A.[34] The phospho-Thr160 occupies the phosphoryl recognition site and Glu162 is exposed. There are small but significant conformational changes in the activation segment from the structure of CDK2/cyclin A. (e) Details of the phospho-recognition site. The color scheme is the same as that defined in Figure 1a. pThr160 contacts three arginines: one (Arg50) from the C-helix, one (Arg126) that is adjacent to the catalytic aspartate, and one (Arg150) from the start of the activation segment. Two arginines (Arg50 and Arg150) form hydrogen bonds to main-chain residues of cyclin A, while Arg126 hydrogen bonds to Tyr180. The phospho site is partially shielded by Ile270 on cyclin A. (f) Details on the interactions of a substrate peptide with the activation segment of pCDK2/cyclin A.[36] The peptide is blue, and the coloring scheme for CDK2 and Cyclin A is as described in Figure 1a. The ribose and triphosphate of the inactive analogue AMPPNP and the Mg^{2+} ion are also shown. The positions of the triphosphate moiety of AMPPNP shift between the CDK2/cyclin A complex and the pCDK2/cyclin A complex and reach their correct conformation for catalysis in the pCDK2/cyclin A/peptide complex. The conformation of the activation segment in the non-phospho CDK2/cyclin A is shown in gray. The shifts in main-chain atoms of Val163 on phosphorylation of Thr160 that converts CDK2/cyclin A to pCDK2/cyclin A create the proline recognition site at the P + 1 position. The substrate serine at P0 is directed toward the γ phosphate of AMPPNP and also hydrogen bonds with the catalytic aspartate Asp127. The lysine at P + 3 contacts the threonine phosphate. (Figures 1-8 were produced with Aesop. Noble, M. E. M. To be published.)

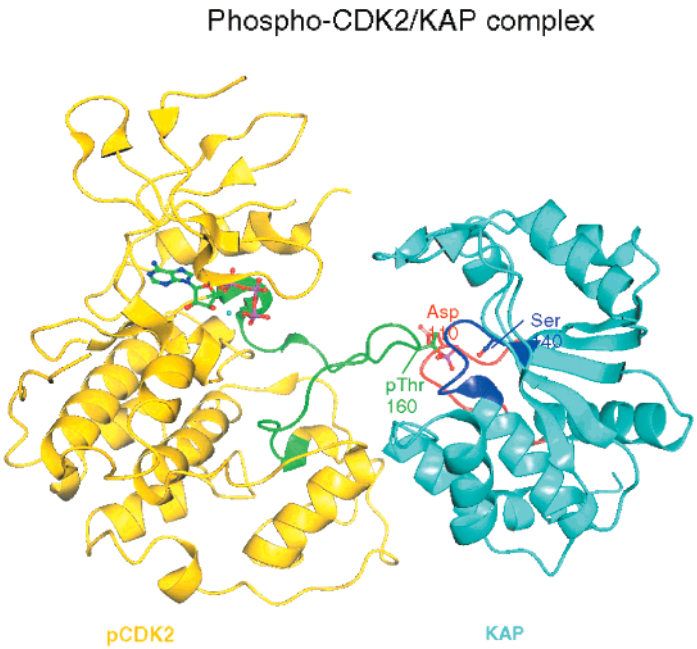


Figure 2. Complex of pCDK2 with KAP.[43] CDK2 is in gold with the activation segment in green and pThr160 marked. KAP is in cyan with the acid/base loop of the phosphatase carrying Asp110 in red and the loop that carries the catalytic cysteine (here mutated to serine, Ser140) in blue. The orientation of CDK2 is rotated with respect to Figure 1 in order to show the interactions with KAP more clearly. The pCDK2/KAP interaction positions the catalytic site of KAP to recognize pThr160. The activation segment is drawn away from the main body of the kinase, and the remainder of the kinase molecule collapses to a conformation that is nearer to the conformation of CDK2 seen in CDK2/cyclin A complexes than that observed for pCDK2. There is a break in the electron density in pCDK2 between residues 34−43 just before the C helix.

rylation of CDK2 in the CDK2/cyclin A complex of 100 000-fold.[37] The phospho and non-phospho forms showed equivalent rates of ATPase activity. Combination of the structures and the kinetic analysis suggests that the role of cyclin A in the CDK2/cyclin A complex is to align residues that form the ATP recognition site, while the role of Thr160 phosphorylation is to align the activation segment to create the correct peptide or protein recognition site to enable phosphoryl transfer. Phosphorylation enhances the rate of phosphoryl transfer (by about 3000) and increases the affinity for substrate binding. In pCDK2/cyclin A, the rate of dissociation of products is rate limiting but the differences in rates between the chemical step and product dissociation are not as extreme as those observed for cAPK[38−40] and phosphorylase kinase (PhK).[41] In pCDK2/cyclin A, the phosphoryl transfer step (k_3) is 22 s^{-1}, which is slower than that observed for cAPK and PhK (k_3 = 500 s^{-1} and >360 s^{-1}, respectively).

Inactivation of pCDK2/cyclin A involves ubiquitin-mediated destruction of cyclin A by the proteasome and dephosphorylation of pThr160. Genetic and biochemical studies implicate kinase-associated phosphatase (KAP) as one of the protein phosphatases responsible for dephosphorylating the activating pThr residue of human CDK2.[42] The enzyme, discovered using a yeast two-hybrid genetic screen as a CDK2-interacting protein, is a member of the dual-specificity phosphatase family and is capable of dephosphorylating artificial protein substrates containing both pThr and pTyr residues. KAP associates with CDK2 in vivo, and dephosphorylation of pThr160 is dependent upon the tertiary structure of CDK2, a feature that distinguishes KAP from the nonspecific PP2A catalytic subunit. The crystal structure of KAP in association with pThr160-CDK2 has recently been solved[43] and provides the first example of a protein phosphatase in complex with its intact protein substrate (Figure 2).

The C-terminal lobe of CDK2 and the C-terminal helix of KAP form the major protein interface between the two molecules, regions remote from the kinase activation segment and the KAP catalytic site. The kinase activation segment interacts with the catalytic site of KAP almost entirely via the phosphate group of pThr160 (Figure 2). The pThr160 phosphate group contacts the main-chain nitrogens of residues Gly142 and Arg146 and the side-chain OG of Ser140, contacts that are similar to those made by phospho substrates in tyrosine phosphatases such as PTB1B.[7] (The KAP catalytic cysteine, Cys140, had been mutated to serine in order to allow a stable complex to be formed.) This interaction requires that the activation segment is drawn away from the kinase molecule, thereby inducing a conformation of pCDK2 to one similar to the activated state observed in the CDK2/cyclin A complex (except for the activation segment conformation). The contact sites on CDK2 for cyclin A and for KAP are nonoverlapping, suggesting that cyclin A and KAP could bind CDK2 simultaneously. However, KAP is unable to dephosphorylate pCDK2/cyclin A and is only effective on free pCDK2.[42] The explanation lies in conformational differences. In the pCDK2/cyclin A complex the pThr160 is firmly bound by the three arginines (Arg50, Arg126, and Arg150) and is not accessible to recognition by KAP. The sequestration of pThr160, the defined conformation of the activation segment, and the localization of the arginine residues in the

pCDK2/cyclin A complex prevent access of KAP to its substrate. In free pCDK2, the activation segment is mobile and pThr160 is accessible for binding to KAP. The second recognition site on CDK2 for KAP, which involves the GDSEID and DYK motifs, is not changed between the pCDK2/KAP and pCDK2/cyclin A structures. It is accessible to KAP in the pCDK2/cyclin A complex, suggesting that it should be possible for KAP to bind to the pCDK2/cyclin A complex but unable to exert phosphatase activity. The structure of the pCDK2/KAP complex reveals how the conformation of a phosphoprotein may be altered following its interaction with a binding partner which recognizes the phosphorylated residue. It will be interesting to see if this is a common feature in other protein/phospho−protein complexes.

CDK2 is inhibited by phosphorylation of Tyr15 in the glycine-rich loop. The phosphorylation by Wee1 kinase and its dephosphorylation by CDC25 phosphatase form important checkpoints in the cell and play a role in response to DNA damage. The crystal structure of pTyr15,pThr160-CDK2 in complex with cyclin A has been determined (Tucker, J.; Endicott, J. A.; Noble, M. E. M. To be published). Tyr15 phosphorylation does not significantly alter the overall structure of the complex nor does it prevent ATP binding. It appears that phosphorylation of Tyr15 may perturb the binding of protein substrate at the catalytic site through steric hindrance, but this has to be explored through further experiments.

MAP Kinases. Members of the MAP kinase family transduce extracellular signals. They phosphorylate and activate transcription factors leading to transformation, proliferation, and other changes in the cell. Multiple parallel MAP kinase pathways are differentially responsive to growth factors and environmental stresses. Each member is activated by a specific MAP/ERK kinase (MEK) and a MAP/ERK kinase kinase (MEKK) cascade, and each phosphorylates a distinct spectrum of cellular substrates.[44,45] The MAP kinase ERK2 is activated in response to growth factors and mitogens by tyrosine kinase receptors, G-protein-linked receptors, and protein kinase C pathways. ERK2 achieves maximum activity only when both Thr183 and Tyr185 from the activation segment are phosphorylated by the upstream kinase MEK. Thr183 is equivalent to Thr160 in CDK2 and Thr197 in cAPK. Singly phosphorylated forms of ERK2 have less than 1% of the activity of the fully phosphorylated enzyme.

The crystal structure of non-phosphorylated ERK2 showed a typical kinase domain in which the two lobes exhibited an open conformation and the peptide binding site was blocked by Tyr185, one of the two residues that are phosphorylated on the activation segment.[46] The activation segment was well ordered with a different conformation than that seen in non-phosphorylated CDK2. ERK2 contains a C-terminal extension in which the chain spans both N-terminal and C-terminal lobes and terminates as an α-helix-labeled L16 that wraps around the C-helix. The structure of the dual-phosphorylated ERK2 (Figure 3) showed a rearrangement of the activation segment bringing the phospho−threonine and the phospho−

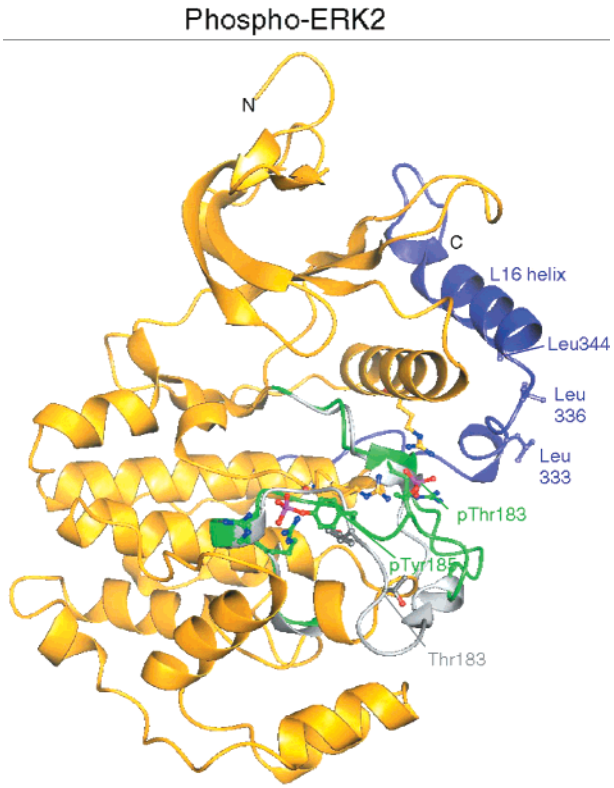


Figure 3. Comparison of non-phospho and phospho-ERK2.[47] ERK2 is in gold with the activation segment (residues 165−195) in green and the C-terminal region (residues 320−359 including the L16 helix) in purple. The L16 helix docks against the C-helix in a similar interaction to that seen with the interaction of cyclin A with CDK2 (Figure 1). The positions of pThr183 and pTyr185 are marked together with the arginine residues with which they interact (Arg68, Arg146, and Arg170 and Arg189 and Arg192, respectively). Note the similarity in conformation of the activation segment to that of pCDK2/cyclin A (Figure 1d). The position of the activation segment in non-phospho-ERK2 is shown in gray. The 9 Å shift in Thr183 is apparent. Also shown are residues Leu333, Leu336, and Leu344 in the loop preceding the L16 helix that form part of the dimerization interface in pERK2 dimers.

tyrosine into alignment with surface arginine residues.[47] Upon phosphorylation, pThr183 turns into the protein, shifting 9 Å to make ionic contacts with N-terminal lobe residues, thereby promoting domain closure. The contacts include Arg68 from the C-helix and Arg146 (preceding the catalytic aspartate) and Arg170 from the activation segment. These contacts are equivalent to those made by pThr160 in CDK2, and like pThr160, pThr183 in ERK2 acts as an organizing center through an extensive network of hydrogen bonds that link it to the C-helix and the C-terminal helix L16 (Figure 3). On phosphorylation, pTyr185 moves from its buried site to a surface site where it interacts with Arg189 and Arg192. To accommodate these changes, the activation segment refolds between residues Gly167 (the Gly of the DFG motif) and Ala187. Ala187 adopts a left-handed conformation (similar to the conformation of Val164 in CDK2) and hydrogen bonds through its main-chain carbonyl to the side chain of Arg192. ERK2, like CDK2, is specific for a proline in the P + 1 position. Modeling the peptide substrate from CDK2 into the catalytic site of pERK2 shows that the same mech-

anism operates for creation of a proline recognition pocket with the same specificity determinant engineered by the left-handed conformation of Ala187.

Thus, in ERK2, the dual phosphorylation promotes domain closure, allowing the catalytic and ATP binding groups to take up their correct conformation for catalysis, and phosphorylation reorganizes the activation segment to create the correct substrate recognition site for the protein substrate. It is interesting that, like CDK2, single phosphorylation is not sufficient to promote these changes; CDK2 requires cyclin A, and ERK2 requires dual phosphorylation.

The dual phosphorylation also promotes changes in the L16 helix. Shifts in L16 are correlated with the movement of the C-helix to its correct conformation. Curiously, the L16 helix appears to form a similar function to cyclin A in CDK2, packing against the C-helix (Figures 1c and 3). The L16 helix is also involved in dimerization. Three leucine residues (Leu333, Leu336, and Leu344) just prior to the L16 helix become more exposed in the activated enzyme. In the crystal, these residues contact their equivalent residues in a symmetry-related molecule forming a hydrophobic leucine zipper. On activation by phosphorylation, ERK2 dimerizes and translocates to the nucleus. The dissociation constant (K_d) for dimers changes from 20 μM for ERK2 to 7.5 nM for 2pERK2. The structural results combined with mutagenesis experiments provide a satisfying explanation for the creation of the dimerization site.[48] They show that such a site is engineered as a direct consequence of the structural changes following phosphorylation, although the phospho residues are not directly involved in the dimerization site.

Structures of three other members of the MAP kinase family have also been solved: P38α in non-phospho state,[49,50] P38γ in dual-phosphorylated state,[51] and JNK3 (a neuronal specific form of JNK) in non-phospho state.[52] The P38 kinases are activated in response to proinflammatory cytokines and by cellular stress. P38α and P38γ exhibit 63% sequence identity. The non-phospho forms of P38α showed a structure similar to ERK2 but with a greater degree of lobe opening and some differences in the conformation of the activation segment, which is six residues shorter in p38 than in ERK2. In p38, Tyr185 is exposed to solvent while it is buried in the non-phospho form of ERK2. The non-phospho form of JNK3 also shows an open conformation leading to misalignment of catalytic residues. The activation segment adopts a different conformation from ERK2, and the equivalent phosphorylatable threonine and tyrosine residues are 16 and 12 Å from their corresponding positions on non-phospho-ERK2. The structure of the dual-phosphorylated p38γ shows a dramatic closure of domains compared with p38α by 20° to bring catalytic residues into their correct orientation. The activation segment, despite its difference in length, adopts a similar conformation to that seen in 2pERK2 with identical contacts for the residues corresponding to pThr183 and pTyr185. These results illustrate a general theme in protein kinase structures. In their inactive state, protein kinase structures are different. In their active state, protein kinase structures converge to a common framework with regard to the correct orientation of catalytic groups and the activation segment. In P38γ, the sequence of amino acid residues in the L16 helix has significant differences from that of ERK2. The three leucine residues of ERK2 become aspartate, arginine, and valine, and the L16 helix has regions of disorder. P38γ does not dimerize on activation by phosphorylation, most likely because of noncomplementary changes in the surface residues in this region.

2. Tyrosine Kinases

Insulin Receptor Tyrosine Kinase. The cytoplasmic region of the insulin receptor contains the tyrosine kinase domain. In response to insulin binding to the extracellular domain, autophosphorylation on three tyrosines (Tyr1158, Tyr1162, and Tyr1163) in the activation segment of the kinase domain stimulates kinase activity toward exogenous substrates such as the insulin receptor substrate (IRS). A number of mutations in the tyrosine kinase region of the gene encoding the insulin receptor have been identified in patients with noninsulin-dependent diabetes.

The structure of the non-phosphorylated insulin receptor tyrosine kinase (IRK)[53] revealed the basic kinase fold in which the two lobes were held open by steric interactions between the glycine-rich loop and the start of the activation segment in the region of the DFG motif. The activation segment traverses the cleft between the N- and C-terminal lobes such that both protein substrate and ATP binding sites are blocked. One of the tyrosines from the activation segment, Tyr1162, is directed into the catalytic site and is an example of active-site-directed intrasteric autoregulation.[54] It is hydrogen bonded to the catalytic aspartate and seemingly poised for *cis*-autophosphorylation. However, catalysis is prevented by the blocking interactions of residues from the activation segment (e.g., Phe1151) that inhibit ATP binding.

The structure of the triply phosphorylated IRK in ternary complex with an inactive ATP analogue, AMPPNP, and a substrate peptide (Figure 4) revealed the significant conformational changes induced by the triple phosphorylation to create a functionally active kinase.[55] On phosphorylation, Tyr1158 shifts 30 Å and occupies a surface position, making no contacts with the protein but possibly offering a phospho-recognition site for other proteins. pTyr1162 is displaced from the catalytic site and hydrogen bonds to Arg1164 on the surface. This arginine is conserved in other tandem-phosphorylated receptor tyrosine kinases. pTyr1163 is the last residue to be phosphorylated but is the most important for conferring activity. It turns in and fills the pocket equivalent to that observed for phospho residues in other protein kinases (e.g., pThr197 in cAPK and pThr160 in CDK2), although the contacts to the phospho residue are less than those observed in the serine/threonine protein kinases. pTyr1163 hydrogen bonds to Arg1155 (from the activation segment equivalent to Arg150 in CDK2), and there are stacking interactions between the aromatic ring

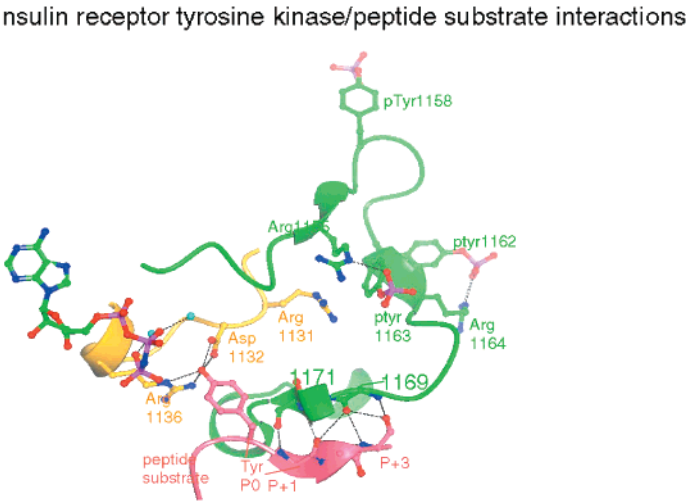


Figure 4. Tyrosine kinase domain of the insulin receptor (IRK)[55] showing interactions of the peptide substrate (in pink) with the activation segment (residues 1150−1179 in green). The catalytic loop (residues 1130−1137) that carries the catalytic aspartate (Asp1132) is shown in gold. The three phospho−tyrosine residues, pTyr1158 that is external, pTyr1162 that contacts Arg1164, and pTyr1163 that contacts Arg1155, are shown. The main chain of the peptide substrate forms a short antiparallel β sheet with the activation segment main-chain atoms of residues 1169−1171. The substrate tyrosine is directed into the catalytic site, and the phenolic OH group is within hydrogen-bonding distance of Asp1132, the catalytic aspartate, and directed toward the γ phosphate of AMPPNP, the inactive analogue of ATP.

of the tyrosine and the aliphatic carbons of the arginine. There is no direct interaction to Arg1131, the arginine that precedes the catalytic aspartate, although this arginine is in the vicinity. There are no contacts to residues from the C-helix, and the residue equivalent to Arg50 in CDK2 or His87 in cAPK is a glutamate. Nevertheless, phosphorylation stimulates significant domain closure by 21° and the C helix rotates by 35° so that the glutamate (E1047) becomes closer to the ATP binding lysine (Lys1030).

The peptide substrate (sequence of GDYMNM) binds with the tyrosine directed into the catalytic site close to but not in contact with the γ phosphate of AMPPNP (Figure 4). The substrate aspartate makes a contact through water to a lysine, while the two methionine residues C-terminal to the tyrosine occupy nonpolar pockets. Crucially the peptide is held in place by hydrogen bonds characteristic of an antiparallel β sheet between the P + 1 and P + 3 main-chain groups and activation segment residues Leu1171 and Gly1169, respectively, thus demonstrating the crucial role for phosphorylation in facilitating the correct conformation of the activation segment to create the peptide substrate recognition site. Likewise the correct orientation of the activation segment is critical for phosphorylase kinase peptide substrate recognition, where there is also antiparallel β sheet interactions between the substrate and the activation segment,[23] and in CDK2 where the activation segment creates the substrate prolyl and basic residue recognition sites.[36]

The fact that the pTyr residue is anchored rather less strongly than the corresponding phosphoryl groups in CDK2 or in cAPK may facilitate access to protein phosphatases that switch off the insulin signal. Recent structural studies[56] on the tyrosine protein phosphatases (PTP1B) in complex with triply and doubly phosphorylated peptides from the insulin receptor indicate that the major specificity is for pTyr1162 but that the specificity is enhanced by the tandem phosphorylation. The pTyr1163 makes specific contacts away from the catalytic site. The results indicate a hierarchical dephosphorylation process.

The structure of the kinase domain of the vascular endothelial growth factor receptor 2 (VEGFR2) has been solved in the phosphorylated state[57] and compared with the structure of the homologous (55% sequence identity) platelet-derived growth factor receptor (PDGFR) kinase domain.[58] Curiously, the VEGFR2 kinase domain, although phosphorylated on Tyr1059 (equivalent to Tyr1163 in IRK or Thr160 in CDK2), exhibits a similar conformation to the nonphosphorylated inactive PDGFR kinase domain. Most of the activation segment is disordered, and a portion of it occupies a position inhibitory to substrate binding. The VEGFR2 kinase domain has kinase activity, but the enhancement of activity for the phospho versus the non-phospho forms of the enzyme is only about 10-fold, compared with the 1000-fold enhancement of activity observed on phosphorylation of CDK2 in the CDK2/cyclin A complex. VEGFR2 kinase domain can additionally be phosphorylated on Tyr1054 (equivalent to Tyr1158 in IRK), and it may be that VEGFR2 kinase requires this phosphorylation or other factors for full activation or it may be that VEGFR kinase domain exists in a dynamic equilibrium involving several conformations that include the inactive conformation observed in the crystal. In this example, single phosphorylation has not been sufficient to establish the definite active conformation, just as single phosphorylation of free CDK2 is insufficient to activate.

Src Family Kinases (Src, Hck, Lck). The Src family of tyrosine kinases are a closely related group of nonreceptor kinases which are involved in signaling pathways that control growth and differentiation in cells in response to the activation of cell surface receptors by growth factors and cytokines. Src family members differ in cellular expression and localization. Src is expressed in a wide range of tissues:

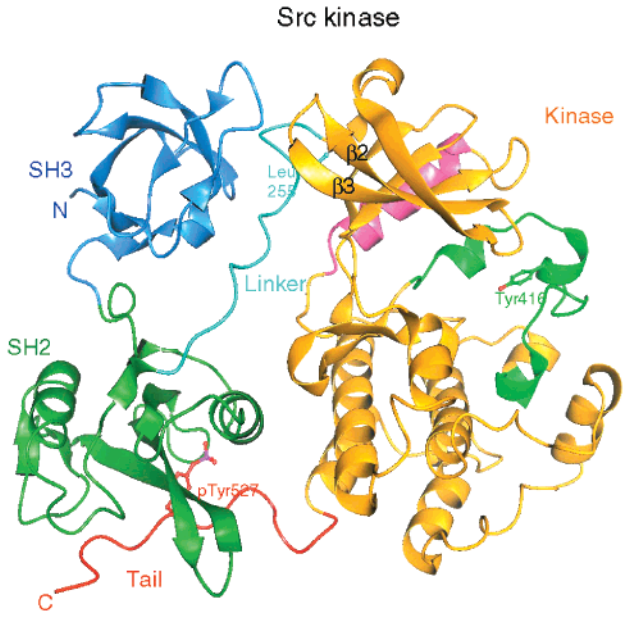


Figure 5. Structure of Src kinase (coordinates taken from ref 69). The N- and C-termini are marked. The SH3 domain is in blue, the SH2 domain in dark green, the linker in cyan, the kinase domain in gold with the C-helix (residues 304−317) in magenta and the activation segment (residues 404−432) in green, and the C-terminal tail in red. Also shown are the positions of Leu255, which is important for the linker to kinase interaction, Tyr416 in the activation segment that is phosphorylated for activation, and pTyr527 that interacts with the SH2 domain and holds Src in the inactive conformation.

haematopoietic cell kinase (Hck) is expressed in lymphoid and myeloid cells and lymphoid cell kinase (Lck) plays a role in T cell signaling. The viral homologue of Src, v-Src, is the gene product responsible for the cell-transforming ability of the Rous sarcoma virus and was the first tyrosine kinase to be identified. Src family members comprise five domains: an N-terminal domain involved in membrane binding and subcellular localization, an SH3 domain, an SH2 domain, a kinase domain (SH1), and a C-terminal tail that includes a tyrosine, Tyr527, whose phosphorylation by another kinase, Csk, leads to down regulation of Src-kinase activity. The importance of the C-terminal tail tyrosine phosphorylation for cellular regulation is demonstrated by v-Src, which lacks this site of phosphorylation and exhibits unregulated activity in cells. Src family members are regulated by autophosphorylation of a tyrosine in the activation segment (Tyr416 in Src).

The crystal structures of the down-regulated forms of human and chicken Src[59,60] and human Hck[61] and the active phospho form of the kinase domain of human Lck[62] have been solved. These structures have revealed a rich variety of control mechanisms for protein kinases both by phosphorylation and by protein/protein interactions. The structures of the SH3−SH2-kinase-C-terminal construct of down-regulated Src and Hck (Figure 5) show a compact structure in which the SH2 and SH3 domains are located on the opposite surface of the kinase domain to the catalytic site, indicating immediately that the explanation for down regulation of kinase activity by these domains is through indirect interactions and not a steric blocking mechanism. The C-terminal pTyr527 interacts with the SH2 domain at a site that is over 40 Å from the catalytic site (Figure 5). In simple terms, the down regulation appears to be effected by two factors: a constraint on the kinase C-helix so that it is displaced from its active conformation and, second, the engagement of the SH2 and SH3 domains at either ends of the two lobes of the kinase in such a way as to impair the correct lobe disposition needed for catalysis. The details have been well summarized in reviews.[63−65] Here we focus on the effects of phosphorylation (inhibitory phosphorylation at Tyr527 and activatory phosphorylation at Tyr416) and show how these phosphorylation control events are correlated with the protein/protein interactions mediated by the domains.

The SH3 domain adopts a similar structure to that observed in the crystal structures of the single domain, as does the SH2 domain. The linker between the SH2 and the kinase domain interacts with residues at the C-terminal end of the kinase C-helix and part of the region of a turn between the β strands β2 and β3 of the N-terminal lobe of the kinase. The linker adopts a conformation similar to a polyproline helix (although it contains only one proline in Src) and forms a docking site for the SH3 domain (Figure 5). These interactions form the first of the components that restrain the kinase in the inactive conformation. The kinase domain has a structure that has many similarities to the inactive conformation of CDK2. The contacts of the SH2 domain to the kinase are mediated by the docking of the C-terminal region pTyr527 to the phospho−tyrosine site of the SH2 domain. The phosphoryl group of pTyr527 in Hck and in chicken Src makes six hydrogen bonds in total to Arg 155, Arg175, Glu178 (main-chain NH), and the OG atoms of Ser177 and Thr179. Arg155 also contributes a positive charge interaction with the π electrons of the phenolic ring of pTyr527. These interactions are similar to the contacts observed in studies of model phospho−peptides bound to SH2 domains and suggest a relatively high-affinity phosphoryl-binding site. However, the sequence surrounding the pTyr527 in both Src and Hck is a nonoptimal sequence for SH2 recognition, and the manner of binding of the surrounding residues in both structures is suggestive of a low-affinity mode. Consistent with this is the ability of high-affinity phospho−tyrosine peptides to displace pTyr527, leading to an increase in kinase activity without phosphorylation on Tyr416.[66]

The contact of the C-terminal tail pTyr527 localizes the SH2 domain, reinforcing the contacts of the SH2−kinase linker region and the contacts of this region to the SH3 domain (Figure 5). Displacement of the SH3 domain by the SH3-binding protein HIV-1 Nef provides maximal activation of Hck,[67] indicating that the restraining interactions mediated by the SH3/linker/kinase N-terminal lobe are critical for maintaining the kinase in the inactive conformation. One interaction, that of Leu255 from the linker region to the N-terminal lobe of the kinase, appears to be an important interaction in the restraint of the kinase.[68] The leucine docks into nonpolar pocket formed by

Trp286 and Tyr326 of the kinase $\beta2/\beta3$ region and $\beta4$ strands, respectively (Figure 5). Mutational studies showed that when replaced by a smaller residue (e.g., Leu255Val), the mutant protein had kinase activity and this activity was not down regulated by phosphorylation of Tyr527. Occupancy of this nonpolar pocket by a bulky hydrophobic group therefore appears to be a key factor in the down regulation, and oblation of this interaction cannot be compensated for by phosphorylation of the C-terminal tail tyrosine.

In the original publications of Src and Hck structures, it was reported that the activation segment was mobile. Later work with different crystal forms, with less glycerol as cryo-protectant, in the presence and absence of ATP analogues, and at higher resolution has shown a defined conformation for the activation segment that provides further explanation for regulation by autoinhibition.[69,70] In the Src numbering, the activation segment running from the DFG to the APE motifs corresponds to residues 404−432. Residues 404−411 adopt a 3_{10} helix similar to the L12 helix in CDK2 (residues 145−152) (Figure 5). Hydrophobic residues from the helix in Src interact with the kinase C-helix, stabilizing the C-helix in its inactive conformation, in a rotated conformation similar to that observed in inactive CDK2. Further on in the activation segment, residues 413−418 form an α-helix such that Tyr416 is directed into to a nonpolar pocket and forms water-mediated hydrogen bonds to residues from the catalytic loop of the kinase. This interaction blocks the catalytic site, and comparison with the substrate-bound active form of IRK shows that the conformation of the activation segment is not in an appropriate conformation to recognize substrate.

Although the Src and Hck structures represent the inactive conformations, much can be learned about the transitions to an active kinase through comparison with the active kinase domain structure of Lck.[62] The construct used for Lck lacks the SH3, SH2 and C-terminal tail. It is comprises only the autophosphorylated (pTyr394) form of the kinase domain. The kinase domain has a structure that is similar to the active forms of other kinases (e.g., pIRK) with the correct disposition of the C-helix and lobe orientations to allow participation of the key binding and catalytic residues in substrate recognition and catalysis. The activation segment (residues 381−402) contains some regions of mobility, but pTyr394 is well localized with the phosphoryl group occupying a similar location to that seen in cAPK, CDK2, and IRK. The phosphoryl group contacts Arg363 (the arginine before the catalytic aspartate corresponding to Arg126 in CDK2) and Arg387 from the activation segment (corresponding to Arg150 in CDK2), but this interaction is mediated by two water molecules. The binding of the phosphoryl group therefore appears to be less strong than in cAPK and CDK2 but is nevertheless sufficient to generate and stabilize an active conformation of the kinase domain.

It has been remarked that Src and CDK2 share similar features in their disposition of the C-helix, which is important for maintenance of the inactive conformation. However, the mechanisms of activation are different. Src is restrained by interactions at the C-terminal end of the C-helix that are relieved by displacement of the SH3 and SH2 domains by a variety of mechanisms, while CDK achieves basal activity through association with cyclin A and the interaction of the cyclin at the N-terminal end of the C-helix (Figures 1c and 5). Once the SH3 and SH2 restraints are relieved in Src, full activation can be achieved by the *trans*-autophosphorylation of the tyrosine Tyr416 in the activation segment increasing activity by at most 50-fold. In contrast, CDK2/cyclin A has only basal activity and requires another kinase to phosphorylate Thr160, thus conferring a 1000-fold increase in activity.

The double SH2-regulated tyrosine phosphatase SHP-2 exhibits a different mechanism for SH2 regulation. In the crystal structure of the inactive form of SHP-2, the N-terminal SH2 domain directly blocks the catalytic site of the phosphatase domain.[71] The interaction appears to alter the phospho−tyrosine peptide binding site on the N−SH2 domain, disrupting its phospho−peptide recognition. It is proposed that when tyrosine−phosphopeptides bind to the two tandem SH2 domains with the C-terminal SH2 domain enhancing the poor affinity of the N−SH2 domain, conformational changes are also induced that release the N−SH2 from the phosphatase catalytic site and allows activation.

B. Phospho-Signaling and Protein/Protein Association

1. STAT Proteins

Further insights into the role of phosphotyrosine/SH2 interactions in regulation have been elucidated with structure determinations of STAT-1[72] and STAT3β[73]. STAT (signal transducers and activators of transcription) proteins are activated in response to extracellular signals. Binding of growth factors or cytokines to their cognate receptors results in intracellular tyrosine phosphorylation. STAT proteins contain an SH2 domain. They are recruited to the receptor intracellular region and then phosphorylated either directly by the receptor kinase or by the receptor-associated JAK kinase. Phosphorylation causes the STAT proteins to dimerize and translocate to the nucleus where they bind to specific promoter sequences in target genes. The JAK/STAT pathway provides one of the most direct signaling pathways from the receptor to the nucleus. The crystal structures of both STATs comprised constructs that lack the N-terminal DNA cooperative binding domain and the C-terminal transcriptional activator domain but included four domains: a four-helix coiled-coil domain, the DNA binding domain that has an immunoglobulin-like fold similar to that of NFκB and p53, a linker region composed of α-helices, and the SH2 domain followed by the C-terminal region containing the phospho−tyrosine. The STAT proteins crystal structures were solved in complex with an 18-mer and a 17-mer duplex of DNA. The structures immediately provided an explanation for the dimerization and DNA binding promoted by the interaction of the phospho−tyrosine with the SH2 domain.

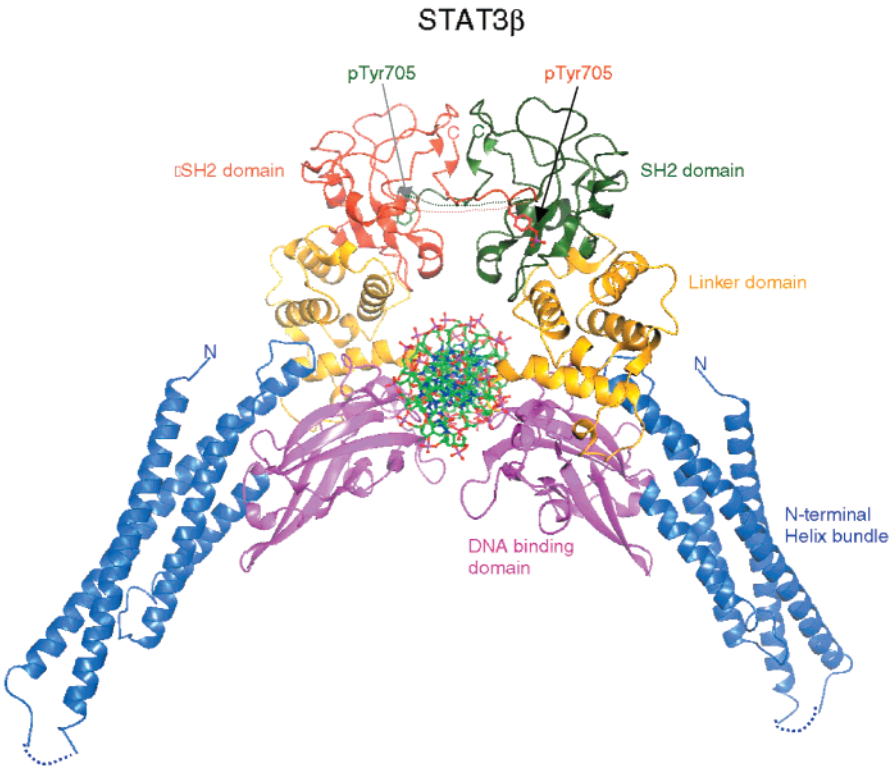


Figure 6. Stat3β homodimer DNA complex.[73] The N-terminal helix bundle (residues 130−320) is shown in blue, the DNA binding domain (residues 320−465) in magenta, the linker domain (residues 465−585) in yellow, and the SH2 domain and phospho−tyrosine-containing region (residues 585−722) in green in the right-hand molecule and red in the left-hand molecule. The disordered regions that link the α1 and α2 helices of the N-terminal helix bundle and the region from 689 to 701 in the phospho−tyrosine-containing loop are shown by dotted lines. The 17 base pair DNA duplex is shown with all atoms drawn. The pTyr705 from one molecule reaches into the SH2 binding site on the other molecule of the dimer. The view is approximately along the DNA dyad axis with the dyad axis of the complex running vertically.

The four domains form a stable tandem structure in which each domain interacts with its neighbors by significant hydrophobic interactions (Figure 6). Two STAT monomers form a dimer in which the monomers are related by a 2-fold crystallographic symmetry axis. The two monomers bind DNA, gripping it like a pair of pliers, in such a way that the two DNA binding domains interact with DNA but not with each other. The phospho−tyrosine of one monomer crosses over and docks into the SH2 recognition site of the other monomer. The C-terminal region then crosses back into its own monomer. The contacts to the phosphoryl group are highly conserved (with the exception that one arginine is a lysine) with those observed for other SH2 interactions, although the sequences of the STAT SH2 domains diverge considerably from those of other SH2 domains. The length of the linker region from the end of the SH2 domain to the phospho−tyrosine is too short to allow the phospho−tyrosine to reach the recognition site of its own monomer. The phosphotyrosine/SH2 interaction provides the major contact in promoting dimerization.

The SH2 domain is in contact with the DNA binding domain through interactions with an α helix of the linker region. Thus, phospho−tyrosine recognition by the SH2 domain promotes dimerization and signals to the dimeric DNA binding site. There are similarities with the way in which the C-terminal tail phospho−tyrosine of Src signals. Both proteins, one through intramolecular (Src) and the other through intermolecular (STAT) recognition, utilize the localization of the SH2 domain by the phospho−tyrosine to signal to other structural elements of the protein. In Src protein/protein interactions mediated by the SH2 hold the kinase in an inactive conformation, and in STAT, protein/protein interactions localize the DNA binding domain.

2. KID Domain of CREB Bound to the KIX Domain of CBP

The cAMP-regulated transcription factor CREB is stimulated for target gene expression by association with the co-activator, the CREB binding protein, CBP. Complex formation between CREB and CBP requires cAPK phosphorylation of Ser133 on CREB. CREB transactivation domain is composed of the kinase-inducible domain (KID) that carries Ser133 and a Q2 domain that stimulates transcription through association with a subunit of TFIID. CBP contains a domain known as KIX with which its binds to the phospho-KID domain of CREB. The NMR-determined structures[74] of pKID (residues 101−160) from CREB in association with KIX (residues 586−666) from CBP have provided a basis for understanding the role of phosphorylation in association. Free pKID exhibits a NMR spectrum indicative of an unstructured peptide. Upon binding to KIX, the pKID peptide folds into two mutually perpendicular helices between the regions 119−145, although outside these regions the peptide remains unstructured. The KIX domain has a well-defined structure of three antiparallel helices with a hydrophobic groove running

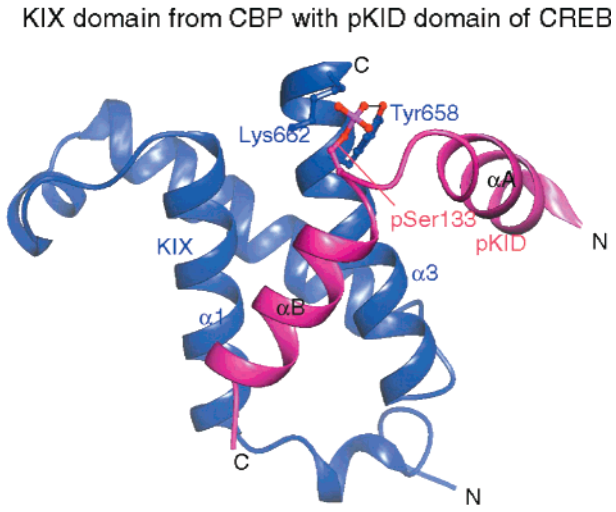


Figure 7. Structure of the KIX domain of CBP bound to the phosphorylated KID transactivation domain of CREB. KIX is in purple and pKID in magenta.[74] The interaction of pSer133(KID) with Tyr658(KIX) is shown. There are extensive interactions between the C-terminal helix of KID and the α1 and α3 helices of KIX.

across helices 1 and 3 into which the second helix (αB) of pKID binds (Figure 7). In addition to the nonpolar interactions between pKID and KIX, there are three charge/charge interactions and the overall change in solvent accessible area (1200 Å^2) indicates a reasonably tight interaction. However, there is no association without phosphorylation of Ser133.

The phosphoryl−serine residue is situated at the start of the second helix (αB) of pKID (Figure 7). The contacts to KIX vary in the family of NMR structures, but the essential interaction, as demonstrated by mutagenesis, is that between the phosphoryl group and Tyr658. A lysine, Lys662, is in the vicinity, but its interaction does not appear to be essential. There is likely to be stabilization of the αB helix through interaction of the phosphoryl group with the helix dipole. Free phosphoryl serine does not bind to KIX (in contrast, phosphoryl tyrosine will bind to SH2 domains). Thus, the mechanism by which phosphorylation confers structure and association of the KID domain appears to utilize mutually cooperative interactions. The folding of pKID through association with KIX is dependent upon the phosphoryl group at the N-terminal end of the developing α-helix and one key interaction of the phosphoryl group to a tyrosine residue.

Phosphoryl binding sites at the start of α-helices are common in enzymes that recognize phospho substrate or cofactors (e.g., triose phosphate isomerase and lactate dehydrogenase). There are now three examples where a phosphoryl amino acid occurs at the start of a 3_{10} or α-helix: glycogen phosphorylase, isocitrate dehydrogenase, and now pKID. To be recognized by their respective kinases (phosphorylase kinase for glycogen phosphorylase and cAPK for KID), the crystal structures of these kinases have shown that the region in the vicinity of the phosphorylatable residue must adopt an extended conformation.[23] Hence, on release from the kinase, a conformational change must take place (extended to helix) and indeed the incentive for this conformational change may well encourage product release from the kinase.

C. Phosphorylation Signals Order/Disorder Transitions and Protein/Protein Dissociation

There are many instances in vivo where phosphorylation on multiple sites causes protein/protein dissociation or prevents protein/protein association. The structures of non-phospho and phospho forms of the inactivation domain of a K^+ channel have provided insight into the basis of these events based on electrostatic repulsion, while two further proteins whose structures have been determined in the non-phospho form (the retinoblastoma tumor-suppressor protein Rb and the neuronal Sec1 protein) have indicated that similar mechanisms may be possible with these systems.

1. Inactivation Gate of a K^+ Channel

In common with many ion channels, the voltage-gated K^+ channel Kv.4 is regulated by phosphorylation. Inactivation of the channel is achieved through the interaction of residues 1−30 (termed the inactivation domain (ID)) which block the pore on the cytoplasmic side. Phosphorylation of the ID by protein kinase C (PKC) on three potential serine residues (Ser8, Ser15, and Ser21) results in reduction of inactivation. The NMR analysis of the non-phosphorylated ID showed a compact structure that had a high affinity for the receptor. Phosphorylation at Ser8 resulted in an increase in the rate of dissociation from the receptor, and phosphorylation of Ser15 and Ser21 resulted in a decrease in the rate of association. The NMR structures of pSer8-ID and pSer15, pSer21-ID showed a loss of overall structural stability.[75] In pSer8-ID, residues 5−15 were ordered (i.e., those around the phospho site) but residues N- and C-terminal were disordered. In pSer15, Ser21-ID, residues 15−25 were ordered (again those regions around the phospho sites) and the remainder of the structure disordered. The phospho sites are close to two glutamate residues, Glu27 and Glu28, and it is likely that the disorder arises from steric and electrostatic repulsion. Whereas in free pCDK2 it is the region carrying the phosphoryl−threonine that becomes disordered because of electrostatic repulsion, in the ID domain of Kv3.4 it is the other parts of the structure.

2. Retinoblastoma Tumor Suppressor

The growth-suppressive action of the tumor-suppressor protein Rb is coupled to the cell cycle through the actions of the CDKs. Inactivation of Rb is a common event in many cancers. In its active state, Rb binds to the transcription factor complex E2F and thereby inhibits expression of genes that are required for the transition from the G1 into the S phase and DNA replication. Rb phosphorylation on multiple sites, first by CDK4/cyclin D and then by CDK2/cyclin E followed by CDK2/cyclin A, leads to dissociation of Rb from E2F and release of transactivation competent E2F. Sixteen potential Ser/Thr-Pro phosphorylation sites have been identified, and at least 10 such

sites have been shown to be phosphorylated in vivo (Ser249, Thr252, Thr373, Ser567, Ser608, Ser780, Ser807, Ser811, Thr821, and Thr826). Some of these are specific for individual CDK/cyclin combinations. Rb also forms complexes with several other proteins including viral oncoproteins. Phosphorylation differentially effects the binding of other proteins.[76]

The crystal structure of the so-called pocket region of Rb comprising the A and B boxes (residues 372−589 and 635−787) has been solved in complex with a peptide from the human papilloma virus E7 oncoprotein.[77] The oncoprotein binds to Rb through a recognition sequence of LxCxE and inactivates Rb causing release of E2F. The A and B boxes are responsible for binding many of the cellular and viral proteins that mediate transcriptional repression. They contain three of the phosphorylation sites (Thr373, Ser567, and Ser780). The other phosphorylation sites are contained in the N-terminal region, the linker between the A and B boxes, and the 15K C-terminal domain. The A and B boxes exhibit the five-helix assembly characteristic of the cyclin fold, and this fold is augmented by eight other helices. Serine567, one of the phosphorylatable serines, is near the interface between the A and B boxes, and its phosphorylation would result in unfavorable contacts. Another phosphorylatable serine, Ser780, is in the middle of a 13 residue carboxyl tail of the B box. It packs against a hydrophobic region but does not make any significant interactions with the rest of the protein. Phosphorylation of Ser780 could result in some restructuring, although the consequences are hard to predict. Two other phosphorylatable residues, Thr821 and Thr826, are outside the construct used for the crystal structure determination. However, they are thought to influence the recognition and binding of proteins containing the LxCxE motif. The Rb fragment structure bound with the E7 LxCxE peptide reveals a six-lysine basic patch on the rim of the LxCxE binding site that may provide a recognition site for the two phosphorylated threonines and allow the phosphorylated segment to bind to and to block the site.[77]

3. Neuronal-Sec1/Syntaxin1a Complex

Syntaxin1a and neuronal Sec1 are two proteins that are involved in vesicle trafficking and membrane fusion. Neuronal Sec1 binds with nanomolar affinity to syntaxin 1a and forms a complex that is essential for vesicle trafficking and membrane fusion and which prevents syntaxin complex formation with other vesicle membrane proteins. The crystal structure of the nSec1/syntaxin1a complex revealed conformational changes in the four-helix syntaxin protein relative to the free syntaxin and that complexed with a SNARE protein.[78] Protein kinase C modulates exocytosis in several different cell types, possibly by regulation of the nSec1/syntaxin interaction. PKC phosphorylates Ser306 and Ser313 of nSec1, which inhibits the formation of the Sec1/syntaxin complex. These sites can no longer be phosphorylated in the complex. Ser313 is near the syntaxin recognition site in a region that caries a negative charge. In the nSec1/syntaxin complex, the negative charge is neutralized by a complementary positive charge region from Sec1. Although a phospho−serine could be accommodated at this site without disruption of the structure, the introduction of a further negative charge by the phosphoryl group could provide electrostatic repulsion and promote conformational changes in nSec1. Thus, phosphorylation of Ser313 is likely to prevent nSec1/syntaxin association. Ser306 however is further away from the nSec1/syntaxin interface, and the consequences of phosphorylation at this site, if any, are not clear.

These examples indicate that electrostatic effects, either electrostatic repulsion (as seen in the inactivation domain of the K^+ channel or that hypothesised for nSec1/syntaxin interactions) or electrostatic attraction (as proposed for Rb pThr820 and pThr826 interactions), are likely to be significant. Both effects are seen in glycogen phosphorylase, where in the non-phospho state the basic peptide region surrounding the phosphorylatable serine, Ser14, docks against an acidic region on the protein surface while on phosphorylation pSer14 shifts to contact a newly created basic site comprising two arginines.[79] The examples also demonstrate the difficulty of predicting the consequences of phosphorylation in the absence of structures of both the phospho and non-phospho forms.

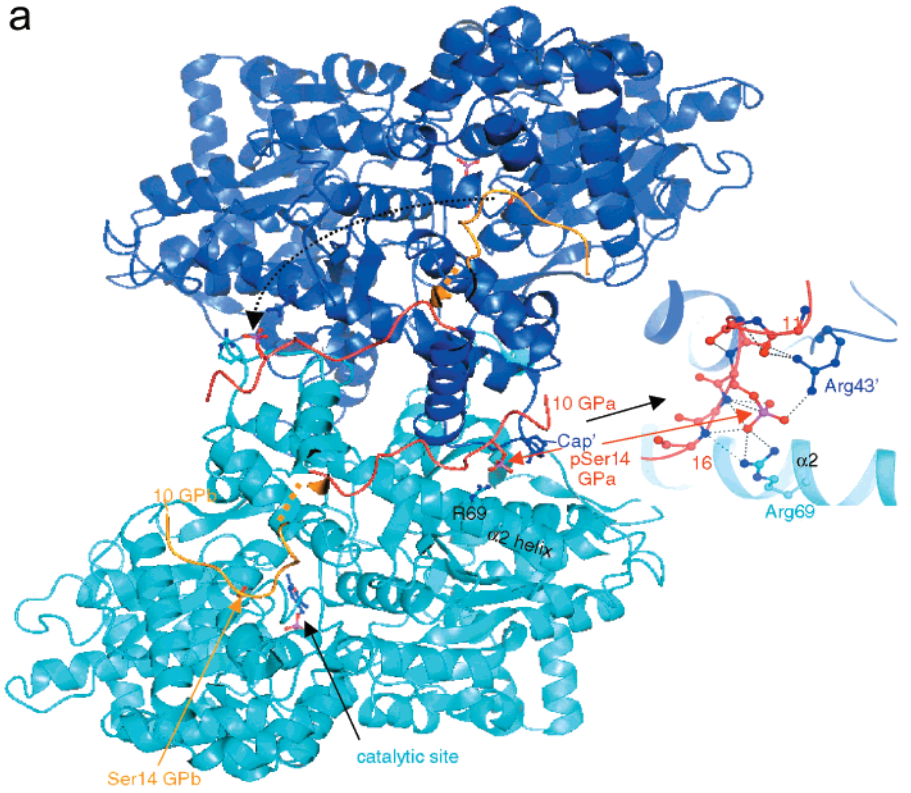
D. Phosphorylase: The Evolution of Control by Phosphorylation

Muscle glycogen phosphorylase, discovered by Carl and Gerty Cori in the late 1930s, was the first enzyme for which control by reversible phosphorylation was established.[80] The enzyme has been the subject of intense biochemical and crystallographic studies. Through structural studies on the non-phospho (rmGPb) and phospho (rmGPa) forms of the rabbit muscle glycogen phosphorylase, the structural basis of control by phosphorylation is well understood (reviewed in ref 11). Some of the features of the response are illustrated in Figure 8a. More recently, the structures of the phosphorylated human liver isozyme (hlGPa) in the inactive and active states,[81] the phospho and non-phospho forms of the yeast phosphorylase (yGPb and yGPa),[82−84] and the *E. coli* maltodextrin phosphorylase[85] have been solved, leading to further insights into the structural basis of control by phosphorylation.

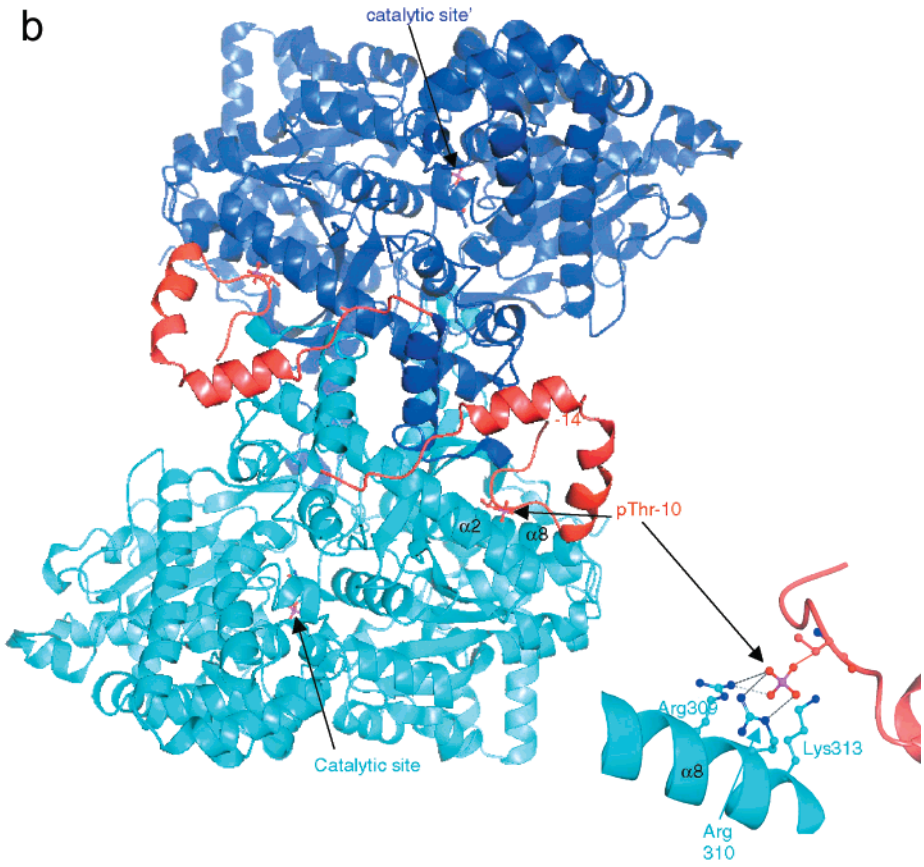
1. Human Liver Glycogen Phosphorylase

Phosphorylases catalyze the phosphorylysis of an α-1,4-glycosidic bond in glycogen or maltodextrins to yield glucose-1-phosphate. In muscle, glucose-1-phosphate is fed into the glycolytic pathway to meet the energy needs of the cell, while in the liver, glucose-1-phosphate is converted to glucose and glucose is output for the benefit of other tissues. The human muscle and liver isozymes exhibit 79% identity in sequence but have slightly different control properties. Both are regulated by phosphorylation on a single serine, Ser14, leading to activation, but the muscle enzyme may be activated by AMP; this activity was inhibited by ATP and glucose-6-phosphate, while the liver enzyme is not sensitive to these

Rabbit muscle glycogen phosphorylase



Yeast glycogen phosphorylase



effectors. Both isozymes are dimers formed from two monomers of 842 residues in muscle and 846 residues in liver. The human liver phosphorylase structure has been solved in its phosphorylated state in an inactive and active conformation.[81] The inactive form was achieved by co-crystallizing in the presence of a potent glucose analogue inhibitor (*N*-acetyl-β-D-glucopyranosylamine).[86,87] The active form was crystallized in the presence of AMP. Thus, although the structures do not allow a direct comparison between phospho and non-phospho forms, they do allow comparison between inactive and active states and further elaborate on the allosteric connection between the phosphoryl−serine site and the catalytic site that are about 45 Å apart.

In the inactive hlGPa, the N-terminal region surrounding Ser14 is disordered, consistent with observations that binding of glucose or glucose-analogue inhibitors at the catalytic site promotes the transition form the R to T state and allows greater accessibility of the phosphoryl−serine to the protein phosphatase PP1. In the active hlGPa, the N-terminal peptide is well ordered, making contacts at the intersubunit interface. The phosphoryl−serine contacts Arg43′ (from the other subunit) and Arg69 from its own subunit, as in rmGPa (Figure 8a). The allosteric transitions are mediated through these intersubunit contacts. In the N-terminal region on one side of the dimer, the phosphoryl−serine contacts tighten the interface and bring the cap region from one subunit (residues 42′-45′, so-called because it caps the AMP recognition site) closer to the α2 helix of the other subunit. At the other side of the dimer, two helices, the tower helices (residues 267−274 and 267′-274′), form intersubunit contacts through helix−helix′ packing. In the inactive hlGPa, the catalytic site is blocked by the 280s loop (residues 280−286) that follows the tower helix, and the 280s loop is held in place by contacts to the glucose analogue and to the 380s loop, as seen in the inactive form of rmGPb. In the active hlGPa, the tower helices are shorter by two turns and the 280s loop folds up in an ordered conformation and contributes an extra turn to a helix, α8 helix. Access to the catalytic site is open, and the 380s loop becomes more disordered. In rmGPa, the 280s loop itself becomes disordered. The structure at the catalytic site and the 280s loop observed with hlGPa is similar to that observed for *E. coli* maltodextrin phosphorylase as described below. The changes observed on the transition from inactive to active hlGPa structures not only allow access to the catalytic site, but also most crucially create the recognition site for the inorganic phosphate substrate. This site is within hydrogen-bonding distance of the 5′-phosphate group of the essential cofactor pyridoxal phosphate (PLP) and is created through the displacement of an acidic residue (Asp283) from the 280s loop and the replacement with a basic residue (Arg569), as observed with the rmGPb and rmGPa comparisons.

As in rmGP, there is a rotation of the hlGPa subunits with respect to each other on activation by about 7°, resulting in hlGPa in a change of surface interface buried from 4200 to 7300 Å^2, a rather larger change than that observed with rmGP. At the AMP recognition site, there is no change in sequence between hl and rmGP but many of the interactions observed for AMP binding to rmGPa are absent in hlGPa as a result of the small differences in subunit orientation and different disposition of side chains. The adenine-recognition loop (residues 315−325) adopts a different conformation. These changes result in the loss of cooperative binding of AMP. Despite sequence conservation between the muscle and liver isozymes, the changes at the subunit interface and not the changes in residues that contact the ligands result in a muscle enzyme that can be allosterically regulated by effectors as well as by phosphorylation and a liver enzyme that can only be activated by phosphorylation. Both can be down regulated by glucose and glucose analogues, a phenomenon that is important in the liver but is not relevant for muscle phosphorylase. In muscle, the effects of the allosteric effectors, AMP, Glc6P, and ATP, can be overridden by phosphorylation.

2. Yeast Phosphorylase

The structures of the non-phospho and phospho forms of the yeast (*S. cerevisiae*) glycogen phosphorylase revealed a subtle modification of the mechanism established for the mammalian phosphorylases. In yeast, phosphorylase is activated during the approach to the stationary phase to enable cells to utilize glycogen when other substrates may be depleted. Yeast glycogen phosphorylase (yGP) has an extra 39 residues at the N-terminal region (numbered −1 to −39) compared with rmGP, and the site of phosphorylation is at Thr-10. Although overall rmGP and yGP exhibit about 46% identity in sequence, there are no similarities in sequence surrounding the sites of phosphorylation. Further, yeast phosphorylase kinase and rabbit muscle phosphorylase kinase are different enzymes and do not cross phosphorylate

Figure 8. Comparison of the structures of the phospho active forms of rabbit muscle glycogen phosphorylase and yeast glycogen phosphorylase. (a) Rabbit muscle phosphorylase a (GPa).[177] The dimeric enzyme is viewed down the 2-fold axis of symmetry. One subunit is in cyan and the other in blue. The N-terminal tail residues 10−23 are in red in both subunits. The N-terminal residues 10−23 in GPb are in yellow in both subunits. The position of Ser14 is marked in the non-phospho form and in the phospho form. The black curved arrow in the top subunit highlights the 50 Å shift in Ser14 on phosphorylation. The other conformational changes on the conversion of GPb to GPa are not shown. The catalytic site is indicated by the position of the pyridoxal phosphate. Entrance to the catalytic site is on the far side of the molecule. The site of interaction of the pSer14 is over 40 Å from the catalytic site. The inset shows the details of the interaction of the pSer14 with Arg69 from the α2 helix and Arg43′ form the cap′ region of the other subunit. (b) Yeast glycogen phosphorylase.[83] The dimeric enzyme is viewed down the 2-fold axis of symmetry. The coloring scheme is similar to that described for part a. The N-terminal residues from −14 to 23 are shown in red for both subunits. The details of the interaction of the pThr-10 with two arginine residues (Arg309 and Arg310) are shown in the inset. Lys313 is in the vicinity but is not in contact. In the non-phospho form of yeast phosphorylase, the N-terminal residues wrap around the other subunit and block entrance to the catalytic site and the pThr-10 site is occupied by glucose-6-phosphate.

rabbit and yeast phosphorylases, respectively. The control sites appear to have evolved through different evolutionary mechanisms in which a phosphorylatable peptide was spliced to a common core of a phosphorylase.

The non-phospho form of yGP was crystallized in the presence of the inhibitor, Glc-6-P. The core of the structure is similar to that of rmGP, but there are some differences in the relative disposition of the two subunits of the dimer.[82,83] The structure of the catalytic site of yGP is similar to that of the activated form rmGPa. The 280s loop is folded up so as to allow access, and the residues that form the inorganic phosphate substrate recognition site are in their active conformation for substrate binding. However, despite this apparent active structure, the non-phospho form of yGP is held in an inhibited state by a blocking mechanism from residues of the extended N-terminal tail. From residue 22 to −39, the tail first wraps around its own subunit, partly mimicking the conformation and position adopted by the N-terminal tail in rmGP. It then continues with residues toward the N-terminus to wrap around the other subunit so that residues −23 to −38 contact the catalytic site entrance and block access. In the non-phospho state, Glc-6-P is bound at the AMP/Glc-6-P recognition site of rmGP and utilizes the same arginine residues for recognition of the phosphate (Arg309 and Arg310) but the contacts of the glucosyl moiety are different than those observed for Glc-6-P binding to rmGPb because of changes in sequence.[88]

In the non-phospho yGP, Thr-10 is surrounded by nonpolar residues. In the phosphoform of yGP, the N-terminal tail undergoes displacement and Thr-10 shifts about 36 Å.[84] Phospho-Thr-10 displaces Glc-6-P and itself occupies the phospho recognition site through interactions with Arg309 and Arg310 (Figure 8b). There are also nonpolar interactions between the nonpolar residues surrounding Thr-10 and those on the core surface of the phosphorylase. These contacts are different than those made by the glucosyl moiety of Glc-6-P, so that there are distinct regulatory responses to the inhibitor (Glc-6-P) and the activator (pThr-10). There are accompanying changes at the subunit interface as in rmGP that include a tightening of contacts at the allosteric effector interface site and opening around the interface leading to the catalytic site. Thus, activation by phosphorylation in yGP involves competition between the phosphoamino acid and an allosteric inhibitor for a phospho-recognition site, relief of steric blocking of the catalytic site, and some reorganization of the two subunits.

3. E. coli Maltodextrin Phosphorylase

E. coli maltodextrin phosphorylase (MalP) is a 796 amino acid dimeric enzyme. Unlike rmGP and yGP, MalP is not regulated by phosphorylation or by allosteric effectors and is constitutively active. Levels of MalP in *E. coli* cells are regulated by gene expression in response to changes in nutritional status. Under conditions of low glucose and high maltose or maltodextrins, the Mal genome pathway is switched on in response to elevated cAMP levels that enable the maltodextrins to be transported into the cell and utilized. Overall MalP is 46% identical in sequence to rmGP, but it lacks 17 residues at the N-terminus and 13 residues at the C-terminus. Residues 18−80 (using the rabbit muscle GP numbering) display minimal sequence similarity between rmGP and MalP. These residues correspond to the first exon in human muscle GP, and it has been suggested that control by phosphorylation in eukaryotic phosphorylases could have been achieved by splicing a phosphorylatable peptide to a common ancestral phosphorylase core.

The structure of the core of MalP is very similar to that of rmGP.[85] At the catalytic site surrounding the cofactor PLP, residues are 100% conserved. Outside the core, the MalP and rmGP structures differ in the subunit/subunit contacts in the region of the cap′/α2 interface and the tower/tower′ interface. Despite the lack of sequence similarity, residues 18−80 in MalP show superficial similarity in structure to rmGP and occupy the same space. The region from 246 to 286 adopts a different conformation. As a result of sequence changes, there is a different organization of residues 254−260 and the tower helices. The 280s loop folds up and adds two turns of α helix to the α-8 helix, allowing access to the catalytic site where the residues are in the correct conformation to promote catalysis.

The structure of *E. coli* MalP has been solved in the ternary complex with inorganic phosphate and pentasaccharide substrate analogue in which the glycosyl oxygen linkage was replaced with a thio linkage.[89] The structure has revealed not only the details of the catalytic residues when poised for catalysis, but also those residues important for the recognition of the oligosaccharide. Tyr280, a residue conserved in all phosphorylases, from the 280s loop plays a crucial role.

Thus, the comparative studies reveal a common structural core for the different phosphorylases, but each exhibits different regulatory properties. In the prokaryotic MalP, there are no effector ligand binding sites and the core exhibits constitutive activity without the need for activation by phosphorylation. In the eukaryotic yeast and mammalian muscle phosphorylases, control by phosphorylation has occurred twice with the splicing of different peptides containing a phosphorylatable group, resulting in different solutions to the same evolutionary challenge.

III. Prokaryotic Signal Transduction

For many years it was generally held that reversible protein phosphorylation was a regulatory mechanism that had developed late in evolution and was thus unique to the demands of signaling within multicellular organisms. Intercellular communication was not understood to be a function of simple, unicellular organisms, since it was assumed that these had evolved with no requirement for sophisticated means of control. It is now well established that regulation by protein phosphorylation/dephosphorylation also occurs in microorganisms and that bacteria do communicate within colonies. Eukaryotes utilize most frequently the hydroxyl side chains of

serine, threonine, and tyrosine residues in kinase cascade signal transduction pathways, and often multiple phosphorylation sites are present in the same protein. Prokaryotes instead favor histidine and aspartic acid, although no cogent argument has been made that explains this difference, and their proteins are most frequently phosphorylated only on single centers. Sequencing of microbial genomes has shown that prokaryotes may encode eukaryotic-like signaling domains within their genomes.[90] For instance, seven putative Ser/Thr kinases and seven hypothetical Ser/Thr and Tyr phosphatases from three distinct phosphatase families have been identified in the genome of the cyanobacterium *Synechocystis* sp. PCC 6803.[91] This is in addition to the ~80 open reading frames (some 2.5% of the genome) encoding proteins of the two-component signal transduction system,[92] the most commonly found signal transduction mechanism in bacteria (for recent reviews of the two-component system, see refs 93−96).

A. Two-Component Signaling

Two-component signaling systems comprise a histidine protein kinase (HK) and a cognate response regulator (RR). RRs are most frequently composed of an N-terminal, conserved, regulatory domain of ~120 amino acids and a C-terminal effector domain, which exhibits a diversity of sequences that relates to the range of biological processes that these molecules regulate: in prokaryotes, for instance, two-component systems mediate metabolic pathways[97] and adaptive responses such as chemotaxis[98] sporulation,[99] virulence,[100] and differentiation.[101] While these systems are prevalent in bacteria and archea, they are found rarely in eukaryotes. While two-component systems regulate the ethylene response of *Arabidopsis thaliana*[102] and osmoregulation in *Saccharomyces cerevisiae*,[103] no two-component systems have been reported in the completed genome sequences of the worm, *C. elegans*, the fruit fly, *D. melanogaster*, or in the available sequence to date (some 1 billion bases) of man. (For a review in greater detail of eukaryotic two-component systems, see the article by Saito in this issue.) The number of two-component proteins found within bacteria can vary significantly; where *B. subtilis*[104] and *E. coli*[92] each have more than 60, none have been found in *Mycoplasma genitalium*.[105] Similarly, there are 0 and 20 in the archaeons *Methanococcus jannaschii* and *Halobacterium* sp. HRC-1, respectively.[106] The presence of two-component systems in pathogenic bacteria such as *S. pneumonia* and *M. tuberculosis* and their absence in man places two-component systems centrally in the future development of novel anti-infective therapies.[107]

In a general reaction, the HK uses ATP to autophosphorylate the Nϵ2 atom of a conserved histidine residue within the kinase. The phosphoryl group is then transferred to a conserved aspartic acid residue in the RR, via a trigonal bipyramid pentavalent phosphorus transition state,[108] a probable intermediate also in the mechanism of dephosphorylation.[109] Divalent cations are essential for these (de)phosphorylation reactions and serve to shield the negative charge on the anions, polarize the P−O bond, and form a template for the transition state. Phosphorylation induces conformational changes in the RR, which can cause both the dissociation (CheB[110]) and association (FixJ[111]) of intramolecular contacts in multidomain RRs and the establishment of new intermolecular interactions in single domain RRs (CheY[112]). The inherent activity of either the RR or its downstream target is thus stimulated to elicit the appropriate biological response. The free energy of hydrolysis of the acyl phosphate of RRs (and proteins of the haloacid dehalogenase superfamily that utilize an acyl phosphate enzyme intermediate[113]) can vary significantly from that measured from the hydrolysis of small molecule model acyl phosphate compounds.[114] Therefore, the energy stored within the acyl phosphate bond may drive the conformational changes in RRs required for activation. Furthermore, the high-energy acyl phosphate bond, created on phosphorylation of aspartic acid, is readily hydrolyzed in aqueous solutions, unlike the stable phosphoester linkages of phospho−serine, −threonine, and −tyrosine, and has precluded structural analysis of proteins phosphorylated on aspartate. As a result, while much has been gleaned from the intense study of activated eukaryotic kinases, it is only recently that the structural consequence of phosphorylation in prokaryotic signal transduction systems has been detailed.

1. Response Regulators

The first structure of a response regulator to be determined (Figure 9) was that of the single-domain chemotaxis factor CheY more than a decade ago.[115] While several other RR domain structures have subsequently been determined, all were similarly in the non-phosphorylated, resting state. The overall fold of all RR domains is a doubly wound $(\beta-\alpha)_5$ barrel, comprising a central core of a five-stranded parallel β-sheet, in the order $\beta2-\beta1-\beta3-\beta4-\beta5$. On one side, this core is flanked by α-helices 1 and 5, whereas α-helices 2, 3, and 4 are found on the opposite face. The lengths of the elements of secondary structure and two important loops are quite similar in all structures determined to date, such that to a first approximation all RR domains are virtually indistinguishable. Despite this conservation in structure, crucial signaling fidelity is maintained even in bacteria that encode over 30 such proteins and over 30 histidine kinases (see below).

The molecular basis of RR activation has remained elusive. RRs spontaneously dephosphorylate, with half-lives of the phosphorylated forms that vary from just a few seconds to several hours. In fact, the half-life of phosphorylated RRs is frequently shorter than that of free phospho−aspartic acid, and it has thus been concluded that RRs harbor an inherent autophosphatase activity. Accordingly, the lability of the phosphorylated form of RRs has made structural analysis inaccessible, until recently. At the time of writing, six structures of activated response regulator domains have been published, those of the sporulation factor Spo0A,[116] FixJ[111] and NtrC[117] (both involved in nitrogen metabolism), and three structures

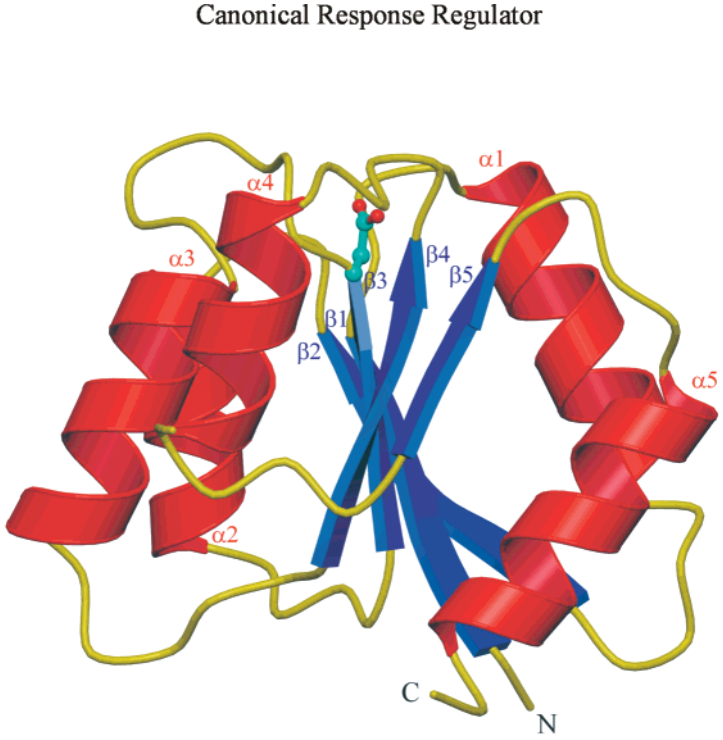


Figure 9. Overall conformation of response regulators. A ribbon diagram of a canonical response regulator to illustrate the fold of the domain, which consists of a doubly wound $(\beta-\alpha)_5$ barrel, with the site of aspartic acid phosphorylation situated at the C-terminal end of the third, middle strand of β-sheet, and drawn as a ball-and-stick model. (This figure was produced by MOLSCRIPT.[178])

of CheY.[118−120] Interestingly, while stable RR activation has been achieved in a variety of different ways, the structures are largely supportive of a general mechanism of two-component RR activation.

For FixJ and NtrC, phosphorylation was achieved by the action of the small molecule phosphodonor acetyl phosphate. Magnesium depletion of FixJ with EDTA stabilized the acyl phosphate prior to crystallization,[111] whereas for NtrC, the sample was maintained in a steady state of phosphorylation/dephosphorylation with a large excess of acetyl phosphate during NMR measurements on multiple samples.[117] The production of the phosphorylated aspartic acid in Spo0A was serendipitous; no effort was made to phosphorylate Spo0A prior to crystallization,[116] and phosphorylation occurred naturally, either by the action of intracellular small molecule phosphodonors, such as acetyl phosphate, or by a noncognate HK found in the *E. coli* expression system. Stabilization of the aspartyl phosphate bond in this instance probably occurs partly as a consequence of the use of protein from a moderate thermophile, *B. stearothermophilus*,[121] and replacement of the obligatory catalytic magnesium ion with calcium, which is likely to perturb reactivity[109] by subtle alterations in metal coordination geometry.

Activation of CheY was accomplished not by deliberate phosphorylation, but with phosphoryl group analogues. In one case, an α-thiophosphonate derivative of a mutant form of CheY was prepared, where the nucleophilic Asp57 had been replaced by cysteine and the structure determined by X-ray crystallography of phosphonomethylated-CheY.[119] In the NMR structure of CheY[118] and crystal structure of CheY in complex with a 16mer peptide from its downstream target, FliM,[120] BeF_3^- was used as the Mg^{2+}-dependent ground state mimic of the phosphoryl group, which persistently activates CheY, whereas the half-life of phospho-CheY is only a few seconds.[122] This analogue is arguably a better mimic of a phosphoryl group than a thiophosphonate moiety (Figure 10), even though at neutral pH the beryllofluoride group is monoanionic whereas a phosphoryl group is dianonic. In the thiophosphonate, an atom equivalent to one carboxylate oxygen atom of aspartic acid is absent and as a result a vital magnesium ligand is not present in phosphono-CheY.[119] Indeed, phosphonomethylation of Asp57Cys CheY is independent of magnesium. Furthermore, the bridging S^{γ} assumes two distinct conformations, possibly because the divalent cation is missing; consequently, the phosphoryl group is directed toward and displaces Lys109. Finally, phosphonomethylation and phosphorylation of CheY do not produce the same chemical shifts, implying that the precise structural changes on covalent modification of Asp57 differ.[119]

The chemical shift differences do correlate very well for beryllofluoride-treated CheY and NtrC against the respective phosphorylated forms of these two RRs,[122] arguing that similar structural changes are likely to occur on activation by either process. Furthermore, in the crystal structure of the FliM peptide complex of CheY−BeF_3^-, the distance between the O^{δ} and Be atoms is 1.5 Å, almost identical to that of the extended O^{δ}−P bond of 1.6 Å in the structures of Spo0A and FixJ. The ions in Spo0A and CheY occupy similar positions to that of the magnesium in the native crystal structure of CheY,[123,124] and the active site architectures are conserved between Spo0A, FixJ, and CheY. In the two structures of activated

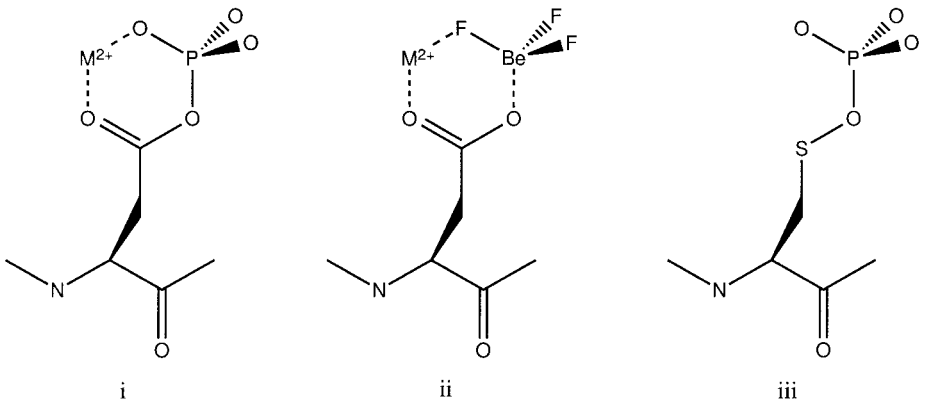


Figure 10. Activation of RR by phosphoryl analogues. Chemical structures of (i) phosphoryl-aspartic acid, (ii) berrylofluoryl aspartic acid, and (iii) thio-phosphonate derivative of Asp57Cys mutant. Note that in the latter case an oxygen atom that is an obligatory magnesium ligand is missing, such that the six-membered chelate ring formed during phosphoryl transfer cannot exist. (This figure was produced by MOLSCRIPT.[178])

The 'aromatic switch' of Spo0A

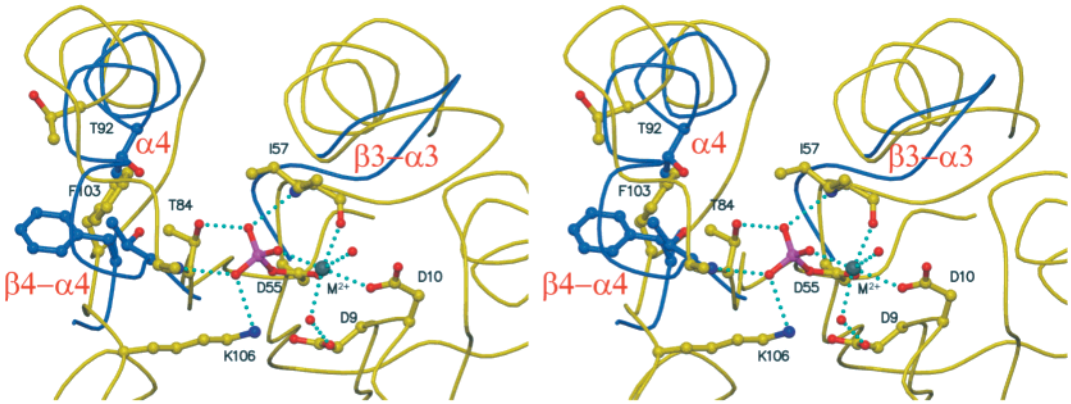


Figure 11. Stereodiagram of the active site of the phosphorylated RR domain of Spo0A.[116] The phosphoryl group and divalent cation draw together the side chains of the five most highly conserved residues within the response regulator family, Asp9, Asp10, Asp55, Thr84, and Lys106. The bound calcium ion displays near-hexavalent, bipyramid coordination geometry that is similar to that observed in other magnesium-bound RR structures. The protein backbone of the phosphorylated RR domain of Spo0A is colored yellow, and the effect of phosphorylation on residues 55−63 (β3-α3 loop) and 83−97 (β4-α4 loop, α4) is illustrated by a comparison to the non-phosphorylated structure of Spo0A, in blue. Note how α4 (top left) has to move away from the main body of the protein to accommodate Phe103; otherwise there would be severe steric clashes with Thr92. In other RR sequences, this amino acid is most commonly branched. For clarity, the seventh calcium ligand, a water molecule, is omitted. Hydrogen bonds drawn as dashed lines.

RR which have divalent cations bound, Spo0A[116] and CheY in complex with the FliM peptide[120] the atoms of the phosphorylated aspartate, $O^{\delta}1$, C^{γ}, $O^{\delta}2$, the metal M^{2+}, and O and P (or Be) from the phosphoryl and beryllofluoride groups form a near-planar, six-membered chelate ring, as originally proposed by Koshland[125] in 1952 (Figure 10).

In all examples studied other than NtrC, activation does not appear to induce drastic and widespread conformational changes within the RR domain. In the following discussion in this paragraph, the residue numbering used is taken from a typical multidomain RR, Spo0A, but is pertinent to all activated RRs so far characterized. Apart from phosphonomethylated CheY,[119] the invariant Lys106 remains static as do the three conserved aspartic acid residues, Asp9, Asp10, and the site of phosphorylation Asp55. The phosphoryl group participates in hydrogen bonds to the backbone amide nitrogens of Ala85 and Ile57, neither of which are situated in α-helices, the N^{ζ} of Lys106, and the adjacent cation. The fifth amino acid that characterizes RRs, Thr84, repositions its hydroxylated side chain from a position pointing away from the active site to a position toward the phosphoryl group, making a hydrogen bond with one of the phosphoryl oxygen atoms.[111,116,118,120] This leaves a void behind the threonine, which is filled by the concerted movement of the side chain of a conserved aromatic residue, Phe103, from a solvent-exposed position to one that is completely buried (Figure 11). To accommodate the bulky, aromatic side chain, α-helix 4, which is loosely tethered to the central β-sheet, moves up and away from the protein core, also causing a structural reorganization of the β4-

α4 loop. This mechanism of activation has been called the "aromatic switch"[116] or "Y−T" coupling.[118]

The rotamer exchange of Tyr106, the equivalent residue to Phe103 in Spo0A, is also important in the binding of CheY to FliM, since the outward resting position of Tyr106 would sterically interfere with FliM binding.[120] Analogously, dimerization of FixJ is mediated solely by residues within the RR domain, caused by structural changes in the α4, *β*5 region by the outward to inward rotamer exchange of Phe101.[111] However, in the structure of the methylesterase CheB,[110] the equivalent residue, Phe104, forms part of the inhibitory interface between the RR and methylesterase domains, and in activation of this RR, the outward to inward movement of the side chain of this amino acid would disrupt the interdomain interface, liberating the methylesterase activity of CheB, the catalytic triad of which is situated some 25 Å away from the site of phosphorylation. The conformation of CheY in the CheY−BeF_3^- and CheY−BeF_3^-:FliM complex is barely altered by binding to the FliM peptide, and therefore, the small, yet significant, conformational changes induced by phosphorylation are sufficient to activate RRs. These slight structural alterations may be consistent with the finding that phosphorylation only increases FliM binding by CheY 20-fold[126] and DNA-binding by Spo0A[127] and FixJ[128] some 20−40-fold.

The structural analysis of the phosphorylated form of NtrC was made difficult by the short half-life of minutes, and thus, multiple NMR data sets with poor signal-to-noise ratios were combined in the structure determination.[117] The effective low resolution of this structure precludes accurate descriptions of the active site stereochemistry and changes in the rotamer conformation of key amino acids. Widespread conformational changes which include a 100° rotation of the inherently flexible α-helix 4 and other changes that extend into α-helix 3 are described, which are not observed in other structures of activated RRs and may either place NtrC in a separate category to CheY, FixJ, and Spo0A or merely represent the difficulty in detailed structural analysis on transient species. Nevertheless, all six recent structures are consistent with the surface of the RR that includes residues from α-helix 4 and *β*-strand 5 as a central component in the activation of RRs, and this same, or an overlapping, surface also defines the site of new intra-[111] and intermolecular[120] interactions requisite in activation (Figure 12).

However, while an apparent general activation mechanism of RRs has been revealed by structural studies of isolated RR domains, interdomain communication in multidomain RRs may be specific to each, individual molecule. While the aromatic switch explains how the methylesterase activity of pCheB is stimulated, how pCheY binds to the flagellar motor, and how pFixJ dimerizes in order to bind to DNA with high affinity, it does not describe how the transcriptional regulatory function of NarL, the only other structure of an intact multidomain RR to have been published,[129] is stimulated. NarL modulates the transcription of several anaerobic electron transport genes, and phosphoryl transfer from its paired his-

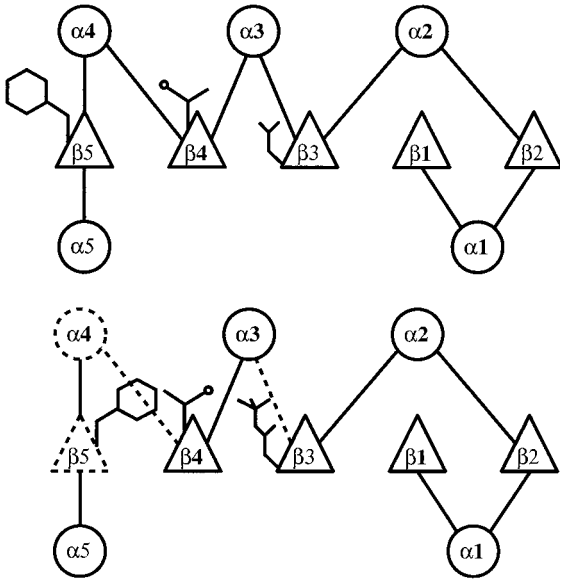


Figure 12. Aromatic switch. A schematic of the mechanism of response regulator activation, where the topology of the canonical response regulator is represented by triangles (*β*-strands) and circles (α-helices). Dashed lines represent regions of structure that appear to undergo significant conformational changes on phosphorylation. On phosphorylation of the aspartic acid on *β*3, the hydroxylated side chain of threonine (or serine) on *β*4 swings toward the newly arrived phosphoryl group, leaving a void which is filled by the outward to inward movement of the aromatic side chain on *β*5 from a solvent exposed to a buried position. To accommodate the aromatic side chain, α4 and the *β*4-α4 loop undergo structural rearrangements, which are central to the activation mechanism. By unmasking hitherto occluded protein surfaces, the response regulator may participate in new inter- or intramolecular interactions or disrupt preexisting inhibitory ones.

tidine kinase enhances its DNA-binding activity. This two-domain RR comprises a typical, CheY-like regulatory domain and a small effector DNA-binding domain composed of a compact four-helical bundle, the middle two of which constitute the helix−turn−helix DNA-binding motif. The "recognition" helix of the HTH packs mostly against the α3−*β*4 loop of the regulatory domain in a manner that would preclude DNA binding. Phosphorylation must induce structural changes to relieve the inhibition of the DNA binding function of NarL but probably not directly through residues in α-helix 4 and *β*-strand 5, as in other RRs. Here, the guanidinium group of Arg75 from the α3−*β*4 loop makes hydrogen-bond contact to the main-chain carbonyl of Gly102, at the base of helix α4. This helix makes few contacts to the main body of the RR domain. Phosphorylation of Asp59 might be transduced to the effector domain through conformational changes in α4, which are propagated to the α3−*β*4 loop and thus the effector domain through this arginine.

Similarly, mutations of Asp75 in the α3−*β*4 loop of the regulatory domain of Spo0A can both increase and decrease the transcription activation properties of Spo0A. A single amino acid deletion of *B. subtilis* Spo0A at position Asp75 renders Spo0A constitutively active.[130] However, mutation of Asp75 to serine confers a sporulation-deficient phenotype on *B. subtilis* cells harboring this mutation, and in vitro

transcription stimulation was drastically reduced in comparison to wildtype.[131] It is thus possible that the α3−β4 loop of the regulatory domain of Spo0A contacts and inhibits the attached effector domain until the inhibition is relieved by phosphorylation some 20 Å away. In the non-phosphorylated domain-swapped dimer structure of the RR domain of Spo0A from *B. stearothermophilus*,[132] the guanidinium group of Arg71 in the α3−β4 loop makes a hydrogen bond to the main-chain carbonyl of Leu98, which, as in NarL, is situated at the C-terminus of α-helix 4. In the structure of pSpo0A, this interaction is not observed at all. Indeed, in two out of the four molecules in the crystallographic asymmetric unit, most of the α3−β4 loop is so badly disordered that it cannot be modeled, and in all four molecules, the side chain of Arg71 is completely solvent-exposed.[116]

Thus, in at least two instances, inhibitory protein/protein contacts between the effector domain and, apparently, residues from the α3−β4 loop in the regulatory domain must be dissociated in the activation process, prior to functional, associative interactions. In *B. subtilis* and *E. coli*, 80% and 70%, respectively, of all response regulators encode either arginine or lysine at the equivalent position to Arg71 (in Spo0A) in the variable α3−β4 loop. The frequency that this residue occurs at this position may imply that for many response regulators, different surfaces are used in inhibition and activation mechanisms. Since most RRs are transcription factors and the majority appear to bind DNA target sequences that occur in pairs close to the promoters of genes they regulate (NarL, FixJ, PhoB, OmpR, Spo0A), dimerization might be required for high-affinity DNA binding utilizing residues from α4/β5/α5 in dimerization. The structural details of these activation mechanisms await structures of intact RRs, both in the resting and active conformations.

2. Histidine Kinases

While much has been learned recently about response regulator function, progress has also been made in the other half of the two-component system, the histidine kinases. HKs are diverse in size, are modular, and exhibit a minimal domain structure of sensing, dimerization, and catalytic domains, although great elaboration of this basic scheme is possible[94] including the presence of phosphatase domains.[133] Many HKs are membrane-bound with periplasmic sensing and cytosolic catalytic domains, but others are purely cytosolic. The sensing domains share little meaningful sequence similarity, presumably since they have to recognize very specific stimulants. However, while little is understood of the sensing mechanism or how the signal is transduced to the catalytic core, recent structures of catalytic domains of CheA[134] and EnvZ[135] have begun to reveal the nature of nucleotide binding and autophosphorylation and highlight differences with other kinases so far studied.

The catalytic kinase core of ∼350 amino acids binds ATP and directs *trans*-autophosphorylation. The overall fold of the HK catalytic domain—the Bergerat fold[136]—is quite unlike those of the eukaryotic protein kinases whose structures are known but belongs to a subfamily of ATP-dependent enzymes which includes DNA gyrase B,[137] Hsp90,[138] and MutL,[139] called the GHKL family[140] (Figure 13). There is little primary sequence homology between the members of the GHKL family; yet the cores of their ATP-binding domains are essentially identical. The Bergerat ATP-binding fold is an α/β bilayered sandwich consisting a four-stranded antiparallel β-sheet flanked on one face by three α-helices with ATP binding between the α-helices and β-sheet. Five noncontiguous conserved regions which characterize histidine kinases, the N, G1, F, G2 and G3 "boxes", are involved in nucleotide binding in the catalytic domain. A long (20−30 residue) flexible loop, the "ATP lid", encloses the ATP-binding site of DNA gyrase in the presence of nucleotide, with the ATP phosphates completely buried,[137] but is disordered when nonnucleotide inhibitors are bound.[141] In EnvZ, the bound nucleotide is exposed since the ATP lid extends away from the rest of the structure[135] and the ATP phosphates point away from the core of the protein. This observation may reflect the requirement of kinases to transfer a phosphoryl group to another protein, which is not a function of the ATPases.

The histidine-substrate H box is situated normally in the dimerization domain, the helical bundle common to HKs, and phosphotransferase domains. To effect *trans*-phosphotransfer, two molecules of the catalytic domain must flank the dimerization domain such that the ATP-binding pocket faces the substrate histidine. The structures of fragments of EnvZ[135,142] do not address the spatial relationship between the H box of the dimerization domain and the catalytic domain, and thus, how phosphoryl transfer occurs from ATP to the H box is unknown. Uniquely in CheA, the substrate histidine is not situated in the dimerization domain, but in another four-helical bundle, the histidine phosphotransfer domain and similarly the spatial relationship between ATP-binding and substrate domains are unknown.[134] The ATP-binding pocket of the catalytic domain of CheA points away from the dimerization domain, and thus, substantial conformational changes must occur to bring together ATP and the H-box to effect phosphotransfer in other kinases. Primary sequence analysis of the GHKL proteins does not completely explain the differences in their functions. A conserved catalytic glutamate in the ATPase members of the GHKL superfamily is replaced with asparagine in most but not all of the histidine kinases. SpoIIAB and RsbW, for instance, encode a glutamate at this locus and show strong further sequence similarity to HKs but are not ATPases and do not phosphorylate on histidine residues but instead on serines and threonines in their substrates.[143,144]

3. Histidine-Containing Phosphotransfer Domains

Details of the mechanism of phosphoryl transfer and how crucial HK−RR fidelity is maintained in bacteria that encode 35 HK−RR pairs have been revealed by the crystal structure determination of the transient complex between Spo0F and Spo0B.[145] Phosphoryl transfer is more complicated in the initiation of sporulation in *Bacilli* since the ultimate RR, Spo0A, is not directly phosphorylated by a HK

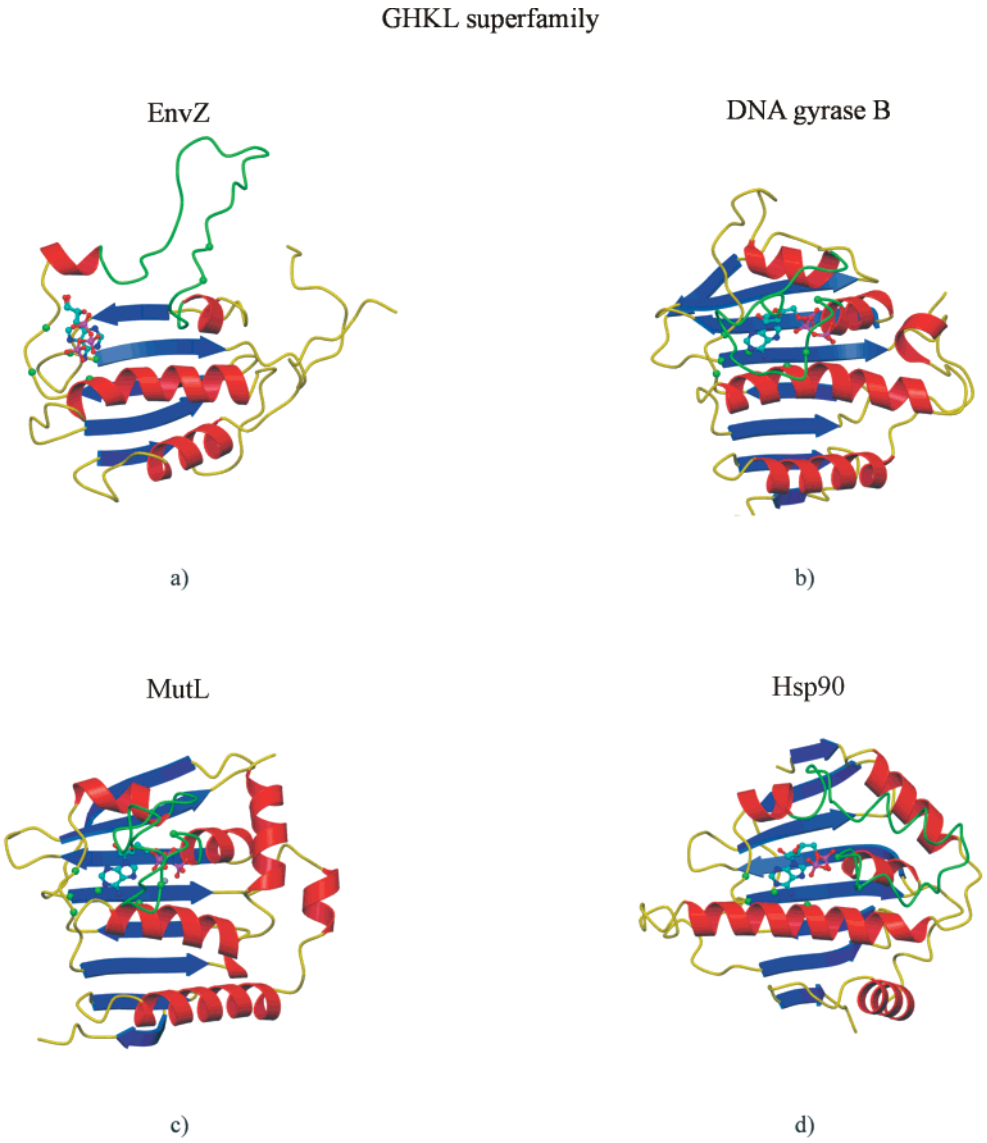


Figure 13. GHKL ATPase/histidine kinase superfamily. Ribbon diagrams of the ATP-binding domains of (a) the histidine kinase EnvZ,[135] (b) the topoisomerase DNA gyrase B,[137] (c) the DNA repair enzyme MutL,[139] and (d) the chaperone Hsp90.[138] The bound nucleotide is shown as a ball-and-stick model, α-helices are colored red, β sheet blue, and coil regions yellow. The ATP lid and Cα atoms from the conserved N, G1, G2, and G3 boxes are colored green. Note how in parts b and c the ATP lid folds over the bound nucleotide. The nucleotide is more exposed in parts a and d. The nucleotide conformations in parts b, c, and d are similar and different than that in part a. The short helices within the ATP-lid of MutL and Hsp90 have been drawn as a random coil for clarity.

but by an expanded phosphorelay including another RR, Spo0F, and a histidine phosphotransferase, Spo0B, which mediates the freely reversible phosphoryl transfer from Spo0F to Spo0A.[146] Dimeric Spo0B is composed of two domains,[147] a kinase-like domain lacking the five signature residues of the HKs that are implicated in ATP binding and a dimerization core, a four-helical bundle that closely resembles the dimerization domains of EnvZ and CheA, and the histidine phosphotransferase domains of ArcB,[148] CheA,[149] and YPD1.[150] The substrate histidines are each located on a solvent-accessible surface of these helical bundles, whereas the substrates of Ser/Thr/Tyr kinases tend to be found in more extended conformations.[151]

Little conformational change takes place on Spo0B binding to Spo0F,[145] consistent with the function of Spo0B in acting as a phosphoryl shuttle. Whereas there are considerable differences in the measurement of the free energy of hydrolysis of phospho−aspartates in proteins in comparison to small molecules, there are negligible differences in the hydrolysis of phospho−histidines within proteins in comparison to small molecules.[152] Large-scale conformational changes are thus unlikely to be driven by hydrolysis of the high-energy phosphoramidate bond but instead the phosphorylation of its target RR. The dimerization helical bundle of Spo0B mediates most of the hydrophobic interactions to Spo0F, which utilizes predominantly residues from the $\beta-\alpha$ loops surrounding the RR aspartyl pocket and α-helix 1. Thus, the four-helical bundle of Spo0B becomes a six-helical bundle by the packing of two molecules of Spo0F per Spo0B dimer. The key residues in phosphoryl transfer, His30 of Spo0B and Asp54 of Spo0F, are ideally situated and orientated to perform such a task (Figure 14).

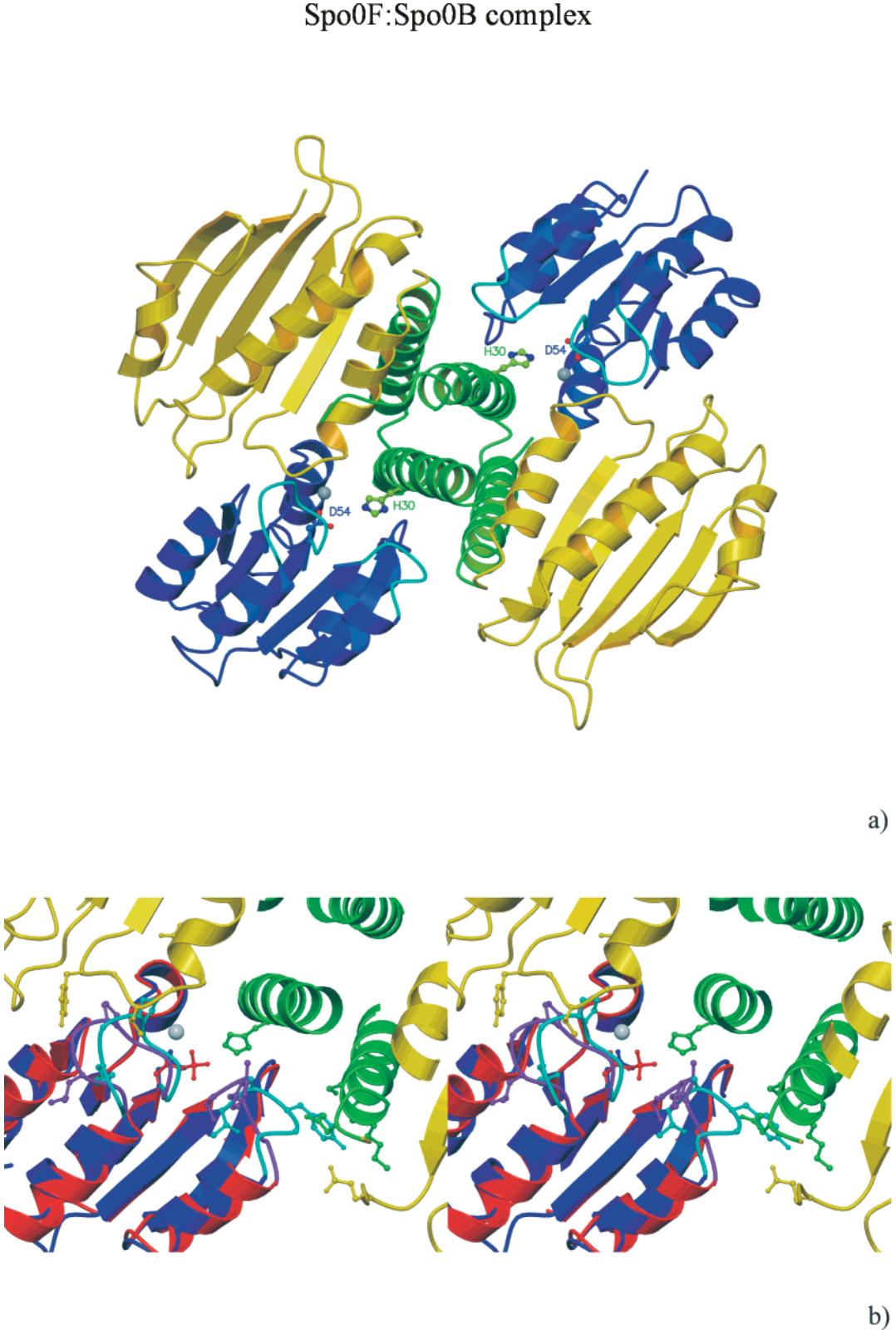


Figure 14. Maintenance of fidelity in two-component signal transduction. The transient complex between Spo0F and Spo0B[145] is drawn as a ribbon diagram (a), with the Spo0B dimer occupying the central area, and molecules of Spo0F, blue, to the left and right. The kinase-like and the histidine phosphotransferase domains of Spo0B are colored yellow and green, respectively. The view is approximately down the molecular dyad and illustrates how His30 of Spo0B and Asp54 of Spo0F are aligned perfectly for phosphoryl transfer, aided by the nearby magnesium ion. (b) Stereodiagram of the details of the interaction of Spo0F with Spo0B and with pSpo0A (red) superimposed on the structure of Spo0F. Some of the residues involved in protein/protein contacts are drawn as ball-and-stick models, and the β3-α3 and β4-α4 loops are colored cyan (Spo0F) and purple (pSpo0A). Phe86 of pSpo0A shields the phosphoryl group from nucleophilic attack, while the equivalent residue in Spo0F, Tyr84, is involved in Spo0F/Spo0B contacts. On phosphorylation, the methylene group of Ala85 of Spo0A clashes with the imidazole ring of His30 of Spo0F, promoting the dissociation of the complex.

While the aromatic residue of the aromatic switch (His101 of Spo0F) is not involved in intermolecular protein/protein contacts, loops $\beta3-\alpha3$ and the hypervariable $\beta4-\alpha4$ of Spo0F make several hydrophobic contacts to Spo0B. The $\beta3-\alpha3$ loop, which stabilizes both the bound metal ion and the phosphoryl group,[153] and the $\beta4-\alpha4$ loop rearrange significantly on phosphorylation of Spo0A and FixJ.[111,116,132] Structural comparisons of the Spo0F/Spo0B complex and pSpo0A predict that these loops are both pulled away from the Spo0B helical bundle on phosphorylation. Furthermore, the rearrangement of the $\beta4-\alpha4$ loop in Spo0A, caused in part by the hydrogen bond between Thr84 and the phosphoryl group, is pushed toward Spo0B such that the methylene side chain of the conserved alanine at position 85 of pSpo0A sterically clashes with the imidazole ring of His30 of Spo0B (Figure 14). Taken together, these movements dissociate the interactions between the RR and phosphotransferase domain, liberating the phosphorylated RR. Residues from the $\beta4-\alpha4$ loop in particular may play dual roles, not only acting as specificity determinants[145] but also influencing the half-life of the phosphorylated form of the RR by shielding the phosphoryl group from nucleophilic attack. Signal transduction fidelity between HK and RR is maintained by subtle variations in sequence in both HK and RR and by differences in helical bundle length and orientation in the HK. Complicated yet understated molecular recognition patterns are crucial in avoiding HK−RR "cross-talk" but await further structural studies of different interacting systems for these rules to be confirmed.

B. The Haloacid Dehalogenase Superfamily

The haloacid dehalogenase (HAD) superfamily of hydrolases includes members such as phosphatases, hydrolases, and dehalogenases, and their shared feature is the utilization of an aspartylester enzyme intermediate, followed by ester transfer. For instance, a phospho ester is part of the reaction catalyzed by the HAD superfamily member phosphoserine phosphatase. In contrast, L-2-haloacid dehalogenase, which catalyses the hydrolytic dehalogenation of L-2-haloalkanoates to the corresponding D-2-hydroxyalkanoates, utilizes a covalent acid ester during its reaction mechanism. Hydrolysis of the enzyme-ester of L-2-haloacid dehalogenase is achieved by nucleophilic attack by water of the aspartyl C^{γ}, whereas as in the RRs, the P-type ATPases and phosphatases are dephosphorylated by nucleophilic attack of water on the phosphorus atom. The catalytic domains of P-type ATPases, which transport cations across membranes across a concentration gradient, are also members of this extended superfamily while displaying little meaningful sequence homology.[154] The P-type ATPases are phosphorylated on a conserved aspartic acid as part of their reaction cycle and couple the free energy of hydrolysis of the acyl phosphate to ion transport. Despite a lack of functional and sequence homology, the HAD and RR superfamilies exhibit a common active site organization.[113] Each comprises a core of a multistranded parallel β-sheet, where the nucleophilic aspartic acid is found at the C-terminus of the central β-strand. Phosphorylation of RRs is dependent upon a conserved lysine and a magnesium ion, which is coordinated by a carbonyl oxygen, two aspartyl carboxylate oxygens, and three water molecules, one of which is held in position by a third aspartate. The conserved lysine and aspartates are present in P-type ATPases and all but two members of the HAD superfamily, and these two enzymes are Mg^{2+}-independent.[113] Since the function of P-type ATPases is regulated by phosphorylation, the recent crystal structure of the Ca^{2+}-bound form of the calcium pump of the sarcoplasmic reticulum[155] will be considered further (Figure 15a).

The calcium pump is an ATP-dependent integral membrane protein which pumps the calcium released by muscle cells on contraction back into the sarcoplasmic reticulum, thus relaxing muscle cells. It comprises 10 transmembrane (TM) spanning helices and three cytoplasmic domains: the N-domain, which binds ATP; the P-domain, which resembles the core of RRs and carries the phosphorylation site (Asp351); and the A-domain, whose function is presumed to involve anchoring domain N. The N-domain forms an open cap above the critical P-domain residue, Asp351, such that considerable relative domain movements must occur for phosphoryl transfer, and calcium binding alone is insufficient to catalyze this. Calcium binding to four TM helices appears to losen the interactions between the three cytoplasmic domains while altering the spatial relationship between the P-domain and two TM helices. The N-domain lid will close over the P-domain and in doing so phosphorylate Asp351 in a manner reminiscent of the transient interaction between RR and HK. An uncharacterized series of conformational changes must then take place to occlude further Ca^{2+} ions from the cytosol and to alter the accessibility of the bound Ca^{2+} to the lumenal spaces. Phosphorylation of the P-domain must somehow be communicated to the calcium binding sites, probably through structural changes in a loop connecting the base of the P domain to one of the TM helices. Consequently, new intramolecular contacts must be made or exisiting ones broken. Once calcium has been released, the phosphoenzyme can be hydrolyzed by water, ready for another cycle of Ca^{2+} transport. One approach to make this conformational change accessible to structural analysis might be the use of phosphoryl group analogues, such as BeF_3^-, as utilized so successfully in the study of RR activation. Residues Thr625, Lys684, Asp703, and Asp707 of the P-domain of the calcium pump have all shown, by a variety of techniques, to be involved in phosphoryl transfer to Asp351. These residues have direct counterparts in the aspartyl pocket of RRs (Thr84, Lys106, Asp9, Asp10), which are crucial in phosphorylation of Asp55 (Figure 15b,c), but there is no aromatic residue in the calcium pump equivalent to Phe103. Despite the close similarity to RRs, phosphorylation of Asp351 in the calcium pump must be propagated to the calcium binding sites in a different and so far uncharacterized manner to signal transduction in the RRs.

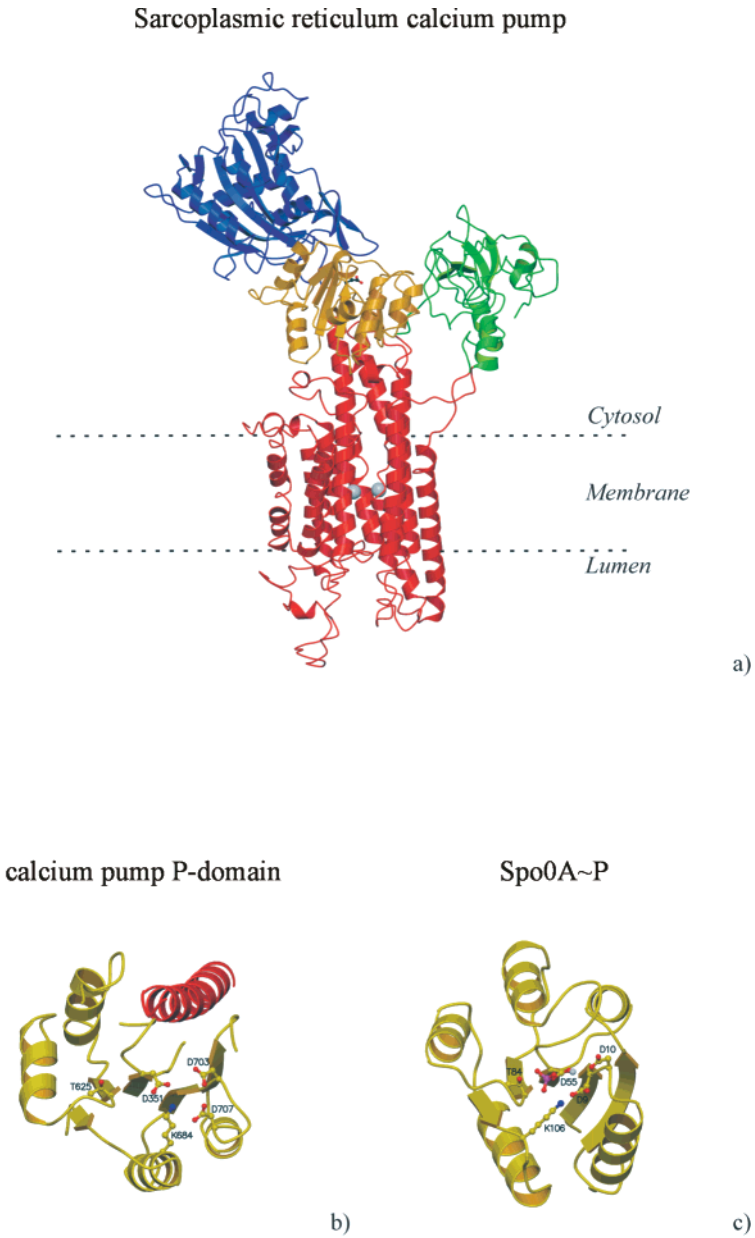


Figure 15. Calcium pump.[155] An overall view of the sarcoplasmic reticulum calcium pump (a), illustrating the 10 TM helices (red), the 2 bound calcium ions (white), and the 3 cytoplasmic domains, N (top left, blue), A (right, green), and P (center, yellow). The approximate position of the boundaries of the lipid bilayer are drawn with straight lines. Phosphorylation at Asp351 in the P-domain is required for transport of the two calcium ions, some 50 Å away, but the nature of the necessary conformational changes is as yet unknown. In part b, the core of the P-domain of the calcium pump is a five-stranded β-sheet flanked by α-helices. One of the TM helices is colored red, and somehow conformational changes to the Ca^{2+}-binding sites are signaled through this. (c) Spo0A phosphoryl site. Although unrelated in sequence to the core of response regulators, the critical aspartate is surrounded similarly by a lysine, two aspartates, and a threonine, which are requisite in phosphorylation.

C. The Sugar Phosphotransferase System

The sugar phosphotransferase system (PTS) plays a central role in the metabolism of many species of bacteria by catalyzing phosphoryltransfer from PEP to metabolic carbohydrates after crossing the cell membrane.[156] Like the sporulation phosphorelay, the PTS relies upon the formation of successive complexes between phosphoryl donors and acceptors. In glycolysing bacteria, ATP levels are high and as a result the histidine-containing phosphocarrier protein, HPr, is phosphorylated on Ser46 by the bifunctional ATP-dependent HPr kinase/phosphatase. Down regulation of the PTS follows: if HPr is phosphorylated on Ser46, His15 cannot accept a phosphoryl group from the upstream, PEP-dependent enzyme I (EI) and phosphoryltransfer down the PTS to the sugar is inhibited. In starving bacteria, the intracellular ATP pool is depleted and HPr does not become phosphorylated on Ser46 by the HPr kinase/phosphatase. HPr can then transfer a phosphoryl group from PEP to the PTS and ultimately to the imported sugar. On another level of control, HPr also influences catabolite repression of sugar uptake. P-Ser46 HPr can form a complex with the catabolite control protein A, thus binding to specific DNA sequences upstream of the promoters of certain genes involved

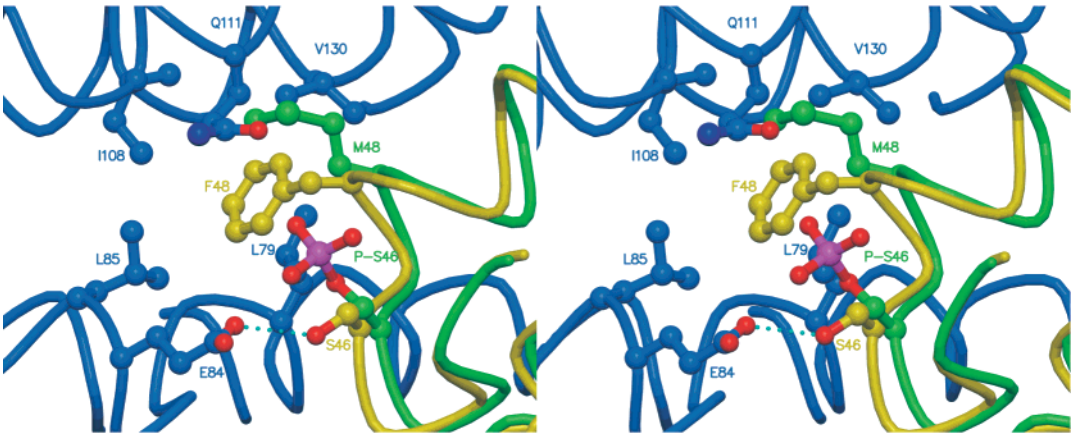


Figure 16. Hpr/EIN interface.[157] A stereoview of the hydrophobic EIN/HPr interface, comprising six residues of EIN (Leu79, Glu84, Leu85, Ile108, Gln111, and Val130, blue) and Phe48 (or Met48) of HPr (yellow). A hydrogen bond between Ser46-HPr and Glu84-EIN mediate, in part, the formation of this complex. However, when Ser46 of HPr is phosphorylated (green), HPr becomes a poor substrate for the transfer of a phosphoryl group, between His15 of HPr and His189 of EIN. Electrostatic repulsion between the negatively charged phospho−serine and glutamate pushes the phospho−serine toward the position of the hydrophobic pocket, displacing the phenylalanine (or methionine, depending on the bacterial source) and disrupting the formation of the HPr/EIN interface.

in carbohydrate degradation, thereby affecting directly their transcription.

Several structures of PTS proteins have been determined in the past decade, but from this review's perspective, details of phosphoryl transfer and the effect on HPr by phosphorylation on Ser46 have most recently been elucidated. The first structure to be solved of a protein-to-protein phosphoryl transfer complex of the PTS, that of HPr and the N-terminal domain of enzyme I (EIN),[157] is complemented by the structure of pSer46 HPr.[158] As is the case with the Spo0F/Spo0B complex, the conformations of HPr and EIN in the free state are barely altered by complex formation and the intermolecular interface requires satisfactory packing of hydrophobic patches in α-helices from both molecules. In the EIN:HPr complex, the side chains of Ser46 of HPr and Glu84 of EIN are in hydrogen-bond contact with one another.[157] The structural effects of phosphorylation of Ser46 are not extensive, and the structure of pSer46 HPr closely resembles the native form.[158] The phosphoryl group is situated at the N-terminal end of an α-helix, and one of the phosphoryl oxygens is in hydrogen-bond distance of the amide nitrogen of residue 49. It appears that the function of phosphorylation of Ser46 is to destabilize the EIN/HPr complex through electrostatic repulsion between the phosphoryl group and Glu84, which pushes phospho-Ser46 into and disrupting a hydrophobic interface which would ordinarily form in the complex of EIN and HPr (Figure 16). The effects of phosphorylation can be mimicked by mutation of Ser46 to aspartic acid.[159] Central in this interface is Phe48 of HPr. This same residue also forms part of the interface between HPr and the adjacent, downstream element, $IIA^{Glucose}$,[160] but in this instance Ser46 of HPr makes no interactions with $IIA^{Glucose}$. The complex between HPr and EIN is sensitive to the phosphorylation state of Ser46 HPr and therefore the levels of the ATP pool as a regulatory mechanism. The downstream component of the PTS, $IIA^{Glucose}$, has no such requirement, and it is impervious to the phosphorylation state of Ser46.

Where EIN is phosphorylated on the Nϵ2 atom of His189,[161] as are the HKs,[94] HPr accepts the phosphoryl group onto atom Nδ1 of His15.[157] In the structure of the EIN/HPr complex, His189 of EIN makes no contact to HPr such that the nitrogen atoms involved in phosphoryl transfer are ∼8 Å apart. These atoms must move closer to effect phosphoryl transfer. The EIN−HPr transition state has thus been modeled,[157] which does not globally affect the conformation of either protein in the complex. As is the case with HK−RR phosphoryl transfer, the transition state comprises a trigonal bipyramid pentacoordinated phosphoryl group with apical donor and acceptor atoms and the three oxygen atoms situated in the equatorial plane. The essential magnesium ion in HK−RR phosphoryl transfer is not required in phosphorylation reactions between EIN and HpR. A conserved lysine (Lys69) and threonine (Thr16) residue and backbone amide groups (Thr16 and Arg17) adjacent to the site of phosphorylation (His15) in HPr would appear to stabilize the transition state of phosphoryl transfer in favor of phospho-HPr relative to phospho-EIN. Similarly, the phosphotransfer transition state between Spo0B (His30) and Spo0F (Asp54) is stabilized by an invariant lysine (Lys104), the threonine (Thr82) of the aromatic switch, and adjoining amide groups from Met55 and Lys56 in Spo0F.[145] His189 of EIN accepts a hydrogen bond from the hydroxyl proton of Thr168, increasing the reactivity of this protonated phosphoramidate,[162] stimulating the appropriate unidirectional phosphoryl flow through the PTS. However, there are no

proton donors sufficiently close to His30 of Spo0B to activate the phosphoramidate, which may explain why Spo0B can equally distribute a phosphoryl group to both Spo0A and Spo0F.[163]

D. Anti-Sigma Factor Antagonist

A final example of the structural consequences of phosphorylation comes in the modulation of sigma factor activity, the best characterized of which is σ^F, a Bacillus sporulation-specific alternative sigma factor. Compartment-specific gene expression is attained by the complicated functions of the anti-sigma factor, SpoIIAB, which can form protein/protein complexes and, although resembling a histidine kinase in its amino acid sequence, acts as serine kinase. In the predivisional and mother cells, SpoIIAB can form a complex with the anti-anti-sigma factor SpoIIAA in a reaction dependent upon ATP.[164] In doing so, SpoIIAB phosphorylates its substrate, SpoIIAA on Ser58, causing dissociation of the SpoIIAA/SpoIIAB complex.[165] The liberated SpoIIAB is then free to form an alternative complex with σ^F, thus inhibiting it from interacting with core RNA polymerase.[166] However, in the developing spore, phosphorylated SpoIIAA is rapidly dephosphorylated by the septum-associated PP2C-type phosphatase,[167] SpoIIE, and thus, SpoIIAA and SpoIIAB form persistent complexes and σ^F is left free to direct specific gene expression.[168]

The NMR structure of the anti-anti-sigma factor, SpoIIAA,[169] bears a passing resemblance to the canonical RR structure, being comprised of a central, multistranded β-sheet flanked on one side by two α-helices. Closer inspection of the structure reveals a different topology to the RR, and the residue which is phosphorylated by SpoIIAB, Ser58, is situated at the N-terminal end of α-helix 2, the side chain of which makes no interactions with the main body of the protein. Presumably, the N-terminal end of the α-helix will stabilize the phosphoserine, through the partial positive charge of the helix dipole. It is possible that phosphorylation either induces drastic conformational changes, which prevents binding to SpoIIAB, or that the phosphoryl group itself merely sterically and electrostatically interferes with complex formation. The recently determined crystal structure at 1.16 Å resolution of phosphorylated SpoIIAA, the highest resolution structure of any phosphoprotein, distinguishes these possibilities.[170] The phosphoryl group makes no contact with any other atom from the same molecule of SpoIIAA and is completely accessible to solvent (Figure 17). The side chains which flank Ser58 (Ser57 and Asp59) reach around behind Ser58, making Ser58 in an accessible conformation for the action of the SpoIIAB kinase. Comparison to the native form of SpoIIAA reveals that phosphorylation imposes only minor conformational changes that are limited to the immediate vicinity of Ser58. As in HPr, mutation of Ser58 to aspartic acid effectively mimics phosphorylation.[166] This lack of structural rearrangement is similar to the consequence of phosphorylation of Ser46 of Hpr in the PTS. It would thus appear that the role of phosphorylation in the establishment of compartment-specific gene expression is to provide a negatively charged steric impediment to protein/protein complex formation between SpoIIAA and SpoIIAB, thus permitting sequestering of σ^F by SpoIIAB.

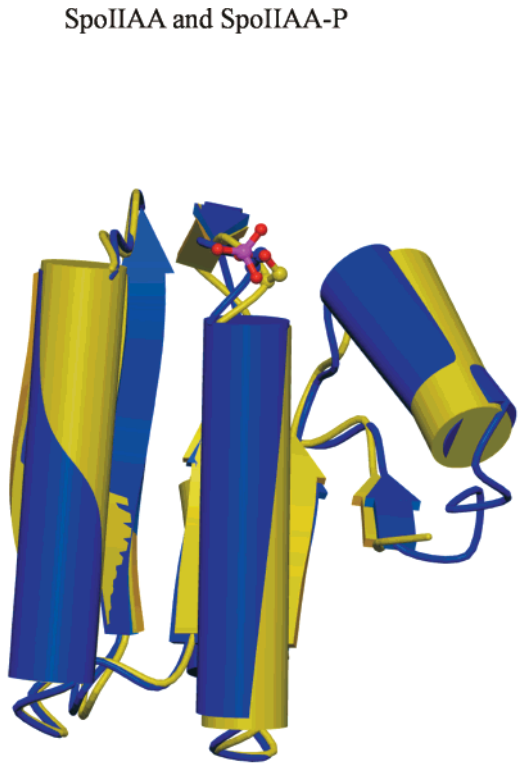


Figure 17. Effects of phosphorylation on SpoIIAA.[170] The structures of SpoIIAA (yellow) and SpoIIAA-P (blue) are compared, with secondary structure elements drawn as arrows (β-strands) and rods (α-helices). The residue that is phosphorylated by the serine kinase SpoIIAB, Ser58, is drawn as a ball-and-stick model in both structures. The phosphoryl group of SpoIIAA-P makes no interaction with any other amino acid and is presented to solvent in such a manner as to exclude binding by SpoIIAB.

One example where multiple phosphorylation sites are found on one protein is in the regulation of the σ^B-dependent stress response of Gram-positive microorganisms, such as *B. subtilis*, in a minikinase/phosphatase cascade of a type so frequently found in higher organisms. Here, σ^B can be sequestered in nonproductive protein complexes with a histidine kinase, RsbW the substrate for which is RsbV.[171] The antagonistic activity of a specific phosphatase, RsbU, brings the system under complete enzymatic control.[172] This mechanism is parallel to that described above for σ^F, except that a further level of regulation exists, another upstream kinase/phosphatase system that transduces other stress signals to σ^B. Here a signal-transducing protein RsbR, which has sequence homology to SpoIIAA, is phosphorylated not on one serine but two threonine residues, 171 and 205, the latter of which is the equivalent to Ser58 of SpoIIAA.[173] One can envisage that phosphorylation at Thr205 abrogates binding of RsbR to the RsbT kinase, analogous to the effects of phosphorylation of SpoIIAA binding to SpoIIAB. In SpoIIAA, the equivalent residue to Thr171 is His23, which is situated at the N-terminal end of the α-helix 1 adjacent to Ser58. The histidine side chain makes no

interaction with Ser58 or pSer58. What the reason for and consequence of phosphorylation on the second site, Thr171, is unknown.

E. Phosphatases

Control over these phosphorylation-dependent activation reactions is achieved by the action of phosphatases, which in the prokaryotic world are poorly characterized structurally. The SpoIIAA-P specific phosphatase, SpoIIE, is a member of the Mn^{2+}-dependent eukaryotic PP2C family of phosphatases, one structure of which is known. In the sporulation phosphorelay, Spo0F is subject to the activity of at least one of the 11 members of the Rap family of phosphatases, which are inhibited by short, specific peptides,[174] whereas Spo0A is dephosphorylated by a completely dissimilar phosphatase, Spo0E. No structural information is available for either of these distinct classes of phosphatase nor any for the unrelated chemotactic phosphatase, CheZ or the phosphatase domain of EnvZ. How these different phosphatases achieve the same end, namely, aspartyl phosphate dephosphorylation, is unknown.

The recently determined structure of phosphoserine phosphatase (PSP) from *M. jannaschii*[175] is similar to the catalytic domain of the HAD superfamily, of which it is a member, and closely resembles the core of response regulators. Where the Ca^{2+} pump[155] could be said to be in an open conformation, PSP is in a closed sate. Asp55, Asp10, Thr84, and Lys106 in Spo0A have direct equivalents in PSP: Asp11, Asp167, Ser99, and Lys144. Furthermore, Thr84 in the RR family is normally followed by a small amino acid, typically Ala, which allows access of the kinase/phosphotransferase domain to the nucleophilic aspartic acid.[145] Similarly, a conserved glycine residue succeeds Ser99 in PSP, presumably to allow access of the substrate to the active site. The magnesium ion in PSP[175] is stabilized by an almost identical hydrogen-bond network, as observed in the structures of metal-complexed response regulators.[116,120,124] A bound phosphate group makes interactions with PSP residues that are similar to those observed between phosphoryl and berrylofluoryl groups to conserved RR residues,[111,116,120] and thus, one might expect the catalytic mechanism of PSP to be similar to that of RR phosphotransfer. Indeed, the half-life of the phosphorylated form of response regulators is affected by the chemical nature of the amino acid +2 from the nucleophilic aspartate, such that short polar residues so located favor a short half-life, which can be extended by substitution with bulky, hydrophobic residues.[153] Some response regulators have been described as harboring auto-phosphatase activities, since the half-life of the phospho form can be extended by denaturing the protein. The phospho forms of CheY and CheB are both short-lived ($t_{1/2} \sim 1$ s) and have Asn and Glu, two residues from the nucleophilic aspartate, respectively. The presence of an aspartic acid at the equivalent position in PSP implies this residue has a key role in catalysis, in addition to a short half-life of the phosphoenzyme ester intermediate. Fascinating parallels in phosphorylation chemistry between PSP and RRs are likely to emerge soon, with the use of BeF_3 as a phosphoryl group mimic in PSP.

IV. Summary and Conclusions

The number of phospho−protein structures has increased considerably since past reviews.[11,176] Likewise the number of mechanisms by which phosphorylation may influence proteins has also increased. Responses to phosphorylation may vary from the 10^5 fold increase in catalytic efficiency with CDK2/cyclin A and $>10^6$ fold inhibition of IDH to a more gentle modulation of 10−40-fold enhancement in FliM binding by CheY or DNA binding by Spo0A. There seems to be no universal response that allows a general mechanism for control by phosphorylation but a series of different responses conferred or induced by the addition of the phosphoryl group. A summary of the structural results is given in Table 1.

Not surprisingly, electrostatic effects dominate the response of proteins to phosphorylation. Repulsive electrostatic effects are observed for a number of systems such as: IDH where the phosphoryl serine blocks access to the catalytic site of the negatively charged substrate isocitrate; the K^+ channel inactivation domain where the phosphorylatable residues are close to two glutamates so that phosphorylation promotes loss of structural stability; CDK2 where the phosphorylatable threonine is close to a glutamate in the non-phospho structure and hence on phosphorylation the segment carrying phospho−threonine becomes disordered; and in the HPr/EIN complex where the phosphorylatable serine, Ser46, is close to a glutamate from EIN so that phosphorylation on Ser46 is likely to lead to electrostatic repulsion and inhibit association of the HPr/EIN complex. Phosphorylation of SpoIIAA by SpoIIAB leads to dissociation of the SpoIIAA/SpoIIAB complex. Electrostatic effects also lead to positive energetic interactions in many systems, for example, mammalian glycogen phosphorylase where the phosphoryl serine contacts two arginines (Arg43′ and Arg69); yeast phosphorylase where the phosphoryl threonine contacts two arginines (Arg309 and Arg310); the protein kinases where in CDK2 and ERK2 the activation segment phospho−threonine interacts with three arginines and in IRK and LCK where the phosphoryl-tyrosine interacts with one or two arginines; and the phospho−tyrosine/SH2 domain interactions in Src and in the STAT proteins where the phosphoryl group interacts with two arginines. In the prokaryotic response regulators there are no arginine clusters but a contact from the phosphoryl group to a single lysine and a metal ion in the phospho forms of Spo0A, FixJ, and NtrC. In the prokaryotic systems the conformational responses induced by phosphorylation are local, in contrast to the eukaryotic systems listed above where a tight binding site appears necessary to produce long-range effects. The participation of the helix dipole in phospho-recognition sites for proteins controlled by phosphorylation are not as numerous as in the catalytic sites of enzymes that utilize phosphate-containing substrates or cofactors, but helix dipole interactions are observed for the phosphoryl groups of glycogen phosphorylase, isocitrate

Table 1. Summary of Structural Results on Phospho−Proteins

protein (PDB code for phosphoprotein)	phosphoryl residue	contacts	mechanism	ref
		eukaryotic systems		
glycogen phosphorylase (1GPA)	Ser14	Arg43′, Arg69	allosteric activation	81, 177
yeast phosphorylase (1YGP)	Thr-10	Arg309, Arg310	allosteric activation and displacement of allosteric inhibitor	88
CDK2/cyclin A (1JST, 1QMZ)	Thr160[a]	Arg50, Arg126, Arg150	phosphoryl group acts as an organizing center to locate residues important in substrate binding	34, 36
ERK2, P38 (2ERK, 1CM8)	Thr183[a]	Arg68, Arg146, Arg170	as in CDK2 pThr183 acts as an organizing center that	47, 51
	Tyr185	Arg189, Arg192	assists creation of correct conformation of the activation segment	
insulin receptor tyrosine kinase (1IR3)	Tyr1158	none	localization of activation segment for substrate recognition	55
	Tyr1162	Arg1164		
	Tyr1163[a]	Arg1155, (Arg1131)		
VEGFR2/PDGFR (1VR2)	Tyr1059[a]	disordered	only 10-fold activation on phosphorylation: activation segment disordered in both non-phospho and phospho forms	57,58
Src family kinases (2SRC, 1QCF, 2PTK)	Tyr527	Arg155, Arg175, Ser177, Glu78(NH), Thr179	interactions with SH2 domain hold Src kinase in inactive state	59−61
LCK (3LCK)	Tyr394[a]	Arg363, Arg387	localizes activation segment in correct conformation	62
STAT proteins (STAT3β numbering) (1BF5, 1BG1)	Tyr705	Lys591, Arg609, Ser611 Ser613	intersubunit contact with the SH2 domain of the other subunit promotes dimerization and DNA binding	72, 73
pKID domain of CREB/KIX domain of CBP (1KDX)	Ser133	Tyr658, (Lys662), helix dipole	phosphorylation promotes association of pKID/KIX	74
K+ channel inactivation domain (1B4G, 1B4I)	Ser8 Ser15 Ser21	none	phosphorylation promotes disorder of domain structure	75
phenylalanine hydroxylase (1PHZ)	Ser16	disordered		10
		prokaryotic systems		
isocitrate dehydrogenase (4ICD)	Ser113	helix dipole	electrostatic and steric blocking of catalytic site	12
Spo0A (1QMP)	Asp55	Thr84, Lys106, Ca^{2+}, Ala85(NH), Ile157(NH)	promotion of dimerization; association with RNA polymerase; and dissociation of inhibitory interactions with effector domain	116
FixJ (1D5W)	Asp54	Thr82, Lys104, Ala83(NH), Arg56(NH)	promotion of dimerization; and dissociation of inhibitory interactions with effector domain	111
CheY (1DJM, 1C4W, F4V)	Asp57	Thr87, Lys109, Mg^{2+}, Ala85(NH), Asn59(NH)	promotion of FliM binding	118−120
NtrC (1DC8)	Asp54	not reported	exposure of residues to interact with effector domain	117
HPr/EIN (3EZB, 1FUO)	His46	Gly49(NH) helix dipole	steric/electrostatic impediment to protein/protein interactions	157, 158
SpoIIAA (1H4X, 1H4Z)	Ser58	none	steric/electrostatic impediment to protein/protein interactions	170

[a] The equivalent residue in cAPK is Thr197.

dehydrogenase, the KID/KIX complex, and HPr, the first example of a helix dipole/phosphate interaction in prokaryotes.

In the tight phosphoryl binding site proteins, the phosphoryl group acts as an organizing center, affecting not only the conformation of regions in the immediate vicinity of the phosphoryl group but also producing longer range effects. In the protein kinases, the phosphoryl residue is about 14−16 Å from the catalytic site but its phosphorylation has profound effects on activity by organizing the activation segment so that protein substrate may be recognized and bound in the correct conformation for catalysis. The nucleophilic aspartate of CheB is 25 Å from the catalytic triad of the methylesterase domain, such that long-range conformational changes must occur to stimulate biological activity. In the Src family kinases, localization of the C-terminal tyrosine phosphoryl group by the SH2 domain assists the interactions of the SH2 linker and the SH3 domain with the kinase, interactions that hold the kinase in its inactive conformation. The catalytic site is 40 Å from the inhibitory phosphoryl tyrosine site. In glycogen phosphorylase, the intersubunit interactions of the

phosphoryl serine promote a quaternary structure change that is coupled with tertiary structural changes at the catalytic site some 50 Å away. In the three systems comprising the Ser/Thr protein kinases, the Src family kinases and glycogen phosphorylase, the phosphoryl group is acting as an allosteric effector, bringing about changes at the catalytic site remote from the site of phosphorylation through stabilization of a particular state of the enzyme. In the multidomain RR, the role of phosphorylation of the nucleophilic aspartate can be likened to a disorganizing center in that inhibitory intramolecular contacts must be broken to elicit a response.

The shifts in protein atoms in response to phosphorylation may be small or large. In IDH, HPr, and SpoIIAA there is no conformational response. The phospho− and non-phospho−proteins have almost identical structures. In SpoIIAA, the phosphoryl group on Ser58 makes no contacts to the protein and it is likely that its role is to provide a negatively charged steric impediment to complex formation. In Spo0A, FixJ, NtrC, and CheY there are small but local changes in the vicinity of the phosphoryl group but these have profound effects for protein/protein association (discussed below). At the other extreme, on phosphorylation the position of Thr183 in ERK2 shifts 9 Å, the position of Thr-10 in yeast phosphorylase shifts 36 Å, and the position of Ser14 in glycogen phosphorylase shifts 50 Å. The incentive for these large changes is a combination of an inhospitable site for a phosphoryl residue in the non-phospho state and a favorable site when the shift is accomplished.

In several of the eukaryotic systems the activatory response cannot be achieved by phosphorylation on a single site. In CDK2, the association with cyclin A and phosphorylation is needed for activation. In MAPK and IRK, double or triple phosphorylation is needed. Yet in other kinases such as cAPK and LCK, kinase domain phosphorylation on a single residue in the activation domain is sufficient. Despite significant homologies in sequences among the kinases and a common core structure, there are differences that govern the pliability of the kinase and the ease with which the active conformation can be stabilized.

Promotion of protein/protein association by phosphorylation is an important property in both eukaryotic and prokaryotic systems. ERK2 exhibits an indirect mechanism whereby activation by dual phosphorylation in the activation segment produces domain closure, contacts to the C-helix, and through the C helix contacts to the loop preceding the C-terminal helix L16. Shifts in the loop preceding L16 create a nonpolar surface that serves to promote dimerization of ERK2 molecules. The response regulators also employ an indirect mechanism for heterodimer formation, and the phosphoryl group is not directly involved in protein/protein association. In RRs, phosphorylation on an aspartate residue results in repositioning of a local threonine that contacts the phosphoryl group. An aromatic residue repositions to fill the void created by the threonine with a corresponding change in a helix and a loop region. This 'aromatic' switch can then create a surface for CheY binding to FliM or FixJ dimerization or inhibitory protein/protein interactions that liberate the methylesterase domain of CheB from its response regulator domain. In the STAT proteins the phosphoryl tyrosine plays a more direct role in dimerization through its association with the SH2 domain of the other subunit. The two subunits, held by the mutual pTyr/SH2 interactions, bring the DNA binding domains to their correct distance and orientation to bind DNA. In the KID/KIX interaction from the CREB and CBP proteins, the phosphoryl group plays a role in promoting the heteroassociation through a direct contact from the phosphoryl−serine on the KID peptide to a tyrosine on the KIX domain. However, this interaction by itself appears to be insufficient to promote association and mutual interactions between protein/protein side chains as well as the direct phosphoryl interaction govern association. In the CDK2/KAP complex, recognition of the phospho−kinase by the phosphatase changes the conformation of the kinase. The notion that phosphoprotein recognition can change the structure of the phospho−protein provides a further mechanism for the ways in which phosphoproteins may signal to other proteins.

The structural studies on protein kinases have shown that the peptide substrate in the vicinity of the phosphorylatable residue must have a defined extended conformation. Therefore, it is not surprising that in some systems (e.g., glycogen phosphorylase, KID) the region of the phosphorylatable residue in the non-phospho protein adopts a mobile structure that can be adapted to fit the kinase catalytic site with little energy cost. However, in the other proteins (e.g., SpoIIAA, HPr, and IDH) the phosphorylatable residues are in regions that are well ordered. In the K+ channel inactivation domain, phosphorylation leads to disorder not in the region around the phosphoryl residue but in regions that are adjacent in space.

Finally, yeast phosphorylase exhibits a mechanism that is so far unique. Here the phosphoryl−threonine displaces an inhibitor, a phosphorylated metabolite (Glc-6-P), and utilizes the same phosphoryl recognition site as the inhibitor. Activation at the catalytic site is achieved through allosteric effects mediated by quaternary and tertiary structural changes.

The structural studies have revealed definite evidence for the mechanism of control by phosphorylation for about 17 proteins and have provided suggestions for several others. Undoubtedly, more mechanisms still await discovery.

V. Abbreviations

cAPK	cyclic AMP-dependent protein kinase
CAK	cyclin-dependent kinase activating kinase
CBP	CREB-binding protein
CDK2	cyclin-dependent protein kinase 2
CREB	cyclic AMP response element binding protein
ERK2	extracellular regulated protein kinase 2
hlGP	human liver glycogen phosphorylase
rmGP	rabbit muscle glycogen phosphorylase
yGP	yeast glycogen phosphorylase
HAD	haloacid dehalogenase
HK	histidine kinase

HTH	helix−turn−helix
ID	inactivation domain of the voltage-gated K^+ channel Kv.4
IRK	insulin receptor tyrosine kinase
KAP	kinase-associated phosphatase
KID	kinase-inducible domain of CREB
KIX	KID binding domain of CBP
MalP	maltodextrin phosphorylase
PDGFR	platelet-derived growth factor receptor
PEP	phosphoenolpyruvate
PSP	phosphoserine phosphatase
PKC	protein kinase C
PTS	phosphosugar transferase system
RR	response regulator
STAT	signal transducers and activators of transcription
TM	transmembrane
VEGFR 2	vascular endothelial growth factor receptor 2

VI. Acknowledgments

We thank Ho Cho, Dave Wemmer, Gerald Audette, Louise Delbaere, and Tony Wilkinson for sharing information and atomic coordinates prior to publication, Martin Noble for use of Aesop for Figures 1−8, and Steve Lee for help with figure reproduction. L.N.J. is a member of the Oxford Centre for Molecular Sciences which is supported by the BBSRC, MRC, and EPSRC. R.J.L. is a Wellcome Trust Research Career Development Fellow.

VII. References

(1) Lander, E. S.; et al. *Nature* **2001**, *409*, 860.
(2) Meng, W.; Sawasdikosol, S.; Burakoff, S. J.; Eck, M. J. *Nature* **1999**, *398*, 84.
(3) Eck, M. J.; Dhepaganon, S.; Trub, T.; Nolte, R. T.; Shoelson, S. E. *Cell* **1996**, *85*, 695.
(4) Kuriyan, J.; Cowburn, D. *Annu. Rev. Biomol. Struct.* **1997**, *26*, 259.
(5) Rittinger, K.; Budman, J.; Xu, J.; Volinia, S.; Cantley, L. C.; Smerdon, S. J.; Gamblin, S. J.; Yaffe, M. B. *Mol. Cell* **1999**, *4*, 153.
(6) Murphy, J. E.; Stec, B.; Ma, L.; Kantrowitz, E. R. *Nat. Struct. Biol.* **1997**, *4*, 618.
(7) Pannifer, A. D. B.; Flint, A. J.; Tonks, N. K.; Barford, D. *J. Biol. Chem.* **1998**, *273*, 10454.
(8) Redinbo, M. R.; Stewart, L.; Kuhn, P.; Champoux, J. J.; Hol, W. G. J. *Science* **1998**, *279*, 1504.
(9) Guo, F.; Gopaul, D. N.; Van Duyne, G. D. *Nature* **1997**, *389*, 40.
(10) Kobe, B.; Jennings, I. G.; House, C. M.; Michell, B. J.; Goodwill, K. E.; Santariero, B. D.; Stevens, R. C.; Cotton, R. G. H.; Kemp, B. E. *Nat, Struct. Biol.* **1999**, *6*, 442.
(11) Johnson, L. N.; Barford, D. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 199.
(12) Stroud, R. M. *Curr. Opin. Struct. Biol.* **1991**, *1*, 826.
(13) Shoji, S.; Titani, K.; Demaille, J.; Fischer, E. H. *J. Biol. Chem.* **1979**, *254*, 6211.
(14) Levin, L. R.; Zoller, M. J. *Mol. Cell Biol.* **1990**, *10*, 1066.
(15) Steinberg, R. A.; Cauthron, R. D.; Symcox, M. M.; Shuntoh, H. *Mol. Cell Biol.* **1993**, *13*, 2332.
(16) Knighton, D. R.; Zheng, J.; Ten Eyck, L. F.; Ashford, V. A.; Xuong, N.-H.; Taylor, S. S.; Sowadski, J. M. *Science* **1991**, *253*, 407.
(17) Knighton, D. R.; Zheng, J.; Eyck, L. F. T.; Xuong, N.; Taylor, S. S.; Sowadski, J. M. *Science* **1991**, *253*, 414.
(18) Knighton, D. R.; Bell, S. M.; Zheng, J.; Eyck, L. F. T.; Xuing, N. H.; Taylor, S. S.; Sodwadski, J. M. *Acta Crystallogr.* **1993**, *D49*, 357.
(19) Bossmeyer, D.; Engh, R. A.; Kinzel, V.; Ponstingl, H.; Huber, R. *EMBO J.* **1993**, *12*, 849.
(20) Hunter, T. *Cell* **1987**, *49*, 1.
(21) Johnson, L. N.; Noble, M. E. M.; Owen, D. J. *Cell* **1996**, *85*, 149.
(22) Owen, D. J.; Noble, M. E.; Garman, E. F.; Papageorgiou, A. C.; Johnson, L. N. *Structure* **1995**, *3*, 467.
(23) Lowe, E. D.; Noble, M. E. M.; Skamnaki, V. T.; Oikonomakos, N. G.; Owen, D. J.; Johnson, L. N. *EMBO J.* **1997**, *16*, 6646.
(24) Hu, S. H.; Parker, M. W.; Lei, J. Y.; Wilce, M. C. J.; Benian, G. M.; Kemp, B. E. *Nature* **1994**, *369*, 581.
(25) Xu, R.-M.; Carmel, G.; Sweet, R. M.; Kuret, J.; Cheng, X. *EMBO J.* **1995**, *14*, 1015.
(26) Longenecker, K. L.; Roach, P.; Hurley, T. D. *J. Mol. Biol.* **1996**, *257*, 618.
(27) Chen, P.; Luo, C.; Deng, Y.; Ryan, K.; Register, J.; Margosiak, S.; Tempczyk-Russell, A.; Nguyen, B.; Myers, P.; Lundgren, K.; Kan, C.-C.; O'Connor, P. M. *Cell* **2000**, *100*, 681.
(28) Mayans, O.; van der Veen, P. F. M.; Wilm, M.; Mues, A.; Young, P.; Furst, D. O.; Willmanns, M.; Gautel, M. *Nature* **1998**, *395*, 863.
(29) Morgan, D. O. *Annu. Rev. Cell. Dev. Biol.* **1997**, *13*, 261.
(30) De Bondt, H. L.; Rosenblatt, J.; Jancarik, J.; Jones, H. D.; Morgan, D. O.; Kim, S.-H. *Nature* **1993**, *363*, 592.
(31) Brown, N. R.; Noble, M. E. M.; Lawrie, A. M.; Morris, M. C.; Tunnah, P.; Divita, G.; N., J. L.; Endicott, J. A. *J. Biol. Chem.* **1999**, *274*, 8746.
(32) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, E.; Hurwitz, J.; Massague, J.; Pavletich, N. P. *Nature* **1995**, *376*, 313.
(33) Brown, N. R.; Noble, M. E. M.; Endicott, J. A.; Garman, E. F.; Wakatsuki, S.; Mitchell, E. P.; Rasmussen, B.; Hunt, T.; Johnson, L. N. *Structure* **1995**, *3*, 1235.
(34) Russo, A.; Jeffrey, P. D.; Pavletich, N. P. *Nat. Struct. Biol.* **1996**, *3*, 696.
(35) Pavletich, N. P. *J. Mol. Biol.* **1999**, *287*, 821.
(36) Brown, N. R.; Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. *Nat. Cell Biol.* **1999**, *1*, 438.
(37) Hagopian, J. C.; Kirtley, M. P.; Stevenson, L. M.; Gergis, R. M.; Russo, A. A.; Pavletich, N. P.; Parsons, S. M.; Lew, J. *J. Biol. Chem.* **2001**, *276*, 275.
(38) Adams, J. A.; Taylor, S. S. *Biochemistry* **1992**, *31*, 8516.
(39) Zhou, J.; Adams, J. A. *Biochemistry* **1997**, *36*, 2977.
(40) Grant, B. D.; Adams, J. A. *Biochemistry* **1996**, *35*, 2022.
(41) Skamnaki, V. T.; Owen, D. J.; Noble, M. E. M.; Lowe, E. D.; Lowe, G.; Oikonomakos, N. G.; Johnson, L. N. *Biochemistry* **1999**, *38*, 14718.
(42) Poon, R. Y. C.; Hunter, T. *Science* **1995**, *270*, 90.
(43) Song, H.; Hanlan, N.; Brown, N. R.; Noble, M. E. M.; Johnson, L. N.; Barford, D. *Mol. Cell* **2001**, *7*, 615.
(44) Cobb, M. H.; Goldsmith, E. J. *J. Biol. Chem.* **1995**, *270*, 14843.
(45) Kyriakis, J. M.; Avruch, J. *J. Biol. Chem.* **1996**, *271*, 24313.
(46) Zhang, F.; Strand, A.; Robbins, D.; Cobb, M. H.; Goldsmith, E. J. *Nature* **1994**, *367*, 704.
(47) Canagarajah, B. J.; Khokhlatchev, A.; Cobb, M. H.; Goldsmith, E. J. *Cell* **1997**, *90*, 859.
(48) Khokhlatchev, A. V.; Canagarajah, B.; Wilsbacher, J.; Robinson, M.; Atkinson, M.; Goldsmith, E.; Cobb, M. H. *Cell* **1998**, *93*, 605.
(49) Wilson, K. P.; Fitzgibbon, M. J.; Caron, P. R.; Griffith, J. P.; Chen, W.; McCaffrey, P. G.; Chambers, S. P.; Su, M. S. *J. Biol. Chem.* **1996**, *271*, 27696.
(50) Wang, Z.; Harkins, P. C.; Ulevitch, R. J.; Han, J.; Cobb, M. H.; Goldsmith, E. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2327.
(51) Bellon, S.; Fitzgibbon, M. J.; Fox, T.; Hsiao, H.-M.; Wilson, K. P. *Structure* **1999**, *7*, 1057.
(52) Xie, X.; Gu, Y.; Fox, T.; Coll, J. T.; Fleming, M. A.; Markland, W.; Caron, P. R.; Wilson, K. P.; Su, M. S. *Structure* **1998**, *6*, 983.
(53) Hubbard, S. R.; Wei, L.; Ellis, L.; Hendrickson, W. A. *Nature* **1994**, *372*, 746.
(54) Kobe, B.; Kemp, B. E. *Nature* **1999**, *402*, 373.
(55) Hubbard, S. R. *EMBO J.* **1997**, *16*, 5572.
(56) Salmeen, A.; Anderson, J. N.; Myers, M. P.; Tonks, N. K.; Barford, D. *Mol. Cell* **2000**, *6*, 1401.
(57) McTigue, M. A.; Wickersham, J. A.; Pinko, C.; Showalter, R.; Parast, C. V.; Tempczyk-Russell, A.; Gehring, M. R.; Mroczkowski, B.; Kan, C.-C.; Villafranca, J. E.; Appelt, K. *Structure* **1999**, *7*, 319.
(58) Mohammadi, M.; Schlessinger, J.; Hubbard, S. *Cell* **1996**, *86*, 577.
(59) Xu, W.; Harrison, S. C.; Eck, M. J. *Nature* **1997**, *385*, 595.
(60) Williams, J. C.; Weijland, A.; Gonfloni, S.; Thompson, A.; Courtneidge, S. A.; Superti-Furga, G.; Wierenga, R. K. *J. Mol. Biol.* **1997**, *274*, 757.
(61) Sicheri, F.; Moarefi, I.; Kuriyan, J. *Nature* **1997**, *385*, 602.
(62) Yamaguchi, H.; Hendrickson, W. A. *Nature* **1996**, *384*, 484.
(63) Sicheri, F.; Kuriyan, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 777.
(64) Williams, J. C.; Wiernega, R. K.; Saraste, M. *TIBS* **1998**, *23*, 179.
(65) Hubbard, S. *Nat. Struct. Biol.* **1999**, *6*, 711.
(66) Liu, X.; Brodeur, S. R.; Gish, G.; Song-Yang, Z.; Cantley, L. C.; Laudano, A. P.; Pawson, T. *Oncogene* **1993**, *8*, 1119.
(67) Moarefi, I.; LeFevre-Bernt, M.; Sicheri, F.; Huse, M.; Lee, C.-H.; Kuriyan, J.; Miller, W. T. *Nature* **1997**, *385*, 650.
(68) Gonfloni, S.; Frischknecht, F.; Way, M.; Superti-Furga, G. *Nat. Struct. Biol.* **1999**, *6*, 760.
(69) Xu, W.; Doshi, A.; Lei, M.; Eck, M. J.; Harrison, S. C. *Mol. Cell* **1999**, *3*, 629.
(70) Schindler, T.; Sicheri, F.; Pico, A.; Gazit, A.; Levitzki, A.; Kuriyan, J. *Mol. Cell* **1999**, *3*, 639.
(71) Hof, P.; Pluskey, S.; Dhe-Paganon, S.; Eck, M. J.; Shoelson, S. E. *Cell* **1998**, 441.

(72) Chen, X.; Vinkemeier, U.; Zhao, Y.; Jeruzalmi, D.; Darnell, J. D.; Kuriyan, J. *Cell* **1998**, *93*, 827.
(73) Becker, S.; Groner, B.; Muller, C. W. *Nature* **1998**, *394*, 145.
(74) Radhakrishnan, I.; Perez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741.
(75) Antz, C.; Bauer, T.; Kalbacher, H.; Frank, R.; Covarrubias, M.; Kalbitzer, H. R.; Ruppersberg, J. P.; Baukrowitz, T.; Fakler, B. *Nat. Struct. Biol.* **1999**, *6*, 146.
(76) Knudsen, E. S.; Wang, J. Y. J. *J. Biol. Chem.* **1996**, *271*, 8313.
(77) Lee, J.-O.; Russo, A. A.; Pavletich, N. P. *Nature* **1998**, *391*, 859.
(78) Misura, K. M. S.; Scheller, R. H.; Weiss, W. I. *Nature* **2000**, *404*, 355.
(79) Johnson, L. N.; Barford, D. *Protein Sci.* **1994**, *3*, 1726.
(80) Fischer, E. H.; Krebs, E. G. *J. Biol. Chem.* **1955**, *216*, 121.
(81) Rath, V. L.; Ammirati, M.; LeMotte, P. K.; Fennell, K. F.; Mansour, M. N.; Danley, D. E.; Hynes, T. R.; Schulte, G. K.; Wasilko, D. J.; Pandit, J. *Mol. Cell* **2000**, *6*, 139.
(82) Rath, V. L.; Fletterick, R. J. *Nat. Struct. Biol.* **1994**, *1*, 681.
(83) Rath, V.; Lin, K.; Hwang, P.; Fletterick, R. J. *Structure* **1996**, *4*, 463.
(84) Lin, K.; Hwang, P. K.; Fletterick, R. J. *Structure* **1997**, *5*, 1511.
(85) Watson, K. A.; Schinzel, R.; Palm, D.; Johnson, L. N. *EMBO J.* **1997**, *16*, 1.
(86) Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Crucianni, G.; Son, J. C.; Bichard, C. J. F.; Fleet, G. W. J.; Oikonomakos, N. G.; Leonidas, D. D.; Kontou, M. *Acta Crystallogr., Sect. D* **1995**, *D51*, 458.
(87) Oikonomakos, N. G.; Kontou, M.; Zographos, S. E.; Watson, K. A.; Johnson, L. N.; Bichard, C. J. F.; Fleet, G. W. J.; Acharya, K. R. *Protein Sci.* **1995**, *4*, 2469.
(88) Lin, K.; Rath, V. L.; Dai, S. C.; Fletterick, R. J.; Hwang, P. K. *Science* **1996**, *273*, 1539.
(89) Watson, K. A.; McCleverty, C.; Geremia, S.; Cottaz, S.; Driguez, D.; Johnson, L. N. *EMBO J.* **1999**, *18*, 4619.
(90) Ponting, C. P.; Aravind, L.; Schultz, J.; Bork, P.; Koonin, E. V. *J. Mol. Biol.* **1999**, *289*, 729.
(91) Zhang, C. C.; Gonzalez, L.; Phalip, V. *Nucleic Acids Res.* **1998**, *26*, 3619.
(92) Mizuno, T. *J. Biochem.* **1998**, *123*, 555.
(93) Stock, A. M.; Robinson, V. L.; Goudreau, P. N. *Annu. Rev. Biochem.* **2000**, *69*, 183.
(94) Stock, J.; Da Re, S. *Curr. Biol.* **2000**, *10*, R420.
(95) Buckler, D. R.; Anand, G. S.; Stock, A. M. *Trends Microbiol.* **2000**, *8*, 153.
(96) Hoch, J. A. *Curr. Opin. Microbiol.* **2000**, *3*, 165.
(97) Iuchi, S.; Weiner, L. *J. Biochem.* **1996**, *120*, 1055.
(98) Manson, M. D.; Armitage, J. P.; Hoch, J. A.; Macnab, R. M. *J. Bacteriol.* **1998**, *180*, 1009.
(99) Hoch, J. A. *Annu. Rev. Microbiol.* **1993**, *47*, 441.
(100) Miller, J. F.; Mekalanos, J. J.; Falkow, S. *Science* **1989**, *243*, 916.
(101) Shapiro, L.; Losick, R. *Science* **1997**, *276*, 712.
(102) Chang, C.; Kwok, S. F.; Bleecker, A. B.; Meyerowitz, E. M. *Science* **1993**, *262*, 539.
(103) Maeda, T.; Wurgler-Murphy, S. M.; Saito, H. *Nature* **1994**, *369*, 242.
(104) Fabret, C.; Feher, V. A.; Hoch, J. A. *J. Bacteriol.* **1999**, *181*, 1975.
(105) Mizuno, T.; Kaneko, T.; Tabata, S. *DNA Res.* **1996**, *3*, 407.
(106) Ng, W. V.; Kennedy, S. P.; Mahairas, G. G.; Berquist, B.; Pan, M.; Shukla, H. D.; Lasky, S. R.; Baliga, N. S.; Thorsson, V.; Sbrogna, J.; Swartzell, S.; Weir, D.; Hall, J.; Dahl, T. A.; Welti, R.; Goo, Y. A.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M. J.; Hough, D. W.; Maddocks, D. G.; Jablonski, P. E.; Krebs, M. P.; Angevine, C. M.; Dale, H.; Isenbarger, T. A.; Peck, R. F.; Pohlschroder, M.; Spudich, J. L.; Jung, K. W.; Alam, M.; Freitas, T.; Hou, S.; Daniels, C. J.; Dennis, P. P.; Omer, A. D.; Ebhardt, H.; Lowe, T. M.; Liang, P.; Riley, M.; Hood, L.; DasSarma, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12176.
(107) Barrett, J. F.; Hoch, J. A. *Antimicrob. Agents Chemother* **1998**, *42*, 1529.
(108) Hershlag, D.; Jencks, W. P. *Biochemistry* **1990**, *29*, 5172.
(109) Lukat, G. S.; Stock, A. M.; Stock, J. B. *Biochemistry* **1990**, *29*, 5436.
(110) Djordjevic, S.; Goudreau, P. N.; Xu, Q.; Stock, A. M.; West, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1381.
(111) Birck, C.; Mourey, L.; Gouet, P.; Fabry, B.; Schumacher, J.; Rousseau, P.; Kahn, D.; Samama, J.-P. *Structure* **1999**, *7*, 1505.
(112) McEvoy, M. M.; Bren, A.; Eisenbach, M.; Dahlquist, F. W. *J. Mol. Biol.* **1999**, *289*, 1423.
(113) Ridder, I. S.; Dijkstra, B. W. *Biochem. J.* **1999**, *339*, 223.
(114) Pickart, C. M.; Jencks, W. P. *Biol. Chem.* **1994**, *259*, 1629.
(115) Stock, A. M.; Mottonen, J. M.; Stock, J. B.; Schutt, C. E. *Nature* **1989**, *337*, 745.
(116) Lewis, R. J.; Brannigan, J. A.; Muchová, K.; Barák, I.; Wilkinson, A. J. *J. Mol. Biol.* **1999**, *294*, 9.
(117) Kern, D.; Volkman, B. F.; Luginbuhl, P.; Nohaile, M. J.; Kustu, S.; Wemmer, D. E. *Nature* **1999**, *402*, 894.
(118) Cho, H. S.; Lee, S.-Y.; Yan, D.; Pan, X.; Parkinson, J. S.; Kustu, S.; Wemmer, D. E.; Pelton, J. G. *J. Mol. Biol.* **2000**, *297*, 543.
(119) Halkides, C. J.; McEvoy, M. M.; Casper, E.; Matsumura, P.; Volz, K.; Dahlquist, F. W. *Biochemistry* **2000**, *39*, 5280.
(120) Lee, S.-Y.; Cho, S. H.; Pelton, J. G.; Yan, D.; Henderson, R. K.; King, D. S.; Huang, L.-S.; Kustu, S.; Berry, E. A.; Wemmer, D. E. *Nat. Struct. Biol.* **2001**, *8*, 52.
(121) Goudreau, P. N.; Lee, P. J.; Stock, A. M. *Biochemistry* **1998**, *37*, 14575.
(122) Yan, D.; Cho, H. S.; Hastings, C. A.; Igo, M. M.; Lee, S. Y.; Pelton, J. G.; Stewart, V.; Wemmer, D. E.; Kustu, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14789.
(123) Stock, A. M.; Martinez-Hackert, E.; Rasmussen, B. F.; West, A. H.; Stock, J. B.; Ringe, D.; Petsko, G. A. *Biochemistry* **1993**, *32*, 13375.
(124) Bellsolell, L.; Prieto, J.; Serrano, L.; Coll, M. *J. Mol. Biol.* **1994**, *238*, 489.
(125) Koshland, D. E. *J. Am. Chem. Soc.* **1952**, *74*, 2286.
(126) Welch, M.; Oosawa, K.; Aizawa, S.; Eisenbach, M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8787.
(127) Trach, K.; Burbulys, D.; Strauch, M.; Wu, J.-J.; Dhillon, N.; Jonas, R.; Hanstein, C.; Kallio, P.; Perego, M.; Bird, T.; Spiegelman, G.; Fogher, C.; Hoch, J. A. *Res. Microbiol.* **1991**, *142*, 815.
(128) Da Re, S.; Bertagnoli, S.; Fourment, J.; Reyrat, J. M.; Kahn, D. *Nucleic Acids Res.* **1994**, *22*, 1555.
(129) Baikalov, I.; Schroder, I.; Kaczor-Grzekkowiak, M.; Grzekowiak, K.; Gunsalus, R. P.; Dickerson, R. E. *Biochemistry* **1996**, *35*, 11053.
(130) Ireton, K.; Rudner, D. Z.; Siranosian, K. J.; Grossman, A. D. *Genes Dev.* **1993**, *7*, 283.
(131) Cervin, M. A.; Spiegelman, G. B. *J. Biol. Chem.* **2000**, *275*, 22025.
(132) Lewis, R. J.; Muchová, K.; Brannigan, J. A.; Leonard, G.; Barák, I.; Wilkinson, A. J. *J. Mol. Biol.* **2000**, *297*, 757.
(133) Zhu, Y.; Qin, L.; Yoshida, T.; Inouye, M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7808.
(134) Bilwes, A. M.; Alex, L. A.; Crane, B. R.; Simon, M. I. *Cell* **1999**, *96*, 131.
(135) 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. *Nature* **1998**, *396*, 88.
(136) Bergerat, A.; de Massy, B.; Gadelle, D.; Varoutas, P. C.; Nicolas, A.; Forterre, P. *Nature* **1997**, *386*, 414.
(137) Wigley, D. B.; Davies, G. J.; Dodson, E. J.; Maxwell, A.; Dodson, G. *Nature* **1991**, *351*, 624.
(138) Prodomou, C.; Roe, S. M.; O'Brien, R.; Ladbury, J. E.; Piper, P. W.; Pearl, L. H. *Cell* **1997**, *90*, 65.
(139) Ban, C.; Yang, W. *Cell* **1998**, *95*, 541.
(140) Dutta, R.; Inouye, M. *TIBS* **2000**, *25*, 24.
(141) Lewis, R. J.; Singh, O. M.; Smith, C. V.; Skarzynski, T.; Maxwell, A.; Wonacott, A. J.; Wigley, D. B. *EMBO J.* **1996**, *15*, 1412.
(142) 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. *Nat. Struct. Biol.* **1999**, *6*, 729.
(143) Najafi, S. M.; Willis, A. C.; Yudkin, M. D. *J. Bacteriol.* **1995**, *177*, 2912.
(144) Alper, S.; Dufour, A.; Garsin, D. A.; Duncan, L.; Losick, R. *J. Mol Biol.* **1996**, *260*, 165.
(145) Zapf, J.; Sen, U.; Madhusudan; Hoch, J. A.; Varughese, K. I. *Structure* **2000**, *8*, 851.
(146) Burbulys, D.; Trach, K. A.; Hoch, J. A. *Cell* **1991**, *64*, 545.
(147) Varughese, K. I.; Madhusudan; Zhou, X. Z.; Whiteley, J. M.; Hoch, J. A. *Mol. Cell.* **1998**, *2*, 485.
(148) Kato, M.; Mizuno, T.; Shimizu, T.; Hakoshima, T. *Cell* **1997**, *88*, 717.
(149) Zhou, H.; Lowry, D. F.; Swanson, R. V.; Simon, M. I.; Dahlquist, F. W. *Biochemistry* **1995**, *34*, 13858.
(150) Xu, Q.; West, A. H. *J. Mol. Biol.* **1999**, *292*, 1030.
(151) Johnson, L. N.; Lowe, E. D.; Noble, M. E. M.; Owen, D. J. *FEBS Lett.* **1998**, *430*, 1.
(152) Stock, J. B.; Stock, A. M.; Mottonen, J. M. *Nature* **1990**, *344*, 395.
(153) Stewart, R. C. *J. Biol. Chem.* **1993**, *268*, 1921.
(154) Aravind, L.; Galperin, M. Y.; Koonin, E. V. *Trends Bioch. Sci.* **1998**, *23*, 127.
(155) Toyoshima, C.; Nakasako, M.; Nomura, H.; Ogawa, H. *Nature* **2000**, *405*, 647.
(156) Herzberg, O.; Klevit, R. *Opin. Struct. Biol.* **1994**, *4*, 814.
(157) Garrett, D. S.; Seok, Y. J.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1999**, *6*, 166.
(158) Audette, G. F.; Engelmann, R.; Hengstenberg, W.; Deutscher, J.; Hayakawa, K.; Quail, J. W.; Delbaere, L. T. *J. Mol. Biol.* **2000**, *303*, 545.
(159) Napper, S.; Andersdon, J. W.; Georges, F.; Quail, J. W.; Delbaere, L. T.; Waygood, E. B. *Biochemistry* **1996**, *35*, 11260.
(160) Wang, G.; Louis, J. M.; Sondej, M.; Seok, Y. J.; Peterkofsky, A.; Clore, G. M. *EMBO J.* **2000**, *19*, 5635.
(161) Garrett, D. S.; Seok, Y. J.; Peterkofsky, A.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1998**, *7*, 789.
(162) Stock, J. B.; Surette, M. G.; Levit, M.; Park, P. Two-component signal transduction systems. structure: function relations and

mechanisms of catalysis. In *Two-Component Signal Transduction*; Hoch, J. A., Silhavy, T. J., Eds.; ASM Press: Washington, DC, 1995.
(163) Hoch, J. A. Control of cellular development in sporulating bacteria by the phosphorelay two-component signal transduction system. In *Two Component Signal Transduction*; Hoch, J. A., Silhavy, T. J., Eds.; ASM Press: Washington, DC, 1995.
(164) Duncan, L.; Losick, R. *Proc. Natl. Acad. Sci.* **1993**, *90*, 2325.
(165) Alper, S.; Duncan, L.; Losick, R. *Cell* **1994**, *77*, 195.
(166) Diederich, B.; Wilkinson, J. F.; Magnin, T.; Najafi, M.; Errington, J.; Yudkin, M. D. *Genes Dev.* **1994**, *8*, 2653.
(167) Duncan, L.; Alper, S.; Arigoni, F.; Losick, R.; Stragier, P. *Science* **1995**, *270*, 641.
(168) Magnin, T.; Lord, M.; Errington, J.; Yudkin, M. D. *Mol. Microbiol.* **1996**, *19*, 901.
(169) Kovacs, H.; Comfort, D.; Lord, M.; Campbell, I. D.; Yudkin, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5067.
(170) Seavers, P. R.; Lewis, R. J.; Brannigan, J. A.; Verschueren, K. H. G.; Murshudov, G. N.; Wilkinson, A. J. *Structure* **2001**, *9*, 605.
(171) Dufour, A.; Haldenwang, W. G. *J. Bacteriol.* **1994**, *176*, 1813.
(172) Kang, C. M.; Vijay, K.; Price, C. W. *Mol. Microbiol.* **1998**, *30*, 189.
(173) Gaidenko, T. A.; Yang, X.; Lee, Y. M.; Price, C. W. *J. Mol. Biol.* **1999**, *288*, 29.
(174) Perego, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8612.
(175) Wang, W.; Kim, R.; J., J.; Yokota, H.; Kim, S.-Y. *Structure* **2001**, *9*, 65.
(176) Johnson, L. N.; O'Reilly, M. *Curr. Opin. Struct. Biol.* **1996**, *6*, 762.
(177) Barford, D.; Hu, S.-H.; Johnson, L. N. *J. Mol. Biol.* **1991**, *218*, 233.
(178) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946.

CR000225S