Title
Intramolecular Signaling and Domain Interactions of the Escherichia coli Response Regulator NarL

Author
Katsir, Galit

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Intramolecular Signaling and Domain Interactions of the *Escherichia coli* Response Regulator NarL

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biochemistry and Molecular Biology

by

Galit Katsir

2012

ABSTRACT OF THE DISSERTATION

Intramolecular Signaling and Domain Interactions of the

Escherichia coli Response Regulator NarL

by

Galit Katsir

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2012

Professor Wayne L. Hubbell, Chair

Phosphoryl transfer from a histidine kinase to a response regulator in two-component signaling systems leads to cellular modifications that enable bacteria to adapt to environmental changes. The response regulator NarL, of the *Escherichia coli* Nar two-component system, becomes phosphorylated by the histidine kinase NarX and regulates genes involved in nitrate respiration. Phosphorylation at the N-terminal "receiver" domain of NarL exposes distant molecular surfaces, including regions of the C-terminal domain that dimerize upon DNA binding. To investigate other alterations in domain surface interactions upon phosphorylation, NarL and its individual domains were examined by analytical ultracentrifugation. Phosphorylation was demonstrated to induce full-length NarL dimerization and tetramerization, and N-terminal domain dimerization. The C-terminal domain was unable to dimerize alone,

suggesting that dimerization of the in-tact protein occurs via receiver domains, while C-terminal domain dimerization is driven by DNA binding. Independent receiver-domain dimerization implicates the widespread α4-β5-α5 surface as the dimerization site, since this region is unavailable in full-length NarL, which was shown to be monomeric. Also, receiver domain dimerization and NarL oligomerization may fulfill binding requirements at low-affinity promoter regions.

Two unphosphorylated NarL crystal structures reveal, previously unreported, distinct equilibrium states with variability in residue positions and polar contacts. The conformations of two activation-associated residues in the receiver domain show that the monoclinic NarL crystal structure represents a semi-activated state. Both structures were analyzed with respect to the interdomain interface and active site in order to gain an understanding of the activation mechanism that leads to domain separation. A solvent-accessibility analysis, along with structural comparisons, confirmed that the conformation of Lys196, in the semi-activated NarL structure, is also representative of activation. Mobility ratios and correlation fluctuation calculations showed that vital interface-loop residues exhibit low mobilities, however can affect the motion of other interface residues, including those that bind DNA. Therefore, interface mutations that lead to constitutively active phenotypes may also result from the interdependence of residue motions. In contrast to the interface loops, the active-site loops were characterized by high mobility ratios. Structural modifications of these flexible loops are expected to coincide with the proposed movements of certain active-site residues.

The dissertation of Galit Katsir is approved.

Carla Koehler

Robert P. Gunsalus

Robert T. Clubb

Wayne L. Hubbell, Committee Chair

University of California, Los Angeles

2012

Acknowledgements

I am extremely grateful to Professor Carla Koehler for her care and help as graduate advisor, which allowed me to reconstitute my committee and resume my graduate studies. Professor Wayne Hubbell kindly assumed the position of committee chair. Professor Rob Gunsalus was extremely generous with his time to discuss ideas and manuscript drafts. Mike Sawaya and Duilio Cascio offered great assistance and expertise in all aspects of crystallography. Professor Richard Dickerson and Mary Kopka provided unwavering support during my time in their research group. Jeffrey Zhang was tremendously helpful with several aspects of the research, as well as being a good friend during the more challenging days of graduate school. Many thanks to Razi and Safrira Katsir (my parents) and other family members and friends who provided support and encouragement, namely Nili Katsir, Brooke Miller, Neil Benin, Natasha Tarasova, and the inspiring women of my morning dance class, especially Cherry Tuck and Prue Delamater.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Acronyms and Initialisms

Å	Angstrom
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
AMPPNP	Adenosine 5'-(β, γ- imido) triphosphate tetralithium
ATPγS	Adenosine 5'-O-(3-thiotriphosphate tetralithium)
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
GMP	Guanosine monophosphate
kDa	Kilodalton
krpm	Kilo-revolutions-per-minute
MALDI	Matrix-assisted laser desorption ionization
MW	Molecular weight
nm	Nanometer
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
TBE	Tris, borate, EDTA buffer
UV	Ultraviolet

Preface

Two-component signal transduction is the principal method for organized cellular signaling in prokaryotes. The two signaling macromolecules, histidine kinase and response regulator, ensure the adaptability of the cell in response to extracellular physical and biological changes. Communication is achieved by a series of phosphoryl transfers through the modular domains of the histidine kinase and response regulator. These domains can be interchanged so long as the reductive elements are present: a phosphorylatable histidine in the first component and a phosphorylatable aspartate in the second. Additional domains bestow functional specificity and lead to cellular modifications, predominantly in the form of transcription.

The chapters herein relate to the structure and function of the response regulator NarL and, to a lesser extent, the structure of its cognate histidine kinase, NarX. The diversity of structure and function of two-component proteins is reviewed in Chapter 1, as well as the consequential diversity of their control. Common regulatory themes and finer distinctions between individual systems are discussed. Chapter 2 consists of a short review of the *E. coli* Nar two-component and gene transcription by the response regulators NarL and NarP. Key features of the x-ray crystal structure of NarL are discussed in relation to biochemical studies. The next two chapters relate to the domain surfaces of NarL as it transitions between phosphorylated and unphosphorylated conformations. In Chapter 3, analytical ultracentrifugation experiments provide evidence for NarL dimerization and oligomerization upon phosphorylation. A model is proposed showing the different possibilities of NarL dimerization at activating or repressing DNA-binding sites. In Chapter 4, two independent crystal structures of NarL, which reveal alternative residue-residue contacts, are compared at the interdomain interface and the active-site. The analysis leads to the identification of key interface

residues, including a marker of activation, and to additional reasons for the severity of specific interface mutations. In addition, active site modifications are proposed that would coincide with phosphorylation. Lastly, the structure of the cytoplasmic portion of NarX is of high interest since NarX possesses elements not present in other histidine kinases that could be important to its specificity as a nitrate sensor. Structural studies of the cytoplasmic region of NarX are presented in Appendix A.

Vita

EDUCATION

1991	Bachelors of Science in Biochemistry, University of California, Los Angeles (UCLA)

RESEARCH

1992	Laboratory Assistant, Sepulveda Veteran's Administration Medical Center
1992-1995	Laboratory Assistant, Jules Stein Eye Institute, UCLA
1996-2005	Research Assistant, Molecular Biology Institute, UCLA
2006	Senior Associate, Amgen, Inc.
2007-2010	Associate Research Scientist, Exelixis Inc.
2011	Research Associate, Genentech Inc.

TEACHING

1995-1996	Teaching Assistant, Biochemistry Lab, UCLA
2001	Teaching Assistant, Introductory and Environmental Chemistry, UCLA
2001-2002	Teaching Fellow, General Chemistry, UCLA
2002	Lecturer, General Chemistry, Loyola Marymount University
2005	Chemistry Instructor, General Chemistry, Santa Monica College
2005	Chemistry Instructor, General Chemistry, San Jose City College

AWARDS

1995-1996	Distinguished Teaching Award, UCLA
1997-1998	Cellular and Molecular Biology Training Fellowship

POSTER PRESENTATIONS

1998	*Crystallization of AT Reading Drugs in the Minor Groove of DNA.* American Crystallographic Association in Arlington, Virginia.
1999	*Crystallization of Magnesium-bound NarL*. The 14th West Coast Protein Crystallography Workshop, Asilomar Conference Center.

Chapter 1

Two-Component Systems

The Two-Component Paradigm

Two-component systems (TCSs) are communication bridges between the external environment of a cell and its internal systems for adaptation (52). To meet the demands of a constantly and rapidly changing environment, TCSs have evolved to allow organisms to adapt to environmental changes such as light, osmolarity, nutrient availability, population density, and virulence, among some examples. These systems exist predominantly in Bacteria and to a relatively lesser extent in the Archaea and Eukaryota domains. Bacteria have more than 50 TCS proteins in an average genome (6). The ubiquity of prokaryotic TCSs has led to their organization in the Prokaryotic TCS database, with a list hitherto containing more than 77,000 two-component genes. Within the Eukaryota domain, TCSs have been detected in fungi and plants, but not in animals (81).

TCSs were first described about twenty years ago where the simplest version of the paradigm was known (52, 65). To signify the need for cellular adaptation, a ligand binds to the sensory domain of the first component, a histidine kinase (HK). This propagates a phosphorylation cascade (Figure 1-1) whereby the HK autophosphorylates on a conserved histidine residue using ATP and then transfers the phosphoryl group to a conserved aspartate residue on the second component, the response regulator (RR). The active-site His of the HK is located on the conserved region called the "kinase core", while the Asp of the RR is located on the conserved N-terminal "receiver" domain (or REC). The intracellular RR, once phosphorylated, mediates a cellular response, usually by employing its more specific C-terminal domain. The C-terminal domain of the RR, called the "output" or "effector" domain, usually functions as a transcription factor that activates or represses genes required to accommodate the

environmental change. Besides transcriptional regulation, the effector can also express output by means of protein-protein interactions and enzymatic activity.

This chapter provides a background of TCSs with an emphasis on the structure and function of its well studied protein families. Although the above paradigm is the essence of TCSs, domains of TCSs are modular, breeding a diversity of mechanistic designs and interesting pathways. A diversity of structure and function (28) may appear in the form of shuffled domains within HKs and RRs, HKs that are cytosolic, systems that involve more than two components, REC domains without an attached effector, and different interdomain arrangements. These and other variations enhance the complexity and depth of this prominent mechanism with which organisms use to survive.

Histidine Kinases

Types of Histidine Kinases

Most HKs are transmembrane bound, pre-existing dimers (53), with their N-terminal sensor domain located in the periplasm (gram-negative bacteria) or the extracytoplasm (gram-positive bacteria) (25). The sensor domain can also exist intracellularly regardless of the whether the HK is attached to the membrane or is cytosolic. Sensor domains among HKs share very little sequence similarity which allows for response specificity to various stimuli encountered by a cell. Usually following the transmembrane or sensor region is a conserved cytoplasmic region of the HK, called the kinase core (Figure 1-1) (reveiwed in 3, 5, 8, 9). The kinase core is essential for HKs to perform the three major functions by which they are defined: autophosphorylation upon an external signal, phosphotransfer to the RR, and dephosphorylation of the RR. The kinase core constitutes two domains: a Dimerization and Histidine

Phosphotransfer (DHp) domain and a Catalytic and ATP-binding (CA) domain. In some HKs additional cytoplasmic signaling domains intervene between the transmembrane region and the kinase core that are thought to help transmit the signal from the periplasm or that may provide structural alignment of the kinase core. Examples of such domains are HAMP, PAS, or GAF domains (86, 88) (see Text Box).

The phosphorylatable His is located on a conserved sequence pattern known as the H-box, however such a region is not only found in DHp domains. Another type of domain called the Histidine Phosphotransfer (Hpt) domain can also become phosphorylated on a conserved His. With two alternative domains acting as phospho-His carriers, we begin to see the possibilities for variations to the TCS archetype. HPt domains are not usually located near CA domains; rather they function more independently and are not part of the kinase core. Thus, the location of the phosphorylatable His relative to the CA domain divides HKs into three main categories: Class I, Class1 hybrid, and Class II (or CheA) (Figure 1-2a,b). In Class I HKs, the invariant histidine is located in the H-box of the DHp domain, which lies adjacent to the CA domain (forming a kinase core). A Class I hybrid HK also contains a kinase core, but in addition contains one or more domains that can participate in phosphoryl transfer, such as a REC or Hpt domain. In Class II HKs, the H box is located in an Hpt domain that is not adjacent to the CA domain, rather one or more domains lie between them. In all classes, autophosphorylation of the HK occurs as a result of the CA domain, upon ATP binding, contacting and phosphorylating the active-site His on either a DHp or Hpt domain.

Four-Helix Bundles

Unlike the kinase core, Hpt domains lack autophosphorylation ability in addition to being unable to dephosphorylate RRs (35). Instead, these domains usually serve as intermediates in phosphorylation relays. For example, the *E. coli* ArcB HK has the same domain architecture and phosphorylation cascade as shown for Class I hybrid kinases in figure 1-2a. In the ArcB system, the ArcB Hpt domain acts as an intermediary between two REC domain in an alternating His-Asp-His-Asp signaling relay that ends with the phosphorylation of the RR ArcA, a global transcriptional regulator involved in respiratory growth and other cellular functions (42). Hpt domains can also participate in phospho-relays where they exist as unattached, stand-alone proteins. The Hpt protein YPD1, involved in osmotic-stress regulation, accepts a phosphoryl group from the REC domain of the hybrid-HK SLN1 and relays it to one of two independent RRs (84). In this system, YPD1 acts as an independent phospho-shuttle that is capable of distinguishing between three different REC domains.

Among the known structures of HPt domains are *S. typhimurium* and *T. Maritima* CheA (48, 76), *E. coli* ArcB (34, 35), *E. coli* YojN (57), *S. cerevisiae* YPD1 (61, 84), and *Zea mays* (maize) ZmHP2 (66). These domains have little sequence identity but possess a remarkably conserved structural architecture. Two-component Hpt domains are entirely helical and exist as monomers. They consist of very similar anti-parallel four-helix bundles, with a prominent helical hairpin formed by the two central helices (Figure 1-3a). The structures differ in the loop areas, and by the presence of additional shorter helices flanking the bundle. The active-site His is consistently located in the second helix of the bundle, with conserved residues in its vicinity identifying an H-box specific to Hpt domains. A conserved Gly near the active site creates an open space and renders the His more easily accessible. In fact, a mutation at this residue reduces

phosphorylation efficiency (66). The overall shape of the Hpt domain is marked by a kink or slight bend, the extent of which is different among various Hpt proteins (Figure 1-3b). ArcB has a pronounced kidney-shape (35), YPD1 is slightly curved (61), and YojN is flat (57). This concavity provides a docking surface for REC domains.

The four-helix bundle of DHp domains is formed by the dimerization of two monomers, as seen in structures of the complete kinase core: *B. subtilis* DesK (1), *T. maritima* HK853 (44), *G. stearothermophilus* KinB (11), and *T. maritima* ThKA (86). Each DHp domain monomer typically consists of two left-handed, anti-parallel alpha-helices that are connected by a hairpin. Two such monomers dimerize to create a four-helix bundle (Figure 1-4a). The four-helix bundle of the DHp domain is stabilized by heptad repeats of hydrophobic residues (1). The active-site His is located in the analogous position as in Hpt domains and is also exposed to solvent for easy access by the RR. However, four-helix bundles created by DHp domains are more or less symmetrical and possess two histidines available for phosphorylation, unlike monomeric Hpt domains which are asymmetrical and contain one phosphorylatable histidine.

The four-helix bundle formed by DHp domains is capable of adopting different orientations and different contacts with the CA domain, both of which are critical to its function (63). The bundle is able to alter its shape through the bending and cogwheeling motions inherent in the helices. Structures of HK853 and DesK support the idea that the shape changes of the central bundle serve to alter its molecular surface in response to each of its three activities-autokinase, phosphotransfer, and phosphatase (Figure 1-4b). For example, a pronounced helical bend is seen to different extents near the phosphorylation site. This bend becomes more pronounced upon phosphorylation (1), allowing for optimal contact between HK and RR when the HK is in phosphotransfer mode. The DHp and CA domains are held by a malleable linker,

allowing for adjustments to the interface between CA and DHp domains according to the HK function. In the presumed phosphatase mode, the CA domain is more rigidly attached to the DHp domain through an extensive binding interface. In order for autophosphorylation to occur, the CA domain must be released from the DHp domain to be able to swing around and contact the active-site His. Thus, structures of unphosphorylated DesK showing a more freely attached CA domain, with little to no interface contacts, are presumed to be the autokinase mode. The orientation of the bundle also dictates whether phosphorylation occurs in *trans* (the CA domain of one monomer phosphorylates the His of the other monomer) as is expected for DesK, EnvZ, and NtrB, or whether phosphorylation occurs in *cis* (each CA domain phosphorylates its attached DHp domain), as is expected for HK853 and PhoR (17).

The four-helix bundle is a conserved motif in two-components signaling, however anomalous structures exist that reveal signs of evolutionary change. The DHp domain of *B. subtilis* Spo0B (72) displays a four-helix bundle containing two exposed, phosphorylatable histidines. However this protein is rather a pseudo-HK in that its CA domain is not a kinase nor can it bind ATP. The role of Spo0B is to execute phosphotransfer as part of a His-Asp-His-Asp phospho-relay that uses four independent proteins. Briefly, a membrane kinase (KinA, KinB, KinC, KinD, or KinE) autophosphorylates on His then transfers the phosphate group to Asp on Spo0F. Spo0F continues the relay to His on Spo0B, and Spo0B phosphorylates Asp on the REC domain of Spo0A, which proceeds to control transcription of bacterial sporulation (73). Thus, Spo0B structurally resembles a kinase core, but behaves like a stand-alone Hpt protein, with a role strictly limited to phosphoryl shuttling (72).

The structure of the *E. coli* CheA kinase core is similar to Class I kinase core structures (12, 86), but like Spo0B, it is a somewhat unusual HK. Its central four-helix bundle is not

formed by dimerization of two DHp domains, but rather by dimerization of a pair of hair-pin helices that lack the active-site His (Figure 1-4c). Thus, the "dimerization" (or P3) domain of CheA resembles a DHp domain, along with the conserved H-box, but is catalytically nonfunctional. Each dimerization domain is attached to a CA domain whose ATP binding site faces in the opposite direction This is not surprising since the CA domain phosphorylates its resident Hpt domain (which then phosphorylates a REC domain). Thus, a CheA dimer is equipped with three four-helix bundles, two of which have phosphorylation sites.

CA Domains

Despite the relatively low sequence identity among HKs, the absence or presence of certain amino acid motifs in the kinase core has led to the classification of 11 HK subfamilies (30, 81, 82, 85). The H-box, as previously mentioned, lies in the DHp domain, while the other conserved regions, the N, D, F, and G boxes (formerly N, G1, F, and G2) are located in the CA domain and are involved in ATP binding and in forming the catalytic site.

In addition to the mentioned kinase core structures, several structures of isolated histidine-kinase CA domains have been elucidated in the apo form and bound to ADP, ATP analogues, or ATP (1, 12, 13, 43, 62, 67, 70). The core of CA domains is characterized by a conserved Bergerat fold (Figure 1-5a), a region composed of a flat four-stranded β-sheet with three α helices on top (22). Two of the helices pack on top of the strands, whereas one is located in a flexible loop that protrudes out of the structure and is characterized by high disorder in crystal structures (22, 67). This loop surrounds the deep ATP-binding pocket and is called the ATP-lid. The ATP-lid changes from an open to closed conformation upon ATP binding. In the open position the loop extends toward the solvent, whereas in the closed position it contacts the β

sheet. Although the ATP-binding pocket is deep, the base of the bound ATP is buried while the phosphate group protrudes out.

Differences in CA domains between different subfamilies are mainly located in the ATP-lid and loop regions. Some HKs, such as CheA (13), have much longer loop regions and also contain additional helices. The flexibility of the loop is attributed mainly to three conserved glycine residues in the G box, while the ATP-lid is anchored by a conserved phenylalanine that defines the F-box. However, not all HKs (subfamily 7) have all three glycines that compose the G-box, nor do some (subfamilies 7 and 8) contain an F-box. Such differences can be seen in the structures of the PhoQ (43) and DesK (70) CA domains (Figure 1-5b). The CA domain of DesK (belonging to subfamily 7) has a shorter loop region because it lacks a helix and an F-box, and is overall less flexible.

The surprising finding that CA domains possess a Bergerat type fold ties them to the otherwise unrelated GHL family, named after the family's representative members: DNA topoisomerase Gyrase B, chaperone protein HSP90, and DNA-repair enzyme MutL (22). GHL recently became GHKL after annexing the Histidine Kinase family. Besides their superimposable ATP-binding domains, HKs have little in common with GHL proteins and share very low sequence similarities, suggesting an old evolutionary ancestry. Structures show that the mode of ATP binding is similar in HKs and GHL proteins but the conformation of the ATP-lid distinguishes them. Differences probably evolved to account for the different functions. The GHL hydrolyze ATP to undergo subunit movements while the HKs bind ATP to catalyze histidine phosphorylation.

Response Regulators

The REC Domain

REC domains undergo magnesium-dependent phosphorylation on an invariant Asp, either as part of a phospho-relay or in RRs where they drive cellular change. Many RRs consist of an N-terminal REC and C-terminal effector domain connected by a flexible linker. Exceptions to this arrangement include intervening domains between the REC and effector domains, or stand-alone REC proteins that lack an output domain (28). Activated REC domains of RRs that are directly involved in regulating cellular adaptation do so by allowing C-terminal output function or, in the case of stand-alone REC domains, by binding to other proteins.

REC domains consists of about 120 amino acids and share about 26% identity among different RRs (28). Of the 200 or so REC domain structures, the majority adopt the signature topology of alternating β strands and α helices, or $(\beta/\alpha)_5$ (Figure 1-6). This consists of a parallel β-sheet, with a 2-1-3-4-5 topology, flanked by a set of 3 and 2 alpha-helices. The internal three β strands (β1, β3, β4) are marked by a series of hydrophobic residues that compose the hydrophobic core (15). At the C-terminal-end of these strands, and at β5, reside conserved residues that form the phosphorylation active site. This active site pocket is recognized by two highly conserved Asp residues (the first is sometimes Glu) following strand β1, an invariant phosphorylatable Asp at the end of β3, a highly conserved Thr/Ser at the end of β4, a somewhat conserved Phe/Tyr in the middle of β5, and a strictly conserved Lys following β5 (yellow ovals in Figure 1-6a,b). The adjacent Asp residues stabilize the phosphate by binding to the Mg^{2+} ion and water molecules. The Thr/Ser and Phe/Tyr are called "switch" residues because they are involved in phosphorylation-associated conformational changes. A small residue usually follows the Thr/Ser residue to allow access to the active site. Finally, the Lys is crucial for making

stabilizing salt bridges with phosphate oxygens and is required for both phosphorylation and dephosphorylation.

The Phosphorylation and Dephosphorylation Reactions

Whether phosphorylation is facilitated by the cognate HK, an Hpt domain, or a small-molecule phosphate donor, a divalent cation, preferably magnesium, is required. The role of the metal ion is to help align the REC domain for imminent phosphorylation (64). Thus, the Mg^{2+} cation is necessary for the phosphorylation chemistry rather than for stability of the active site (37), and can be removed afterwards (14). The octahedral coordination by Mg^{2+} in REC domains involves contacts with the three conserved Asp residues (the first Asp following strand β1-strand coordinates via a water molecule), a backbone carbonyl from the D+2 residue (Figure 1-6a), and two oxygen atoms from water molecules, one of which is replaced by a phosphate oxygen at the time of phosphorylation (15, 64). In this metal-bound configuration the Asp residues are oriented to accommodate the ion, making closer contacts than with a previously placed water molecule. The Mg^{2+} binding pocket can also accommodate other cations such as Mn^{2+} (lee2001), Ca^{2+} (40), or even Sm^{3+} (29), as seen by reported structures where these metals exhibit a similar coordination geometry as that of Mg^{2+}.

RRs are capable of autophosphorylation in the presence of a small molecule phosphate donor (15, 28, 65). Therefore it is thought that RRs actively participate in the enzymatic phosphorylation reactions with their cognate HKs rather than solely being inactive substrates. *In vitro*, RRs can use a high energy phosphate donor such as phosphoramidite, acetyl phosphate, or carbamoyl phosphate to autophosphorylate. The phosphorylation reaction mechanism is thought to begin by a nucleophilic attack by the Asp carboxylate oxygen (the one not coordinated to the

magnesium ion) on the phosphorus atom. The phosphorus atom transitions from a tetrahedral orientation to a trigonal bipyramidal transition-state whereby the attacking nucleophile is in the axial position. Active site residues and the metal cation stabilize the transition state while the leaving group is in line at the opposite axial position. The planar (equatorial) oxygen atoms attached to phosphorus make contacts with the metal ion, the Thr/Ser switch residue, the conserved Lys, and other protein backbone contacts. This mechanism of the phosphorylation reaction is presumed to be the same whether executed by an HK, an HPt domain, or a small molecule, although the rate of phosphoryl transfer from a small molecule is usually slower than from a cognate HK or Hpt domain.

The high energy, labile phospho-aspartyl bond is short-lived, lasting from a few seconds to a few hours regardless of the phosphate source (15). This renders structural endeavors to trap the REC in an activated state difficult and has led to the application of phosphate analogues that mimic phosphorylation and bestow more stabile active sites (18). Beryllofluoride (BeF_3^-) binds non-covalently to the phosphorylatable Asp, and confers transcriptional activation to similar levels as activation by phosphorylation (45, 79, 87). Like its phosphate counterpart, BeF_3^- forms a tetrahedral, however is unable to transition through the trigonal bi-pyramidal structure required by the phosphorus and therefore remains bound to the active site.

RRs are able to undergo magnesium-dependent autodephosphorylation. The mechanism of autodephosphorylation presumably proceeds in the reverse direction of phosphorylation, however the nucleophilic attack is launched by a water-oxygen (15, 33). Dephosphorylation by by a cognate HKs or by an auxiliary phosphatases (such as the dephosphorylation of CheY by CheZ (31)), is thought to occur by these proteins exploiting the inherent phosphatase ability of the RR, that is by helping to position the water for a nucleophilic attack (15).

Available research indicates that certain residues in the active site contribute to aligning the water molecule to initiate autodephosphorylation, or to stabilize the phosphorylated state (14, 15, 58). Different combinations of the D+2 and T/S+2 residues can alter the autophosphatase rate dramatically. They are thought to modify the accessibility of the water molecule to its targeted phosphorus atom. In phosphorylated FixJ, the His T+2 side-chain moves by 9Å compared to the unphosphorylated structure, and this movement is thought to screen the phosphoryl group from the nucleophilic water (14). Paradoxically, this residue may also act as a base to activate the nucleophilic water (32). Proteins that dephosphorylate RRs have been shown to execute phosphatase activity by exploiting the intrinsic autophosphatase mechanism of the RR, for example by using the D+2 residue to help align the attacking water molecule (15).

The conserved residues at the active site of REC domains are similar to the catalytic core of the haloacid dehalogenase (HAD) family. One member of this family also forms a phospho-Asp intermediate with a lifetime in the millisecond range. The equivalent D+2 residue is an Asp and has been proposed to assist in the dephosphorylation reaction. In RRs this position is often Asn, Gln, or Arg which could explain their longer phospho-Asp life spans compared to members of the HAD family (58). In Spo0F, changing the D+2 Lys to Asp increased autophosphatase activity while in CheB the equivalent residue, Glu, changed to Lys decreased autophosphatase activity (14).

Activated REC Domains

Comparisons of phosphorylated or BeF_3^--activated RRs show similar structural modifications and indicate equivalent forms of activation (3, 14, 15, 28, 32, 37, 40) . In general, activation results in no real change in secondary structure, however, displacements of β-strands

and α-helices occur, and to different extents among RRs. This is due to movements in the active site.

The crystal structure of *E. coli* $CheY^{(BF)\dagger}$ (37) illustrates a typically networked active site (Figure 1-7a). One fluoride forms a hydrogen bond with the conserved Thr, one forms a salt bridge with the conserved Lys, and another coordinates with the Mn^{2+} cation. The phosphoryl group is reinforced by contacts with main-chain amides of the D+2 (Asn59), D+1 (Trp58), and Thr87 residues. The conserved Asp residues and Lys further stabilize the active site.

In the phosphorylation process, the most distinguishable movement occurs as the Thr/Ser switch residue changes from being exposed to being buried. This movement is essential for making the hydrogen bond with a phospho-oxygen (or fluoride if activated by beryllofluoride) and effectively moves the β4-α4 loop by a few angstroms. The other switch residue, Phe/Tyr on strand β5, undergoes a rotomeric conformation from being exposed to being buried, occupying the space left behind by the Thr/Ser. The Lys on β5 has a slight repositioning, and makes contacts with the first of the conserved Asp (or Glu) residues.

In $CheY^{(BF)}$ both switch residues are oriented towards the active site upon phosphorylation. In addition, upon transitioning from an exposed to a buried rotameric state, the hydroxyl group of the switch residue Tyr106 makes a hydrogen bond with the main-chain carbonyl-oxygen of the T+2 residue (Glu89) (Figure 1-7b). This not only serves to anchor the Tyr but also helps stabilize the β4-α4 loop. This bond is also seen in other RRs with a flipped aromatic Tyr switch residue (2–4, 37, 68, 69). (Although this bond is absent in RRs containing a Phe in this position, the equivalent Phe101 residue in *S. meliloti* $FixJ^{Np}$ is stabilized by van der

† Superscripts are used to represent different domain regions or protein states: "C" denotes the C-terminal domain, "N" denotes the N-terminal domain, "p" refers to phosphorylated, and "BF" refers to beryllofluoride activated.

Waals contacts with the side-chain of its D+1 residue (14)). Prior to phosphorylation, the β4-α4 loop, in general, is flexible and often accompanies poor electron density and higher crystallographic B factors (37, 68). Upon phosphorylation, it becomes more rigid due to the hydrogen bonds associated with the conserved Tyr (if present), the conserved Ser/Thr, and the invariant Lys.

$CheY^{(BF)}$ and $FixJ^{Np}$ show overall similar structural modifications, namely movements along helix α4, the β4-α4 loop, and strand β5 (Figure 1-7c). $CheY^{(BF)}$ has a closer bond between its conserved Lys and first conserved Asp residue, leading to an additional slight displacement of its β5-α5 loop (though this may be due to the lack of a metal cation in the $FixJ^{Np}$ structure). The main result of repositioning residues in the active-site is to extend and stabilize the β4-α4 loop and to reposition helix α4 and sometimes also helix α5. The prevalence of this theme among activated RRs has lead to this region being named the "α4-β5-α5 face".

RRs such as $Spo0A^{Np}$ (40) and cyanobacterial $RcpA^{p}$ (9), exhibit modest changes upon activation, whereas *S. typhimurium* $NtrC^{Np}$ (32, 36) and *E. coli* $PhoB^{Np}$ (3) demonstrate more dramatic changes in α4-β5-α5 face upon activation. $PhoB^{Np}$ and $NtrC^{Np}$ experience a rearrangement of α4, the β4-α4 loop, and the α4-β5 loop. In $PhoB^{Np}$ this causes an extension of helix α4, rotating it by almost 100º, which in turn exposes hydrophobic residues that contribute to a dimer interface (Figure 1-7c). In $NtrC^{Np}$, helix α4 partially unwinds so that some residues become part of the β4-α4 loop while two residues of the α4-β5 loop incorporate into the C-terminal-end of helix α4.

Several RRs have crystallized as "active-like" without the presence of an activating agent, as indicated by a certain dimeric state or by the switch residues observed in the buried conformation (8, 10, 69). Also, overexpression of some unphosphorylatable mutant RRs can

bestow transcription and compensate for a deletion phenotype (27). These observations can be explained by recent evidence suggesting that phosphorylation of RRs does not induce a conformational change, but rather serves to drive an active-inactive equilibrium towards the active conformation. NMR studies of $NtrC^{N}$ in phosphorylated, unphosphorylated, and partially active states, demonstrate that the active and inactive forms coexist (75). About 2-10% of active species is present in the unphosphorylated wild-type protein. This newer paradigm is one of NtrC toggling between two states, rather than adopting random states, with phosphorylation serving to shift a preexisting equilibrium rather than to convert NtrC to a new conformation.

Histidine Kinase and Response Regulator Complexes

Structures have been reported of HKs or phosphotransferases bound to their cognate RR, with the RR being in activated or inactivated states (17, 74, 83, 85, 86, 89, 91). The overall architectural fold of the proteins remains the same and the REC domains are positioned below the phosphorylatable His (Figure 1-8a). Contacts between the partners are not extensive, which is consistent with a passing interaction. A salient feature common to these complexes is that the phosphorylatable His and phosphorylatable Asp are directly facing each other and are optimally aligned for phosphoryl transfer (Figure 1-8b). Structures of complexes containing activated verses inactivated REC domains have modest differences, mainly displaying tighter binding between the two proteins and possibly better alignment (74, 91).

The structure of $Spo0F^{(BF)}$ with Spo0B (*B. subtilis*), of $Trra^{(BF)}$ with ThkA (*T. maritima*), and of $RR468^{(BF)}$ with HK853 (*T. maritima*) show a 2:2 HK:RR ratio with the RRs located at opposite sides of the DHp domain. Common modes of interaction exist in spite of the HKs belonging to different classes. In the three complexes, helix α1 and the β5-α5 loop of the RR

surround helix α1 of the DHp domain (containing the phosphorylatable His). The loop connecting the two DHp helices also makes contact with helix α1 of the RR (Figure 1-8a, right). In addition, the RR β4-α4 loop makes contacts with helix α2' of the neighboring HK protomer, so that each RR contacts three of the four helices in the DHp four-helix bundle. A surprising feature in the HK853-RR468 structure is that the RR also made contacts with the CA domain and its connecting linker to the DHp domain (17). In general, this structure is consistent with biochemical data that implicate certain residues as being important for the intermolecular contacts. Seven nonconserved residues in HKs found to be critical for selecting the correct RR are shown to be involved in HK853-RR468 complex interactions. Similarly, a few nonconserved residues in helix α1 of RRs that are thought to be important for choosing the correct HK are consistent with $RR468^{(BF)}$ interactions with HK853.

The structure of $SLN1^{(BF)}$ bound to YPD1 (*S. cereviciae*) shows a 1:1 structure where $SLN1^{(BF)}$ contacts the first three of four conserved helices in YPD1, this being accomplished also via helix α1 and the loop regions of the RR (91). The two proteins also have strong surface complementarity (Figure 1-8c). This glove-like fit by which HKs and Hpt domains dock with their cognate RRs explains their faster rates of phosphorylation compared to REC-domain phosphorylation by a small molecule (7). Surface complementarity, therefore, could provide another level of control for binding to the correct RR.

Types of Output Domains and their Functions

Unlike REC domains which are mostly redundant, effector domains obtain a variety of functions specific to each TCS. N-terminal REC domains control at least 60 types of C-terminal effector domains (27, 28), some of which are listed in Table 1-1. The predominant output is

DNA binding (~63%), having the effect of regulating a few to several operons. A smaller group is relegated to enzymatic function (~13%), protein or ligand binding (~3%), and RNA binding (~1%). The remaining (~17%) serve as stand-alone proteins.

There are three main subfamilies in the family of DNA-binding RRs: OmpR/PhoB, NarL/FixJ, and NtrC/DctD (26). Structures of individual domains exist from all three subfamilies, and structures of full-length RRs and DNA-bound structures exist from the OmpR/PhoB and NarL/FixJ subfamilies. The largest subfamily, OmpR/PhoB, employ a winged-helix motif for DNA binding, the NarL/FixJ members utilize a four-helix helix-turn-helix motif, and the NtrC/DctD subfamily members use a Fis (factor for inversion stimulation) type of helix-turn-helix binding (26, 28). Recently, a new subfamily, LytR, was discovered. This unusual DNA-binding domain consists mostly of β strands and DNA recognition is achieved through the residues in the loops between the strands (60).

Enzymatic output domains can appear in the form of methylesterases, diguanylate cyclases, cyclic-diguanylate-specific phosphodiesterases, and histidine kinases, among others. Those with available structures have been characterized, such as CheB (19) whose REC domain prevents the C-terminal methylesterase domain from access to its substrate until phosphorylation releases the inhibition. PleD is an enzymatic RR with two REC domains followed by a diguanylate cyclase (GGDEF) domain. Upon phosphorylation, two PleD monomers, each bound to GMP, dimerize to create cyclic diguanylate (c-di-GMP) (25, 77). Output domains that function as c-di-GMP-specific phosphodiesterases are EAL and HD-GYP domains (names based on conserved residues), represented by RRs PyrR and RpfG, respectively (25, 26).

Bacteria utilize a substantial amount of stand-alone REC proteins, and in Archea they comprise about 50% of RRs. (15). Some stand-alone REC proteins regulate output by binding to

other proteins, such as the chemotaxic protein CheY which regulates motility by binding to the flagellar switch protein FliM, consequentially transitioning bacterial motion from smooth swimming to tumbling (39). Alternatively stand-alone REC proteins are involved in phosphorelays, such as the sporulation RR Spo0F, which acts as a phosphoryl intermediate between Spo0B and Spo0A (74), and is analogous to REC domains contained in hybrid kinases.

Not surprisingly, RR output domains are weakly conserved, allowing for their diversity of functions. Many domains have been identified as a consequence of being attached to REC domains but their mechanisms need more investigation. In Archaea, some RRs are combined with PAS or GAF domains and the function of these RRs remains elusive (26). The list of eclectic domains that can be attached to REC domains to control cellular physiology will most likely continue to expand as more RRs are discovered.

A Diversity of Regulation

The effect of REC domains equipped with diverse output domains is a diversity of regulatory mechanisms. REC domains exercise control by positive regulation (directly allowing output) or negative regulation (releasing an inhibitory affect that subsequently allows output). An interface between the REC and output domains implies negative regulation since the REC domain often precludes the output domain from carrying out its function. RRs such as FixJ, NarL, CheB, and Spo0A, are negatively regulated since phosphorylation is predicted to loosen the domain interface and enable the, otherwise fastened, C-terminal domain to implement its function (5, 14, 19, 41). However, the recent torrent of activated REC domain structures shows that regulation can be both positive and negative, whereby upon release of the output domain the REC domains are able to dimerize. Examples include phosphorylated FixJ and members of the

OmpR/PhoB subfamily (69). Interdomain interfaces are not always present in RRs, making the task of deciphering their regulatory mechanisms more difficult, however these RRs are thought to be governed by positive regulation whereby activation allows the formation of suitable dimers (27). Furthermore, where an interface exists, it does not necessarily mandate the same regulatory mechanism even within subfamily members.

Biochemical and structural evidence show that positive or negative regulation via intra- and intermolecular domain surface contacts often involve or foretell involvement of the α4-β5-α5 face. Since activation often divulges or buries certain residues at the α4-β5-α5 face, this region is often manipulated to allow for protein-protein interactions, dimerization, or intramolecular binding.

The OmpR/PhoB Subfamily

Positive regulation by dimerization is ostensibly the predominant method of mechanistic control in the OmpR/PhoB subfamily where structures of several activated REC domain show a common dimer at the α4-β5-α5 surface (2, 3, 7, 10, 68, 69) (Figure 1-9). Conserved residues strictly in the OmpR/PhoB subfamily are responsible for sustaining the dimer interface (68), suggesting that the motive of activation in this subfamily is to drive symmetric REC-domain dimerization and oligomerization. Active dimers and a flexible linker are thought to facilitate the asymmetric binding of the output domains to tandem DNA sites (3, 46, 69).

Structures of full-length inactivated RRs from the OmpR/PhoB subfamily, however, also suggest that regulation can be negative, and the variety seen in domain interfaces precludes a canonical regulatory mechanism. Where an interface exists, it consistently involves the α4-β5-α5 surface (Figure 1-10). *M. tuberculosis* PrrA (49) and *M. tuberculosis* MtrA (24) both have

extensive interfaces involving the α4-β5-α5 region. These interfaces both obstruct the recognition helix- although using different contacts- and suggest a negative regulatory mechanism. PrrA and MtrA, along with *T. maritima* DrrB (56), also anchor their C-terminal domain with a hydrogen bond to the Tyr switch residue. In all three cases the hydrogen bond is in a distinct residue-equivalent position, but common to all is that this bond must be broken for proper REC dimerization to occur (Figure 1-10a-c). Although DrrB has an extensive domain interface, which implies negative regulation, it paradoxically has a solvent accessible recognition helix. *T. maritima* DrrD (16) also has an exposed recognition helix, however lacks a domain interface (Figure 1-10d). The influence of the REC domain in DrrB and DrrD remains unclear, but could entail positive regulation by the formation of active dimers for DrrD and both positive and negative regulation for DrrB. Similar to DrrD, the full-length structure of inactivated *M. tuberculosis* PhoP (46) shows an exposed recognition helix and an absence of an interface (Figure 1-10e). This RR crystallized as a dimer at the α4-β5-α5 face. The proposed effect of phosphorylation is to shift and strengthen the pre-existing dimer and enhance DNA binding affinity by bringing the effector domains into closer proximity. Thus, although members of the OmpR/PhoB subfamily seem to depend on the same mode of dimerization, additional specific controls are implemented through different interdomain relationships.

Variations in domain interfaces could possibly alter the equilibrium of populations possessing different conformational states. Proteins with looser interfaces may be prone to adopt active-like conformations and give rise to a basal level of DNA binding. Several inactivated OmpR/PhoB REC domains crystallized as "active-like" dimers where positions of residues in the active site somewhat resemble the activated state (2, 68, 69). High protein concentrations during crystallization may drive the equilibrium toward the active-like state, which is supported by the

observation that some unphosphorylated OmpR/PhoB members can bind DNA at high concentrations (46, 69). These quasi-activated dimers may stimulate transcription at low-affinity binding sites, whereas high-affinity binding sites, which are often indicative of RR oligomerization, require phosphorylation.

The NtrC/DctD Subfamily

The NtrC/DctD subfamily relies on oligomerization to induce transcription. Members are composed of three domains: an N-terminal REC domain, a central AAA+ ATPase domain, and a C-terminal DNA binding domain. Upon phosphorylation, the central domain forms a hexameric or heptameric ring assembly that contacts the sigma-54 subunit of RNA polymerase and uses ATP hydrolysis to make a polymerase complex suitable for transcription initiation (8, 27, 51). No full-length structures from this subfamily are available, however structures of individual domains and of combined REC and central domains have been elucidated (8, 21, 32, 39, 51) (Figure 1-11a).

Despite the strong sequence identity among these proteins, regulatory mechanisms differ (27). In $NtrC^{Np}$, hydrophobic residues that become exposed on helix α4 are thought to interact with the central domain (of another NtrC molecule) and allow oligomerization, while the movement of the switch Tyr helps create a binding surface (32). Thus, the central domain of NtrC is intrinsically incapable of assembly and requires positive activation by the REC domain. This is in contrast to the central domains of NtrC1, NtrC4, and DctD, which are intrinsically active but the REC domain of each inhibits the central domain by forming inactive dimers (8, 21, 38, 51) (Figure 1-11b). In these cases, and in contrast to OmpR/PhoB members, dimerization at α4-β5-α5 represents the inactive form. Upon phosphorylation, the REC domains dimerize akin

to phosphorylated $FixJ^N$ (from *S. meliloti*) (14), using helix α4 and strand β5 (Figure 1-11c,d). (Concomitantly, the aromatic switch residue becomes buried.) The changed dimer interface disrupts the inhibition of the central domain and permits domain assembly (51). Even further control within the subfamily ensues due to the weaker dimer interface of inactive $NtrC4^N$ compared to that of $NtrC1^N$ and $DctD^N$, rendering NtrC4 a weaker transcriptional inhibitor (8).

The NarL/FixJ Subfamily

Besides $FixJ^{Np}$ no other RR in this subfamily demonstrates a direct use of the α4-β5-α5 region. This could, however, be due to the hitherto scarcity of activated structures. Some members of this subfamily do however exhibit unusual linker regions compared to other RR subfamilies, which could be a source of regulation. The structure of the activated *S. pneumonia* $spr1814^N$, shows dimerization at the linker region of each monomer. Furthermore, this linker region is positioned within the α4-β5-α5 face (50) (Figure 1-12a).

As seen in the OmpR/PhoB subfamily, full length structures in the NarL/FixJ subfamily show diversity in domain interfaces and possible modes of regulation. Full-length *E. coli* NarL has an extensive interface (4, 5) that is predominantly formed by loop and linker regions (Figure 1-12b). This interface obstructs the recognition helix, which putatively becomes unfastened upon phosphorylation and domain separation. In conjunction, the linker region must relocate to allow output-domain dimerization upon DNA binding.

Full-length StyR from *P. fluorescens*, unlike NarL, lacks a domain interface (Figure 1-12c). Instead of a separate linker helix, as seen in NarL, TraR, and $RcsB^C$, StyR's linker helix is merged with the canonical helix α5 of the REC domain to create a remarkably elongated helix that separates the two domains by more than 16Å (47). (This is notably reminiscent of the

elongated α5 in other RRs, such as NtrC4.) Since the recognition helix of StyR is exposed, the idea of released inhibition by phosphorylation is negated, and instead leads to the proposition that phosphorylation assists in the formation of a dimer that enhances the DNA binding, similar to the activation induced dimerization proposed for FixJ and some members of the OmpR/PhoB subfamily. However, the nature of the dimer interface is unknown.

Full-length DosR, has an extensive interface (Figure 1-12d) and its helix α10 shows a different conformation from that seen in the DNA-bound structure. The DosR helix α10 packs against α1 of the N-terminal domain (80). One of its residues makes a hydrogen bond with the phosphorylatable Asp54. The recognition helix also makes contacts with the N-terminal domain and is proposed to become exposed upon phosphorylation, thus also indicative of negative regulation. Furthermore, the REC domain of this protein has an anomalous $(\beta/\alpha)_4$ topology and the residues located in β5 and α5 of canonical REC domains, shift into the linker helix α5 in DosR. There are two alpha helices in the linker of DosR, which make extensive contacts with the REC domain. In the inactive dimer, the two-helical linkers join to form a four-helix bundle. Phosphorylation is proposed to release helix α10, induce a conformational change that would restore the canonical REC domain, expose the recognition helix, and create a functional dimer along α10 that is capable of binding DNA. This proposed model would also allow dimerization of the REC domains at the α4-β5-α5 face.

The interdomain region of RcsB has a linker helix that is very similar to helix α6 of NarL (20, 55) (Figure 1-12b,e). In NarL and RcsB this linker is expected to move so as to facilitate proper DNA binding (5, 55). This same type of linker and release mechanism is also predicted for VraR (20). The analogous linker in TraR adopts different conformations in each monomer, creating an asymmetric dimer that binds DNA (71, 90). Interestingly, one monomer has almost

no contacts between the N- and C-terminal domains, whereas the other monomer maintains several interdomain contacts (90). Thus the linker is flexible, often disordered in x-ray structures, and allows mobility of the N- and C-terminal domains upon activation and DNA binding.

Other Response Regulators

Positive and negative regulation through manipulation of the α4-β5-α5 face is also demonstrated in other RRs. The full length structure of *S. typhimurium* CheB (19) confirms an intricate domain interface between the REC and output domains at the α4-β5-α5 surface, and the Phe switch residue participates in interdomain contacts. Upon phosphorylation, blocked residues required for substrate binding become exposed (Figure 1-13a). $CheY^{(BF)}$ binds to the flagellar-switch protein, FLiM, and regulates bacterial motion (39). Binding occurs through the α4-β5-α5 face and movement of the switch Tyr to a buried position is a crucial prerequisite to creating a proper binding surface, to which the Tyr also contributes contacts (Figure 1-13b). CheY also uses the α4-β5-α5 region to bind the P2 domain of CheA (78), and to a fragment of CheZ (31). Lastly, PleD, an RR of which a full-length, activated structure exists, exercises positive and inhibitory regulation depending on its mode of inter- and intramolecular contacts. At activation, the two REC domains adjust their pre-existing α4-β5-α5 dimer to a better aligned, tighter interface, and, as seen in other RRs, this movement is concomitant with the inward motion of the switch Phe residue. The catalytically competent dimer confers proper interactions between the effector domains, leading to catalysis (77).

Conclusions

Two component systems are an exemplary model of the way in which nature can take advantage of modular domains to fulfill a variety of different cellular needs that are required for adaptation to a fluctuating environment. Each system uses a common foundation but adds rearrangements or new domain structures to create specificity for a particular system. HPt, CA, REC, and Hpt are common domains to all TCSs but can be fused together in a myriad of arrangements. Added specificity is incurred by slight differences at binding sites, surface area complementarity between proteins, different domain interfaces, dimerization and oligomerization, or by the fusion to other signaling domains. The α4-β5-α5 region of RRs points to a common site for allosteric modifications, giving rise to a diversity of control through layers of inter- and intra-molecular interactions upon activation.

Text Box - Domains Discussed in Text

AAA+ ATPase
A conserved ATPase domain of the AAA+ superfamily. In RRs, this domain forms a ring structure that contacts sigma-54 of RNA polymerase and, upon ATP hydrolysis, alters the promoter complex to initiate transcription (8, 51).

CA (Catalytic and ATP-binding)
The conserved ATP-binding domain of the kinase core (28).

DHp (Dimerization and Histidine phosphotransfer)
The domain in the kinase core that houses the phosphorylatable His and dimerizes with another DHp domain to create a four-helix bundle (28).

EAL
Named after conserved residues by which it was identified. A phosphodiesterase domain that hydrolyzes cyclic diguanylate (c-di-GMP) (26, 28).

Effector (or Output)
A domain of a response regulator that is attached to REC domain and is responsible for executing the functions of the response regulator upon phosphorylation (28).

Fis helix-turn-helix
A helix-turn-helix DNA-binding motif represented by the transcriptional regulator, Fis (factor for inversion stimulation) protein (27).

GAF (cyclic GMP, Adenylyl cyclase, Fhla)
A cytoplasmic sensory and signaling domain found in a variety of enzymes that can bind several small ligands and is structurally similar to PAS (25, 28, 54).

GGDEF
A diguanylate cyclase domain that synthesizes cyclic diguanylate (c-di-GMP) from two GTP molecules (25, 28).

HAMP (histidine kinase, adenylyl cyclase, methyl-accepting chemotaxis proteins and phosphatase)
A signaling domain comprised of two amphipathic helices separated by a loop region. This domain is usually found in the linker region of class I HKs, located in between the second transmembrane helix and the kinase core (28, 44).

Hpt (Histidine phosphotransfer)
A domain (or stand-alone protein) constituting a four-helix bundle that has a phosphorylatable His but does not dimerize. This domain is often found in phosphorelays (28).

PAS (Per-Arnt-Sim)
Named after the proteins with which it is associated. A cytoplasmic sensory and signaling domain that can bind several small ligands and is structurally similar to GAF (28, 54).

REC (Receiver)
A domain of a RR or a stand-alone RR, with a conserved $(\beta/\alpha)_5$ fold and a phosphorylatable Asp that becomes phosphorylated by a DHp or Hpt domain (15, 28).

Figure 1-1. An archetype of the two-component system. Upon stimulation by a signal (brown), the HK sensor domain (beige) sends a signal to the cytoplasmic region that reaches the kinase core, which is composed of the DHp domain (red) and CA (cyan) domains. Autophosphorylation on an invariant histidine (H) ensues using ATP. The HK then phosphorylates the RR on an invariant aspartate (D) in the N-terminal REC domain (blue), rendering its C-terminal effector domain (yellow) competent to implement a response. The phosphoryl group is represented by a yellow diamond, and arrows above the diagram indicate the direction of phosphoryl transfer.

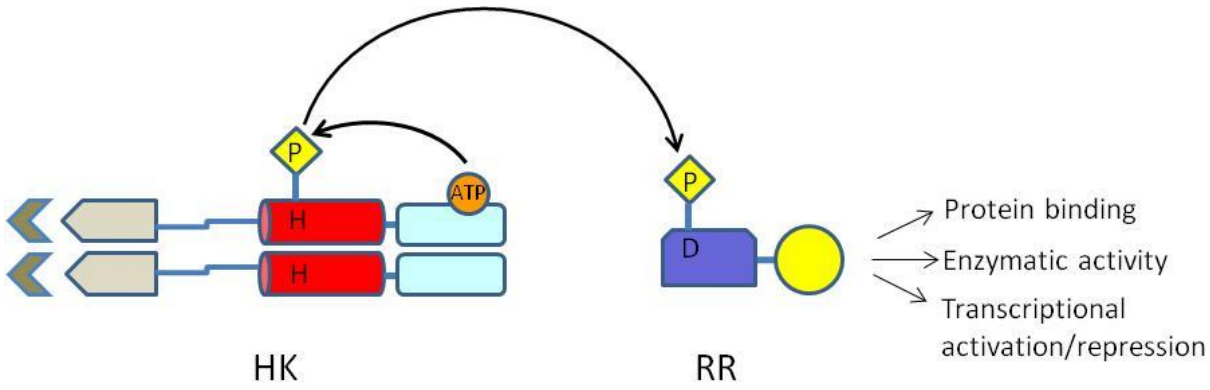
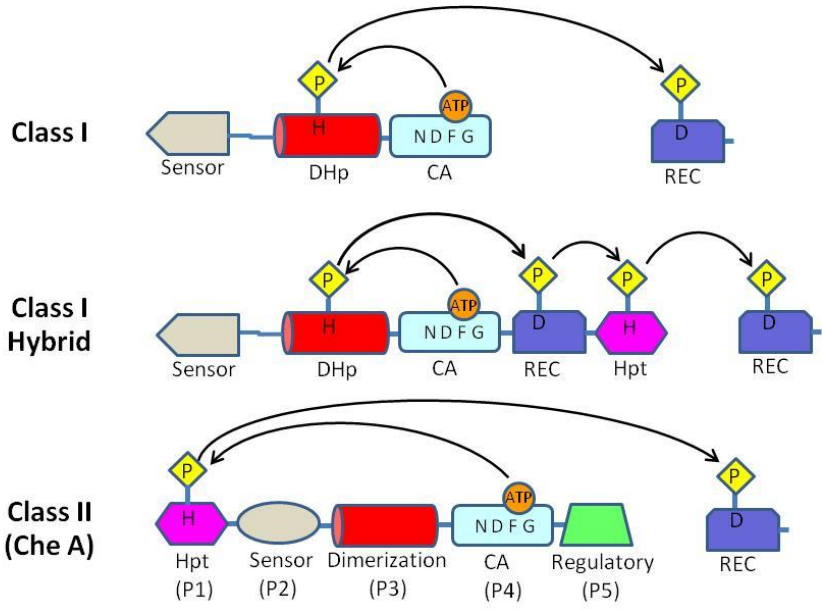


Figure 1-2. Types of HKs and domains of TCSs. **(a)** A schematic representation of the types of histidine kinases and direction of phosphoryl transfer in TCSs. HKs are divided based on the location of the conserved His compared to the CA domain. In class I and class I hybrid HKs, the conserved His is located in a domain that is adjacent to the CA domain, though they differ in that class I hybrid is capable of an internal phospho-relay. Class II kinases (such as CheA) have one or more domains in between the CA domain and the domain carrying the active-site His. The phosphoryl group is represented by a yellow diamond, and arrows indicate the direction of phosphoryl transfer. Note: sensor domains are not necessarily membrane bound. **(b)** Ribbon representations of conserved domains in TCSs with site of phosphorylation shown in yellow stick model: the kinase core (PDB ID: 2C2A) consisting of the DHp (red) and CA (cyan) domains, respectively, with the CA domain containing a stick model of ATP; the Hpt domain (magenta, PDB ID: 2A0B); the REC domain (blue, PDB ID: 2CHF). These images, and all subsequent structure pictures in this chapter, were created with Pymol (59).

(a)



(b)

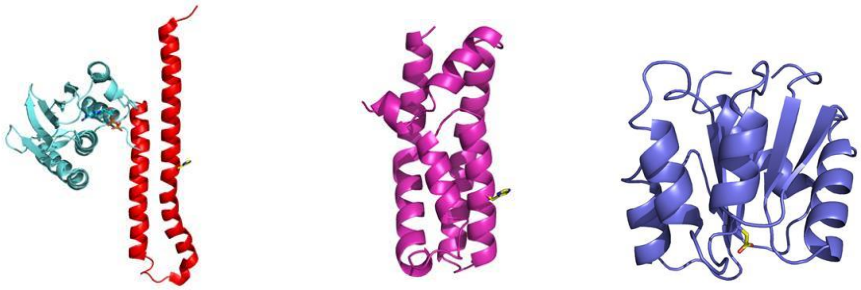
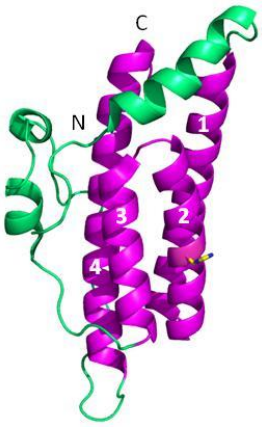


Figure 1-3. The Hpt domain. **(a)** *S. cerevisiae* YPD1 (PDB ID: 1C02), a stand-alone Hpt protein. The conserved four-helix bundle is shown in magenta with the helices labeled 1-4, while nonconserved regions are shown in green. The conserved hairpin is formed by helices α2 and α3. The active-site His (yellow stick) protrudes out of helix α2 and is solvent exposed. **(b)** Comparison of the helical tilt between *S. cerevisiae* YPD1 (PDB ID: 1C02), *E. coli* ArcB (PDB ID: 2A0B), and *E. coli* YojN (PDB ID: 1SR2, part of N-terminal loop not shown). The color scheme is the same as in (a).

(a)



(b)

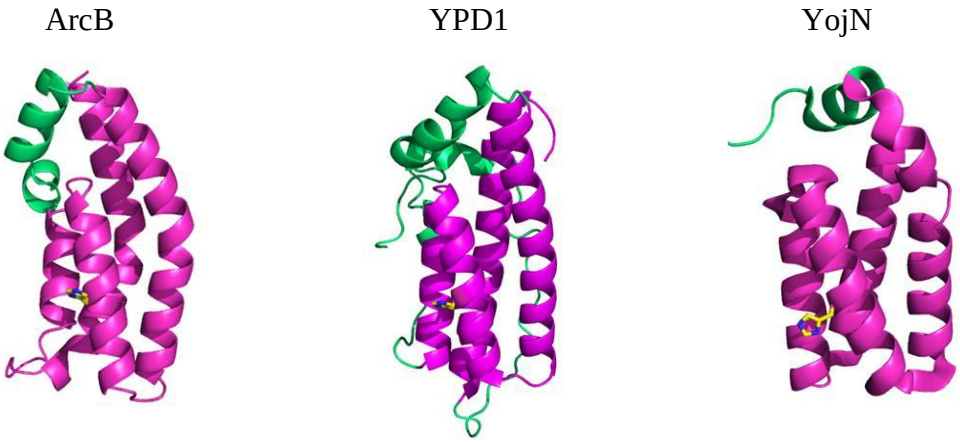
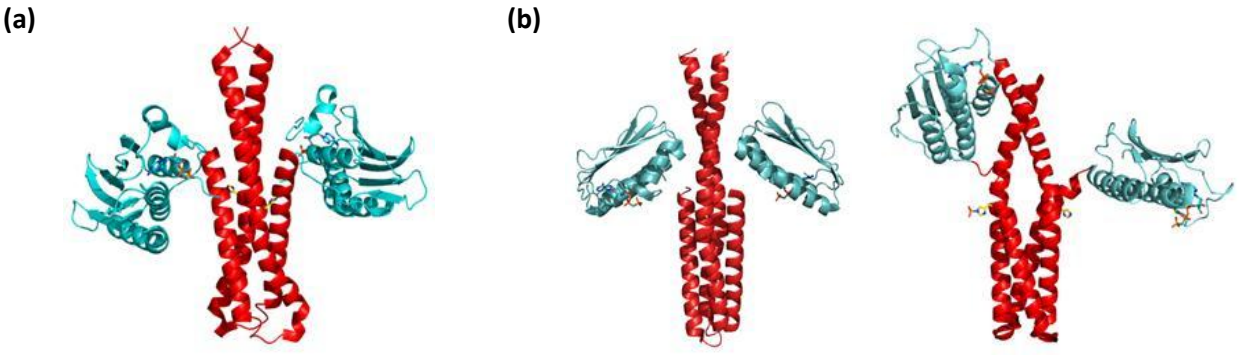


Figure 1-4. The kinase core. **(a)** The HK853 kinase core (PDB ID: 2C2A). The helical hairpin of each monomer forms the four-helix bundle and extended central helices (red). The CA domain (cyan) is held by a flexible linker and in this conformation makes extensive contacts with the open hairpin segment of the DHp domain. The His phosphorylation site and ATP molecule are shown in stick representation. **(b)** DesK kinase cores (same color scheme) in the presumed kinase competent mode (left, PDB ID: 3GIE) and phosphorylated state (right, PDB ID: 3GIG). The CA domain can adopt different orientations with respect to the four-helix bundle based on the working function of the HK. Both states show little to no interdomain contacts, unlike the phosphatase mode structure (not shown), which resembles the HK853 structure and has an extensive interdomain interface. The asymmetric phosphorylated DesK structure (where only one His is phosphorylated) has a pronounced helical bend at the His phosphorylation site and the CA domain of either protomer makes little to no contacts with the four helix bundle. **(c)** A cartoon of the cytoplasmic HK CheA in its dimeric form. The CA domain phosphorylates only the Hpt domains while the central DHp-like domains are strictly used for dimerization. This picture is a reproduction from a published version (23).



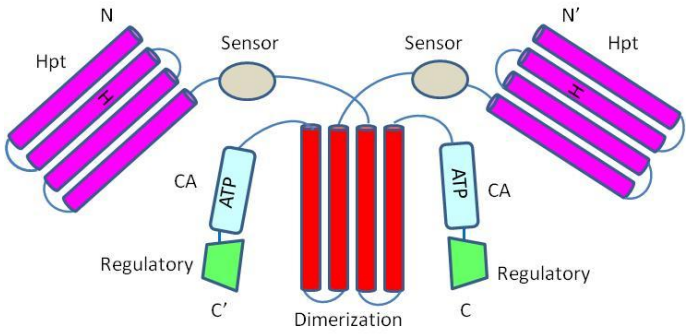
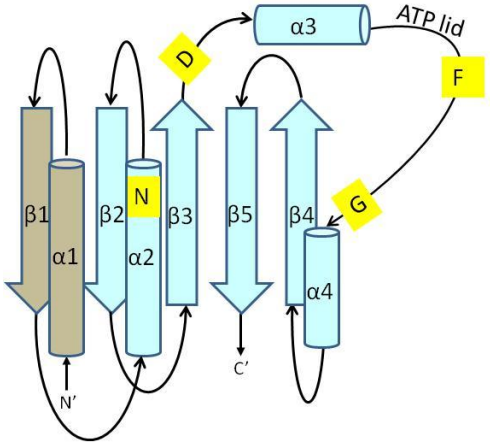


Figure 1-5. The CA domain. **(a)** General topology of the CA domain in HKs. Secondary structures not part of the Bergerat fold are shown in beige. Alpha helices are represented by cylinders, beta strands as arrows. The conserved regions (N, D, F, G) of the ATP-binding domain are indicated by yellow boxes. **(b)** The CA domains of PhoQ (PDB ID: 1ID0) and DesK (PDB ID: 3EHG) with the same color scheme as in (a). The two structures have slight differences in their ATP lids, which classifies them in different HK subfamilies. The ATP lid of DesK is shorter than that of PhoQ and also lacks a helix and an F box.

(a)



(b) **PhoQ** **DesK**

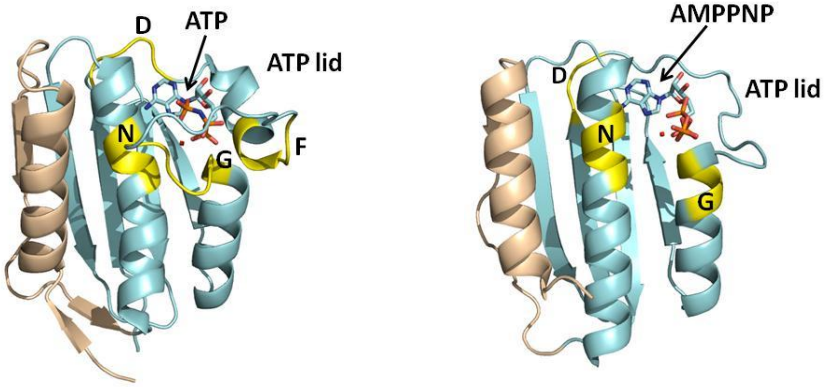
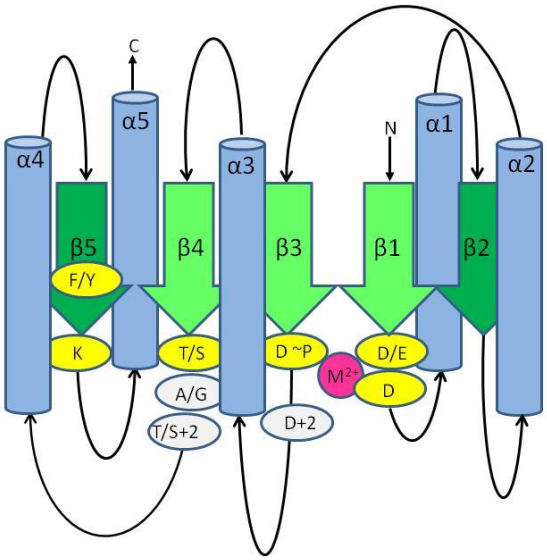


Figure 1-6. The REC domain. **(a)** The conserved topology of the response regulator REC domain (picture reproduced from a published version (15)). Beta strands are illustrated as green arrows, alpha helices are illustrated as blue cylinders. Lighter green beta-strands emphasize the hydrophobic core. Highly conserved residues in the active site are labeled in yellow ovals and other residues of interest are labeled in gray ovals. The location of the divalent metal cation (magenta) is shown near the three highly conserved Asp residues, one of which is the phosphorylatable Asp (D ~P). The residue that occupies the position two amino-acids after the phosphorylation site is referred to as "D+2," and similarly designated is the "T/S +2" residue. **(b)** The inactivated structure of *E. coli* CheY, a stand-alone RR, is shown (PDB ID: 2CHF) with the same color scheme and orientation as in (a). The site of Asp phosphorylation is displayed in yellow stick model.

(a)



(b)

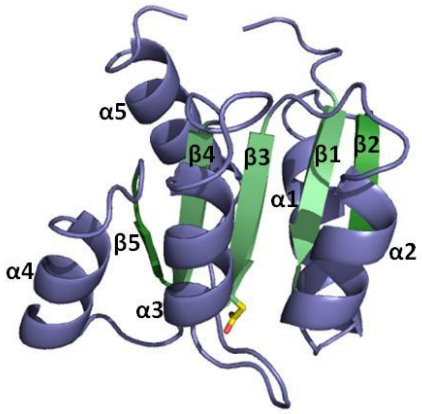
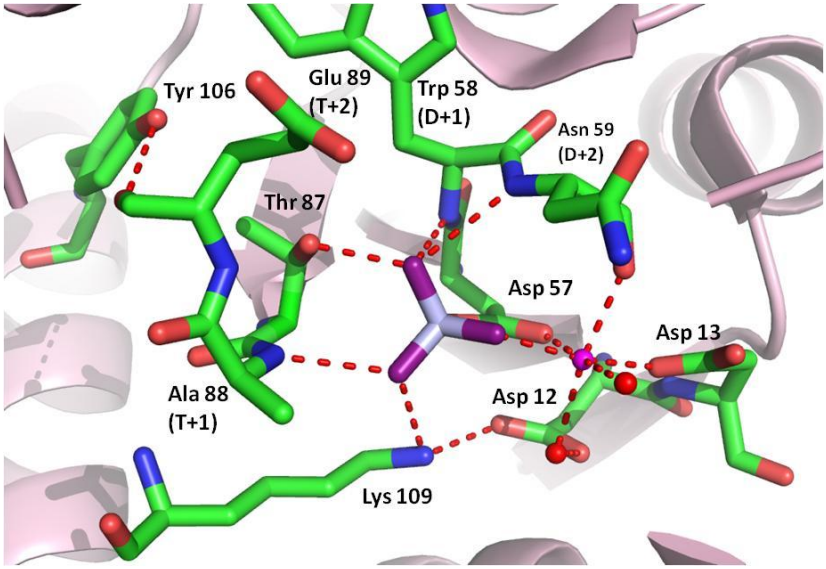
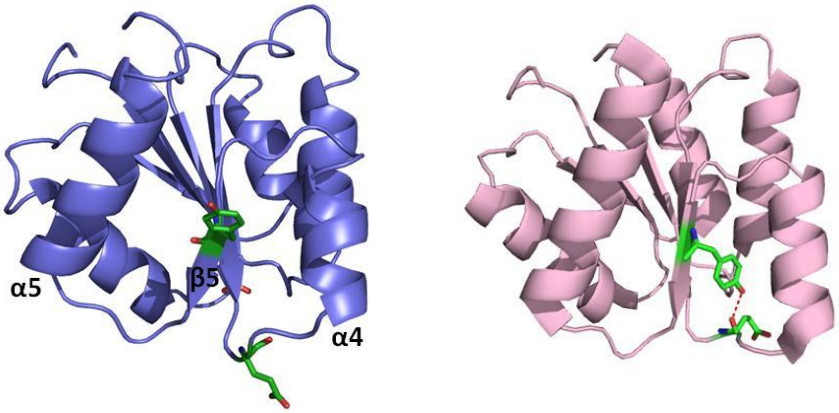


Figure 1-7. Activated REC domains. **(a)** A close up of the networked active-site of the *E. coli* $CheY^{(BF)}$ structure (PDB ID: 1FQW). Hydrogen bonds and salt bridges are shown as red dotted lines, the manganese ion is depicted by a magenta sphere, and water molecules as red spheres. BeF_3^- , shown in violet and purple, is noncovalently bound to the phospho-accepting Asp 57. **(b)** The exposed Tyr106 switch residue in inactivated (left, PDB ID: 2CHF) verses activated (right) (PDB ID: 1FQW) CheY. In the activated structure, this residue is stabilized by a hydrogen bond between the Tyr106 hydroxyl group and the main-chain carbonyl of Glu89 (the T+2 residue). **(c)** Overlays of the inactivated (blue) and activated (pink) structures of CheY (PDB IDs: 1CHN, 1FQW), $FixJ^{Np}$ (PDB IDs: 1DCK, 1D5W), and $PhoB^{N(BF)}$ (PDB IDs: 1B00, 1ZES), all in the same orientation though structural features are only labeled in CheY. Structural perturbations can be seen to different extents at the α4-β5-α5 surface, with the change in helix α4 of PhoB being the most drastic.

(a)



(b)



(c) CheY FixJ PhoB

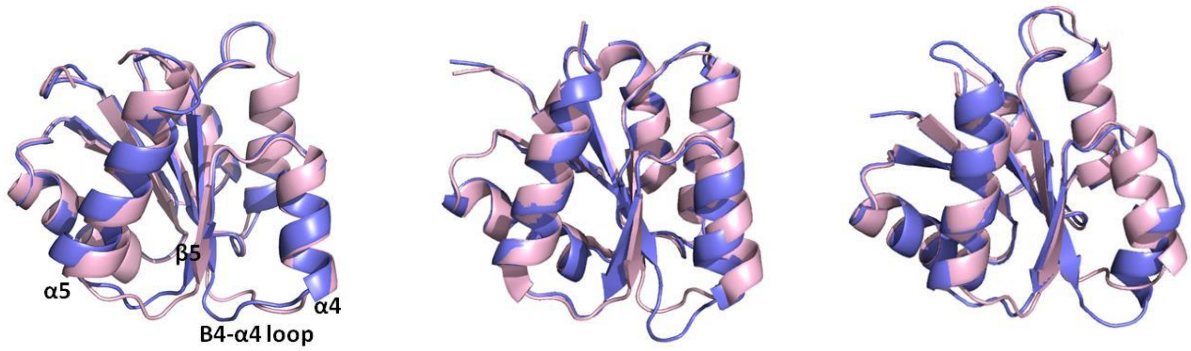
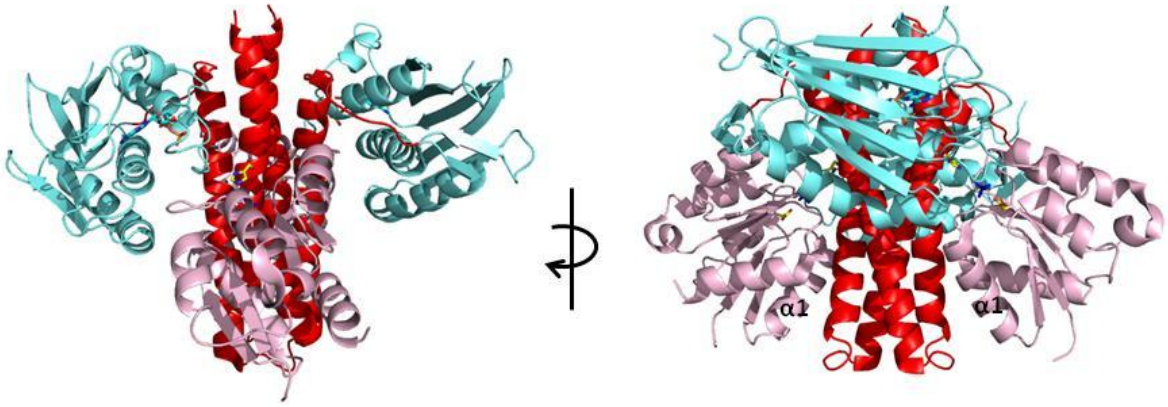
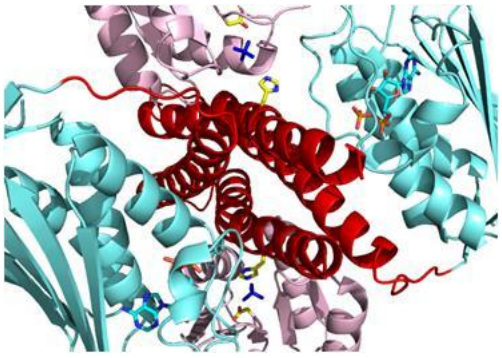


Figure 1-8. HK- and Hpt-REC domain complexes. **(a)** Two side views of the HK853-RR468 complex (PDB ID: 3DGE). Two RR468 molecules (pink) are bound on opposite sides of the kinase core of HK853 (DHp domain in red, CA domain in cyan). Active site His and Asp are shown in stick representations. RR468 has a sulfate ion (blue stick model) in the active site which mimics a phosphoryl group in location and active-site chemistry, rendering RR468 akin to the activated form. **(b)** Top view of the HK853-RR468 complex showing the close proximity of the RR468 active site to the phospho-accepting His. **(c)** *S. cereviciae* YPD1-SLN1$^{(BF)}$ complex structure (PDB ID: 2R25). The molecular surface of YPD1 (magenta) and activated SLN1 (pink) emphasize the strong surface complementarity that aligns the active-site residues (shown in stick representation).

(a)



(b)



(c)

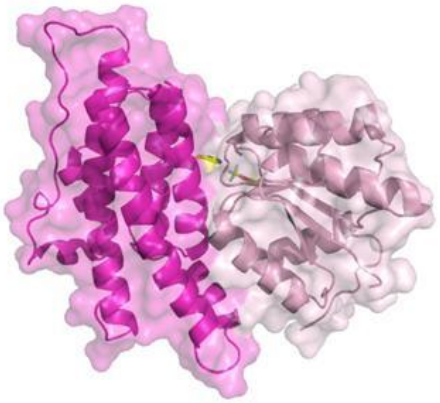


Figure 1-9. REC dimerization at the α4-β5-α5 face. A view above the active site of the $ArcA^{N(BF)}$ (PDB ID: 1XHF). This mode of α4-β5-α5 dimerization is common among members of the OmpR/PhoB subfamily. Each activated monomer is colored pink with the α4-β5-α5 face highlighted in light blue; the activated Asp, beryllofluoride, and flipped Tyr are shown in stick models.

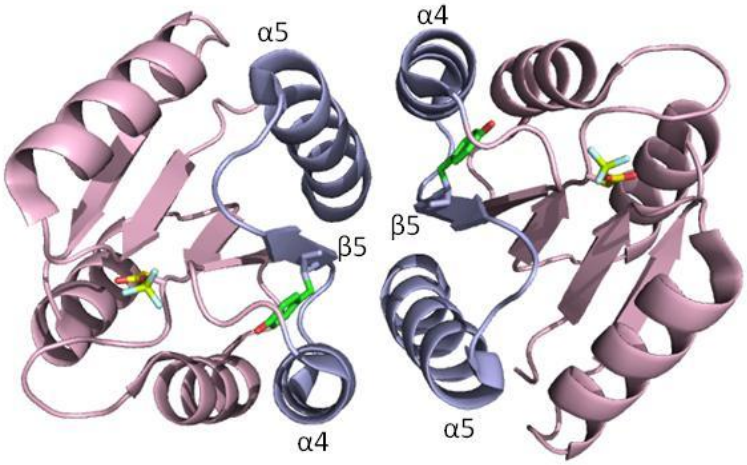


Figure 1-10. Differences in interfaces and exposure of the recognition helix among OmpR/PhoB members with full-length structures. The α4-β5-α5 region is highlighted in light blue and marks the domain interface between REC (blue) and effector (yellow) domains where one exists. The recognition helix of the winged helix-turn-helix binding motif is colored orange. **(a)** MtrA (PDB ID: 2GWR), with the Tyr102 and Asp190 hydrogen bond (stick models), **(b)** PrrA (PDB ID: 1YS6), with the Tyr105 and Asn198 hydrogen bond (stick models), **(c)** DrrB (PDB ID: 1P2F,), with the Tyr97 and Asp131 hydrogen bond (stick models), **(d)** DrrD (PDB ID: 1KGS), and **(e)** one monomer of PhoP (PDB ID: 3R0J).

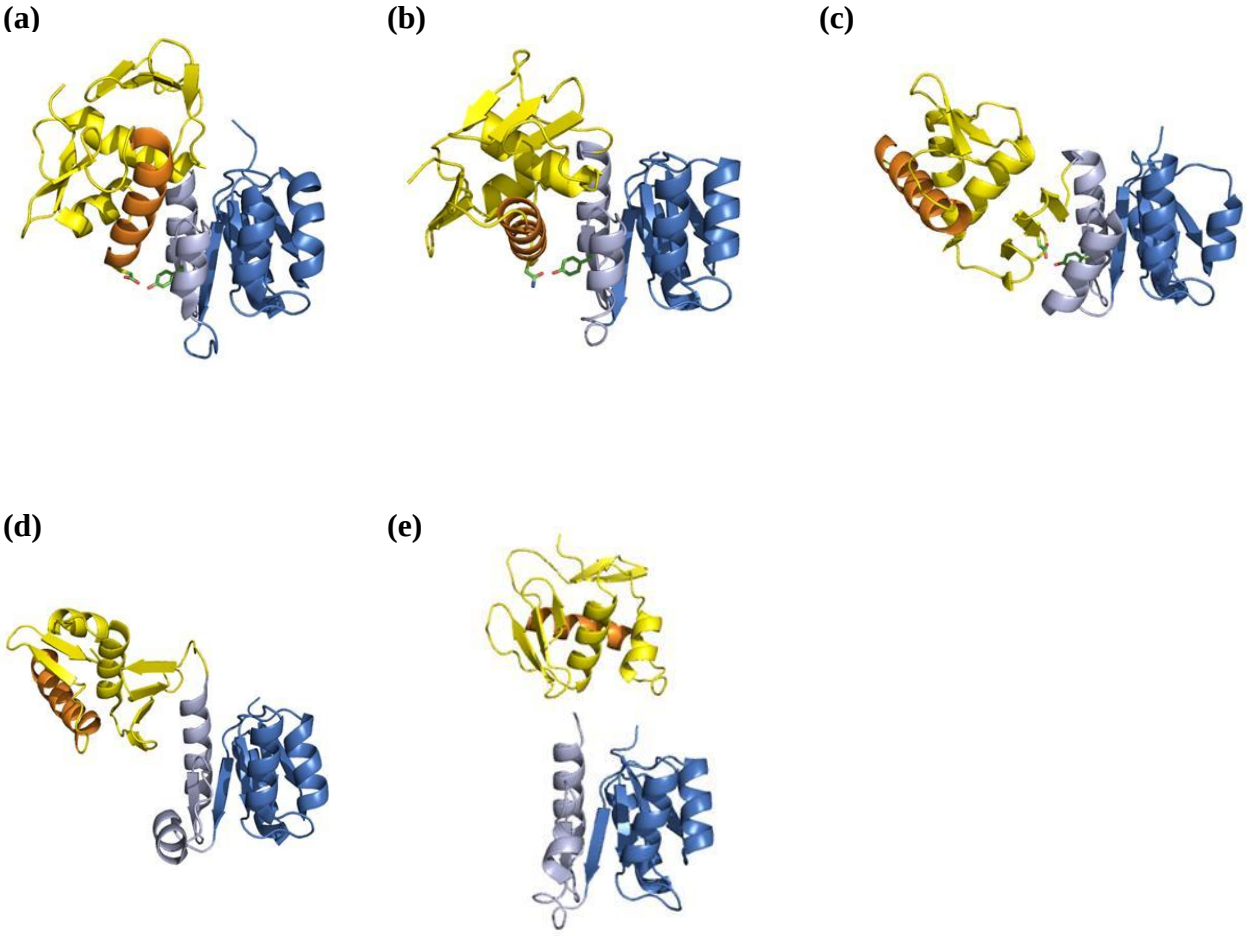


Figure 1-11. Different modes of REC domain dimerization. **(a)** The inactive dimer of NtrC1 (PDB ID: 1NY5) with the receiver (blue) and central (red and green) domains. Upon phosphorylation, the REC domain releases the inhibition on the central domain and enables its oligomerization. **(b)** The dimerization interface of the REC domains shown in (a) (PDB ID: 1NY5, but only REC domains shown). Dimerization occurs at the α4-β5-α5 face with the switch Tyr residue (green stick model) pointed outward. **(c)** Activated $NtrC1^{N}$ (PDB ID: 1ZY2). Upon activation, helix α5 repositions and the dimerization interface transforms to a tighter dimer between the α4-β5 regions. This transition is accompanied by the rotation of the switch Tyr residue (green stick model) to a buried conformation. **(d)** Dimerization of $FixJ^{Np}$ (PDB ID: 1D5W) is also between α4 and β5 of each monomer, with the switch Phe residue (green stick model) buried. Views for (b), (c), and (d) are above the active site with the activated-Asp (c and d) in stick model. Inactivated REC domains are colored blue, activated REC domains are colored pink, and the α4-β5-α5 region in light blue.

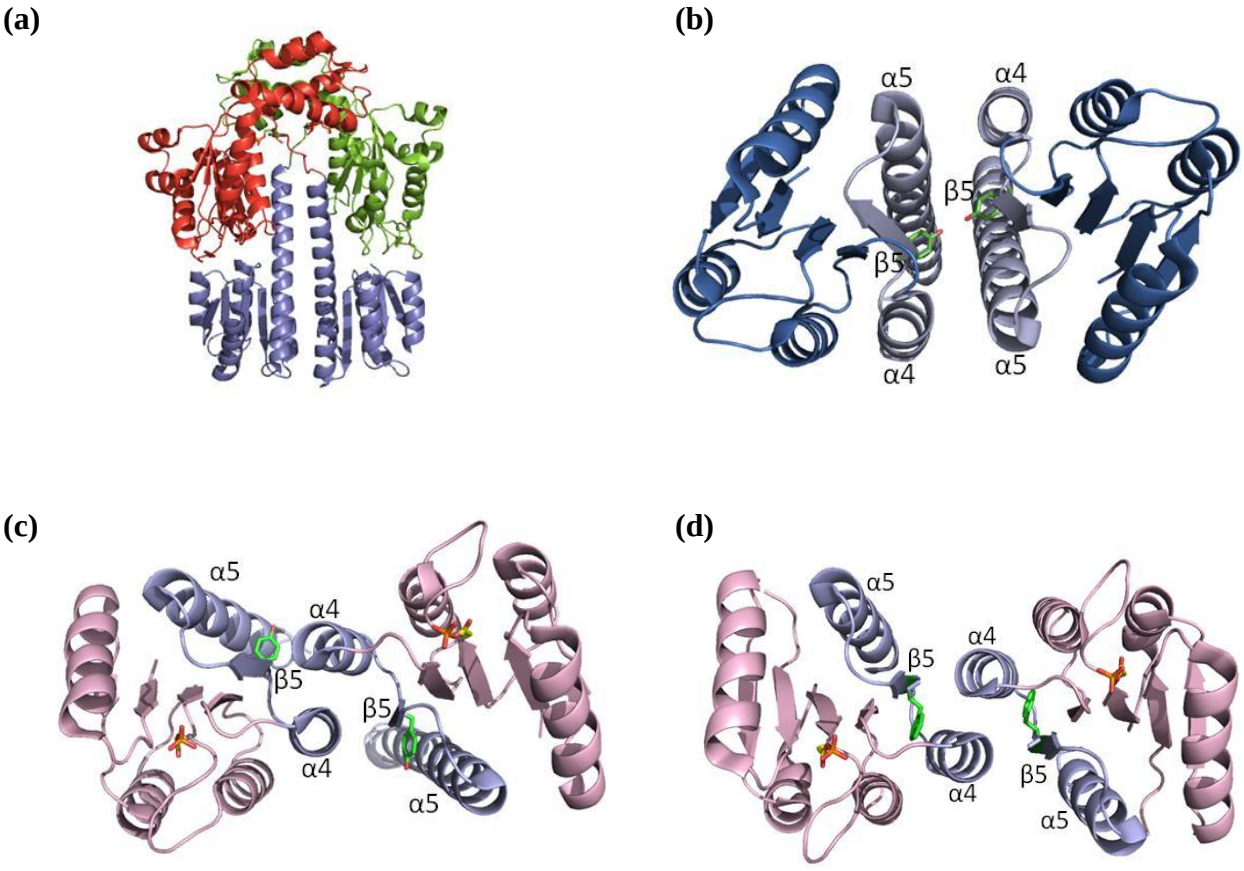


Figure 1-12. Structures of the NarL/FixJ subfamily. **(a)** The activated spr1814 REC domain (PDB ID: 4E7P) colored pink with the α4-β4-α5 region highlighted in light blue, and linker region in red. Asp-BeF_3^- and switch Tyr are shown in stick model. **(b)** Full-length NarL (PDB ID: 1RNL). **(c)** Full-length StyR (PDB ID: 1YIO). **(d)** Full-length DosR (PDB ID: 3C3W). **(e)** The effector domain of RcsB (PDB ID: 1P4W). All REC domains of full-length structures (blue) are shown in the same orientation. The helix-turn-helix effector domains are shown in yellow, with the recognition helix colored orange. Linker regions are colored red. The linker regions are pronouncedly different between the three full-length RR structures; however the helix in the linker of NarL has notable similarity to the analogous helix of RcsB.

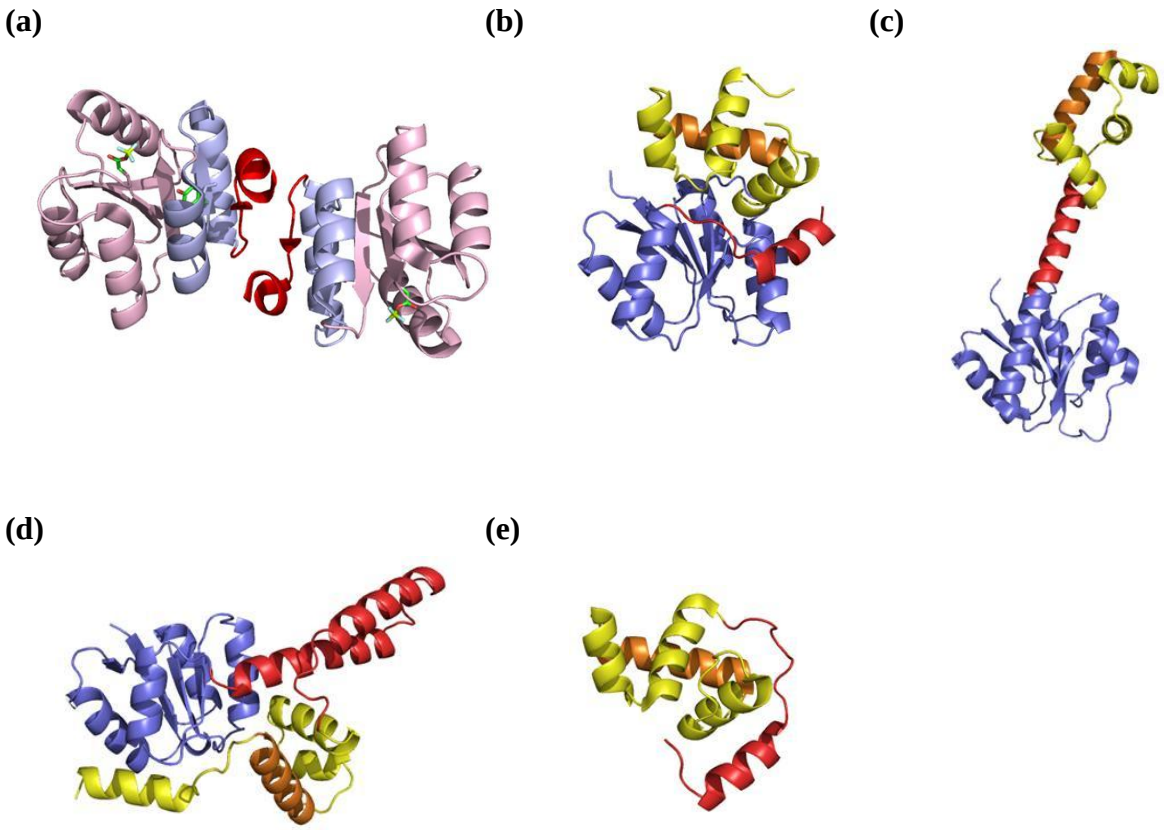


Figure 1-13. The α4-β5-α5 face in CheB and CheY. **(a)** Structure of CheB (PDB ID: 1A2O) showing the extensive domain interface between the effector (yellow) and REC domains (blue). The α4-β5-α5 face (light blue), with the switch Phe (green stick model) in the outward position, makes contacts with, and inhibits, substrate binding elements of the output domain. (b) Activated CheY (pink) binding to an N-terminal peptide of FliM (gold) (PDB ID; 1F4V) using the α4-β5-α5 region (light blue). The inward movement of the Tyr switch residue (green stick model) is necessary for binding.

(a) **(b)**

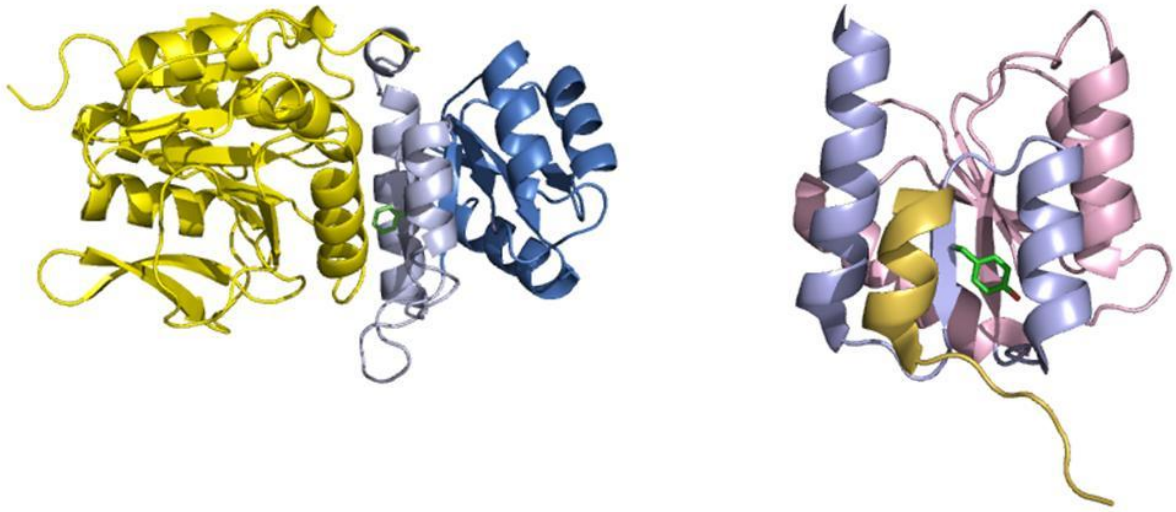


Table 1-1. Examples of different types of effector domains (table adopted from published versions (27, 28)).

RR Representative	Classification by structure or function
DNA Binding (63%)	
OmpR/PhoB	winged helix-turn-helix
NarL/FixJ	helix-turn-helix
NtrC/DctD	AAA+ ATPase and Fis
LytR	LytTR
RNA Binding (1%)	
AmiR/NasR	ANTAR
Enzymatic (13%)	
CheB	Methylesterase
PleD	Diguanylate cyclase (GGDEF)
PyrR	c-di-GMP phosphodiesterase (EAL)
RpfG	c-di-GMP phosphodiesterase (HD-GYP)
Stand alone (17%)	
Spo0F	Phosphotransfer
CheY	Chemotaxis
Protein binding (3%)	
CheV*	CheW-like (chemotaxis adapter)
PhyR*	RpoE type anti-anti-sigma factor
Other (3%)	

*RRs with REC domain in the C-terminus.

References

1. **Albanesi, D., M. Martín, F. Trajtenberg, M. C. Mansilla, A. Haouz, P. M. Alzari, D. de Mendoza, and A. Buschiazzo**. 2009. Structural plasticity and catalysis regulation of a thermosensor histidine kinase. Proceedings of the National Academy of Sciences of the United States of America **106**:16185–90.

2. **Bachhawat, P., and A. M. Stock**. 2007. Crystal structures of the receiver domain of the response regulator PhoP from Escherichia coli in the absence and presence of the phosphoryl analog beryllofluoride. Journal of bacteriology **189**:5987–95.

3. **Bachhawat, P., G. V. T. Swapna, G. T. Montelione, and A. M. Stock**. 2005. Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states. Structure (London, England : 1993) **13**:1353–63.

4. **Baikalov, I., I. Schröder, M. Kaczor-Grzeskowiak, D. Cascio, R. P. Gunsalus, and R. E. Dickerson**. 1998. NarL dimerization? Suggestive evidence from a new crystal form. Biochemistry **37**:3665–76.

5. **Baikalov, I., I. Schröder, M. Kaczor-Grzeskowiak, K. Grzeskowiak, R. P. Gunsalus, and R. E. Dickerson**. 1996. Structure of the Escherichia coli Response Regulator NarL. Biochemistry **35**:11053–61.

6. **Barakat, M., P. Ortet, and D. E. Whitworth**. 2011. P2CS: a database of prokaryotic two-component systems. Nucleic acids research **39**:D771–6.

7. **Barbieri, C. M., T. R. Mack, V. L. Robinson, M. T. Miller, and A. M. Stock**. 2010. Regulation of response regulator autophosphorylation through interdomain contacts. The Journal of biological chemistry **285**:32325–35.

8. **Batchelor, J. D., M. Doucleff, C. Lee, K. Matsubara, S. De Carlo, J. Heideker, M. H. Lamers, J. G. Pelton, and D. E. Wemmer**. 2008. Structure and Regulatory Mechanism of Aquifex aeolicus NtrC4 : Variability and Evolution in Bacterial Transcriptional Regulation. Journal of Molecular Biology. Elsevier Ltd **384**:1058–1075.

9. **Benda, C., C. Scheufler, N. Tandeau de Marsac, and W. Gärtner**. 2004. Crystal structures of two cyanobacterial response regulators in apo- and phosphorylated form reveal a novel dimerization motif of phytochrome-associated response regulators. Biophysical journal **87**:476–87.

10. **Bent, C. J., N. W. Isaacs, T. J. Mitchell, and A. Riboldi-tunnicliffe**. 2004. Crystal Structure of the Response Regulator 02 Receiver Domain , the Essential YycF Two-Component System of Streptococcus pneumoniae in both Complexed and Native States. Journal of bacteriology **186**:2872–2879.

11. **Bick, M. J., V. Lamour, K. R. Rajashankar, Y. Gordiyenko, C. V Robinson, and S. A. Darst**. 2009. How to switch off a histidine kinase: crystal structure of Geobacillus stearothermophilus KinB with the inhibitor Sda. Journal of molecular biology. Elsevier Ltd **386**:163–77.

12. **Bilwes, A. M., L. A. Alex, B. R. Crane, and M. I. Simon**. 1999. Structure of CheA, a signal-transducing histidine kinase. Cell **96**:131–41.

13. **Bilwes, A. M., C. M. Quezada, L. R. Croal, B. R. Crane, and M. I. Simon**. 2001. Nucleotide binding by the histidine kinase CheA. Nature structural biology **8**:353–60.

14. **Birck, C., L. Mourey, P. Gouet, B. Fabry, J. Schumacher, P. Rousseau, D. Kahn, and J. P. Samama**. 1999. Conformational changes induced by phosphorylation of the FixJ receiver domain. Structure (London, England : 1993) **7**:1505–15.

15. **Bourret, R. B.** 2010. Receiver domain structure and function in response regulator proteins. Current opinion in microbiology. Elsevier Ltd **13**:142–9.

16. **Buckler, D. R., Y. Zhou, and A. M. Stock**. 2002. Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from Thermotoga maritima. Structure **10**:153–64.

17. **Casino, P., V. Rubio, and A. Marina**. 2009. Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell. Elsevier Ltd **139**:325–36.

18. **Chabre, M.** 1990. Aluminofluoride and beryllofluoride complexes: new phosphate analogues in enzymology. Trends Biochem Sci **15**:6–10.

19. **Djordjevic, S., P. N. Goudreau, Q. Xu, A. M. Stock, and A. H. West**. 1998. Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. Proceedings of the National Academy of Sciences of the United States of America **95**:1381–6.

20. **Donaldson, L. W.** 2008. The NMR structure of the Staphylococcus aureus response regulator VraR DNA binding domain reveals a dynamic relationship between it and its associated receiver domain. Biochemistry **47**:3379–88.

21. **Doucleff, M., B. Chen, A. E. Maris, D. E. Wemmer, E. Kondrashkina, and B. T. Nixon**. 2005. Negative regulation of AAA + ATPase assembly by two component receiver domains: a transcription activation mechanism that is conserved in mesophilic and extremely hyperthermophilic bacteria. Journal of molecular biology **353**:242–55.

22. **Dutta, R., and M. Inouye**. 2000. GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci **25**:24–8.

23. **Dutta, R., L. Qin, and M. Inouye**. 1999. MicroReview Histidine kinases : diversity of domain organization. Molecular Microbiology **34**:633–640.

24. **Friedland, N., T. R. Mack, M. Yu, L. Hung, and C. Thomas**. 2007. Domain Orientation in the Inactive Response Regulator Mycobacterium tuberculosis MtrA Provides a Barrier to Activation. Biochemistry **46**:6733–43.

25. **Galperin, M. Y., a N. Nikolskaya, and E. V Koonin**. 2001. Novel domains of the prokaryotic two-component signal transduction systems. FEMS microbiology letters **203**:11–21.

26. **Galperin, M. Y.** 2010. Diversity of Structure and Function of Response Regulator Output Domains. Current opinion in microbiology **13**:150–159.

27. **Gao, R., T. R. Mack, and A. M. Stock**. 2007. Bacterial response regulators: versatile regulatory strategies from common domains. Trends in biochemical sciences **32**:225–34.

28. **Gao, R., and A. M. Stock**. 2009. Biological insights from structures of two-component proteins. Annual review of microbiology **63**:133–54.

29. **Gouet, P., B. Fabry, V. Guillet, C. Birck, L. Mourey, D. Kahn, and J. P. Samama**. 1999. Structural transitions in the FixJ receiver domain. Structure (London, England : 1993) **7**:1517–26.

30. **Grebe, T. W., and J. B. Stock**. 1999. The histidine protein kinase superfamily. Advances in microbial physiology **41**:139–227.

31. **Guhaniyogi, J., V. L. Robinson, and A. M. Stock**. 2006. Crystal structures of beryllium fluoride-free and beryllium fluoride-bound CheY in complex with the conserved C-terminal peptide of CheZ reveal dual binding modes specific to CheY conformation. Journal of molecular biology **359**:624–45.

32. **Hastings, C. A., S.-Y. Lee, H. S. Cho, D. Yan, S. Kustu, and D. E. Wemmer**. 2003. High-resolution solution structure of the beryllofluoride-activated NtrC receiver domain. Biochemistry **42**:9081–90.

33. **Huynh, T. N., and V. Stewart**. 2011. Negative control in two-component signal transduction by transmitter phosphatase activity. Molecular microbiology **82**:275–286.

34. **Ikegami, T., T. Okada, I. Ohki, J. Hirayama, T. Mizuno, and M. Shirakawa**. 2001. Solution structure and dynamic character of the histidine-containing phosphotransfer domain of anaerobic sensor kinase ArcB from Escherichia coli. Biochemistry **40**:375–86.

35. **Kato, M., T. Mizuno, T. Shimizu, and T. Hakoshima**. 1997. Insights into multistep phosphorelay from the crystal structure of the c-terminal Hpt domain of ArcB. Cell **88**:717–23.

36. **Kern, D., B. F. Volkman, P. Luginbühl, M. J. Nohaile, S. Kustu, and D. E. Wemmer**. 1999. Structure of a transiently phosphorylated switch in bacterial signal transduction. Nature **402**:894–8.

37. **Lee, S. Y., H. S. Cho, J. G. Pelton, D. Yan, E. a Berry, and D. E. Wemmer**. 2001. Crystal structure of activated CheY. Comparison with other activated receiver domains. The Journal of biological chemistry **276**:16425–31.

38. **Lee, S. Y., A. De La Torre, D. Yan, S. Kustu, B. T. Nixon, and D. E. Wemmer**. 2003. Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domain. Genes & Development **17**:2552–63.

39. **Lee, S.-Y., H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L. Huang, S. Kustu, E. A. Berry, and D. E. Wemmer**. 2001. Crystal structure of an activated response regulator bound to its target. Nature structural biology **8**:52–6.

40. **Lewis, R. J., J. a Brannigan, K. Muchová, I. Barák, and a J. Wilkinson**. 1999. Phosphorylated aspartate in the structure of a response regulator protein. Journal of molecular biology **294**:9–15.

41. **Lewis, R. J., S. Krzywda, J. a Brannigan, J. P. Turkenburg, K. Muchová, E. J. Dodson, I. Barák, and a J. Wilkinson**. 2000. The trans-activation domain of the sporulation response regulator Spo0A revealed by X-ray crystallography. Molecular microbiology **38**:198–212.

42. **Liu, X., and P. De Wulf**. 2004. Probing the ArcA-P modulon of Escherichia coli by whole genome transcriptional analysis and sequence recognition profiling. The Journal of biological chemistry **279**:12588–97.

43. **Marina, A., C. Mott, A. Auyzenberg, W. a Hendrickson, and C. D. Waldburger**. 2001. Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. The Journal of biological chemistry **276**:41182–90.

44. **Marina, A., C. D. Waldburger, and W. a Hendrickson**. 2005. Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. The EMBO journal **24**:4247–59.

45. **Matte, A., L. W. Tari, and L. T. Delbaere**. 1998. How do kinases transfer phosphoryl groups? Structure (London, England : 1993) **6**:413–9.

46. **Menon, S., and S. Wang**. 2011. Structure of the response regulator PhoP from Mycobacterium tuberculosis reveals a dimer through the receiver domain. Biochemistry **50**:5948–57.

47. **Milani, M., L. Leoni, G. Rampioni, E. Zennaro, P. Ascenzi, and M. Bolognesi**. 2005. An active-like structure in the unphosphorylated StyR response regulator suggests a

phosphorylation- dependent allosteric activation mechanism. Structure (London, England : 1993) **13**:1289–97.

48. **Mourey, L., S. Da Re, J. D. Pédelacq, T. Tolstykh, C. Faurie, V. Guillet, J. B. Stock, and J. P. Samama**. 2001. Crystal structure of the CheA histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis. The Journal of biological chemistry **276**:31074–82.

49. **Nowak, E., S. Panjikar, P. Konarev, D. I. Svergun, and P. a Tucker**. 2006. The structural basis of signal transduction for the response regulator PrrA from Mycobacterium tuberculosis. The Journal of biological chemistry **281**:9659–66.

50. **Park, A. K., J. H. Moon, K. S. Lee, and Y. M. Chi**. 2012. Crystal structure of receiver domain of putative NarL family response regulator spr1814 from Streptococcus pneumoniae in the absence and presence of the phosphoryl analog beryllofluoride. Biochem Biophys Res Commmun **421**:403–7.

51. **Park, S., M. Meyer, A. D. Jones, H. P. Yennawar, N. H. Yennawar, and B. T. Nixon**. 2002. Two-component signaling in the AAA+ ATPase DctD: binding Mg 2+ and BeF 3- selects between alternative dimeric states of the receiver domain. The FASEB Journal **16**:1964–66.

52. **Parkinson, J. S., and E. C. Kofoid**. 1992. Communication modules in bacterial signaling proteins. Annual review of genetics **26**:71–112.

53. **Perry, J., K. Koteva, and G. Wright**. 2011. Receptor domains of two-component signal transduction systems. Molecular bioSystems **7**:1388–98.

54. **Podust, L. M., A. Ioanoviciu, and P. R. Ortiz de Montellano**. 2008. 2.3 Å X-ray structure of the heme-bound GAF domain of sensory histidine kinase DosT of Mycobacterium tuberculosis. Biochemistry **47**:12523–31.

55. **Pristovsek, P., K. Sengupta, F. Löhr, B. Schäfer, M. W. von Trebra, H. Rüterjans, and F. Bernhard**. 2003. Structural analysis of the DNA-binding domain of the Erwinia amylovora RcsB protein and its interaction with the RcsAB box. The Journal of biological chemistry **278**:17752–9.

56. **Robinson, V. L., T. Wu, and A. M. Stock**. 2003. Structural Analysis of the Domain Interface in DrrB , a Response Regulator of the OmpR / PhoB Subfamily. Journal of bacteriology **185**:4186–94.

57. **Rogov, V. V., F. Bernhard, F. Löhr, and V. Dötsch**. 2004. Solution structure of the Escherichia coli YojN histidine-phosphotransferase domain and its interaction with cognate phosphoryl receiver domains. Journal of molecular biology **343**:1035–48.

58. **Schnell, R., D. Agren, and G. Schneider**. 2008. 1.9 Å structure of the signal receiver domain of the putative response regulator NarL from Mycobacterium tuberculosis. Acta crystallographica. Section F, Structural biology and crystallization communications **64**:1096–100.

59. **Schrödinger, L.** The PyMOL Molecular Graphics System. Version 1.3.

60. **Sidote, D. J., C. M. Barbierie, T. Wu, and A. M. Stock**. 2008. Structure of the Staphylococcus aureus AgrA LytTR Domain Bound to DNA Reveals a Beta Fold with a Novel Mode of Binding **16**:727–35.

61. **Song, H. K., J. Y. Lee, M. G. Lee, J. M. K. Min, J. K. Yang, and S. W. Suh**. 1999. Insights into eukaryotic multistep phosphorelay signal transduction revealed by the crystal structure of Ypd1p from Saccharomyces cerevisiae. Journal of molecular biology **293**:753–61.

62. **Song, Y., D. Peisach, A. a Pioszak, Z. Xu, and A. J. Ninfa**. 2004. Crystal structure of the C-terminal domain of the two-component system transmitter protein nitrogen regulator II (NRII; NtrB), regulator of nitrogen assimilation in Escherichia coli. Biochemistry **43**:6670–8.

63. **Stewart, R. C.** 2010. Protein Histidine Kinases: Assembly of Active Sites and Their Regulation in Signaling Pathways. Curr Opin Microbiol. **13**:133–141.

64. **Stock, A. M., E. Martinez-Hackert, B. F. Rasmussen, A. H. West, J. B. Stock, D. Ringe, and G. A. Petsko**. 1993. Structure of the Mg(2+)-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. Biochemistry **32**:13375–80.

65. **Stock, A. M., V. L. Robinson, and P. N. Goudreau**. 2000. Two-Component Signal Transduction. Annual review of biochemistry **69**:183–215.

66. **Sugawara, H., Y. Kawano, T. Hatakeyama, T. Yamaya, N. Kamiya, and H. Sakakibara**. 2005. Crystal structure of the histidine-containing phosphotransfer protein ZmHP2 from maize. Protein Science **14**:202–8.

67. **Tanaka, T., S. K. Saha, C. Tomomori, R. Ishima, D. Liu, K. I. Tong, H. Park, R. Dutta, L. Qin, M. B. Swindells, T. Yamazaki, A. M. Ono, M. Kainosho, M. Inouye, and M. Ikura**. 1998. NMR structure of the histidine kinase domain of the E. coli osmosensor EnvZ. Nature **396**:88–92.

68. **Toro-Roman, A., T. R. Mack, and A. M. Stock**. 2005. Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face. Journal of molecular biology **349**:11–26.

69. **Toro-Roman, A., T. I. Wu, and A. M. Stock**. 2005. A common dimerization interface in bacterial response regulators KdpE and TorR. Protein Science **14**:3077–88.

70. **Trajtenberg, F., M. Graña, N. Ruétalo, H. Botti, and A. Buschiazzo**. 2010. Structural and enzymatic insights into the ATP binding and autophosphorylation mechanism of a sensor histidine kinase. The Journal of biological chemistry **285**:24892–903.

71. **Vannini, A., C. Volpari, C. Gargioli, E. Muraglia, R. Cortese, R. De Francesco, P. Neddermann, and S. Di Marco**. 2002. The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. The EMBO journal **21**:4393–401.

72. **Varughese, K. I., Madhusudan, X. Z. Zhou, J. M. Whiteley, and J. A. Hoch**. 1998. Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. Molecular cell **2**:485–93.

73. **Varughese, K. I.** 2005. Conformational Changes of Spo0F along the Phosphotransfer Pathway. Journal of bacteriology **187**:8221–7.

74. **Varughese, K. I., I. Tsigelny, and H. Zhao**. 2006. The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. Journal of bacteriology **188**:4970–7.

75. **Volkman, B. F., D. Lipson, D. E. Wemmer, and D. Kern**. 2001. Two-state allosteric behavior in a single-domain signaling protein. Science (New York, N.Y.) **291**:2429–33.

76. **Vu, A., D. J. Hamel, H. Zhou, and F. W. Dahlquist**. 2011. The structure and dynamic properties of the complete histidine phosphotransfer domain of the chemotaxis specific histidine autokinase CheA from Thermotoga maritima. Journal of biomolecular NMR **51**:49–55.

77. **Wassmann, P., C. Chan, R. Paul, A. Beck, H. Heerklotz, U. Jenal, and T. Shirmer**. 2007. Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure **15**:915–27.

78. **Welch, M., N. Chinardet, L. Mourey, C. Birck, and J. P. Samama**. 1998. Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. Nature structural biology **5**:25–9.

79. **Wemmer, D. E., and D. Kern**. 2005. Beryllofluoride Binding Mimics Phosphorylation of Aspartate in Response Regulators. Journal of bacteriology **187**:8229–8230.

80. **Wisedchaisri, G., M. Wu, D. R. Sherman, and W. G. J. Hol**. 2008. Crystal structures of the response regulator DosR from Mycobacterium tuberculosis suggest a helix rearrangement mechanism for phosphorylation activation. Journal of molecular biology **378**:227–42.

81. **Wolanin, P. M., P. A. Thomason, and J. B. Stock**. 2002. Protein family review Histidine protein kinases : key signal transducers outside the animal kingdom. Genome Biology **3**:1–8.

82. **Wolanin, P. M., D. J. Webre, and J. B. Stock**. 2003. Mechanism of phosphatase activity in the chemotaxis response regulator CheY. Biochemistry **42**:14075–82.

83. **Xu, Q., S. W. Porter, and A. H. West**. 2003. The yeast YPD1/SLN1 complex: insights into molecular recognition in two-component signaling systems. Structure **11**:1569–81.

84. **Xu, Q., and A. H. West**. 1999. Conservation of structure and function among histidine-containing phosphotransfer (Hpt) domains as revealed by the crystal structure of YPD1. Journal of molecular biology **292**:1039–50.

85. **Yamada, S., and Y. Shiro**. 2008. Structural Basis ofthe Signal Transduction in the Two-Component Signal output. Bacterial Signal Transduction: Networks and Drug Targets. Edited by R. Utsumi **Chapter 3**:22–39.

86. **Yamada, S., H. Sugimoto, M. Kobayashi, A. Ohno, H. Nakamura, and Y. Shiro**. 2009. Structure of PAS-linked histidine kinase and the response regulator complex. Structure (London, England : 1993). Elsevier Ltd **17**:1333–44.

87. **Yan, D., H. S. Cho, C. a Hastings, M. M. Igo, S. Y. Lee, J. G. Pelton, V. Stewart, D. E. Wemmer, and S. Kustu**. 1999. Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. Proceedings of the National Academy of Sciences of the United States of America **96**:14789–94.

88. **Yang, X., J. Kuk, and K. Moffat**. 2008. Crystal structure of Pseudomonas aeruginosa bacteriophytochrome: photoconversion and signal transduction. Proceedings of the National Academy of Sciences of the United States of America **105**:14715–20.

89. **Zapf, J., U. Sen, Madhusudan, J. a Hoch, and K. I. Varughese**. 2000. A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. Structure (London, England : 1993) **8**:851–62.

90. **Zhang, R., K. M. Pappas, J. L. Brace, P. C. Miller, T. Oulmassov, J. M. Molyneaux, J. C. Anderson, J. K. Bashkin, S. C. Winans, and A. Joachimiak**. 2002. Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. Nature **417**:971–4.

91. **Zhao, X., D. M. Copeland, S. Soares, Alexei, and A. H. West**. 2008. Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. Journal of molecular biology **375**:1141–1151.

Chapter 2

The *Escherichia coli* Nar Two-Component System

Introduction

Signal transduction by phosphoryl transfer is a prevalent means of intracellular communication in bacteria. The simplest paradigm, the two-component systems (see Chapter 1), is critical to organisms for adapting to changes in their environment. This mechanism involves phosphoryl transfer between a histidine kinase (HK) that detects an environmental change (nutrient deficiency, anaerobic conditions, etc.) and a response regulator (RR) that initiates cellular action. In *E. coli*, a change from aerobic to anaerobic conditions triggers the Nar two-component system (Figure 2-1a) (17, 32). During oxygen deprivation, nitrate (followed by nitrite) is the hierarchically preferred alternative as a terminal electron-acceptor for respiration. NarX and NarQ are two HKs that detect extracellular nitrate or nitrite. When bound by these signals, NarX and NarQ autophosphorylate and then transfer the phosphoryl group to the RRs NarL and NarP, which proceed to regulate several genes involved in anaerobic respiration. In this short review nitrate respiration in *E. coli* is summarized, with a focus on the well-studied genes regulated by NarL and NarP. In addition, crystallographic and other structural studies of NarL are described.

Anaerobic Respiration and Nitrate Reduction

The switch to anaerobiosis in *E. coli* mandates the alteration of the cellular machinery to accommodate other terminal electron acceptors besides oxygen. Nitrate or nitrite is second to oxygen in order of preference followed by dimethyl sulfoxide (DMSO), trimethylamine-N-oxide (TMAO), and fumarate (17).

The conversion of nitrate to nitrite to ammonium involves the transcription of genes in the nitrate and nitrite respiratory chains. The membrane-bound nitrate respiratory chain is

composed of the iron-molybdoenzymes formate dehydrogenase-N (formate is the preferred electron donor for nitrate), and nitrate reductase (11, 33). Formate dehydrogenase-N and nitrate reductase are encoded by the *fdnGHI* and *narGHJI* operons, respectively. Nitrite is reduced either by a formate-utilizing cytoplasmic nitrite-reductase inscribed by the *nrfABCDEFG* operon, or by a periplasmic NADH-utilizing nitrite-reductase inscribed by the *nirBDC operon*.

The dispatch of oxidoreductases that utilize nitrate or nitrite necessitates a nexus of transcriptional regulation, the simplified purpose of which is to induce genes encoding for nitrate and nitrite reductases while simultaneously inhibiting the transcription of genes encoding for reductases lower in the hierarchy (17). The presence of nitrate, for example, leads to the induction of the *narG* and *fdnG* operons, and to the repression of the fumarate reductase (*frdABCD*) and DMSO reductase (*dmsABC*) operons.

Nar, a Dual Two Component System

One level of general transcriptional regulation during anaerobiosis is implemented by the global transcription factor Fnr (Regulator of Fumarate and Nitrate Reduction) (27, 34). Upon sensing oxygen depletion, Fnr induces or inhibits transcription of pertinent genes by interacting with RNA polymerase. The Fnr binding-site location (whether at -41.5 of the transcription start site, or farther upstream) specifies different contacts between Fnr and RNA polymerase.

Another, more specific, level of transcriptional regulation of nitrate- and nitrite-responsive operons is executed by the Nar two-component system (17, 33). The presence of nitrate or nitrite triggers a phosphorylation cascade whereby the HKs NarX and NarQ autophosphorylate on conserved His in their cytoplasmic domain (Figure 2-1a). Subsequently either NarX or NarQ can phosphorylate RRs NarL and NarP. Upon aspartate-phosphorylation of

the receiver (REC) domain, NarL and NarP regulate transcription of genes involved in nitrate (or nitrite) adaptation. Most of these Nar-responsive operons also require Fnr (35).

Either NarX or NarQ is sufficient to generate a cellular response to nitrate or nitrite in conjunction with NarL and NarP, however this dual system is not symmetrical (26) (Figure 2-1b). Although either NarX or NarQ can phosphorylate both NarL and NarP, their preferences are not equal. NarX exhibits a preference for phosphorylating NarL over NarP, while NarQ shows relatively equal phosphotransfer ability to NarL and NarP. Specific residues in the DHp domain (see Chapter 1) of NarX are tailored for NarP specific binding but do not show the same specificity in NarQ. NarX also has a preference for dephosphorylating NarL (10). In the presence of nitrate, NarX primarily phosphorylates NarL, but in the presence of nitrite has a differential response to dephosphorylating NarL. Therefore, higher levels of $NarL^{p\dagger}$ exist in the presence of nitrate as opposed to nitrite. In contrast, NarQ has a uniform response to NarL and NarP in the presence of nitrate or nitrite.

Nar Regulated Operons

The dual NarX/Q and NarL/P TCS regulates nitrate and nitrite metabolism in a complex and intertwined fashion (8, 32). NarL and NarP are each capable of activating or repressing in total at least 50 operons, either acting in alliance with or against the affects of other transcription factors. Several *E. coli* Nar operons have been studied in detail, examples of which are shown in Figure 2-2 and listed in Table 2-2.

[†] Superscripts are used to represent different domain regions or protein states: "C" denotes the C-terminal domain, and "p" refers to phosphorylated.

NarL and NarP share the DNA-binding consensus sequence 5'-TACYYMT-3' (Y=C or T, M=A or C) but their binding affinities are not always congruent and can change based on the presence of nitrate or nitrite (9). Single or multiple TACYYMT consensus sites (or heptamers) are found in different operons with different arrangements (Figure 2-2). A common binding motif is two inverted heptamers, allowing a dimer formation, with a 2 base-pair spacer (called a "7-2-7" arrangement). Phosphorylated NarP seems to only recognize this binding mode while phosphorylated NarL can bind other arrangements and also single heptamers (10).

The *narG* operon, not having any 7-2-7 sites, is solely controlled by NarL in the presence of nitrate or nitrite. Although there are several heptameric sites, only the -89 and -195 seem to be essential for $NarL^p$ induction. At higher concentrations $NarL^p$ is thought to bind the flanking heptamers through cooperative binding (39).

$NarL^p$ also shows cooperative binding in the *fndG* operon (10). In the presence of growing nitrate concentrations, this operon is strongly induced by $NarL^p$ but only weakly induced or even antagonized by $NarP^p$ (42). This is because full induction requires that single heptamers be bound in addition to the 7-2-7 sites, and $NarP^p$ only binds 7-2-7 repeats. DNAse I footprinting experiments show that $NarL^p$ binds to the 7-2-7 sites first and then occupies the single sites in a cooperative manner.

Both $NarL^p$ and $NarP^p$ induce the *nirB* promoter in the presence of nitrate or nitrite, with nitrate being the more potent signal (41). In the presence of either anion, $NarL^p$ is a better activator of *nirB*, and $NarP^p$ even mildly antagonizes the effects of $NarL^p$ at low nitrate concentrations. $NarL^p$ has a very strong affinity for the *nirB* promoter, enabling it to confer activation even under nitrite growth when it is present in low amounts (10).

The *nrfA* operon, like *nirB*, has only one 7-2-7 site, but the presence of additional single heptameric sites creates a more complex system (10, 35). Under nitrite growth, both $NarL^p$ and $NarP^p$ can activate *nrfA* expression. But in the presence of high nitrate concentrations, when $NarL^p$ levels rise, $NarL^p$ is able to repress this operon by binding the nearby lower-affinity -22 and -50 single heptamers. One study also showed that very high nitrite concentrations can also stimulate NarL repression at this promoter (41). Consistently, and not surprisingly, $NarP^p$ does not repress the *nrfA* operon under any conditions, since it cannot bind to the single heptamer to repress transcription (10, 35, 41).

The *napF* operon is also induced by nitrite and low levels of nitrate, but inhibited by high nitrate levels (11, 35, 43). This promoter is different, however, from the *nrfA* control region in two major ways: the Fnr site is placed farther upstream than in most other operons, and activation of *napF* is dependent on NarP. $NarL^p$ binds the 7-2-7 site but does not activate transcription. It competes with $NarP^p$ for binding, reducing the level of transcription induced by $NarP^p$ and Fnr by over 50% (11, 35).

Added layers of complexity to these regulatory mechanisms arise from the binding of other transcription factors, such as modE (molybdate-responsive transcription factor), Fis (factor for inversion stimulation), IHF (Integration Host Factor), and Crp (cyclic-AMP Receptor Protein). Activation of the narG operon by $NarL^p$ is dependent on the presence of IHF, which presumably bends the DNA and allows $NarL^p$ at distant binding sites to gain closer proximity to the activation site (30). Induction of the *napF* operon is strongly compromised without ModE. This transcription factor responds to molybdate, binding to a site upstream of Fnr and NarP (25). The *nrfA* operon, as with the *nirB* operon, is regulated by multiple Fis and IHF binding sites (5–7). $NarL^p$ or $NarP^p$ can counteract the affects of IHF by binding to their heptameric sites and

changing the DNA curvature to allow Fnr activation. However, Fis binding to the *nrfA* promoter region overrides $NarL^p$, $NarP^p$, and Fnr.

Although most Nar-regulated operons are co-regulated with Fnr, some are Fnr-independent (22). The *yeaR-yoaG* operon is activated by $NarL^p$ in the presence of nitrate or nitrite, but is Fnr independent. However, recent discoveries show that this operon is also controlled by the global repressor, NsrR. NsrR responds to toxic nitric-oxide created as a by-product of nitrate metabolism. Overlapping NsrR and NarL control is also implicated in the *napF* and *nrfA* operons (16).

In certain operons, NarL works against the activity of Fnr. The *ydhY-T*, *dmsABC*, and *frdA* operons encoding for alternative oxidoreductases are repressed by $NarL^p$ and activated by Fnr (4, 27). In all three cases $NarL^p$ binds multiple sites over a large genomic span, indicative of cooperative binding, and presumably blocks Fnr and RNA polymerase from binding. In contrast, a single NarL 7-2-7 binding site is sufficient for inhibition of the *ynfE-I* operon (45), an operon also activated by Fnr. Contrary to initial predictions, recent evidence indicates that $NarP^p$ can also act as a transcriptional repressor, also regardless of Fnr (8, 27).

This mechanistic labyrinth to control genes involved in *E. coli* anaerobic respiration does not necessarily exhibit redundancy between families of the Gammaproteobacteria class. For example NarL can activate the *napF* promoter in *H. influenzae* (34) but not in *E. coli*. In fact, not all families of Gammaproteobacteria employ both Nar systems (26). A dual NarX/Q and NarL/P system exists predominantly in Enterobacteriaceae, giving these organisms an added advantage for adaptation to the constant fluctuations of nitrate and nitrite concentrations.

NarL Structure and Function

Full Length NarL

Two full length structures of the 24 kDa response regulator, NarL (1, 2), have been elucidated, each confirming an N-terminal REC domain composed of the signature $(\beta/\alpha)_5$ fold found in the CheY superfamily (1, 2, 38), and a C-terminal output domain constituting four helices (α7-α10) that form the helix-turn-helix DNA-binding motif (Figure 2-3a). The two domains are connected by a linker region composed of the α5- α6 loop, helix α6, and a short 13-residue tether that could not be fully resolved in either of the two crystal structures. In either unphosphorylated structure, the two domains of NarL are oriented in a manner that prevents the effector domain from binding DNA. The inactivated REC domain acts as a repressor by precluding the output domain from accessing DNA. Phosphorylation ostensibly unbridles the recognition helix and allows DNA binding. In the process, the domain interface is expected to sever.

The NarL/FixJ subfamily is defined by the helix-turn-helix binding motif of the effector domain. Several structures in this subfamily have been elucidated that provide insight to this structural motif (12, 13, 18, 21, 28, 44). In general, the helix-turn-helix fold is held together by a hydrophobic core (12), where the scaffold (α8) and recognition (α9) helix compose the helix-turn-helix motif. A conserved Gly secures a proper turn angle between helices α7 and α8 (2, 28, 44), and a conserved salt bridge between helices α7 and α9 determines the relative orientation of helices α8 and α9 and also provides stability to the fold (13, 36).

The NarL active-site pocket contains the distinguishing slew of conserved residues typically seen in REC domains: Asp 13, Asp14, Asp59 (phosphorylation site), Lys 109, and

Ser87. The Asp59 residue must be phosphorylated to confer DNA binding to promoter regions; a D59N mutation results in a null phenotype (14). As reported for other RRs, the phosphorylated Asp is short lived and NarL cannot remain in an activated state for very long. Phosphorylation of NarL is maintained for only up to 30 minutes in the presence of NarX or NarQ (29, 40). In the absence of these sensor kinases, however, $NarL^{p}$ can be sustained for at least three hours at room temperature (29).

The NarL C-terminus Bound to DNA

Three crystal structures of $NarL^{C}$ bound to 7-2-7 heptamers reveal that the overall fold of this domain does not change in comparison to the full-length structure (23, 24). Two $NarL^{C}$ monomers bind in an anti-parallel fashion and are strengthened by dimerization between helices α10 (Figure 2-3b). Dimerization serves to orient the recognition helix (α9) into the major groove, where three amino acids (Val189, Lys192, Lys188) make base contacts. Other, nonspecific, contacts include van der Waals and water-mediated contacts, facilitated by the bendability of the DNA. The structures contest the importance for helix α10, in addition to the recognition helix, to become exposed at some point prior to DNA binding (24). The full-length NarL structure shows that helix α10 is unavailable for dimerization since it contacts helices α6 and α4. This implies that phosphorylation leads to the disruption of the hydrophobic contacts that binds these three helices, in conjunction with the relocation of helix α6 to allow C-termini dimerization.

Other structures of dimerized helix-turn-helix effector domains in the NarL/FixJ subfamily, whether alone (13, 18) or bound to DNA (36, 44, 47), also show a dimer interface formed by helices α10. These proteins also arrange, or are expected to arrange, in an anti-

parallel fashion ready to bind inverted consensus sequences. Therefore, not surprising is the conserved make up of the hydrophobic patch that sustains this dimer (18, 44). Structures of GerE, $NarL^C$, $TraR^C$, and $DosR^C$ show three to four common residues in the dimer interface, which are usually polar or hydrophobic, with one central residue being a conserved Val.

The effector domains in the NarL/FixJ family contain a rich amount of positive charge, and bind DNA with similar orientations (18, 24, 44). Binding specificity between different RRs, however, is achieved through variation in residues that contact DNA and their positions along the recognition helix, by slight angle variations between helices, and by differences in the inflicted curvature on the DNA.

The Open and Closed Forms of NarL

The absence of an activated structure of NarL has led to several solutions studies that have provided insight into the nature of activated NarL. An NMR study of separated NarLN- and C-terminal domains demonstrated a rather weak interface that can be disrupted upon phosphorylation (15). Phosphorylation reduced the affinity of the two domains by 100 fold. The linker region showed not to be required for signal transmission but rather for concentrating the domains. This implies that the weak interface is heavily counteracted by a forced proximity due to the linker. Similar to the studies of NtrC that demonstrate an equilibrium between the active and inactive states (37), NarL is thought to be in an equilibrium between its "open" (unattached domains) and closed states (domains attached at the interface), with the linker making the closed state more predominant (15). A small portion of NarL proteins, estimated to be 0.1-1%, exist in the open state, and phosphorylation serves to shift this population to the majority. This equilibrium model is corroborated by the observation that high concentrations of inactivated

NarL can bind DNA (RP Gunsalus, unpublished), and also by certain residues adopting active-like conformations in one of the NarL structures (discussed in Chapter 4) (1).

Regulating the open and closed forms of NarL is critically dependent on the domain interface. The results of the NMR (15) experiments, along with EPR (46) and a series of other biochemical studies (3, 19) have highlighted regions and specific residues in NarL that are critical to maintaining the integrity of the interface. These are residues located in the α7-α8 loop, helices α6, α9, and α10, and in the loops of α3-β4, α4-β5, and α5-α6. EPR experiments of full-length NarL also confirmed movements of specific interface residues upon phosphorylation (46). The overall fold of the C-terminus was not altered upon phosphorylation, but rather certain residues between the N- and C- terminal domains separated (one pair by at least 11Å), validating the "open" and "closed" model of NarL. Zhang et al. (46) proposed one model of activated NarL, where the two domains separate by rotation along a hinge that is located somewhere C-terminal to Gly127 (the beginning of the α5-α6 loop).

Conclusions

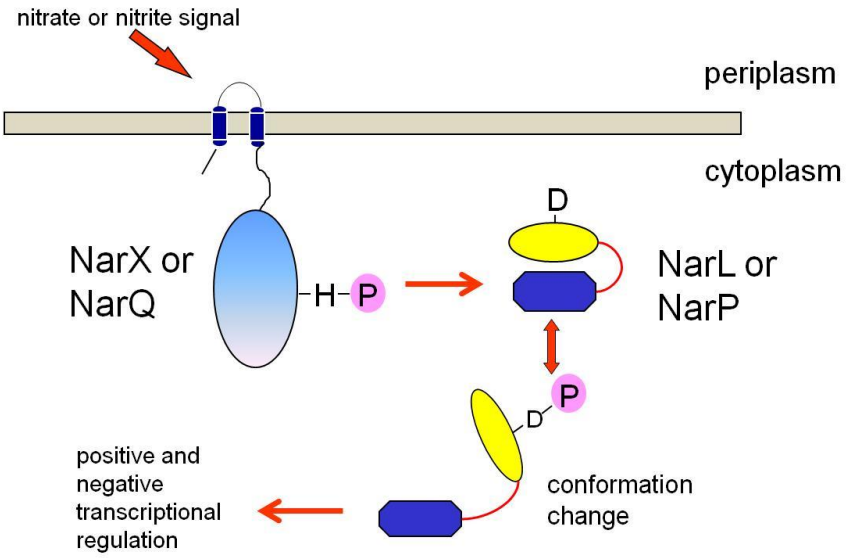
The dual Nar TCS in *E. coli* is an example of a redundant adaptive response mechanism to enable anaerobic survival in the presence of nitrate or nitrite. This system is complex with several layers of control affecting transcriptional regulation by NarL and NarP. These include the asymmetric preference for the RR by the HK, promoter site affinity, the proximity and distance of the binding sites, the ability to bind 7-2-7 sites and single heptamers, and the presence or absence of Fnr and other transcription factors.

Structures of full-length NarL and $NarL^{C}$ bound to DNA, along with biochemical studies, support the mechanism of a shift from a closed to an open conformation upon phosphorylation.

In the process, two regions of the output domain are expected to become exposed and available for binding; these are located on the recognition (α9) and dimerization (α10) helices. As a result of helix α10 becoming available for dimerization, helices α4 and α6 will likely also experience modifications since these three helices form a hydrophobic pocket. In addition, the loop residues of the N-terminal domain also expected to become exposed. The exact mechanism of activation is unknown, however phosphorylation of Asp59 leads to allosteric conformational changes that alter the interface region and possibly other domain surfaces, revealing new binding sites vital to NarL function.

Figure 2-1. The Nar two-component system. (a) Nitrate, or nitrate, signal the autophosphorylation of membrane bound NarX/NarQ), and the subsequent phosphoryl transfer to an aspartate residue in the N-terminal domain (blue) of NarL/NarP. The resulting conformation change of the response regulator disrupts the domain interface and allows the C-terminal domain (yellow) to bind DNA. (Diagram provided by the Gunsalus lab, UCLA). **(b)** Asymmetric regulation in the Nar TCS. Upon stimulation by nitrate or nitrite, NarX and NarQ autophosphorylate using ATP (producing ADP) and subsequently phosphorylate NarL and NarP (denoted with a superscript "p"). Dashed arrows indicate slower reactions. $NarL^p$ and $NarP^p$ regulate operons involved in nitrate and nitrite respiration (representative examples shown). When no signal is present NarX and NarQ dephosphorylate $NarL^p$ and $NarP^p$. The asymmetry results from NarX having a preference for NarL whereas NarQ has an equal response to NarL and NarP. (This diagram is a reproduction, with slight modification, of a published version (26)).

(a)



(b)

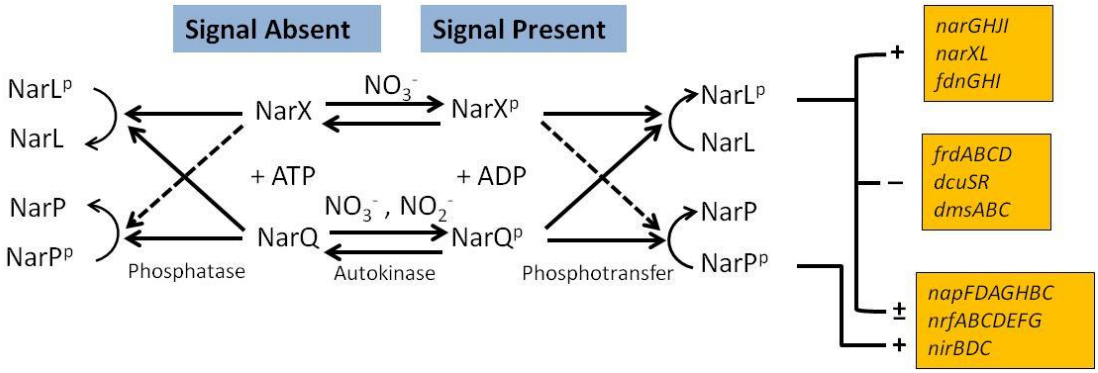


Figure 2-2. Promoter regions of *nrfA*, *nirB*, *napF*, *fdnG*, *narG*, and *yeaR* operons. Binding sites are depicted. Black inverted arrows: NarL and NarP 7-2-7 heptamers; white arrows: NarL single heptamer; inverted gray arrows: FNR; blue box: Fis; dark-red box: IHF; dark-green box: NsrR; orange box: modE; white box: Crp. The scale denotes nucleotides, and an arrow at the +1 nucleotide represents the transcription start site. Note: not all binding elements are shown. Information used to generate this image was taken from the EcoCyc database (20).

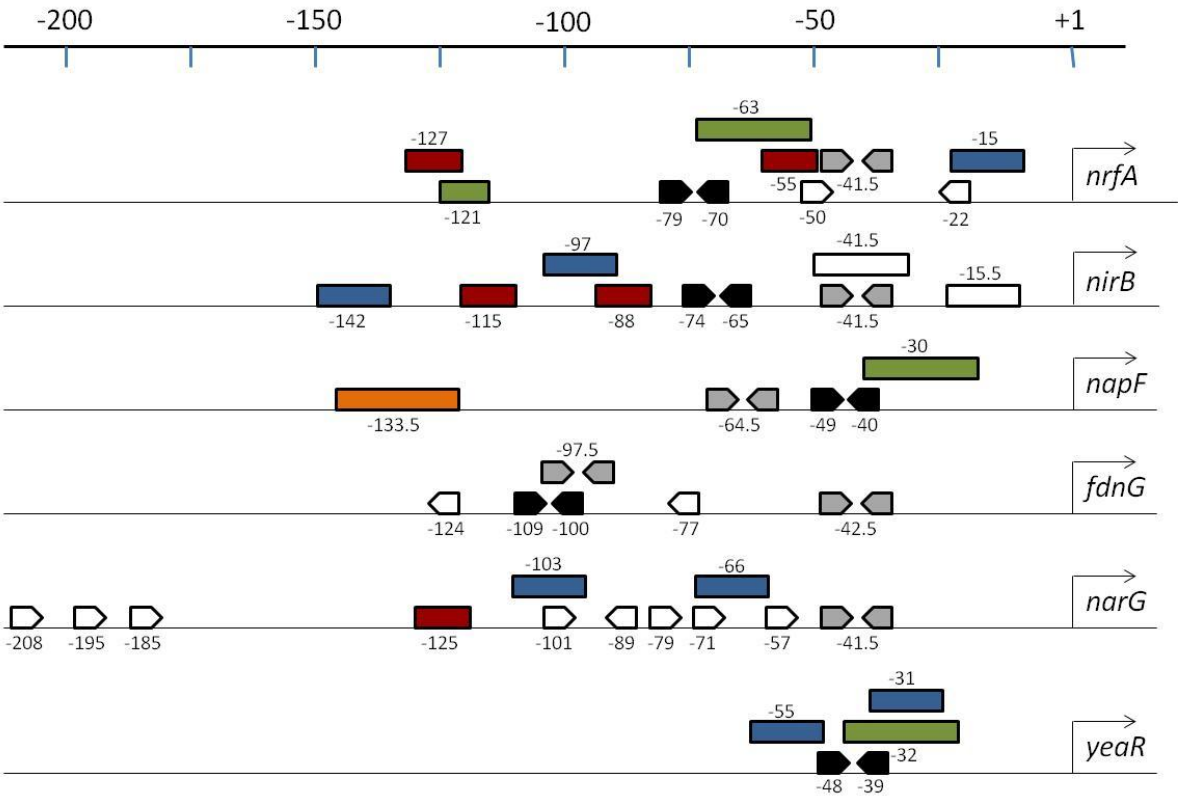
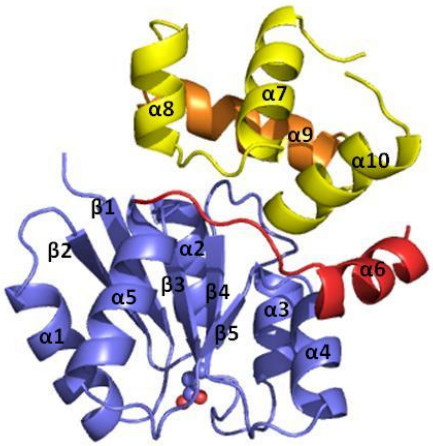


Figure 2-3. Structures of full-length NarL and NarL^{C} bound to DNA. **(a)** NarL (PDB ID: 1RNL) with REC domain (blue), effector domain (yellow), and linker region (red). The region between helix α6-α7 was too disordered to be resolved crystallographically, and is therefore missing. In this inactivated form, NarL cannot bind DNA because its recognition helix (orange) is obstructed by the REC domain. Phosphorylation at Asp59 (spheres) is expected to cause allosteric structural modifications that lead to domain separation. **(b)** NarL^{C} is bound to an artificially created palindromic 7-2-7 site (left, PDB ID: 1JE8), each heptamer being from the -74 promoter region of *nirB*. Two NarL^{C} monomers bind in an antiparallel fashion and dimerize through contacts made by the α10 helices. Three residues (sphere models) of the recognition helix (orange) contact the DNA bases while other residues make phosphate-backbone contacts. These images were created with Pymol (31).

(a)



(b)

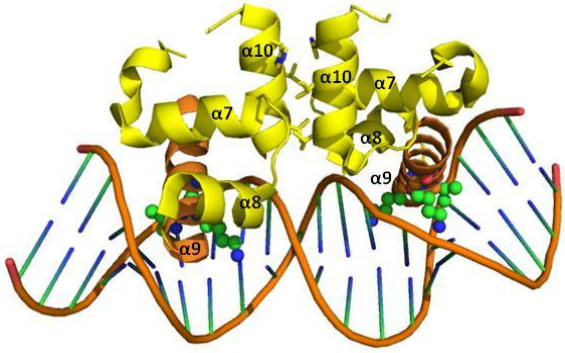


Table 2-1. Regulation of representative operons by NarL and NarP.

Operon	Function
Activated by NarL	
fdnGHI	Formate dehydrogenase-N
narGHJI	Cytoplasmic membrane nitrate reductase
narK	Nitrite transporter
narX	Sensor Histidine Kinase
yeaR-yoaG	Unknown
Repressed by NarL	
pfl	Pyruvate-formate lyase
adhE	Alcohol dehydrogenase
frdABCD	Fumarate reductase
dmsABC	DMSO/TMAO reductase
ynfEFGHI	Selenate Reductase
Activated by NarL and NarP	
nirBDC	Periplasmic NADH-nitrite reductase
Repressed by NarL and NarP	
ydhYVWXUT	Putative oxidoreductase
Activated by NarP and NarL but repressed by NarL (at high nitrate levels)	
napFDAGHBC	Periplasmic nitrate reductase
nrfABCDEFG	Cytoplasmic formate-nitrite reductase

References

1. **Baikalov, I., I. Schröder, M. Kaczor-Grzeskowiak, D. Cascio, R. P. Gunsalus, and R. E. Dickerson**. 1998. NarL dimerization? Suggestive evidence from a new crystal form. Biochemistry **37**:3665–76.

2. **Baikalov, I., I. Schröder, M. Kaczor-Grzeskowiak, K. Grzeskowiak, R. P. Gunsalus, and R. E. Dickerson**. 1996. Structure of the Escherichia coli Response Regulator NarL. Biochemistry **35**:11053–61.

3. **Bartkowski, W.** 2010. Defining the interdomain interface of the Escherichia coli response regulator NarL. UCLA, PhD Dissertation.

4. **Bearson, S. M. D., J. a Albrecht, and R. P. Gunsalus**. 2002. Oxygen and nitrate-dependent regulation of dmsABC operon expression in Escherichia coli: sites for Fnr and NarL protein interactions. BMC microbiology **2**:13.

5. **Browning, D. F., J. a Cole, and S. J. W. Busby**. 2004. Transcription activation by remodelling of a nucleoprotein assembly: the role of NarL at the FNR-dependent Escherichia coli nir promoter. Molecular microbiology **53**:203–15.

6. **Browning, D. F., D. C. Grainger, C. M. Beatty, A. J. Wolfe, J. a Cole, and S. J. W. Busby**. 2005. Integration of three signals at the Escherichia coli nrf promoter: a role for Fis protein in catabolite repression. Molecular microbiology **57**:496–510.

7. **Browning, D. F., D. J. Lee, A. J. Wolfe, J. a Cole, and S. J. W. Busby**. 2006. The Escherichia coli K-12 NarL and NarP proteins insulate the nrf promoter from the effects of integration host factor. Journal of bacteriology **188**:7449–56.

8. **Constantinidou, C., J. L. Hobman, L. Griffiths, M. D. Patel, C. W. Penn, J. A. Cole, and T. W. Overton**. 2006. A Reassessment of the FNR Regulon and Transcriptomic Analysis of the Effects of Nitrate , Nitrite , NarXL , and NarQP as Escherichia coli K12 Adapts from Aerobic to Anaerobic Growth. Journal of Biological Chemistry **281**:4802–4815.

9. **Darwin, A. J., J. Li, and V. Stewart**. 1996. Analysis of nitrate regulatory protein NarL-binding sites in the fdnG and narG operon control regions of Escherichia coli K-12. Molecular Microbiology **20**:621–632.

10. **Darwin, A. J., K. L. Tyson, S. J. Busby, and V. Stewart**. 1997. Differential regulation by the homologous response regulators NarL and NarP of Escherichia coli K-12 depends on DNA binding site arrangement. Molecular microbiology **25**:583–95.

11. **Darwin, A. J., E. C. Ziegelhoffer, P. J. Kiley, and V. Stewart**. 1998. Fnr , NarP , and NarL Regulation of Escherichia coli K-12 napF (Periplasmic Nitrate Reductase) Operon Transcription In Vitro. Journal of bacteriology **180**:4192–4198.

12. **Donaldson, L. W.** 2008. The NMR structure of the Staphylococcus aureus response regulator VraR DNA binding domain reveals a dynamic relationship between it and its associated receiver domain. Biochemistry **47**:3379–88.

13. **Ducros, V. M.-A., R. J. Lewis, C. S. Verma, E. J. Dodson, G. Leonard, J. P. Turkenburg, G. N. Murshudov, A. J. Wilkinson, and J. A. Brannigan**. 2001. Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in Bacillus subtilis. Journal of molecular biology **306**:759–71.

14. **Egan, S. M., and V. Stewart**. 1991. Mutational analysis of nitrate regulatory gene narL in Escherichia coli K-12. Journal of bacteriology **173**:4424–4432.

15. **Eldridge, A. M., H.-S. Kang, E. Johnson, R. Gunsalus, and F. W. Dahlquist**. 2002. Effect of phosphorylation on the interdomain interaction of the response regulator, NarL. Biochemistry **41**:15173–80.

16. **Filenko, N., S. Spiro, D. F. Browning, D. Squire, T. W. Overton, J. Cole, and C. Constantinidou**. 2007. The NsrR regulon of Escherichia coli K-12 includes genes encoding the hybrid cluster protein and the periplasmic, respiratory nitrite reductase. Journal of bacteriology **189**:4410–7.

17. **Gunsalus, R. P.** 1992. Control of electron flow in Escherichia coli: coordinated transcription of respiratory pathway genes. Journal of bacteriology **174**:7069–74.

18. **Hobbs, C. a, B. G. Bobay, R. J. Thompson, M. Perego, and J. Cavanagh**. 2010. NMR solution structure and DNA-binding model of the DNA-binding domain of competence protein A. Journal of molecular biology. Elsevier Ltd **398**:248–63.

19. **Jarvis, M. R.** 1999. Functional analysis of the Escherichia coli nitrate response regulator, NarL. UCLA, PhD Dissertation.

20. **Keseler, I. M., J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñiz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp**. 2011. EcoCyc: a comprehensive database of Escherichia coli biology. Nucleic acids research **39**:D583–90.

21. **Kurashima-ito, K., Y. Kasai, K. Hosono, K. Tamura, S. Oue, M. Isogai, Y. Ito, H. Nakamura, and Y. Shiro**. 2005. Solution Structure of the C-Terminal Transcriptional Activator Domain of FixJ from Sinorhizobium meliloti and Its Recognition of the fixK Promoter. Biochemistry **44**:14835–14844.

22. **Lin, H.-Y., P. J. Bledsoe, and V. Stewart**. 2007. Activation of yeaR-yoaG operon transcription by the nitrate-responsive regulator NarL is independent of oxygen-responsive regulator Fnr in Escherichia coli K-12. Journal of bacteriology **189**:7539–48.

23. **Maris, A. E., M. Kaczor-Grzeskowiak, Z. Ma, M. L. Kopka, R. P. Gunsalus, and R. E. Dickerson**. 2005. Primary and secondary modes of DNA recognition by the NarL two-component response regulator. Biochemistry **44**:14538–52.

24. **Maris, A. E., M. R. Sawaya, M. Kaczor-Grzeskowiak, M. R. Jarvis, S. M. D. Bearson, M. L. Kopka, I. Schröder, R. P. Gunsalus, and R. E. Dickerson**. 2002. Dimerization allows DNA target site recognition by the NarL response regulator. Nature structural biology **9**:771–8.

25. **Mcnicholas, P. M., and R. P. Gunsalus**. 2002. The Molybdate-Responsive Escherichia coli ModE Transcriptional Regulator Coordinates Periplasmic Nitrate Reductase (napFDAGHBC) Operon Expression with Nitrate and Molybdate Availability. Journal of bacteriology **184**:3253–59.

26. **Noriega, C. E., H.-Y. Lin, L.-L. Chen, S. B. Williams, and V. Stewart**. 2010. Asymmetric cross regulation between the nitrate-responsive NarX-NarL and NarQ-NarP two-component regulatory systems from Escherichia coli K-12. Mol Microbiol. **75**:394–412.

27. **Partridge, J. D., D. F. Browning, M. Xu, L. J. Newnham, C. Scott, R. E. Roberts, R. K. Poole, and J. Green**. 2008. Characterization of the Escherichia coli K-12 ydhYVWXUT operon: regulation by FNR, NarL and NarP. Microbiology **154**:608–618.

28. **Pristovsek, P., K. Sengupta, F. Löhr, B. Schäfer, M. W. von Trebra, H. Rüterjans, and F. Bernhard**. 2003. Structural analysis of the DNA-binding domain of the Erwinia amylovora RcsB protein and its interaction with the RcsAB box. The Journal of biological chemistry **278**:17752–9.

29. **Schröder, I., C. D. Wolin, R. Cavicchioli, and R. P. Gunsalus**. 1994. Phosphorylation and dephosphorylation of the NarQ, NarX, and NarL proteins of the nitrate-dependent two-component regulatory system of Escherichia coli. Journal of bacteriology **176**:4985–92.

30. **Schröder, I., S. Darie, and R. P. Gunsalus**. 1993. Activation of the Escherichia coli Nitrate Reductase (narGHJI) Operon by NarL and Fnr Requires Integration Host Factor. Journal of Biological Chemistry **268**:771–4.

31. **Schrödinger, L.** The PyMOL Molecular Graphics System. Version 1.3.

32. **Stewart, V.** 1994. Dual interacting two-component regulatory systems mediate nitrate- and nitrite-regulated gene expression in Escherichia coli. Research in microbiology **145**:450–4.

33. **Stewart, V.** 1993. Nitrate regulation of anaerobic respiratory gene expression in Escherichia coii. Molecular Microbiology **9**:425–434.

34. **Stewart, V., and P. J. Bledsoe**. 2005. Fnr-, NarP- and NarL-Dependent Regulation of Transcription Initiation from the Haemophilus influenzae Rd napF (Periplasmic Nitrate Reductase) Promoter in Escherichia coli K-12. Journal of bacteriology **187**:6928–35.

35. **Stewart, V., and P. J. Bledsoe**. 2003. Synthetic lac Operator Substitutions for Studying the Nitrate- and Nitrite-Responsive NarX-NarL and NarQ-NarP Two-Component Regulatory Systems of Escherichia coli K-12. Journal of bacteriology **185**:2104–2111.

36. **Vannini, A., C. Volpari, C. Gargioli, E. Muraglia, R. Cortese, R. De Francesco, P. Neddermann, and S. Di Marco**. 2002. The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. The EMBO journal **21**:4393–401.

37. **Volkman, B. F., D. Lipson, D. E. Wemmer, and D. Kern**. 2001. Two-state allosteric behavior in a single-domain signaling protein. Science (New York, N.Y.) **291**:2429–33.

38. **Volz, K.** 1993. Structural conservation in the CheY superfamily. Biochemistry **32**:11741–53.

39. **Walker, M., and J. DeMoss**. 1994. NarL-phosphate must bind to multiple upstream sites to activate transcription the narG promoter of Escherichia coli. Molecular Microbiology **14**:633–41.

40. **Walker, M., and J. DeMoss**. 1993. Phosphorylation and Dephosphorylation Catalyzed in Vitro by Purified Components of the Nitrate Sensing System, NarX and NarL. Journal of Biological Chemistry **268**:8391–93.

41. **Wang, H., and R. P. Gunsalus**. 2000. The nrfA and nirB Nitrite Reductase Operons in Escherichia coli Are Expressed Differently in Response to Nitrate than to Nitrite. Journal of Bacteriology **182**:5813–5822.

42. **Wang, H., and R. P. Gunsalus**. 2003. Coordinate Regulation of the Escherichia coli Formate Dehydrogenase fdnGHI and fdhF Genes in Response to Nitrate , Nitrite , and Formate : Roles for NarL and NarP. Journal of bacteriology **185**:5076–85.

43. **Wang, H., C. Tseng, and R. P. Gunsalus**. 1999. The napF and narG Nitrate Reductase Operons in Escherichia coli Are Differentially Expressed in Response to Submicromolar Concentrations of Nitrate but Not Nitrite. Journal of bacteriology **181**:5303–08.

44. **Wisedchaisri, G., M. Wu, A. E. Rice, D. M. Roberts, D. R. Sherman, and W. G. J. Hol**. 2005. Structures of Mycobacterium tuberculosis DosR and DosR-DNA complex involved in gene activation during adaptation to hypoxic latency. Journal of molecular biology **354**:630–41.

45. **Xu, M., S. J. W. Busby, and D. F. Browning**. 2009. Activation and repression at the Escherichia coli ynfEFGHI operon promoter. Journal of bacteriology **191**:3172–6.

46. **Zhang, J. H., G. Xiao, R. P. Gunsalus, and W. L. Hubbell**. 2003. Phosphorylation triggers domain separation in the DNA binding response regulator NarL. Biochemistry **42**:2552–9.

47. **Zhang, R., K. M. Pappas, J. L. Brace, P. C. Miller, T. Oulmassov, J. M. Molyneaux, J. C. Anderson, J. K. Bashkin, S. C. Winans, and A. Joachimiak**. 2002. Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. Nature **417**:971–4.

Chapter 3

The *Escherichia coli* Response Regulator NarL Dimerizes and Oligomerizes upon Phosphorylation

Introduction

Response regulators (RRs) of two-component systems (13) are often composed of two domains: an N-terminal receiver (REC) domain and a C-terminal "output" domain (see Chapter 1). The two domains are kept in proximity by a flexible linker and are also, but not always, bound together through an interface. Phosphorylation of an invariant aspartate on the REC domain disrupts the interface and separates the two domains, with the linker region keeping the protein intact. Upon phosphorylation, the unbound domains become available to bind other molecular surfaces. Output domains are able to carry out a diversity of functions in their unbound state. The majority of output domains bind specific DNA sites in transcription promoter regions, as seen in RRs of the FixJ/NarL, OmpR/PhoB, and NtrC/DctD subfamilies (13). The functions of an unbound REC domain are less certain, though its ability to dimerize suggests that, in RRs that regulate transcription, dimerization may enhance DNA binding or support transcriptional activation, or both (14).

The domains of the *E. coli* RR NarL have independent functions, which together serve to regulate transcription. Two crystal structures (3, 4) depict an N-terminal REC domain and a C-terminal output domain held together by a linker and an extensive interface (Figure 3-1a). The interface masks the DNA recognition sites of the output domain (reviewed in Chapter 2). Phosphorylation of the REC domain propagates a signal that severs the domain interface; the inhibition placed on the output domain becomes relieved and DNA binding is enabled. Once liberated from the interface, the output domain binds high-affinity DNA recognition sites as a dimer (Chapter 2, Figure 2-3b) (24, 25). Inverted heptamers flanking two base pairs, called "7-2-7" sequences, are optimal binding sites, though NarL is able to bind other heptameric arrangements (Figure 3-1b, and reviewed in Chapter 2) (11). Suggested roles for the unbound

NarL REC domain include stabilizing the output domain, binding to RNA polymerase or other proteins involved in transcription, and facilitating cooperative binding at certain promoters (23, 30, 35). Whether the NarL REC domain dimerizes as a result of phosphorylation is unknown. However, since several activated REC-domain structures are dimeric (1, 6, 8, 27, 33), often supported by biochemical evidence as well, the NarL REC domain is also likely to dimerize.

The numerous structures of activated RRs, and the two structures of full-length unphosphorylated NarL, provide possibilities as to how NarL may dimerize. Several structures of phosphorylated REC domains, or those activated with beryllofluoride (38), show that dimerization occurs partly or entirely through contacts made at the α4-β5-α5 region, a common binding site in activated RRs (reviewed in (14) and Chapter 1). REC domains from the OmpR/PhoB subfamily, such as *E. coli* $ArcA^{N(BF)\dagger}$, all dimerize using the entire α4-β5-α5 surface (33) (Figure 3-2a). Structures from members of the NtrC/DctD subfamily (6, 27), and the structure of *S. meliloti* $FixJ^{Np}$ (8) from the FixJ/NarL subfamily, show a dimer interface at α4-β5 (Figure 3-2a). The response regulator $spr1814^{N(BF)}$ from *S. pneumonia* (26) putatively belongs to the NarL/FixJ subfamily and dimerizes via its linker region that lies adjacent to the α4-β5-α5 surface (Figure 3-2a). In NarL, the α4-β5-α5 surface is partially blocked by regions of the linker, specifically it is blocked by loop α5-α6 and helix α6 (Figure 3-1a). If helices α6 and α4 of NarL separate upon phosphorylation, the α4-β5-α5 face would become exposed and the above modes of dimerization would be possible. The separation of helices α6 and α4 is plausible since helix α4 is a common target for activation-induced structural changes in many RRs (2, 8, 16, 20).

† Superscripts are used to represent different domain regions or protein states: "C" denotes the C-terminal domain, "N" denotes the N-terminal domain, numbers in parentheses indicate a protein-residue range, "p" refers to phosphorylated, and "BF" refers to beryllofluoride activated.

An alternative location for NarL dimerization is at helix α1. Two crystal structures of NarL show the ability of NarL to dimerize at helix α1 in the unphosphorylated state (3) (Figure 3-2b); therefore, this helix may also serve as a binding region in phosphorylated NarL. Furthermore, the effect of mutations in helix α1 in the response regulator UhpA- also of the NarL/FixJ subfamily- has led to the proposal that this helix is involved in UhpA oligomerization along the DNA (37).

To investigate the ability of full-length NarL and its individual domains to dimerize, analytical ultracentrifugation was used to determine the oligomeric states of full-length NarL in native and phosphorylated states, of $NarL^N$ and parts of linker in native and phosphorylated forms, and of two $NarL^C$ constructs that contain different linker lengths. Phosphorylation was found to induce oligomerization of NarL and further drive $NarL^N$ towards dimerization. $NarL^C$ was determined to exist primarily as a monomer. The results suggest that in addition to liberating the C-terminal domain, phosphorylation of NarL serves to create REC domain dimers and NarL oligomers. REC domain dimerization is proposed to occur at the α4-β5-α4 region. Possible binding modes of wild-type $NarL^p$ and of the NarL S80R mutant protein, as deduced by the possible role of the receiver domain at different promoter regions, are also presented.

Materials and Methods

Protein Constructs. Full-length NarL (29), $NarL^N$ (12), and the $NarL^C$ constructs (25) were expressed and purified by the Gunsalus lab (UCLA) as previously described.

Sample Preparation for Sedimentation Equilibrium Experiments. $NarL^p$ was prepared by incubating 246 uM NarL with phosphorylation buffer (40 mM KCl, 50 mM TRIS, pH 7.5, 40 mM $MgCl_2$, 2.5% glycerol) and 295 mM acetyl phosphate (1200:1 ratio of acetyl phosphate to

NarL) for 10 minutes at room temperature. $NarL^{Np}$ was prepared the same except using 242 uM $NarL^{N}$ and 291mM acetyl phosphate (1200:1 ratio of acetyl phosphate to $NarL^{N}$). Immediately following the reaction, the samples were placed in microdialysis buttons (Hampton), with a 1,000 molecular-weight-cutoff membrane (Spectrum), and dialyzed against several changes of buffer: 25 mM Tris pH 7.5 (or 8.5, which proved to make no difference) and 500 mM NaCl. $NarL^{p}$ however was first dialyzed into the buffer solution containing 750 mM NaCl before lowering it to 500 mM NaCl since it is more soluble in high salt and sometimes precipitates if taken directly to 500mM NaCl. After dialysis the proteins were diluted to the desired concentration, 60uM or 160uM. About 2ug of $NarL^{p}$ was withdrawn and run on a 20% native gel using the Phast Gel System (GE Healthcare) in order to confirm a successful reaction. This was not done for $NarL^{Np}$ since, unlike $NarL^{p}$, $NarL^{Np}$ does not produce a shifted band relative to its band in the unphosphorylated state. The yield of $NarL^{p}$ was measured to be 75% according to readings from the AlphaImager densitometer (Alpha Innotech Corp.).

$NarL^{C\ (147\text{-}261)}$, $NarL^{C\ (126\text{-}214)}$, NarL, and $NarL^{N}$ were also dialyzed using the same protocol and buffer solution as mentioned above and were diluted to 60uM or 160uM after dialysis. $NarL^{C\ (147\text{-}261)}$ at 60uM, however, was dialyzed once in a buffer solution containing 150mM NaCl and another time in a buffer solution containing 500mM $(NH_4)_2SO_4$. Both samples were tested and showed that different salt concentrations produced no significant difference in their sedimented molecular weights.

Sedimentation Equilibrium. Sedimentation equilibrium runs were performed at 4°C in a Beckman Optima XL-A analytical ultracentrifuge using absorption optics. Samples were examined in 3 mm double sector, 12 mm double sector and 12 mm six sector cells at an appropriate wavelength (240, 280, or 295 nm) to ensure the absorbance was sufficient to give a

good signal-to-noise ratio and the maximum absorbance was within the linear range of the instrument (less than 1.35 OD). Sedimentation equilibrium profiles were measured at 11,000, 15,000 and 18,000 rpm for NarL and $NarL^{p}$ and 15000, 18,000 and 22,000 rpm for $NarL^{C\ (147-216)}$, $NarL^{C\ (126-216)}$, $NarL^{N}$ and $NarL^{Np}$. For low wavelength (240 nm) scans a baseline was determined by pelleting the protein at 50,000 rpm. The data were initially fit with a nonlinear, least-squares exponential for a single ideal species using the Beckman Origin-based software (Version 3.01) to give a weight average molecular weight of all species in solution. When concentration and speed dependence of the molecular weights indicated association behavior, multiple runs (at least 4, including two concentrations and two different speeds) were analyzed using the "multifit" option of the Beckman global analysis software. The monomeric sequence molecular weight and various models (monomer-dimer, monomer-tetramer, and so on) were tested to see which would give the best fit to the data. Partial specific volumes of 0.740 for NarL, 0.733 for $NarL^{N}$, 0.734 for $NarL^{C\ (147-216)}$ and 0.739 for $NarL^{C\ (126-216)}$ calculated from the amino acid composition and corrected to 4˚C were used (9, 19). The calculated effect of the phosphate group on the partial specific volume was negligible, and thus ignored.

Structural Representations. All structural images were created with Pymol (31).

Results and Discussion

Analytical ultracentrifugation was applied to determine the oligomeric state of full-length NarL in its native and phosphorylated form, of $NarL^{N}$ (including part of the linker) in its native and phosphorylated form, and of two $NarL^{C}$ constructs, $NarL^{C\ (147-216)}$ and $NarL^{C\ (126-216)}$, containing different linker lengths. The constructs used for these studies are shown diagrammatically in Figure 3-3, and the results are summarized in Table 3-1.

NarL^{C} is Predominantly a Monomer

The constructs NarL$^{C\ (147\text{-}216)}$ and NarL$^{C\ (126\text{-}216)}$ were found to be primarily monomeric in solution at concentrations of about 60uM and 160uM (Table 3-1). The molecular weight of the shorter construct, NarL$^{C\ (147\text{-}216)}$, which lacks helix α6 and has a sequence molecular weight of 9.6 kD, was determined to be 9.6 kD at 60uM and 22,000 rpm. No salt conditions were found to promote a greater amount of dimerization of this construct. At 160uM, however, the molecular weight of NarL$^{C\ (147\text{-}216)}$ was determined to be 11.4 kD at 22,000 rpm. The concentration dependence of the weight-average molecular-weight suggested that weak dimerization might be present.

At 22,000 rpm the molecular weight of NarL$^{C\ (126\text{-}216)}$, which contains loop α5- α6 and helix α6, was determined to be 12.7 kD at 60uM (Figure 3-4a) and 12.4 kD at 160uM. Compared to the sequence molecular weight of 11.8 kD, NarL$^{C\ (126\text{-}216)}$ is mainly monomeric with a small amount of dimer present. Residuals from the exponential fitting were small, indicating little molecular weight heterogeneity. For both NarL^{C} constructs, additional data was taken at speeds of 15,000 rpm and 18,000 rpm. This data gave similar results regarding molecular weights and residuals to that taken at 22,000 rpm, confirming that NarL^{C}, with or without the linker region, is predominantly monomeric in solution. In contrast, NarL$^{C\ (147\text{-}216)}$ bound to DNA in crystal structures dimerizes at helix α10 (Chapter 2, Figure 2-3b) (24, 25). This suggests that the presence of DNA drives NarL^{C} dimerization.

NarL^{P} Forms Dimers and Higher-Order Oligomers

Full-length, unphosphorylated NarL, was monomeric with a measured molecular weight of 24.0 kD, at 11,000 rpm, with small fitting residuals (Table 3-1 and Figure 3-4b). Its sequence-

determined molecular weight is 23.9 kD. When full-length NarL was phosphorylated using acetyl phosphate and examined at 11,000 rpm, the oligomeric state changed drastically with a weight-average molecular-weight of 48.3 kD, indicative of a dimer. At the higher protein concentration of 160uM and the same ultracentrifugation speed, the measured molecular weight of NarL was 23.9 kD (i.e. monomer) while that of $NarL^P$ rose dramatically to 70.0 kD (i.e. dimer and tetramer). At both concentrations, $NarL^P$ was heterogeneous as seen by the non-random residuals, especially at the speed of 18,000 rpm. However, using data taken at 11,000, 15,000 and 18,000 rpm and the two concentrations, $NarL^P$ fit best to a monomer-dimer-tetramer equilibrium. Densitometry measurements of the 160uM run showed that $NarL^P$ was initially about 75% phosphorylated (Figure 3-5, lane 5). The level of phosphorylation is expected to have been largely maintained since our small-scale electrophoretic studies of $NarL^P$ have demonstrated its stability for up to one week at 4˚C (data not shown).

Thus, the full-length NarL data shows that unphosphorylated NarL is mainly a monomer in solution, with a small amount of dimerization at higher salt concentrations (data not shown; though this occurred only at 60uM). Upon phosphorylation, the molecular weight radically increases, especially at higher concentrations where it seems to be capable of forming oligomers. This behavior occurred consistently in repetitious runs using the same or improved (80% yield, Figure 3-5, lane 6) phosphorylation procedures. $NarL^P$ appears to be a dimer in low concentrations and at equilibrium between monomer, dimer, and tetramer at higher concentrations. Unlike detached $NarL^C$ that dimerizes only when bound to DNA, $NarL^P$ forms dimers and tetramers prior to DNA binding. The ability of $NarL^P$ to tetramerize was also suggested in a recent *in vivo* study, whereby $NarL^P$ dimers, bound to synthetic 7-2-7 sites, presumably formed a tetramer (32) to repress transcription.

Both NarL^{N} and NarL^{Np} are able to Dimerize

NarL^{N} used for the ultracentrifugation studies included the regulatory domain and part of the linker region (Figure 3-3), giving a sequence molecular weight of 16.5 kD (Table 3-1 and Figure 3-4c). At the lower 60uM concentration of NarL^{N}, speed-dependent molecular weights of 26.1 kD (15,000 rpm), 24.9 kD (18,000 rpm), and 23.4 kD (22,000 rpm) were found and best fit by a monomer and dimer. The same respective speeds for NarL^{Np} gave molecular weights of 30.8 kD, 30.2 kD (Figure 3-4c), and 29.0 kD and were best fit by a population that is largely in the dimeric form. Raising the concentration to 160uM gave similar results. NarL^{N} had apparent molecular weights of 25.9 kD (15,000 rpm) and 23.0 kD (22,000 rpm), while those for NarL^{Np} were 30.4 kD (15,000 rpm) and 27.4 kD (22,000), with good to decent looking residuals. Again this suggests the presence of a monomer and dimer for NarL^{N} and a primarily dimeric species for NarL^{Np}. The slightly lower-than-expected measured molecular weights for NarL^{Np} were surprising, though may be due to the interference of a loosened helix α6 (discussed later), or due to a slight loss of phosphorylation. Nonetheless, the drive towards NarL^{N} dimerization is apparent. This data suggests that phosphorylation not only releases the inhibition of the recognition helix, α9, but additionally enables the formation of NarL REC-domain dimers.

Neither the Linker Region nor Helix α1 is Suggestive of a Dimerization Site

Both NarL^{C} proteins analyzed here exist mainly in the monomeric form in a solution devoid of DNA. The main difference between the two constructs is the presence of the α5-α6 loop and helix α6 in NarL$^{C\ (126\text{-}214)}$ that is absent in NarL$^{C\ (147\text{-}216)}$. Therefore, no role in dimer formation was attributed to the α5-α6 loop and helix α6 in these constructs, since an even smaller amount of dimerization was detected in NarL$^{C\ (126\text{-}216)}$ compared to NarL$^{C\ (147\text{-}216)}$. If the α5-α6

loop and helix α6, on their own, did contribute to dimerization we would expect a predominantly dimeric species for $NarL^{C\ (126\text{-}216)}$ in contrast to the predominantly monomeric species for $NarL^{C\ (147\text{-}216)}$. These two portions of the linker region are also present in the $NarL^{N}$ construct and are also unlikely to contribute to its dimerization (discussed below).

The sedimentation results presented here also discredit NarL dimerizing at helix α1, or other dimerization regions observed in the NarL monoclinic crystal structure. At the concentrations tested, unphosphorylated NarL is a monomer despite these regions being fully exposed. In addition, the $NarL^{N}$ and $NarL^{Np}$ sedimentation results imply that dimerization occurs in a region that is normally inaccessible. These results, however, do not negate the possibility of helix α1 being involved in multimerization. In fact, the crystal structure of $FixJ^{Np}$ shows three molecules in the asymmetric unit with the center molecule dimerizing at α4-β5 on one side (as shown in Figure 3-2a), and at α1-β2 on the other (8). Similarly, helix α1 dimerization of NarL may occur when multiple $NarL^{p}$ molecules are required to bind cooperatively along continuous DNA sites, as has been proposed for UhpA (37).

NarL Dimerization at the α4-β5-α5 Region is a Strong Possibility

In the absence of the output domain, the apparent molecular weight of the N-terminal domain, $NarL^{N}$, increases so that approximately half the population is a monomer and half is a dimer. The presence of an unphosphorylated $NarL^{N}$ dimer was surprising given that unphosphorylated NarL was a monomer. This observation can be explained by two possibilities. The first is that during the purification process some $NarL^{Np}$ co-purified with $NarL^{N}$. This was tested using the mass spectrometry technique MALDI, which determined that both phosphorylated and unphosphorylated $NarL^{N}$ species were present in the sample (data not

shown). However we believe that it is unlikely that the smaller $NarL^{Np}$ peak detected by MALDI, relative to the larger $NarL^{N}$ peak, would account for such a large amount of dimerization. The second possibility is that the absence of the C-terminal domain, which might normally impede dimerization, now allows $NarL^{N}$ to more readily dimerize even without phosphorylation. This implies that dimerization of $NarL^{N}$ occurs in a region that is unexposed in the full-length protein.

The α4-β5-α5 region is a strong candidate for a region being inaccessible in a closed NarL but accessible in the open form of NarL, and perhaps partially accessible in detached $NarL^{N}$. In the unphosphorylated, full-length structure of NarL, helix α6 is secured by hydrophobic interactions with helix α10 of the C-terminal domain and by hydrophobic interactions with helix α4 of the N-terminal domain. The absence of helix α10 in $NarL^{N}$ may render helix α6 less secure such that it is susceptible to swaying and exposing the α4-β5-α5 surface. This would allow a population of $NarL^{N}$ molecules to form dimers at this region, of which could take the form of the dimerization schemes shown in Figure 3-2a. Phosphorylation of $NarL^{N}$ would further drive the release of α6 (severing its contacts with helix α4) and give rise to a larger dimer population.

Of the dimerization schemes shown in Figure 3-2a, that involving the linker region (such as that of $spr1814^{N(BF)}$) is the least plausible for NarL. First, the α5-α6 loop and helix α6 region did not dimerize while attached to $NarL^{C}$, and second, the possibility that this region may dimerize when structured against the α4-β5-α5 surface is also unlikely. In the latter case, larger dimer populations would be expected for $NarL^{N}$ and $NarL^{Np}$ than was observed. In contrast, if dimerization occurred directly at the α4-β5-α5 region, a loosened helix α6 and preceding loop may be a slight impediment to dimerization, which is more consistent with our data of partial and

predominant dimers for $NarL^{N}$ and $NarL^{Np}$, respectively. Furthermore, dimerization at the α4-β5-α5 region, not involving the linker region, would model after the structure of activated, full-length, *A. tumefaciens* TraR (34, 39). TraR and NarL both belong to the LuxR-type family that contains a helix-turn-helix DNA-binding motif. In the structure of activated, full-length TraR bound to DNA, the linker region serves to facilitate domain rearrangements rather than contribute to dimerization. NarL, which needs to bind different heptamer arrangements, is more apt to utilize its linker in this manner, as is also expected for members of the OmpR/PhoB subfamily (2, 14). Thus, if $NarL^{N}$ dimerizes at the α4-β5-α5 region, the dimerization scheme is anticipated to be similar to that of $ArcA^{N(BF)}$ or $FixJ^{Np}$ (Figure 3-2a).

The Roles of the NarL N- and C-terminal Domains

The analytical ultracentrifugation results indicate that the N- and C-terminal domains of NarL have different roles in transcriptional regulation as molecular surface contacts are altered during activation (summarized in Figure 3-6). Dimerization of the C-terminal domain seems to rely on the presence of DNA (Figure 3-6a,b), and serves to facilitate binding to high-affinity 7-2-7 sites (24, 25). Detached $NarL^{C}$ is able to bind certain 7-2-7 sites with similar affinities as $NarL^{p}$ (23, 25), and in some cases activates transcription, however it cannot activate transcription at promoters that are notorious for exercising cooperativity (23). This suggests several roles for the N-terminal domain, which is able to dimerize once separated from the C-terminal domain (Figure 3-6c). Dimerization of the NarL REC domain may stabilize the output domain and increase the affinity of $NarL^{p}$ for DNA, especially at lower-affinity binding sites (Figure 3-6d,e). These are non 7-2-7 sites that have a diversity of arrangements and often require cooperativity, such as in the *narG* operon (Figure 3-1b) (Chapter 2 and (11, 35)). Such roles for the N-terminal

domain have also been proposed for other members of the NarL/FixJ family (15, 21, 37). $DosR^C$, for example, is unable to bestow full induction of transcription, presumably due to lack of cooperativity or protein-protein interactions that would be available in the presence of the N-terminal domain. UhpA requires phophorylation to bind cooperatively to its low-affinity binding sites, and the components required for UhpA cooperativity point to its REC domain.

In vivo expression studies of the NarL S80R mutant confirm the required presence of the NarL N-terminal domain for activation of the *narG* promoter (17). Located in the central α3-β4 loop of the NarL interface, Ser80 is a vital residue for maintaining interface integrity (5, 17). An Arg substitution at this location renders the protein constitutively active; probably due to the introduced steric hindrance that pries the domains apart and exposes the DNA-recognition helix. NarL S80R mutation analysis shows that this protein constitutively represses transcription at the *frdA* promoter in a phosphorylation-independent manner, however requires phosphorylation to constitutively activate the *narG* promoter (17). Therefore, phosphorylation presumably exposes surfaces that facilitate binding and activation of the *narG* operon that are not required at the *frdA* operon. One possible explanation for these observed phenotypes is that NarL REC-domain dimerization (or $NarL^p$ oligomerization) may be required at the *narG* promoter but is not obligatory at the *frdA* promoter. The *frdA* promoter contains one high-affinity site and three single sites crowded near the transcription start site (Figure 3-1b) (22). Since the role of $NarL^p$ at repression sites is to occlude RNA polymerase (7, 22, 23, 28, 36), perhaps this may be achieved at the *frdA* promoter without REC domain dimerization (or $NarL^p$ oligomerization). Thus, the NarL S80R mutant may fulfill the *frdA* promoter binding requirement solely via an exposed recognition helix. At the high-affinity 7-2-7 dimer site, the C-terminal domains may possibly dimerize on the DNA while the REC domains remain undimerized (Figure 3-7). The single sites

need not necessarily be bound, depending on the minimum requirements for repression, but if so may also not require REC dimerization. In contrast, the *narG* promoter is comprised only of low-affinity binding sites (Figure 3-1b) that rely on cooperativity for their occupancy and for full activation (10, 11, 35). This promoter region is also expected to bend significantly upon binding of Integration Host Factor (30) to allow $NarL^{p}$ at upstream sites to contact the transcription initiation region. Thus, the binding requirements at the *narG* promoter are more complex and REC dimerization and $NarL^{p}$ oligomerization is likely to facilitate cooperativity. Once NarL S80R is phosphorylated and NarL REC-domain dimerization ensues, conceivably at the α4-β5-α5 region or an otherwise site that becomes exposed upon phosphorylation, binding to the lower affinity *narG* sites (11, 35) would become attainable (Figure 3-7).

The required presence of the NarL N- and C-terminal domains vary at different promoter regions (23), and seem to depend on heptamer affinity and the regulatory goal (whether activation or repression). The output domain is certainly required for DNA binding, however the role of the N-terminal REC domain may vary according to the specific need of a promoter region. The N-terminal domain appears to have versatile roles, which likely include stabilizing the output domain and engaging in protein-protein interactions. In NarL, phosphorylation-induced dimerization is almost certainly an essential part of these processes.

Conclusions

Analytical ultracentrifugation was applied to full-length NarL and to different NarL domains. The results show that phosphorylation induces dimerization of the receiver domain, and induces dimerization and multimerization of full-length NarL. Both $NarL^{C}$ constructs, which differed by linker length, were predominantly monomeric, suggesting that output-domain

dimerization occurs upon DNA binding. The part of the linker region examined, namely the α5-α6 loop and helix α6, did not dimerize when attached to $NarL^C$. This does not necessarily eliminate a scheme where this region orients against the α4-β5-α5 surface and provides surface contacts for dimerization, however the more plausible dimerization mode proposed for NarL involves direct dimerization at α4-β5-α5 with the linker region maintaining proximity between the two domains as they maneuver to bind different heptamer arrangements along promoter regions. Phosphorylation of NarL S80R fulfills a requirement at the *narG* promoter, which is not required at the *frdA* promoter. This could be the formation of active REC dimers that would facilitate binding to lower-affinity heptamers. NarL seems to have different binding requirements at different promoter sites, these based on DNA affinity and whether its role at a particular operon is to activate or repress transcription. Dimerization, where pertinent, is expected to be integral to its function.

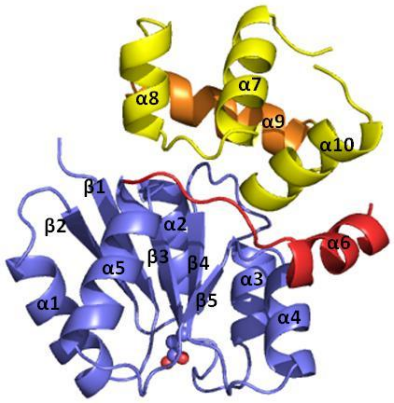
These results advance the current paradigm of NarL activation. Phosphorylation has a dual purpose: to release the inhibition on the output domain *and* to induce receiver-domain dimerization. In unphosphorylated NarL, the N- and C-terminal domains are bound by a mutual interface. Once liberated, each domain dimerizes and forms alternative surface interactions, allowing the protein as a whole to regulate transcription.

Acknowledgements

Martin Phillips (for help with all aspects of the analytical ultracentrifugation studies), Woytek Bartkowski, Michael R. Jarvis, Ann E. Maris, Mary L. Kopka, Imke Schroeder, Richard E. Dickerson, Wayne L. Hubbell, and Robert P. Gunsalus.

Figure 3-1. The structure of NarL and its mode of binding at the *frdA* and *narG* operons. **(a)** Ribbon diagram of NarL (PDB ID: 1RNL). The REC domain is colored blue, the linker region in red, and the output domain in yellow. The "tether" region between helix α6-α7 was too disordered to be resolved crystallographically, and is therefore missing. The DNA recognition helix (α9) is colored orange for emphasis, and the site of phosphorylation, Asp59, is shown in spheres. **(b)** NarL heptameric binding sites along the *frdA* and *narG* operons. Black inverted arrows represent high-affinity, 7-2-7 binding arrangements; white arrows represent non 7-2-7 sites or single heptameric sites; the dark-red box represents the Integration Host Factor binding site. The scale denotes nucleotides, and an arrow at the +1 nucleotide represents the transcription start site. Information used to generate this image was taken from the EcoCyc database (18) and references therein.

(a)



(b)

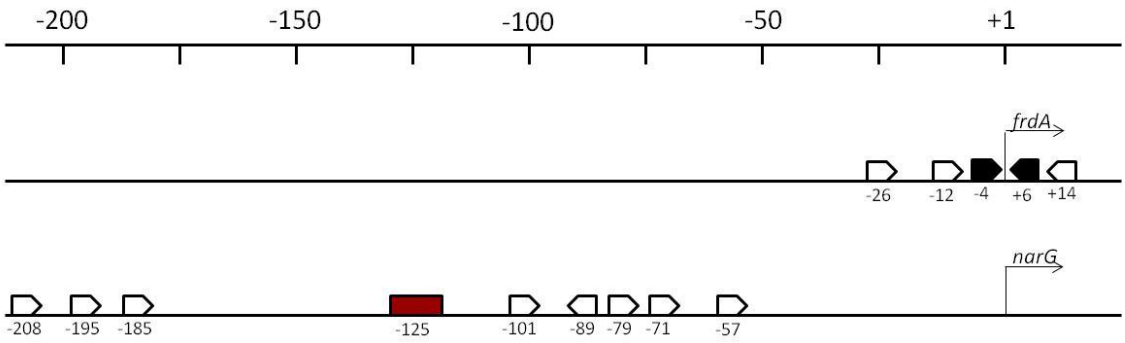


Figure 3-2. Dimerization modes of REC domains. **(a)** Dimerization of $FixJ^{Np}$ (PDB ID: 1D5W), $ArcA^{N(BF)}$ (PDB ID: 1XHF), and $spr1814^{N(BF)}$ (PDB ID: 4E7P). REC domains are colored pink, the α4-β5-α5 region in light blue, and, in $spr1814^{N(BF)}$, the linker region is colored red. The activated aspartate residue is shown in stick model to provide orientation. The REC domains show different versions of dimerization at the α4-β5-α5 region. **(b)** NarL dimerization as seen in its two crystal structures (left is PDB ID: 1RNL, and right is PDB ID: 1A04). The REC domain is colored blue with helices α1 in green, and the output domain is colored yellow. Both structures show dimerization between helices α1 (though at different orientations).

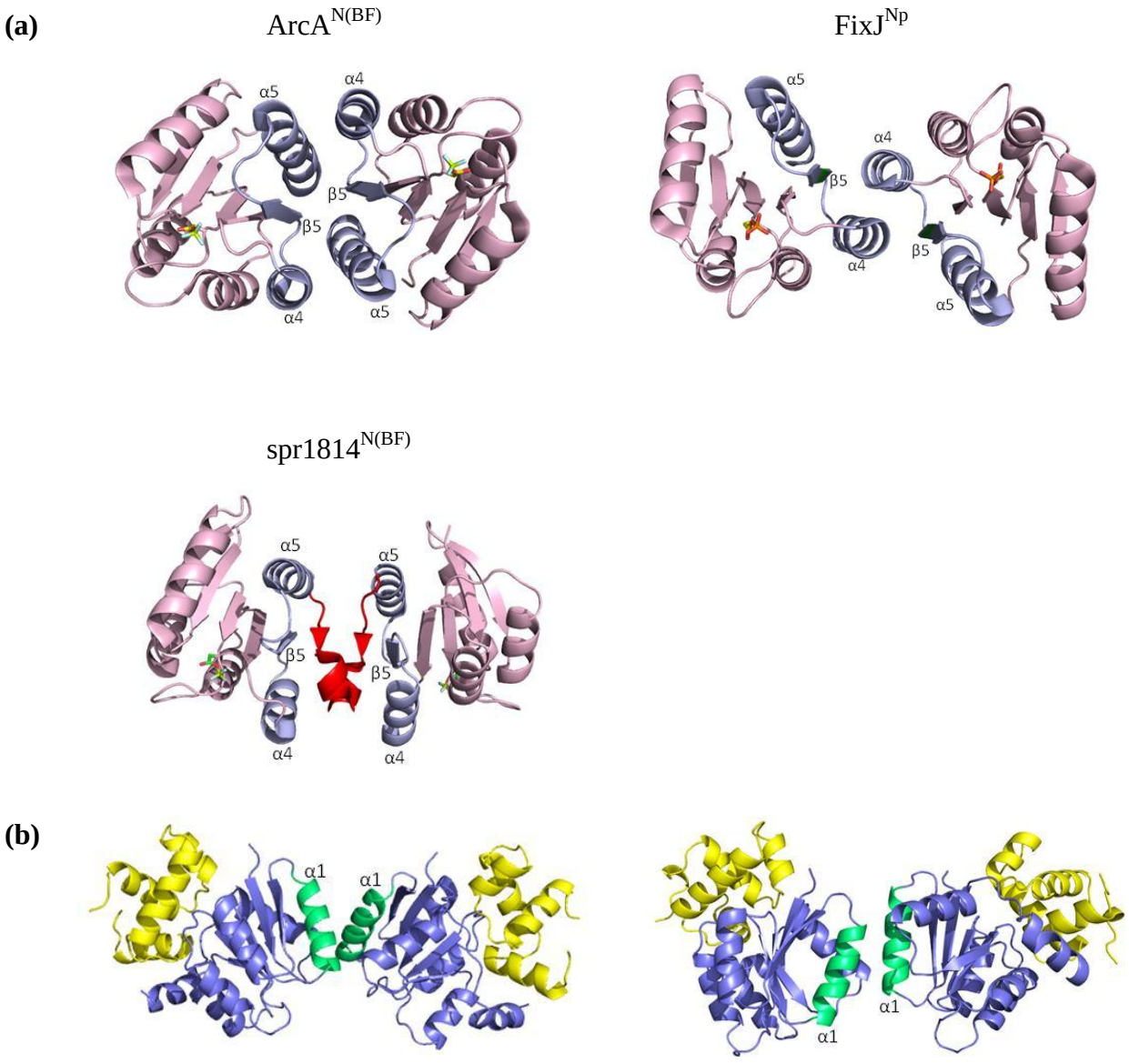
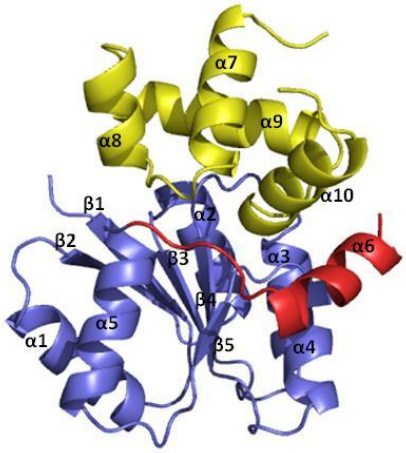


Figure 3-3. Constructs used for analytical ultracentrifugation experiments. The secondary and tertiary structures of NarL (PDB ID: 1RNL) are shown using the same color scheme: the REC domain in blue, the linker region in red, and the output domain in yellow. Arrows in the secondary structure diagram represent β-strands and cylinders represent α-helices. The constructs used for analytical ultracentrifugation experiments are designated by colored bars and residue range is indicated.



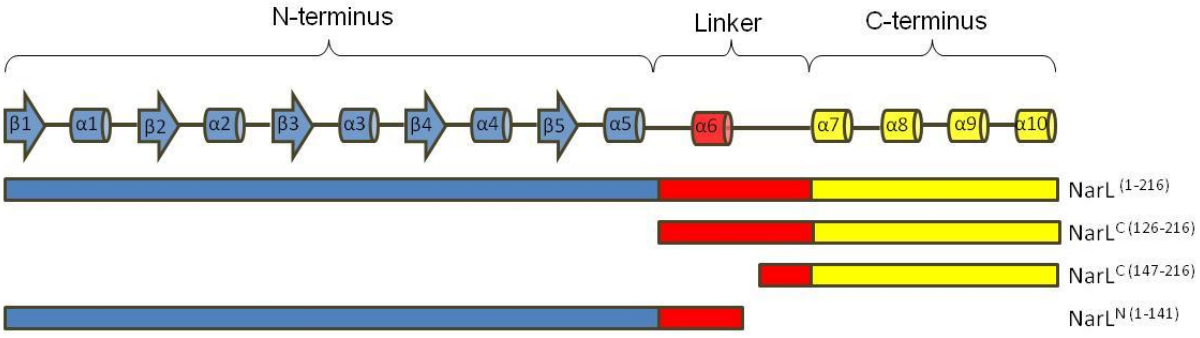


Figure 3-4. Analytical ultracentrifugation studies of NarL constructs and the effect of phosphorylation. Plots of optical density of protein versus distance from the center of rotation are shown. These are single exponential fits with the residuals shown in the upper plot of each panel. The weight-average molecular-weights are indicated. Triangles represent unphosphorylated proteins, circles indicate phosphorylated proteins. Samples were at 4°C in 500mM NaCl and 25mM Tris pH 7.5-8.5, except NarL which was in 100mM NaCl and 25mM Tris pH 8.5. **(a)** The $NarL^{C\ (126\text{-}216)}$ construct, which includes the α5-α6 loop and helix α6, after reaching equilibrium at 22,000 rpm and 280 nm. The monomeric molecular weight determined from the sequence is 11.8 kD; the experimental molecular weight shown is 12.8 kD, indicating that $NarL^{C\ (126\text{-}216)}$ is mostly a monomer. **(b)** NarL and $NarL^{p}$ after reaching equilibrium at 11,000 rpm and 280 nm. The monomeric molecular weight of NarL determined from the sequence is 23.9 kD, the measured molecular weight is 24 kD, suggesting a monomer. The measured molecular weight shown for $NarL^{p}$ is 48.3 kD, suggesting a dimer. **(c)** $NarL^{N}$ and $NarL^{Np}$ after reaching equilibrium at 15,000 rpm and 240 nm. The monomeric molecular weight of $NarL^{N}$ determined from the sequence is 16.5 kD, the measured molecular weight is 26.1 kD, suggesting a monomer-dimer equilibrium. The measured molecular weight for $NarL^{Np}$ shown is 30.8 kD suggesting it is almost entirely a dimer.

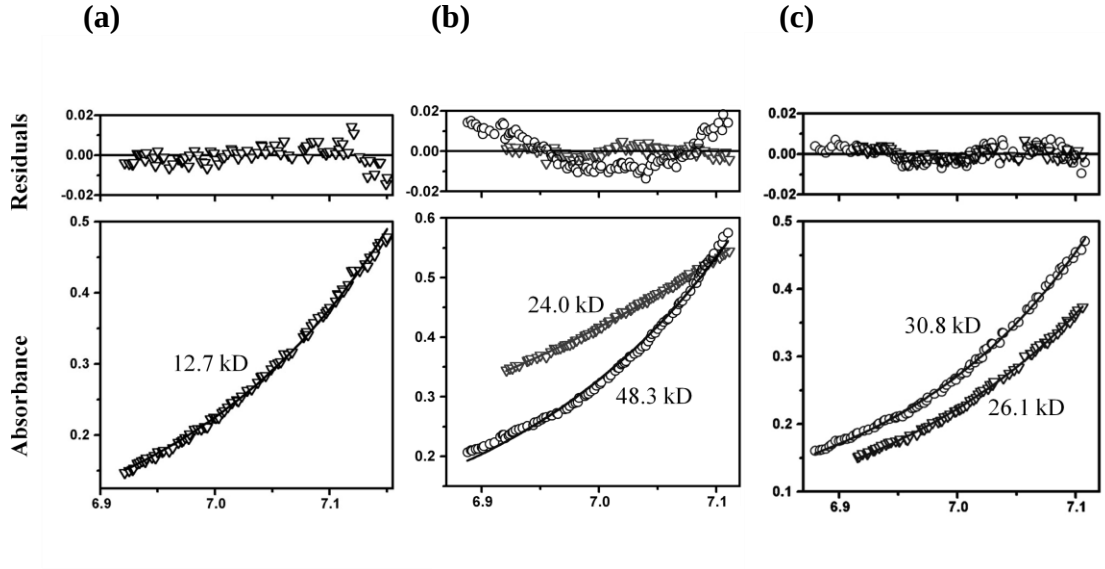


Figure 3-5. The NarL phosphorylation reaction. NarL was phosphorylated using different ratios of acetyl phosphate (AP) to NarL. Lane 1- NarL only; Lane 2- $NarL^P$ at 400:1 AP:NarL; Lane 3- $NarL^P$ at 600:1 AP:NarL; Lane 4- $NarL^P$ at 800:1 AP:NarL; Lane 5- $NarL^P$ at 1200:1 AP:NarL; Lane 6- $NarL^P$ at 400:1 AP:NarL and 175mM additional KCl. Lanes 5 and 6 are 76.7% and 80% phosphorylated, respectively, as determined from an AlphaImager densitometer. Lane 5 represents the reaction used for the sedimentation equilibrium experiments, lane 6 represents a slightly improved version by adding additional KCl and lowering the AP:NarL ratio.

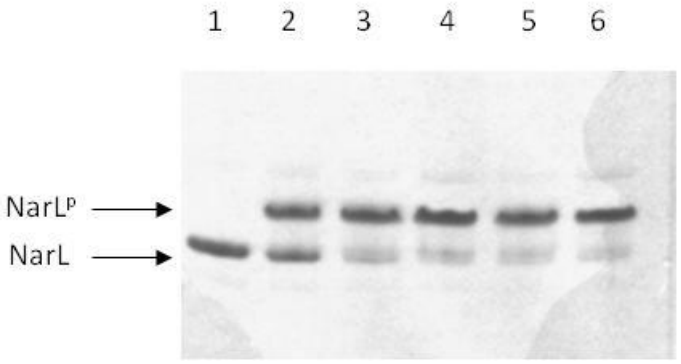


Figure 3-6. The N- and C-terminal domains of NarL dimerize independently of each other and may have different roles in DNA binding. NarL is represented by a blue box for the N-terminal domain (labeled "N"), a yellow circle for the C-terminal domain (labeled "C"), and the linker region in red with helix α6 represented by a cylinder. A pink box represents the transition to an activated, phosphorylated REC domain with the phosphoryl group represented by a green dot. Black inverted arrows represent a 7-2-7 site; white arrows represent lower-affinity heptameric binding-sites that are non 7-2-7 sites. The DNA is represented by an orange band shown to bend upon binding. **(a)** The $NarL^{C\ (147\text{-}216)}$ construct is predominantly monomeric, however **(b)** this construct is able to dimerize in the presence of a high affinity 7-2-7 DNA binding-site (25). **(c)** $NarL^{N}$ is able to dimerize in the absence of the C-terminal domain. **(d)** Proposed mode of $NarL^{p}$ activation and DNA binding at a high-affinity 7-2-7 site. Phosphorylation severs the NarL interdomain interface and allows REC domain dimerization. Dimerization of the C-terminal domains would follow upon DNA binding. **(e)** Proposed mode of $NarL^{p}$ activation and DNA binding at lower affinity binding sites. NarL REC domain dimerization is likely to enhance and enable C-terminal domain binding to different, non 7-2-7, binding sites (examples of two possible types of binding arrangements are shown).

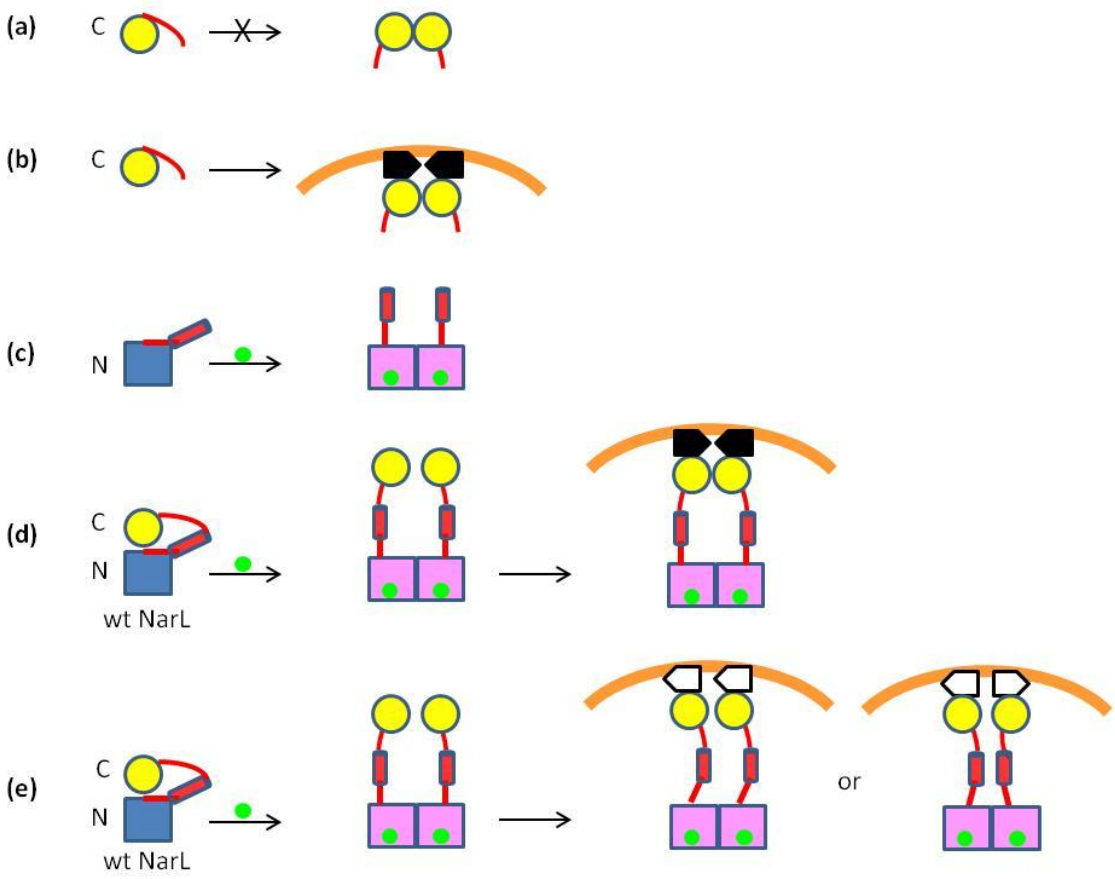


Figure 3-7. Proposed modes of NarL S80R binding at the *frdA* and *narG* promoter regions. The NarL S80R protein is represented by a pink box for the N-terminal domain (labeled "N"), a yellow circle for the C-terminal domain (labeled "C"), and the linker region in red with helix α6 represented by a cylinder. Black arrows represent a 7-2-7 site; white arrows represent lower-affinity heptameric binding-sites. The DNA is represented by an orange band shown to bend upon binding. Initially, the domains of S80R are already separated and the output domain is proposed to dimerize at the 7-2-7 site of the *frdA* promoter without the requirement for REC dimerization. Efficient binding of NarL S80R at *narG* does not occur without phosphorylation (green dot), perhaps due to a requirement of REC dimerization, which would enhance binding to the lower-affinity, non 7-2-7, sites at this promoter. The linker region is shown to facilitate the domain movement as NarL binds to different heptameric arrangements.

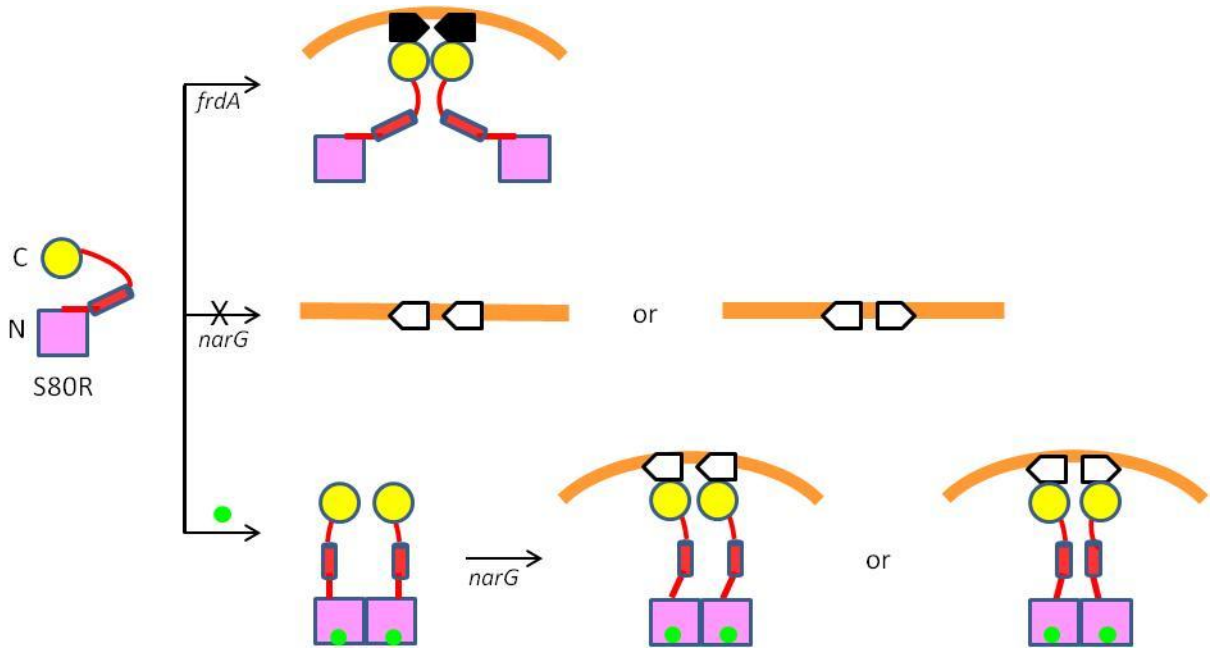


Table 3-1. Summary of sedimentation equilibrium results. Sedimentation equilibrium experiments were carried out using at least two different speeds and with two different concentrations of each protein, 60uM and 160uM. The measured molecular weight compared to the sequence derived molecular weight is shown for each run. Certain constructs had additional data collected at other speeds which are not shown in the table but were included in determining the best oligomeric fit (discussed in text). All runs shown were carried out in a buffer solution containing 25mM Tris, pH 7.5-8.5 and 500mM NaCl, except for two cases: $NarL^{C\ (147\text{-}214)}$ at 60uM contained 500mM $(NH_4)_2SO_4$, and NarL at 60uM contained 100mM NaCl.

Construct	Sequence MW (kD)	Speed (krpm)	Measured MW at 60uM (kD)	Measured MW at 160uM (kD)	Oligomeric State
$NarL^{C\ (147\text{-}216)}$	9.6	15	9.7	12.3	Predominantly
		22	9.6	11.4	Monomer
$NarL^{C\ (126\text{-}216)}$	11.8	15	14.1	13.6	Predominantly
		22	12.7	12.4	Monomer
NarL	23.9	11	24.0	23.9	Monomer
		15	22.6	23.0	
$NarL^{P}$	24.0	11	48.3	70.0	Monomer/Dimer/
		15	37.0	49.4	Tetramer
$NarL^{N}$	16.5	15	26.1	25.9	Monomer/Dimer
		22	23.4	23.0	
$NarL^{Np}$	16.6	15	30.8	30.4	Predominantly
		22	29.0	27.4	Dimer

References

1. **Bachhawat, P., and A. M. Stock**. 2007. Crystal structures of the receiver domain of the response regulator PhoP from Escherichia coli in the absence and presence of the phosphoryl analog beryllofluoride. Journal of bacteriology **189**:5987–95.

2. **Bachhawat, P., G. V. T. Swapna, G. T. Montelione, and A. M. Stock**. 2005. Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states. Structure (London, England : 1993) **13**:1353–63.

3. **Baikalov, I., I. Schröder, M. Kaczor-Grzeskowiak, D. Cascio, R. P. Gunsalus, and R. E. Dickerson**. 1998. NarL dimerization? Suggestive evidence from a new crystal form. Biochemistry **37**:3665–76.

4. **Baikalov, I., I. Schröder, M. Kaczor-Grzeskowiak, K. Grzeskowiak, R. P. Gunsalus, and R. E. Dickerson**. 1996. Structure of the Escherichia coli Response Regulator NarL. Biochemistry **35**:11053–61.

5. **Bartkowski, W.** 2010. Defining the interdomain interface of the Escherichia coli response regulator NarL. UCLA, PhD Dissertation.

6. **Batchelor, J. D., M. Doucleff, C. Lee, K. Matsubara, S. De Carlo, J. Heideker, M. H. Lamers, J. G. Pelton, and D. E. Wemmer**. 2008. Structure and Regulatory Mechanism of Aquifex aeolicus NtrC4 : Variability and Evolution in Bacterial Transcriptional Regulation. Journal of Molecular Biology. Elsevier Ltd **384**:1058–1075.

7. **Bearson, S. M. D., J. a Albrecht, and R. P. Gunsalus**. 2002. Oxygen and nitrate-dependent regulation of dmsABC operon expression in Escherichia coli: sites for Fnr and NarL protein interactions. BMC microbiology **2**:13.

8. **Birck, C., L. Mourey, P. Gouet, B. Fabry, J. Schumacher, P. Rousseau, D. Kahn, and J. P. Samama**. 1999. Conformational changes induced by phosphorylation of the FixJ receiver domain. Structure (London, England : 1993) **7**:1505–15.

9. **Cohn, E., and J. Edsall**. 1943. Density and apparent specific volume of proteins. Proteins, Amino Acids and Peptides as Ions and Dipolar Ions (Edited by Cohn e.J. and Edsall J.T.). Reinhold Publishing Corporation, New York 370–81.

10. **Darwin, A. J., J. Li, and V. Stewart**. 1996. Analysis of nitrate regulatory protein NarL-binding sites in the fdnG and narG operon control regions of Escherichia coli K-12. Molecular Microbiology **20**:621–632.

11. **Darwin, A. J., K. L. Tyson, S. J. Busby, and V. Stewart**. 1997. Differential regulation by the homologous response regulators NarL and NarP of Escherichia coli K-12 depends on DNA binding site arrangement. Molecular microbiology **25**:583–95.

12. **Eldridge, A. M., H.-S. Kang, E. Johnson, R. Gunsalus, and F. W. Dahlquist**. 2002. Effect of phosphorylation on the interdomain interaction of the response regulator, NarL. Biochemistry **41**:15173–80.

13. **Gao, R., and A. M. Stock**. 2009. Biological insights from structures of two-component proteins. Annual review of microbiology **63**:133–54.

14. **Gao, R., and A. M. Stock**. 2010. Molecular Strategies for Phosphorylation-Mediated Regulation of Response Regulator Activity. Curr Opin Microbiol. **13**:160–7.

15. **Gautam, U. S., S. Chauhan, and J. S. Tyagi**. 2011. Determinants outside the DevR C-terminal domain are essential for cooperativity and robust activation of dormancy genes in Mycobacterium tuberculosis. PloS one **6**:e16500.

16. **Hastings, C. A., S.-Y. Lee, H. S. Cho, D. Yan, S. Kustu, and D. E. Wemmer**. 2003. High-resolution solution structure of the beryllofluoride-activated NtrC receiver domain. Biochemistry **42**:9081–90.

17. **Jarvis, M. R.** 1999. Functional analysis of the Escherichia coli nitrate response regulator, NarL. UCLA, PhD Dissertation.

18. **Keseler, I. M., J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñiz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp**. 2011. EcoCyc: a comprehensive database of Escherichia coli biology. Nucleic acids research **39**:D583–90.

19. **Laue, T., B. Shah, T. Ridgeway, and S. Pelletier**. 1992. Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins. Analytical Ultracentrifugation in Biochemistry and Polymer Science, S.E. Harding, A.J. Rowe and J.C. Horton ed. Cambridge, Great Britain, The Royal Society of Chemistry 90–125.

20. **Lee, S. Y., H. S. Cho, J. G. Pelton, D. Yan, E. a Berry, and D. E. Wemmer**. 2001. Crystal structure of activated CheY. Comparison with other activated receiver domains. The Journal of biological chemistry **276**:16425–31.

21. **Leoni, L., P. Ascenzi, A. Bocedi, G. Rampioni, L. Castellini, and E. Zennaro**. 2003. Styrene-catabolism regulation in Pseudomonas fluorescens ST: phosphorylation of StyR induces dimerization and cooperative DNA-binding. Biochemical and Biophysical Research Communications **303**:926–931.

22. **Li, J., S. Kustu, and V. Stewart**. 1994. In Vitro Interactions of Nitrate-responsive Regulatory Protein NarL with DNA Target Sequences in the fdnG, narG, narK, and frdA Operon Control Regions of Escherichia coli K-12. Journal of molecular biology **241**:150–65.

23. **Lin, A. V, and V. Stewart**. 2010. Functional roles for the GerE-family carboxyl-terminal domains of nitrate response regulators NarL and NarP of Escherichia coli K-12. Microbiology (Reading, England) **156**:2933–43.

24. **Maris, A. E., M. Kaczor-Grzeskowiak, Z. Ma, M. L. Kopka, R. P. Gunsalus, and R. E. Dickerson**. 2005. Primary and secondary modes of DNA recognition by the NarL two-component response regulator. Biochemistry **44**:14538–52.

25. **Maris, A. E., M. R. Sawaya, M. Kaczor-Grzeskowiak, M. R. Jarvis, S. M. D. Bearson, M. L. Kopka, I. Schröder, R. P. Gunsalus, and R. E. Dickerson**. 2002. Dimerization allows DNA target site recognition by the NarL response regulator. Nature structural biology **9**:771–8.

26. **Park, A. K., J. H. Moon, K. S. Lee, and Y. M. Chi**. 2012. Crystal structure of receiver domain of putative NarL family response regulator spr1814 from Streptococcus pneumoniae in the absence and presence of the phosphoryl analog beryllofluoride. Biochem Biophys Res Commmun **421**:403–7.

27. **Park, S., M. Meyer, A. D. Jones, H. P. Yennawar, N. H. Yennawar, and B. T. Nixon**. 2002. Two-component signaling in the AAA+ ATPase DctD: binding Mg 2+ and BeF 3-selects between alternative dimeric states of the receiver domain. The FASEB Journal **16**:1964–66.

28. **Partridge, J. D., D. F. Browning, M. Xu, L. J. Newnham, C. Scott, R. E. Roberts, R. K. Poole, and J. Green**. 2008. Characterization of the Escherichia coli K-12 ydhYVWXUT operon: regulation by FNR, NarL and NarP. Microbiology **154**:608–618.

29. **Schröder, I., C. D. Wolin, R. Cavicchioli, and R. P. Gunsalus**. 1994. Phosphorylation and dephosphorylation of the NarQ, NarX, and NarL proteins of the nitrate-dependent two-component regulatory system of Escherichia coli. Journal of bacteriology **176**:4985–92.

30. **Schröder, I., S. Darie, and R. P. Gunsalus**. 1993. Activation of the Escherichia coli Nitrate Reductase (narGHJI) Operon by NarL and Fnr Requires Integration Host Factor. Journal of Biological Chemistry **268**:771–4.

31. **Schrödinger, L.** The PyMOL Molecular Graphics System. Version 1.3.

32. **Stewart, V., and P. J. Bledsoe**. 2008. Substitutions at Auxiliary Operator O3 Enhance Repression by Nitrate-Responsive Regulator NarL at Synthetic lac Control Regions in Escherichia coli K-12. Journal of bacteriology **190**:428–33.

33. **Toro-Roman, A., T. R. Mack, and A. M. Stock**. 2005. Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face. Journal of molecular biology **349**:11–26.

34. **Vannini, A., C. Volpari, C. Gargioli, E. Muraglia, R. Cortese, R. De Francesco, P. Neddermann, and S. Di Marco**. 2002. The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. The EMBO journal **21**:4393–401.

35. **Walker, M., and J. DeMoss**. 1994. NarL-phosphate must bind to multiple upstream sites to activate transcription the narG promoter of Escherichia coli. Molecular Microbiology **14**:633–41.

36. **Wang, H., and R. P. Gunsalus**. 2000. The nrfA and nirB Nitrite Reductase Operons in Escherichia coli Are Expressed Differently in Response to Nitrate than to Nitrite. Journal of Bacteriology **182**:5813–5822.

37. **Webber, C. A., and R. J. Kadner**. 1997. Involvement of the amino-terminal phosphorylation module of UhpA in activation of uhpT transcription in Escherichia coli. Molecular Microbiology **24**:1039–48.

38. **Yan, D., H. S. Cho, C. a Hastings, M. M. Igo, S. Y. Lee, J. G. Pelton, V. Stewart, D. E. Wemmer, and S. Kustu**. 1999. Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. Proceedings of the National Academy of Sciences of the United States of America **96**:14789–94.

39. **Zhang, R., K. M. Pappas, J. L. Brace, P. C. Miller, T. Oulmassov, J. M. Molyneaux, J. C. Anderson, J. K. Bashkin, S. C. Winans, and A. Joachimiak**. 2002. Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. Nature **417**:971–4.

Chapter 4

Analysis of Two NarL Structures Reveal Insights into the Domain Interface and Activation Mechanism

Introduction

Phosphorylation of response regulators (RRs) of two-component signal-transduction systems occurs on an invariant aspartate located in the receiver domain (Chapter 1 and (15)). This "activated" state is accompanied by movements in the vicinity of the active site and in distant protein regions, the latter usually leading to specific RR function.

Receiver domain activation is often marked by the movements of two conserved "switch" residues near the active-site: a Ser/Thr located at the end of strand β4, and the more distant Tyr/Phe, in strand β5 (Chapter 1, and (9)). The Ser/Thr shifts towards the active site in order to form a hydrogen bond between its hydroxyl group and a phosphate-oxygen of the phosphorylated aspartate. This occurs with the simultaneous flip of the Tyr/Phe residue from an exposed to a buried rotamer. The motion of this aromatic switch residue affects RR function by facilitating dimerization or enhancing binding to other proteins (23, 33). In line with the movements of the switch residues, the β4-α4 loop becomes displaced (18, 22).

The changes in the switch residues and β4-α4 loop have given rise to different allosteric models associated with receiver domain activation. The widely accepted "two-state equilibrium" model (31, 35) involves the pre-existence of a small population of activated proteins within the larger population of inactivated protein molecules, where phosphorylation serves to shift the equilibrium to the activated conformation. This description is the simplest form, as a growing number of structures indicate that besides the designated "active" or "inactive" states, other subpopulations also exist. Structures of inactivated *E. coli* CheY, either alone or peptide-bound, show variable conformational combinations of the switch residues (10, 17, 23, 30). Structures from the OmpR/PhoB family show active-like properties of receiver domains, such as dimerization and a flipped aromatic switch residue, in the absence of a phosphorylating agent (6,

7, 34). Receiver domains in these forms are thought to represent an intermediate, or "meta-active", state between active and inactive conformations. The number of intermediates and the extent to which they are populated is unknown and may differ between different response regulators.

Two crystal structures of the *E. coli* response regulatory NarL were elucidated from two distinct crystal forms: one of an orthorhombic lattice (4) in the I222 space group (called $NarL_O$) and the other of a monoclinic lattice (3) having space group C121 (called $NarL_M$) . Both full-length structures have the same protein fold, with an RMSD of 0.574 (28) and corroborate the necessity of a severed domain interface to allow the C-terminal domain to bind DNA. However, the two structures show previously unreported variations in residue positions and rotameric conformations, including in the switch residues of the N-terminal receiver domain. In $NarL_O$, the switch residues, Ser87 and Tyr106, are exposed, while in $NarL_M$ these two residues are buried (Figure 4-1a). In addition, the β4-α4 loop in $NarL_M$ is slightly shorter (Figure 4-1b), and has a greater resemblance to the equivalent loop in activated CheY (22) than to the same loop in $NarL_O$. Tyr106 is also secured by a hydrogen bond with Ser89, as in the hydrogen-bond comprised of the equivalent residues in activated-CheY. A fully activated NarL would entail a separation of the N- and C-terminal domains (4, 12, 37). Therefore, $NarL_M$ as a whole is not activated, but represents an intermediate state, while $NarL_O$ represents the inactivated form or an intermediate closer to it.

In the unphosphorylated form of NarL, the DNA-recognition helix is fastened to the receiver domain by an extensive interface, which involves loops in the N-terminal domain that are located 20Å away from the active-site loops (Figure 4-2a). A network is formed between residues from the C-terminal domain (mainly lysines) and the protruding interface-loop residues,

these being Asp 52, Asp54, Ser78, and Ser80 (Figure 4-2b). Despite this nexus of contacts, a single mutation at the interface can pry the domains apart. This has been demonstrated by isothermal titration calorimetry (ITC) (5) where certain mutations compromised or precluded the separated N- and C-terminal domains from joining, and by *in vivo* expression analysis where certain single mutations resulted in constitutively-active phenotypes (19). This was postulated to be caused by compromised alignment or a loss of too many residue-residue contacts. In addition, dissociation constants calculated by NMR (12) and ITC (5) showed that the interface is rather weak. A weak interface makes sense for a protein that needs to open and close at any given moment. But what makes the domain interface structurally weak in the first place? Bartkowski (5) thoroughly detailed the disrupted residue contacts that would be incurred by individual mutations, concluding that the interface is maintained by a relatively small number of charge-to-charge interactions and the absence of contacts created by one mutation weakens the interface to the point of no return. The two NarL structures show alternative interface contacts, implying that the interface has plasticity and toggles between different equilibrium states. Thus, besides the loss of immediate contacts, the effect of residue motions may also play a role in the mutational sensitivity of the interface.

NarL activation involves at least two molecular regions, the active-site and the interdomain interface. In the first part of this study, a computational analysis of both NarL interfaces is compared to biochemical findings. Residues vital to sustaining the interface are identified based on the extent to which they become buried upon interface formation. These matched experimental findings, and also identified other potentially important residues, including a marker for activation in the C-terminal domain. In addition, residue mobilities and fluctuation correlations were used to explain the interface sensitivity observed in mutation

studies. In the second part, further movements in the active-site loop residues of $NarL_M$ are proposed that would accompany full activation.

Methods

Domain Interface Properties. The domain interface of both NarL structures (PDB IDs: 1RNL and 1A04) was analyzed with the Protein Interfaces, Surfaces, and Assemblies (PISA) web-server (21) using input files where the N- and C-terminal domains (residues 5-142 and 155-216, respectively, in $NarL_O$; residues 5-142 and 150-216, respectively, in $NarL_M$) were treated as separate proteins.

Mobility Ratios and Fluctuation Correlations. Missing residues in both NarL structures were rebuilt using SWISS-MODEL Workspace (1); these are residues 142-154 in $NarL_O$ (PDB ID: 1RNL) and residues 143-149 in $NarL_M$ (PDB ID: 1A04). The rebuilt structures were used to calculate residue mobility using the program Vibe (27). Residue mobility is determined based on residue center-of-mass and inter-residue contacts. A mean-squared fluctuation is calculated for each residue and then normalized to the average mean-squared fluctuation of residues in the N-terminal domain; this value is the residue "mobility ratio."

Structural Representations. All structural images were created with Pymol (29).

Results and Discussion

The NarL Interdomain Interface

The Interface Properties of $NarL_O$ and $NarL_M$ are Similar

Both the orthorhombic and monoclinic NarL crystal structures were analyzed with the Protein Interfaces, Surfaces, and Assemblies (PISA) web-server (21) using input files where the

N- and C-terminal domains were treated as separate proteins. The results showed properties of the interface, and associated residues, as the domains conjoined according to the x-ray crystal structures (Table 4-1). In both structures, 60 residues comprise the interface, burying an area of 1026Å^2 in NarL$_O$ and 967 Å^2 in NarL$_M$. Both areas lie within the standard for a heterocomplex (2, 20), with the interface of NarL$_M$ being slightly smaller. The joining of the domains results in a negative energy of solvation (which is a result of hydrophobic contacts): -6.3 kcal/mol for NarL$_O$ and -7.8 kcal/mol for NarL$_M$. These negative, and thus more favorable, values is in agreement with ITC studies (5) which showed that the uniting of domains is favorable and entropically driven. The structure of NarL$_M$ seems to have a slightly higher affinity with respect to hydrophobic interactions. NarL$_O$, however, has more polar contacts, as seen in the number of hydrogen bonds and salt bridges (Table 4-1).

Overall, the interfaces between NarL$_O$ and NarL$_M$ are very similar, although the smaller interface in NarL$_M$ and fewer polar contacts are consistent with being closer to an activated state. The small differences between NarL$_O$ and NarL$_M$ are expected in an allosteric model where the protein is sampling different conformational states until activation shifts the equilibrium to one particular state (13). In this paradigm, initial and final states of a molecule or domain may be similar, and therefore so would the intermediates.

Identification of Important Interface Residues in NarL$_O$ and NarL$_M$

The percent buried-surface-area (% BSA), or the extent to which a residue becomes buried as a consequence of domain binding (i.e. interface formation), identified residues that are likely to be imperative to sustaining the interface in NarL$_O$ and NarL$_M$. Residues with larger %BSA values are thought to contribute more to domain interface interactions (20). In the C-

terminal domain, these are located in the α7- α8 loop, along helix α9, and the end of the α9-α10 loop leading into the middle of helix α10 (Figure 4-3). In the N-terminal domain, they are expectedly located in the three interface loops (α2-β3, α3-β4, and α4-β5) and in one surprising outlier residue, Ala7 in strand β1. The %BSA values for the α5-α6 loop and helix α6 should be judged with caution since they were included in a detached N-terminal domain having the structure shown in inactivated NarL. In reality, these regions, which are part of the "linker region" that holds the two domains together during activation, are likely to separate from the N-terminal domain during activation and therefore the calculated %BSA for this region is probably not accurate.

The interface regions identified by PISA, in both $NarL_M$ and $NarL_O$, are consistent with those identified by NMR (12). They are also consistent with specific residues identified by other biochemical experiments (5, 19, 37), these being Asp52, Asp54, Ser78, Ser80, Arg82, Leu99, Arg174, Lys188, Lys196, and Val204. With the exception of Val204, these residues, when mutated, either render the protein constitutively active, or compromise and even preclude the two domains from joining in ITC studies (5). Val204 was shown by EPR studies (37) to experience a large increase in mobility upon phosphorylation, in agreement with its obliged transition from a residue involved at the interdomain interface to one being involved in a dimer interface between the two C-termini (25). Based on the %BSA values or their location in the NarL interface, other potentially vital interface-residues that are worth experimental validation include Ala7, residues in loop α4-β5 (Gly102, Ala103, Asp104), residues in the α7-α8 loop (such as Gly170), Val191, Ser202, Arg203, and Val191. Gly102 and Ala103 are especially interesting since they are strongly conserved in the CheY superfamily (32, 36).

$NarL_O$ and $NarL_M$ show a comparable pattern of deeply buried residues (Figure 4-3), and therefore both interfaces rely on the same group of integral residues of which they are comprised. Again, this reflects the resemblance between different protein populations in equilibrium. A few residues, however, show a remarkable discrepancy in the %BSA between the $NarL_O$ and $NarL_M$ structures, which may provide insight into the NarL activation mechanism (discussed below).

Lys196 and Other Possible Markers of Activation

A few residues show striking contrast in their %BSA between $NarL_O$ and $NarL_M$, salient examples being Glu184 and Lys196 (Figure 4-3). The change in solvent accessibility of these residues may be associated with $NarL_M$ being in an intermediate state prior to activation.

Glu184 experiences a significant increase in its %BSA, from 5.2% in $NarL_O$ to 58.3% in $NarL_M$. In both structures, this residue forms a salt bridge with Lys174, though uses different side-chain aspartate-oxygens. In the structure of the detached C-terminal domain ($NarL^C$) bound to DNA, however, this salt bridge is absent, and is replaced with a salt bridge between Glu184 and Lys188 (24), the latter being a DNA-binding residue. Though Glu184 does not form a salt bridge with Lys188 in neither NarL structure, the Lys174 residue in which it stabilizes does show significant changes. Lys174 experiences a 1Å shift accompanied with a change in %BSA from 60% in $NarL_O$ to 37% buried in $NarL_M$ (Figure 4-4a). In the process, its molecular contacts change. Lys174 loses a salt bridge to Asp54 in $NarL_O$, but in $NarL_M$ it makes a new hydrogen bond between its side-chain NZ and the main-chain carbonyl of Glu5. For this to happen, Glu5 undergoes a rotameric flip. The significance of these changes in relation to NarL activation is not apparent and remains to be determined. The role of Glu184 in $NarL_M$ may be to stabilize the overall conformational changes in this region as the protein prepares for activation.

In contrast, the movements of Lys196, located at the N-terminal end of the recognition helix, are clearly indicative of a state closer to activation. In $NarL_O$, the Ser78 hydroxyl, along with the main-chain carbonyls of Glu76 and Lys77, hydrogen bond to NZ of Lys196 (Figure 4-4b, left). In $NarL_M$, however, all three hydrogen bonds are absent (Figures 4-4b, right). Lys196 also experiences a 1Å shift between the two structures and its %BSA changes considerably, from being 47% buried in $NarL_O$ to being 7% buried in $NarL_M$. This movement creates a distance too far to make contacts with residues in the α3-β4 loop. Lys196 and Ser78 were targets of EPR studies (37), which demonstrated that the two residues separate by at least 20Å upon phosphorylation. Thus, the movement of Lys196 is suggestive of a conformation closer to activation. To confirm this hypothesis, three $NarL^C$-DNA structures (24) were overlaid with the C-terminal domain from $NarL_O$ and $NarL_M$, and the conformations of all the lysines were compared (Figure 4-5). There is a resounding consensus in the position of Lys196 in $NarL_M$ and the $NarL^C$ structures, but not with Lys196 from $NarL_O$. Lys196 in $NarL_M$, which does not make contacts with the N-terminal domain, is shifted by almost 2Å relative to the location of Lys196 in $NarL_O$. If not for the repositioning, it would clash with the DNA backbone. Thus, the movement of Lys196 in $NarL_M$, compared to its position $NarL_O$, is indicative of a conformation associated with activation.

Residue Mobilities and Fluctuation Correlations may also Influence the Phenotypes caused by NarL Interface Mutations

Mutation analyses demonstrate that certain interface substitutions result in a constitutively active NarL in expression assays or compromise the N- and C-terminal domains from rejoining in ITC (5, 19). The most impactful interface residues were shown to be Asp52,

Asp54, Ser78, Ser80, and Arg82 of the N-terminal interface-loops, and Lys174 and Lys188 of the C-terminal domain. A thorough look into the lost hydrogen bonds, salt bridges, and van der Waals interactions upon different substitutions has been detailed (5). But since the interface is dynamic and toggles at least between two equilibrium states, residue fluctuations and correlations may also play a role in the mutational phenotypes observed.

Mobility ratios, determined by residue center-of-mass and inter-residue contacts, of residues in the N-terminal domain of both NarL structures were calculated using the program Vibe (27) (Figure 4-6) . At the α2-β3 loop, which includes Asp52 and Asp54, all residues are weakly fluctuating, as indicated by their mobility ratios being below 0.7 (26, 27). The α3-β4 loop has some residues that are moderately fluctuating, having mobility ratios between 0.7-1, such as Ser78, which is consistent with it being located in a less networked region (Figure 4-4b). That most of the interface residues surrounding the recognition helix have low to medium mobility expresses the denser network in this region. The other extreme is shown by the much higher mobilities of the active site loops (β3-α3, β4-α4, and β5-α5), which are not bound by an interface.

Several of the residues in the mutation studies were substituted with alanine or glycine. Glycine followed by alanine are the two most fluctuating residues found in loop regions, with the most propensity to elicit protein disorder (26). Indeed, in the interface loops of NarL, Gly81, Gly102, and Gly105 are highly fluctuating, having mobility ratios greater than one, while Ala103 is moderately fluctuating. A glycine or alanine substitution in place of a key interface residue, such as Asp52, would increase fluctuations at that site in addition to abolishing several polar contacts. The profoundness of the effects of increased residue mobility would depend on solvent accessibility and the number of neighbors in the region (27).

In addition to changes in residue mobilities, mutations can also impact the movements of correlated residues, which may not be in direct contact with the mutated amino-acid. This was evaluated by using Vibe to determine residue fluctuation correlations. Table 4-2 shows fluctuation correlations in $NarL_M$ (similar correlations were seen for $NarL_O$) of sample residues that lead to drastic phenotypes in mutation studies. These residues are correlated to several other interface residues or to the DNA binding residues, and sometimes to both. A mutation in Asp52, for example, affects the motions of downstream residues, residues in the neighboring α3-β4 loop, and residues in the recognition helix (Figure 4-7a). Similarly, mutations in Ser80 affect the movements of several residues in the recognition helix (Figure 4-7b). Arg82 mutations, though also constitutively active, were found to be less severe. This may be because Arg82 does not correlate with any DNA binding residues (Lys188, Val189, and Lys192). The K174A mutation, which did not completely abolish domain binding in ITC studies, may be explained by Lys174 being correlated to only two interface residues (Table 4-2 and Figure 4-7c). However this mutant behaves in a constitutive manner at the *frdA* promoter (Gunsalus *et al.*, unpublished) perhaps because it is correlated to Lys188. These data demonstrate that mutations of certain interface residues not only abolish immediate polar contacts, but may also pose consequences to the motions of correlated residues that contribute to interface destabilization. In addition, the interdependence of residue motions likely plays a role in the activation mechanism.

Proposed Residue Movements in the Active Site upon Activation

Further movements are expected to lead $NarL_M$ from a partially activated to a fully activated state. These conformational changes may be predicted by comparing the active site of both NarL structures, along with structures of other activated response regulators.

The β1-α1 Loop

The active site of NarL contains a slew of highly conserved residues found in canonical receiver domains (9), of which include two highly conserved aspartates (Asp13 and Asp14 in NarL) at the end of strand β1. Activated receiver-domain structures show that the two aspartates must face towards the phosphorylatable aspartate in order to carry out their role of stabilizing the Mg^{2+} ion that is required for activation (9, 22). In both NarL structures Asp13 and Asp14 are facing away from the active site (Figure 4-8a), and therefore are expected to rotate. These movements are likely to be broadly impactful since correlation data show that these aspartates influence the movements of residues throughout the N-terminal domain as well as the recognition helix (not shown).

The β3-α3 Loop

Asn61, located in the β3-α3 loop, experiences a rotameric conformational change between the two NarL structures, concomitant with several polar contact changes (Figure 4-8b). The rotameric flip in Asn61, looking from $NarL_O$ to $NarL_M$, abolishes the three hydrogen bonds made between its side-chain amide and oxygens from Phe86, Ser87, and Asp59 (the site of phosphorylation). Instead, new hydrogen bonds are formed between the Asn61 side-chain carbonyl and the backbone amide of Val88, between the Asn61 side-chain carbonyl and the side-chain NZ of Lys109, and between the Asn61 side-chain ND2 and the hydroxyl side-chain of Ser87. These three hydrogen bonds are reminiscent of those formed by the oxygens (fluorides) of the phosphoryl (beryllofluoride) group (22), suggesting that Asn61 may act as a temporary place-holder for the phosphate group before activation. Consequentially, access to the nucleophilic Asp59 oxygen is hindered. Either upon metal binding or upon phosphorylation,

Asn61 in its $NarL_M$ conformation, will likely relocate. This could be facilitated by the relatively high mobility of downstream loop residues, such as Pro63, Gly64, and Met65 (Figure 4-6).

The β4-α4 Loop

In structures of activated receiver domains, the Ser/Thr switch-residue, located at the beginning of the β4-α4, bonds to a phosphate-oxygen (or fluoride) of the phosphorylated (or beryllofluoride-activated) aspartate (8, 22). Although the switch Ser87 residue of NarL shows movement between $NarL_O$ and $NarL_M$, it is still too far away from Asp59 compared to this distance in other activated receiver domains. In beryllofluoride-activated CheY (22) the distance between the hydroxyl oxygen of its switch Thr87 residue and the OD1 of its phospho-aspartate (Asp57) is 4.2 Å (Figure 4-8c). This distance in $NarL_M$, being 6.8Å, is more similar to that in meta-activated CheY (30), which is 6.5Å. Therefore, Ser87 is likely to move closer to Asp59 upon phosphorylation. In doing so, additional movements to the β4-α4 loop are expected, and possibly also to strand β5, of which Ser87 has two hydrogen bonds. The combined movements may lead to the disruption of the hydrophobic region created by helices α4 and α6 and to the subsequent availability of the α4-β5-α5 region, a common region for receiver domain dimerization (Chapter3, and (16).

In general, the repositioning of the β4-α4 loop plays a large role in receiver domain activation. Crystallographic structures and molecular dynamic studies point to a coupled motion between the β4-α4 loop and the Tyr/Phe switch residue, rather than between the Ser/Thr and Tyr/Phe switch residues (10, 14). In these studies, the β4-α4 loop was shown to have conformational flexibility, which is congruent with the high mobility-ratios of residues in this loop (Figure 4-6). Perhaps the flexibility the β4-α4 loop facilitates the toggling between

different equilibrium states (in an equilibrium-shift model of allostery). Upon phosphorylation, the Ser/Thr serves to stabilize this loop in the activated conformation by its hydrogen bond to a phosphate-oxygen (14, 22). In NarL, this idea is supported by a V88A NarL mutant that is constitutively active *in vivo* (11). Val88 immediately follows the Ser87 switch residue, and in most RRs this position is usually occupied by a small residue in order to facilitate the movement of Ser/Thr (9, 36). Val88 is already highly fluctuating (Figure 4-6) and an alanine substitution would increase the mobility at this site, thus facilitating the Ser87 turn more readily and perhaps rendering the activated conformation preferable.

Conclusions

A comparison of the two NarL structures was presented with a focus on the domain interface and active site. The domain interfaces between the two structures have similar physical properties, supporting the idea that proteins toggle between closely related states. Differences in percent buried-surface-area confirmed or identified important interface residues, and Lys196 was identified as a marker of activation. Some residues in the N-terminal interface loops and recognition helix become highly buried upon domain binding and are engaged in a close network of polar contacts. This is confirmed by the low relative mobilities of loop residues at the interface. Mutating residues in key network locations not only abrogates polar contacts but also alters residue mobility and affects the motions of correlated residues. Thus, residues have propagating effects that may contribute to the mutational sensitivity of the interface and to activation in general.

Full activation of $NarL_M$ is predicted to be accompanied by further movements in the active site loops. Asp13 and Asp14, in the β1-α1 loop will be required to flip towards the active

site. Asn61, which seems to be a temporary place holder for the phosphoryl group, will likely reposition along with other β3-α3 loop residues. Ser87 is expected to shift closer to phospho-Asp59, which will probably affect residues in loops β3-α3 and β4-α4, both of which contain highly mobile residues. Should these combined movements occur they may also induce fluctuations in correlated residues, with the signal eventually resulting in a severed interface.

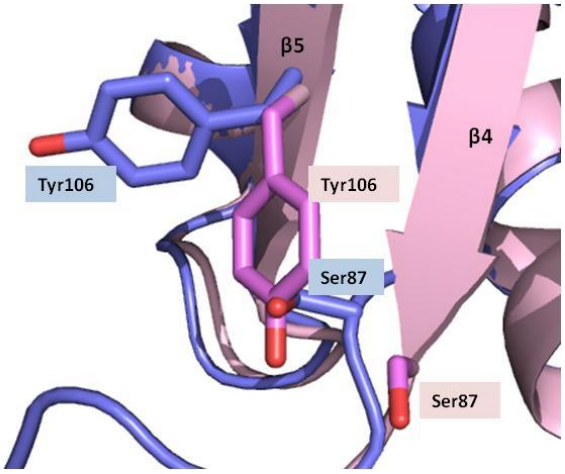
The domains of NarL work together in a fascinating interplay. The controlled architectural transition begins with active-site phosphorylation and propagates to the interface. Based on other activated receiver domains, cation binding followed by phosphorylation would cause the flexible, acidic active-site near Asp59 to become more networked, ordered, and the loops near the active site to be more stabilized. In contrast, the interface becomes more disordered as the unbridled C-terminal domain exposes several basic residues, and the N-terminal interface loops become unbound. Each domain exercises an equally important role that ultimately leads to controlled DNA binding and transcriptional regulation.

Acknowledgements

Jeffrey H. Zhang, Anatoly Ruvinsky, Michael R. Sawaya, Woytek Bartkowski, Wayne L. Hubbell, and Robert P. Gunsalus.

Figure 4-1. The switch residues of the $NarL_M$ receiver domain are indicative of activation. **(a)** Ser87 and Tyr106 change from being exposed in $NarL_O$ (blue, PDB ID: 1RNL) to being buried in $NarL_M$ (pink, PDB ID: 1A04). Note that the conformation of Ser87 of $NarL_O$ would clash with Tyr106 of $NarL_M$, suggesting that Tyr106 cannot flip to a buried conformation unless Ser87 moves out of the way. **(b)** The two NarL structures (same color scheme) overlapped with activated CheY (purple, PDB ID: 1FQW). The structure of the $NarL_M$ β4-α4 loop shows more similarity to the same loop in activated CheY than in $NarL_O$. In addition, both $NarL_M$ and activated CheY have a hydrogen bond made between an equivalent residue in the β4-α4 loop and the switch Tyr residue.

(a)



(b)

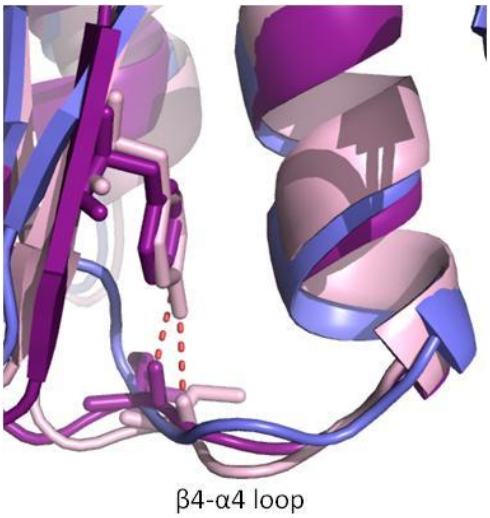
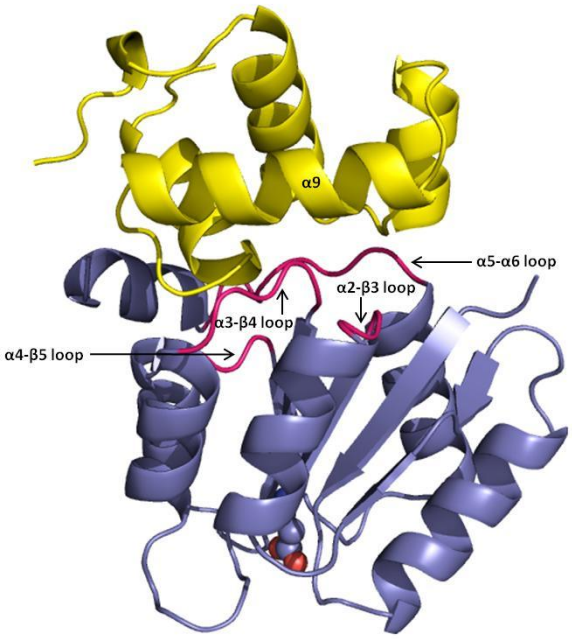


Figure 4-2. The interdomain interface of NarL. **(a)** Full-length NarL (PDB ID: 1A04) showing the N-terminal (blue) loop regions (magenta) in close contact to the C-terminal domain (yellow), especially at the recognition helix (α9). The phosphorylatable aspartate is shown in spheres. **(b)** The same color scheme emphasizing the residues that secure the DNA recognition helix.

(a)



(b)

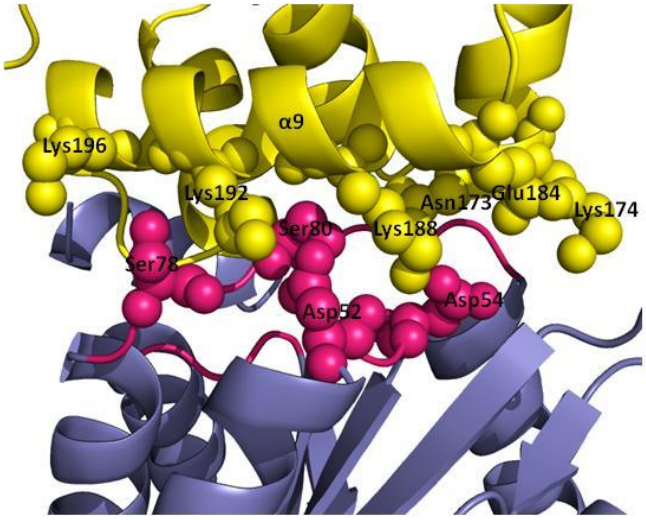


Figure 4-3. The percent buried-surface-area of interface residues in $NarL_M$ and $NarL_O$. Secondary structure regions are labeled in boxes. Residues with a higher percent buried-surface-area are indicative of being important to sustaining the interface.

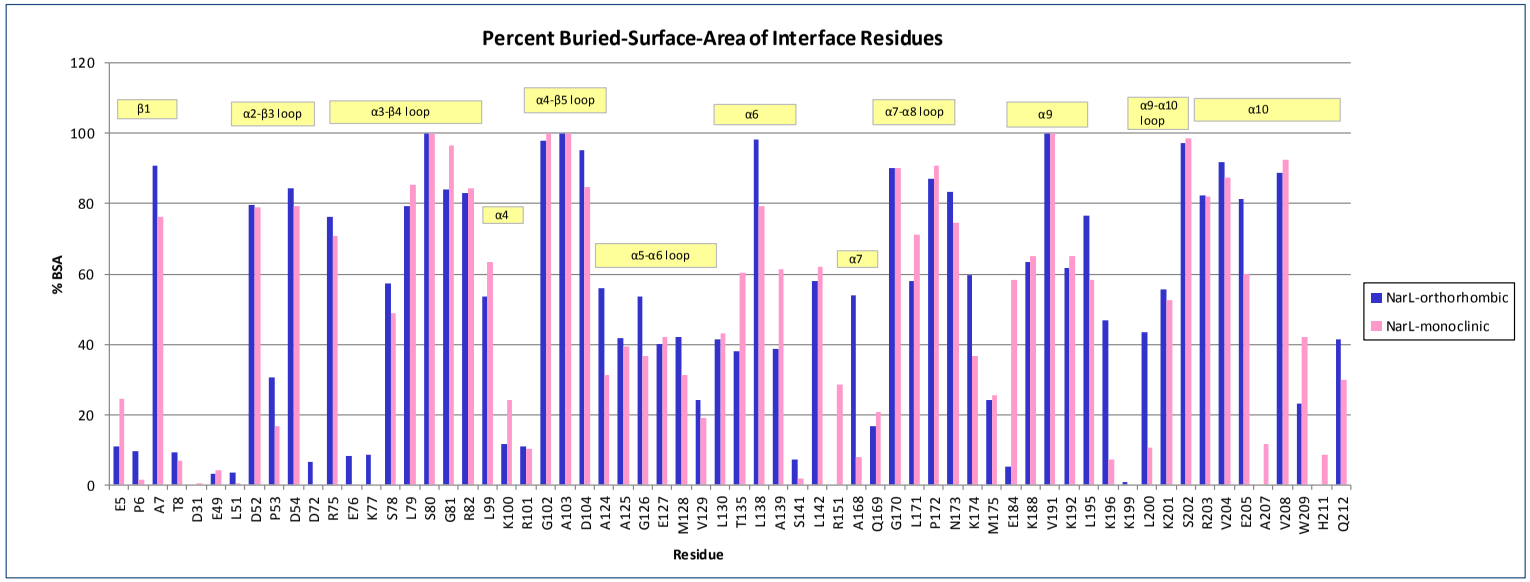
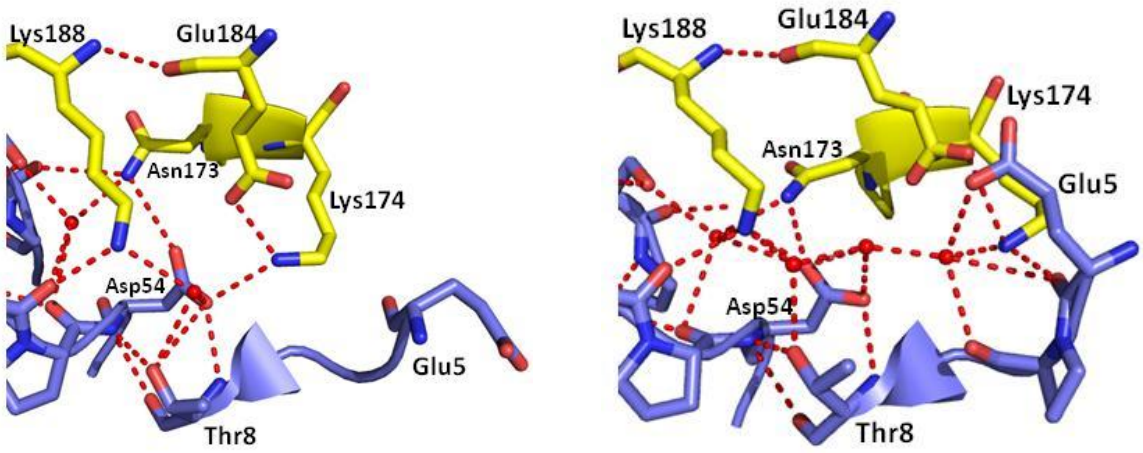


Figure 4-4. Changes in the solvent accessibility and polar contacts of Glu184 and Lys196. **(a)** The polar contact modifications as Glu184 loses solvent accessibility from $NarL_O$ (left, PDB ID: 1RNL) to $NarL_M$ (right, PDB ID: 1A04). Lys174 shifts and Glu5 experiences a rotameric flip. N-terminal residues are colored blue, and C-terminal residues are colored yellow. For all stick representations, oxygen atoms are red and nitrogen atoms are dark blue. Water molecules are represented by red spheres and polar contacts by red dashed lines. **(b)** Changes in hydrogen bonding (same color scheme) made by Ser78 and Lys196 in $NarL_O$ (left, PDB ID: 1RNL) and $NarL_M$ (right, PDB ID: 1A04), which serve to increase the solvent accessibility of Lys196 in $NarL_M$.

(a) $NarL_O$ $NarL_M$



(b) $NarL_O$ $NarL_M$

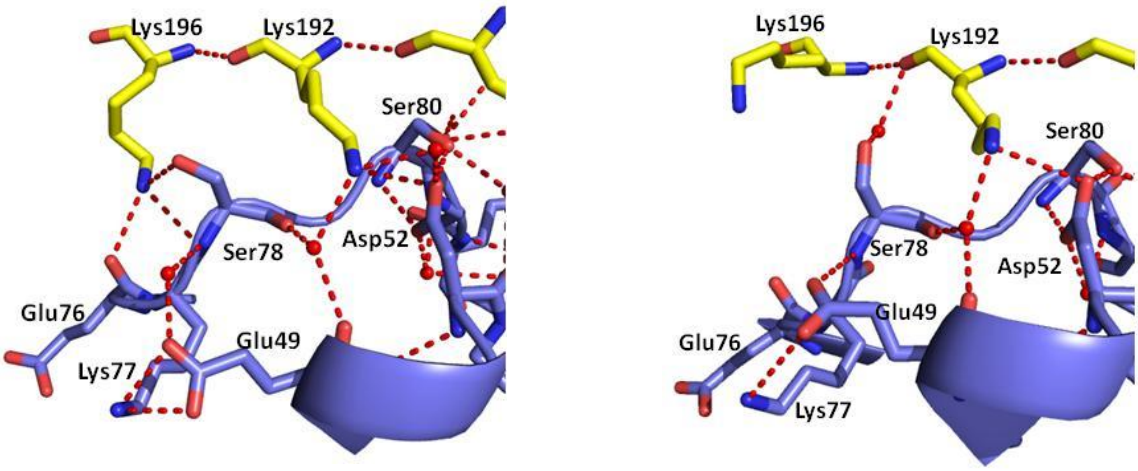


Figure 4-5. The conformation of Lys196 in $NarL_M$, as compared to its position in $NarL_O$, is a marker of activation. Four $NarL^C$ molecules from the structure of $NarL^C$ bound to the *nirB* -74/-74 promoter (PDB ID: 1JE8) were overlapped with $NarL^C$ monomers from two other $NarL^C$-DNA structures (PDB IDs: 1ZG1, 1ZG5), and with $NarL^C$ from $NarL_O$ (pink, PDB ID: 1RNL) and $NarL_M$ (yellow, PDB ID: 1A04). There is a consensus of the rotameric conformation of all the Lys196 residues except for the striking contrast of this residue in $NarL_O$ (circled for emphasis). In its current conformation, Lys196 of $NarL_O$ would clash with the DNA (orange backbone).

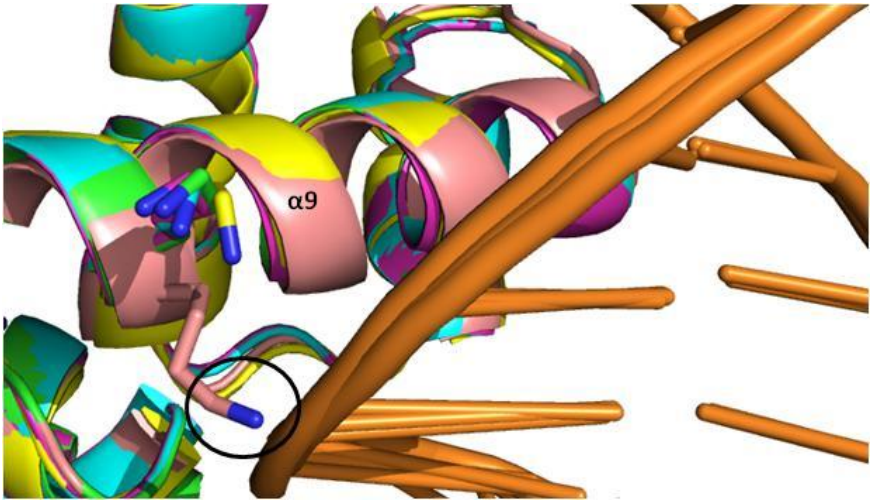


Figure 4-6. The mobility ratios of loop residues in the receiver domains of $NarL_M$ and $NarL_O$. Loop regions are labeled in boxes above the bars. Interface loops (orange boxes) are α2-β3, α3-β4, and α4-β5. Active-site loops (light yellow boxes) include β3-α3, β4-α4, and β5-α5. The site of phosphorylation (D59) is emphasized with a green dot.

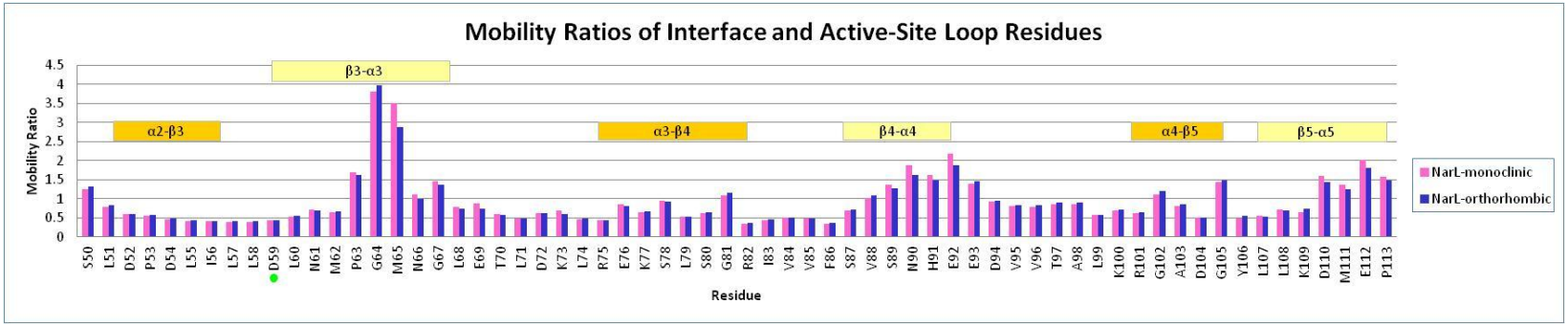
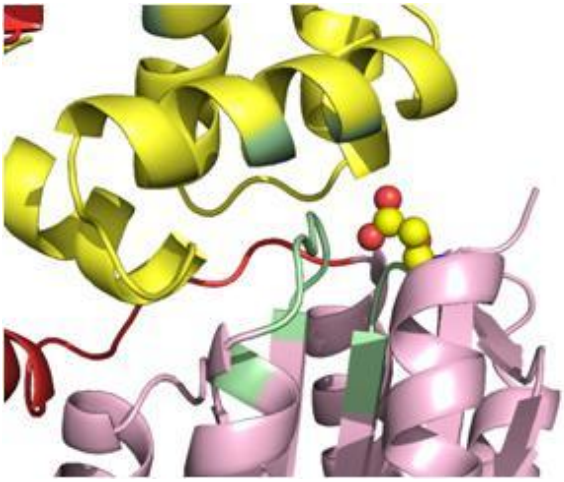
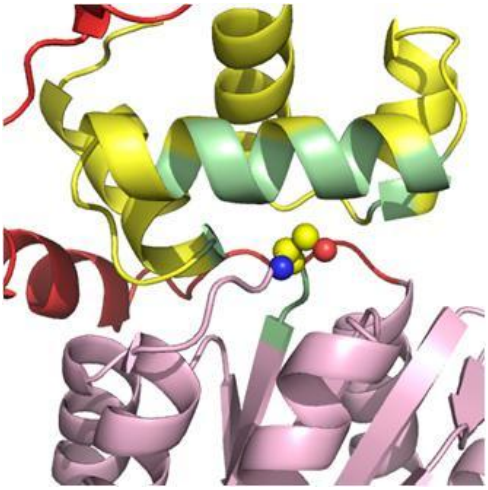


Figure 4-7. Correlation fluctuations of different interface residues. **(a)** Asp52, **(b)** Ser80, and **(c)** Lys174. In each picture, the receiver domain is colored pink, the C-terminal domain in yellow, and linker region in red. The residue of interest is shown in spheres and the residues correlated to it are highlighted in light green.

(a)



(b)



(c)

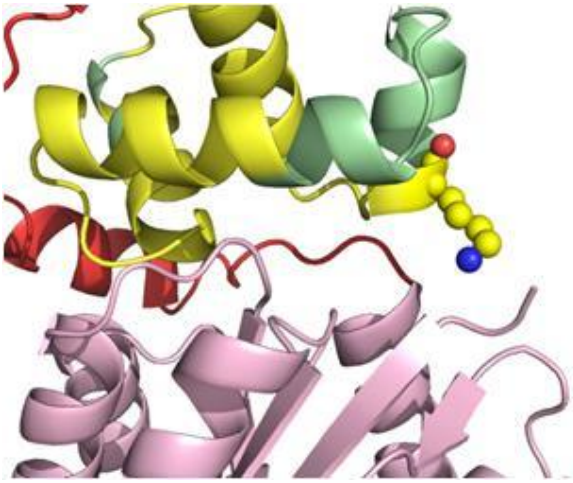
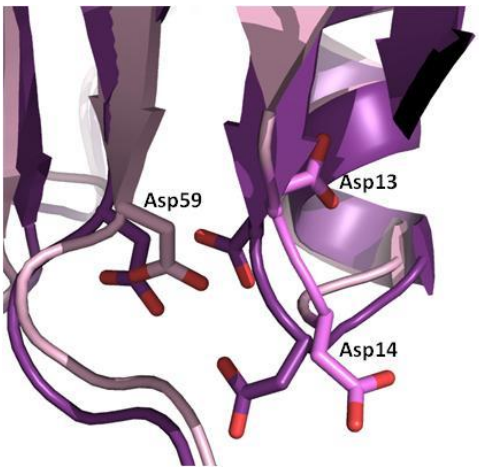
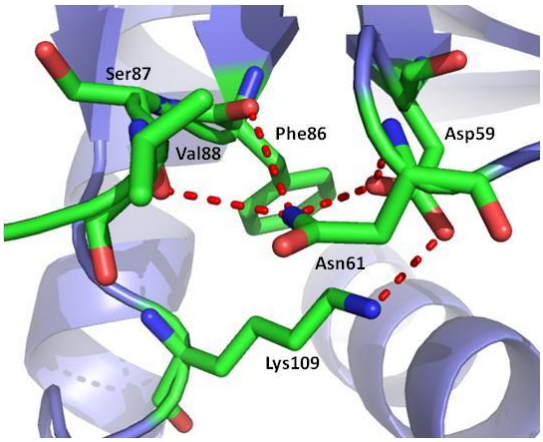


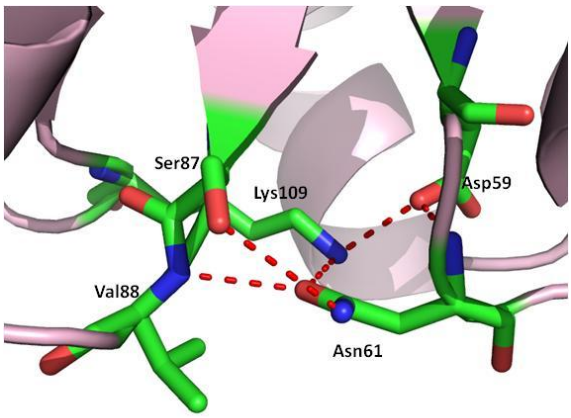
Figure 4-8. Proposed movements in $NarL_M$ upon full activation. **(a)** The conserved Asp residues (stick models) of the β1-α1 loop in $NarL_M$ (pink) and in activated CheY (purple). Asp13 and Asp14 (NarL numbering) of $NarL_M$ will need to shift towards the active-site, Asp59, as in the equivalent Asp residues in activated CheY. **(b)** The contacts in Asn61 in $NarL_O$ (blue, left) as opposed to $NarL_M$ (pink, right). This residue will likely be displaced upon phosphorylation, as will be the β3-α3 loop in which it is located. **(c)** The distance between Ser87 and Asp59 in $NarL_M$ (pink), and equivalent residues in activated CheY (purple) and meta-active CheY (cyan). Upon phosphorylation, Ser87 in $NarL_M$ is expected to move closer to Asp59 by at least 2Å. (PDB IDs: $NarL_O$, 1RNL; $NarL_M$, 1A04; beryllofluoride-activated CheY, 1FQW; and meta-activated CheY, 1JBE).

(a)



(b)





(c)

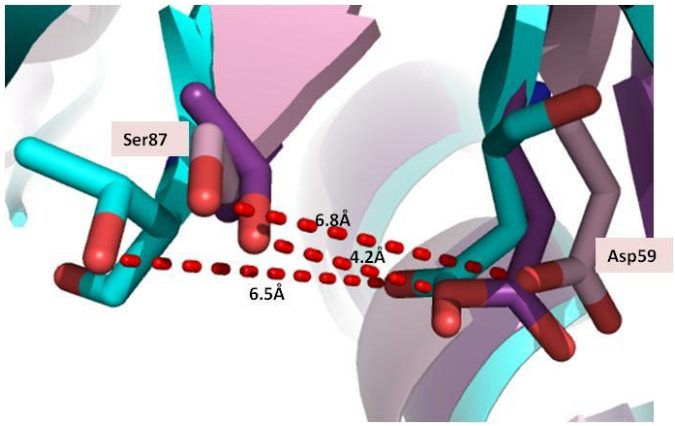


Table 4-1. Properties of the $NarL_O$ and $NarL_M$ interfaces.

	$NarL_O$	**$NarL_M$**
Number of Residues	60	60
Surface Area (Å^2)	1025.7	966.7
Gain in Solvation Energy of Complex (kcal/mol)	-6.3	-7.8
Number of Hydrogen Bonds	12	7
Number of Salt Bridges	3	2

Table 4-2. Residue correlation fluctuations of interface residues that have been mutated in ITC or *in vivo* expression studies. Mutations listed led to the inability or weakened ability of the domains to conjoin in ITC (5), or led to constitutive activity in transcriptional regulation (19). Correlations shown occur in $NarL_M$; positive (negative) correlations indicate that movement of the designated residue elicits a movement in the same (opposite) direction of the correlated residue. Correlated residues are divided between interface and non interface, and those in bold are directly involved in DNA-binding.

Residue	Mutation(s)	Phenotype	Positive Correlations	Negative Correlations
Asp52	D52A/R	Constitutively active[19]	Interface: Pro53, Asp54, Arg75, Lys77, Ser78, Leu79, Ser80, Gly81, Arg82, **Lys188**, **Lys192** Non Interface: Leu55, Ile56	Non Interface: Arg161
Asp54	D52A/R	Constitutively active[19]	Interface: Leu79, Ser80, Gly81, Arg82, Asp104, Ala124, Ala125, Met128, Pro172, Asn173, Glu184, **Lys188**, Arg203 Non Interface: Leu55, Ile56, Ile83, Gln122, Ala123,	Non Interface: Arg161, Val153, Gln155, Leu156, Thr157, Pro158, Arg159, Arg161, Asp162
Ser78	S78G	Precluded domain binding[5]	Interface: Leu79, Ser80, **Lys188**, Val191, **Lys192**, Leu195, Lys196	
Ser80	S80R/A/T	Constitutively active[19]	Interface: Gly81, Arg82, Asn173, Glu184, **Lys188**, Val191, **Lys192**, Leu195, Arg203 Non Interface: Ser185, Val187, **Val189**, His193	
Arg82	R82A/D/E	Constitutively active[19] (though less severe)	Interface: Ala103, Asp104, Ala124, Ala125, Gly126, Glu127, Met128, Val129, Gly170, Leu171, Pro172, Asn173, Ser202, Arg203, Val204 Non Interface: Ile83, Val84, His121, Gln122, Ala123, Ile167	Non Interface: Val153, Val154
Lys174	K174A	Weakened domain binding[5]	Interface: Glu184, **Lys188** Non Interface: Met175, Ile176, Ala177, Arg178, Arg179, Leu180, Asp181, Ile182, Thr183, Ser185, Thr186, Val187,	Interface: Trp209 Non Interface: Glu213

References

1. **Arnold, K., L. Bordoli, J. Kopp, and T. Schwede**. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics **22**:195–201.

2. **Bahadur, R. P., and M. Zacharias**. 2008. The interface of protein-protein complexes : Analysis of contacts and prediction of interactions. Cellular and molecular life sciences **65**:1059–72.

3. **Baikalov, I., I. Schröder, M. Kaczor-Grzeskowiak, D. Cascio, R. P. Gunsalus, and R. E. Dickerson**. 1998. NarL dimerization? Suggestive evidence from a new crystal form. Biochemistry **37**:3665–76.

4. **Baikalov, I., I. Schröder, M. Kaczor-Grzeskowiak, K. Grzeskowiak, R. P. Gunsalus, and R. E. Dickerson**. 1996. Structure of the Escherichia coli Response Regulator NarL. Biochemistry **35**:11053–61.

5. **Bartkowski, W.** 2010. Defining the interdomain interface of the Escherichia coli response regulator NarL. UCLA, PhD Dissertation.

6. **Batchelor, J. D., M. Doucleff, C. Lee, K. Matsubara, S. De Carlo, J. Heideker, M. H. Lamers, J. G. Pelton, and D. E. Wemmer**. 2008. Structure and Regulatory Mechanism of Aquifex aeolicus NtrC4 : Variability and Evolution in Bacterial Transcriptional Regulation. Journal of Molecular Biology. Elsevier Ltd **384**:1058–1075.

7. **Bent, C. J., N. W. Isaacs, T. J. Mitchell, and A. Riboldi-tunnicliffe**. 2004. Crystal Structure of the Response Regulator 02 Receiver Domain , the Essential YycF Two-Component System of Streptococcus pneumoniae in both Complexed and Native States. Journal of bacteriology **186**:2872–2879.

8. **Birck, C., L. Mourey, P. Gouet, B. Fabry, J. Schumacher, P. Rousseau, D. Kahn, and J. P. Samama**. 1999. Conformational changes induced by phosphorylation of the FixJ receiver domain. Structure (London, England : 1993) **7**:1505–15.

9. **Bourret, R. B.** 2010. Receiver domain structure and function in response regulator proteins. Current opinion in microbiology. Elsevier Ltd **13**:142–9.

10. **Dyer, C. M., and F. W. Dahlquist**. 2006. Switched or not?: the structure of unphosphorylated CheY bound to the N terminus of FliM. Journal of bacteriology **188**:7354–63.

11. **Egan, S. M., and V. Stewart**. 1991. Mutational analysis of nitrate regulatory gene narL in Escherichia coli K-12. Journal of bacteriology **173**:4424–4432.

12. **Eldridge, A. M., H.-S. Kang, E. Johnson, R. Gunsalus, and F. W. Dahlquist**. 2002. Effect of phosphorylation on the interdomain interaction of the response regulator, NarL. Biochemistry **41**:15173–80.

13. **Feher, V. A., and J. Cavanagh**. 1999. Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. Nature **400**:289–93.

14. **Formaneck, M. S., L. Ma, and Q. Cui**. 2006. Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). Proteins **63**:846–67.

15. **Gao, R., T. R. Mack, and A. M. Stock**. 2007. Bacterial response regulators: versatile regulatory strategies from common domains. Trends in biochemical sciences **32**:225–34.

16. **Gao, R., and A. M. Stock**. 2010. Molecular Strategies for Phosphorylation-Mediated Regulation of Response Regulator Activity. Curr Opin Microbiol. **13**:160–7.

17. **Guhaniyogi, J., V. L. Robinson, and A. M. Stock**. 2006. Crystal structures of beryllium fluoride-free and beryllium fluoride-bound CheY in complex with the conserved C-terminal peptide of CheZ reveal dual binding modes specific to CheY conformation. Journal of molecular biology **359**:624–45.

18. **Hastings, C. A., S.-Y. Lee, H. S. Cho, D. Yan, S. Kustu, and D. E. Wemmer**. 2003. High-resolution solution structure of the beryllofluoride-activated NtrC receiver domain. Biochemistry **42**:9081–90.

19. **Jarvis, M. R.** 1999. Functional analysis of the Escherichia coli nitrate response regulator, NarL. UCLA, PhD Dissertation.

20. **Jones, S., and J. M. Thornton**. 1996. Principles of protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America **93**:13–20.

21. **Krissinel, E., and K. Henrick**. 2007. Inference of macromolecular assemblies from crystalline state. Journal of molecular biology **372**:774–97.

22. **Lee, S. Y., H. S. Cho, J. G. Pelton, D. Yan, E. a Berry, and D. E. Wemmer**. 2001. Crystal structure of activated CheY. Comparison with other activated receiver domains. The Journal of biological chemistry **276**:16425–31.

23. **Lee, S.-Y., H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L. Huang, S. Kustu, E. A. Berry, and D. E. Wemmer**. 2001. Crystal structure of an activated response regulator bound to its target. Nature structural biology **8**:52–6.

24. **Maris, A. E., M. Kaczor-Grzeskowiak, Z. Ma, M. L. Kopka, R. P. Gunsalus, and R. E. Dickerson**. 2005. Primary and secondary modes of DNA recognition by the NarL two-component response regulator. Biochemistry **44**:14538–52.

25. **Maris, A. E., M. R. Sawaya, M. Kaczor-Grzeskowiak, M. R. Jarvis, S. M. D. Bearson, M. L. Kopka, I. Schröder, R. P. Gunsalus, and R. E. Dickerson**. 2002. Dimerization allows DNA target site recognition by the NarL response regulator. Nature structural biology **9**:771–8.

26. **Ruvinsky, A. M., T. Kirys, A. V Tuzikov, and I. A. Vakser**. 2012. Structure Fluctuations and Conformational Changes in Protein Binding. Journal of bioinformatics and computational biology **10**:1–18.

27. **Ruvinsky, A. M., and I. A. Vakser**. 2010. Sequence composition and environment effects on residue fluctuations in protein structures. The journal of chemical physics **133**:155101.

28. **Schrödinger, L.** The PyMOL Molecular Graphics System. Version 1.3.

29. **Schrödinger, L.** The PyMOL Molecular Graphics System. Version 1.3.

30. **Simonovic, M., and K. Volz**. 2001. A Distinct Meta-active Conformation in the 1 . 1-Å Resolution Structure of Wild-type ApoCheY. Journal of Biological Chemistry **276**:28637–40.

31. **Stock, A. M., and J. Guhaniyogi**. 2006. A new perspective on response regulator activation. Journal of bacteriology **188**:7328–30.

32. **Swanson, R. V, D. F. Lowry, P. Matsumura, M. M. McEvoy, M. I. Simon, and F. W. Dahlquist**. 1995. Localized perturbations in CheY structure monitored by NMR identify a CheA binding interface. Nature structural biology **2**:906–10.

33. **Toro-Roman, A., T. R. Mack, and A. M. Stock**. 2005. Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face. Journal of molecular biology **349**:11–26.

34. **Toro-Roman, A., T. I. Wu, and A. M. Stock**. 2005. A common dimerization interface in bacterial response regulators KdpE and TorR. Protein Science **14**:3077–88.

35. **Volkman, B. F., D. Lipson, D. E. Wemmer, and D. Kern**. 2001. Two-state allosteric behavior in a single-domain signaling protein. Science (New York, N.Y.) **291**:2429–33.

36. **Volz, K.** 1993. Structural conservation in the CheY superfamily. Biochemistry **32**:11741–53.

37. **Zhang, J. H., G. Xiao, R. P. Gunsalus, and W. L. Hubbell**. 2003. Phosphorylation triggers domain separation in the DNA binding response regulator NarL. Biochemistry **42**:2552–9.

Appendix A

Crystallization and X-ray Diffraction Studies of the *Escherichia coli* Histidine Kinase NarX

Introduction

Histidine kinases (HKs) begin the phosphorylation cascade of two-component signal transduction systems (Chapter 1, and (18)). The Nar two-component system in *E. coli* enables cellular adaptation to anaerobiosis in the presence of the preferred oxygen substitute, nitrate (Chapter2, and (5)). In this dual system, histidine kinases NarX and NarQ detect nitrate (or nitrite) in the periplasm and autophosphorylate on a conserved histidine in the cytoplasm. The phosphoryl group is then transferred to a second component, NarL or NarP, cytoplasmic response regulators that control transcription relating to nitrate and nitrite reduction.

Structural information regarding NarX is limited, although sequence and biochemical data have revealed a topology comprised of four main modules (Figure A-1) (16, 17). The detection of nitrate (or nitrite) begins at the Transmembrane Signaling region and propagates to the cytoplasmic portion of the kinase, eventually reaching the Transmitter module. The periplasmic region, and the site of nitrate (or nitrite) binding, is flanked by two transmembrane helices. This region of NarX highly resembles another class of bacterial signaling proteins called methyl-accepting chemotaxis proteins (MCPs). Structural studies indicate that NarX and MCPs, such as Tar, utilize the same signaling mechanism. The crystal structure of the periplasmic domain of NarX reveals a dimer comprised of two four-helix bundles, where the nitrate ligand binds to a conserved region (P box). Upon nitrate binding, helical displacements in the bundle are thought to transmit the signal across the transmembrane helices to the HAMP signaling domain (Chapter 1, text box and (3)), which follows the second transmembrane helix.

The approximately 50-residue HAMP domain of the Signal Conversion module marks the beginning of the cytoplasmic portion of the molecule (Figure A-1) (16, 17). This region, also common to MCPs, consists of two helices connected by a linker. The HAMP domain is

presumably responsible for being a signaling bridge between the Transmembrane module and the downstream cytoplasmic domains. Its predicted structure is that of coiled-coils with heptad repeats. The second helix of HAMP overlaps with the signaling helix (S helix), a short helical region that is common to several other signaling proteins. This region is postulated to work together with HAMP in transmitting the signal across the membrane border to the Central module and ultimately to the Transmitter module for catalysis. The S helix has also shown to be important for preventing constitutive output activity by the kinase.

In some HKs, the signal from the periplasm proceeds directly to the Transmitter module (or "kinase core") (Chapter 1 and (1, 8)). In contrast, NarX contains an intervening Central module between the transmembrane region and the kinase core (Figure A-1) (16, 17). This region is of unknown function and so far seems to be unique to NarX and NarQ, among which also shows variability. The region stretches about 50 amino-acids longer in NarX, and contains a cysteine cluster that is conserved among NarX proteins from different genera, but is inconsistent or missing in NarQ proteins.

The goal of HK intramolecular signaling is to reach the Transmitter module, which is involved in executing all three functions of the HK: autophosphorylation, phosphoryl transfer to a response regulator, and dephosphorylation of the response regulator. The Dimerization and Histidine Phosphotransfer (DHp) domain (4, 16, 17) houses the site of histidine phosphorylation and is the docking site for response regulators. It also contributes to passing the signal, via the motions of its four-helix bundle, to the Catalytic and ATP-binding (CA) domain (1, 15). The CA domain contains an ATP-binding pocket that catalyzes the autophosphorylation reaction of the HK. NarX contains the defined regions of the kinase core, including an X box that is not strongly conserved but exists in some HKs. Adjacent to the highly conserved phospho-His site

(in the H-box) is the recently discovered DxxQ motif that is critical for phosphatase activity in HKs belonging to subfamily 7 (6). Also characteristic of HK7 subfamily members, as observed in the sequence of NarX, are missing elements in the CA domain, such as the F box, that result in a less flexible ATP lid (Chapter1 and (21)).

Other than the structure of the periplasmic domain, no other structures exist of NarX. The structure of the cytoplasmic portion of NarX ($NarX^C$) would divulge the nature of the Central module and provide insight into its function. Furthermore, it may shed light on the method of intramolecular signaling as compared to other HKs. In this study, a $NarX^C$ construct was crystallized and a data set was collected to 3.5Å. Structure determination trials using heavy-atom soaks are described. None of the data sets resulting from these crystals gave rise to a viable structure, however strategies to obtain better quality crystals are ongoing.

Materials and Methods

Protein Cloning, Expression, and Purification. The $NarX^C$ construct (residues 219-598, with a 6 His C-terminal tag) was prepared by Woytek Bartowski in the laboratory of Robert P. Gunsalus (UCLA). The pTrx1 plasmid containing His-tagged $NarX^C$ was transformed into BL21/DE3 pLysE *E. coli* cells and overexpressed. Cell lysates were loaded onto a HiTrap Chelating column charged with nickel (GE Healthcare), washed with high salt and eluted with an imidazole gradient. $NarX^C$ fractions, as identified by PAGE, were subsequently loaded onto a butyl sepharose column (GE Healthcare) and eluted with a decreasing ammonium sulfate gradient. At this point forward, all buffer solutions contained 40mM DTT. $NarX^C$ eluted in a buffered solution without ammonium sulfate. This fraction was loaded onto a HiTrap Q column (GE

Healthcare), and eluted with a sodium chloride gradient. Purified $NarX^C$ was in a final solution of 50mM Tris 7.5, 300mM NaCl, 40mM DTT, and 2% glycerol.

Crystallization. $NarX^C$ crystallization trials were performed using the hanging drop vapor diffusion method using an equal volume of 9.0 mg/ml $NarX^C$ and reservoir solution, against a 1mL outer reservoir. Crystallization screens were incubated at room temperature. Reservoir solutions were taken from Crystal Screens 1 and 2 by Hampton Research, and more recently, also from the JCSG+ Screen by Qiagen.

Data Collection. All in-house data sets were collected using a Rigaku Raxis4++ under a liquid-nitrogen stream. Data was processed using the HKL package (13). Crystals were cryoprotected prior to mounting using a solution of 1.2M Formate, 0.05M MES pH 6.5, and 20% glycerol. Synchrotron data sets were collected at Lawrence Berkeley National Laboratory using the Advanced Light Source, beamline 8.2.2.

Heavy Atom Screening. $NarX^C$ was mixed with 17 different heavy metal compounds, in a 2:1 millimolar ratio, and the presence of a shifted band compared to the $NarX^C$ alone band indicated binding of the heavy atom. 2ug of $NarX^C$ alone or soaked with a heavy metal compound was loaded onto a 12.5% native Phast gel (GE Healthcare) and run on a PhastSystem (Amersham Biosciences). The gels were stained with Coomasie Blue.

Results and Discussion

$NarX^C$ Crystallization

The $NarX^C$ construct (residues 219-598, Figure A-1), beginning at the end of the HAMP linker region, gave rise to the most successful expression and purification and was used for all the crystallization studies described here. Initial crystals of $NarX^C$ were grown in 1.0M sodium

formate and 0.05M sodium acetate pH 5.6. The diffraction pattern resulting from these crystals showed reflections extending to about 7Å. After subsequent crystal screens the conditions were refined by replacing the volatile sodium-acetate buffer with MES pH 6.5 and lowering the protein concentration from 10mg/mL to 7-9 mg/ml (Figure A-2a). In general, $NarX^C$ is sensitive to pH and will only form crystals in the pH range of 5.0 to 6.5. Reservoir solutions below pH 5.0 form precipitations (which can be cleared by adding a high pH buffer) while those above pH 7.0 remain clear and preclude crystal formation. The crystals must be frozen within one week of growth since three-week old crystals produce no diffraction pattern. Crystals with visible hexagonal points and sharp edges also produced better diffraction patterns. Such modification improved the quality of diffraction (Figure A-2b).

Structure Determination Trials

A 4.1Å data set of $NarX^C$ was determined to have a space group of either $P6_122$ or $P6_522$, with unit cell dimensions 158.32Å x 158.32 Å x 160.73Å. An asymmetric unit of two $NarX^C$ molecules, or twelve dimers per unit cell, yields a Matthews coefficient of 3.32 $Å^3$/Dalton, which is reasonable. This construct possesses eight methionine residues that can be replaced by seleno-methionine (Se-Met) for use in solving the structure by the multiple anomalous dispersion (MAD) method. Several attempts were made to synthesize Se-Met $NarX^C$ however all proved unsuccessful. Instead, a synchrotron data set of this crystal, called narx92b (Table A-1), was collected to 3.5Å.

Besides trying to overexpress and purify Se-Met $NarX^C$, we also aimed to optimize the diffracting power of the crystals. Several trials were conducted to improve the data. Dehydrating the crystals under different oils, using different additives, changing the temperature

of crystal growth, and experimenting with different reducing agents are some examples. No usable crystals grew in the presence of two non-hydrolyzable ATP analogues, ATPγS and AMP-PNP. The same shaped crystals of apo- $NarX^C$ also grew in two other reservoir solutions, in 0.5M ammonium sulfate and in 0.33M sodium citrate pH 5.6, 0.17M ammonium sulfate, and 0.3M lithium sulfate, however these new crystals produced no change in the data quality.

We first tried to obtain phase information, from the 3.5Å narx92b data set, using the molecular replacement program Phaser (9). Individual domains and intact proteins from all known structures of HKs, GHL ATPases, and other structural homologues were used as models, but resulted in inadequate Z-scores or unit cell arrangements with visible clashes of the atoms. One trial using the two helices of the dimerization domain of CheA (2) resulted in a high Z-score of 9.36 however no pdb file was generated due to too many clashes. The ATP-binding domains of PhoQ (7) and EnvZ (19) had decent Z-scores but did not show credible unit cell packing.

The next attempt to obtain phase information was using heavy atom derivatives. Native gels were used to identify heavy atom binding. The mercury (Hg) compound Mersalyl (Figure A-3, lane 3) showed a clear shift, and potassium gold cyanide showed a present but weaker shift (Figure A-3, lane 2). $NarX^C$ crystals were washed in mother liquor to remove the DTT and then soaked in either Mersalyl, gold acetate, or lead nitrate for different times and concentrations (lead nitrate, although was not tested on a gel shift, also appeared to have soaked into the crystal based on diffraction data). The most successful soaks were with 0.5mM-1mM heavy atom for 15-30 minutes. Data sets of the heavy-atom soaked crystals were collected in-house and also at a synchrotron radiation source, which produced a 4.0Å data set of a lead-nitrate soaked crystal, called narx183pb. Table A-1 shows a summary of the most promising data sets collected from native and heavy atom soaked crystals.

Difference Patterson maps were calculated for the lead (Pb) and Hg data sets using coefficients derived from isomorphous differences, anomalous differences, and a combination of isomorphous and anomalous differences. Maps were generated using XPREP and XtalView (10) software. All three sets of coefficients were used as input in SHELXD (14) to find the heavy atom sites. The xyz coordinates of the heavy atom sites were converted to uvw and evaluated for correlation with difference Patterson peaks on three Harker sections, w = 0.500, 0.333, and 0.167. The four Hg peaks found by SHELXD appear to fit the isomorphous difference Patterson map for narx173b at 4.8 Å (Figure A-4). Note that an even number of heavy atom sites are predicted to be present since the Matthews coefficient indicates that the asymmetric unit is a dimer. Similarly, in narx183pb, the first two (of three) Pb sites from SHELXD mostly overlapped with the actual peaks on the isomorphous and anomalous difference Patterson maps. Sites found using anomalous differences matched the sites found using isomorphous differences (Figure A-5). This led us to believe that the predicted lead sites were promising.

Phases were calculated for the narx173b and narx183pb data sets using programs from the ccp4 suite (22), which included MLPHARE (12) that was used to refine the heavy atom positions and calculate phases. Phases were further improved by solvent flattening and an electron density map was then calculated for each hand, $P6_122$ and $P6_522$. Looking at the bones of the electron density to aid in the visualization, no real secondary structure could be recognized from the narx183pb data from either space group. The best map resulted from narx173b in the space group $P6_122$. However, attempts to model in secondary structure from known crystal structures, such as CheA (2) and PhoQ (7), were unsuccessful.

The phases from narx183pb were obtained by running MLPHARE using only the first of two predicted lead site from SHELXD. A self-anomalous and self-isomorphous difference

Fourier, in both hands of the space group, was calculated to find the position of the second lead site. The second peak was found from the self-anomalous difference Fourier and coincided with the second lead site predicted by SHELXD. Figure A-6 shows that peaks 1 and 2 from the original SIRAS difference Patterson are located in the same places as peaks 0 and 3 in the difference Fourier (highlighted in yellow). Thus we believe that the first two coordinates of the lead atoms predicted by SHELXD are real since they were confirmed by a difference Fourier and by the SIRAS difference Patterson maps.

We used the phases from narx183pb (calculated using both lead sites predicted by SHELXD) to calculate a cross-isomorphous and cross-anomalous Fourier, in both hands of the space group, to find and confirm the positions of the mercury atoms. This proved unsuccessful as the predicted peaks from the cross Fourier did not correspond to the actual peaks of the narx173b difference Patterson map. The same procedure was carried out to find the lead atom positions from the mercury, narx173b, data set. Again, the heavy atom positions predicted by this calculation did not consistently match the peaks of the difference Patterson maps nor did they consistently match the predicted peaks found by SHELXD.

Using the native data set, narx92b, and the two heavy atom data sets, narx183pb and narx173b, the program SOLVE (20) was also used to calculate phases, predict the heavy atom positions, and create a pdb file. The Hg and Pb positions found by SOLVE, however, did not consistently coincide with peaks on the Patterson maps and were evaluated to be incorrect.

A three wavelength MAD experiment was attempted with narx183pb. The peak wavelength was 0.9494 Å, the inflection wavelength was 0.9506 Å, and the high remote wavelength was 0.8856 Å. Anomalous and dispersive difference Patterson maps were calculated. Four-sigma peaks were seen on the Harker sections for the anomalous map while the

dispersive map was flat. The poor quality of the latter map could be due to an incorrect choice of inflection wavelength.

Conclusions

Attempts to gain phase information using more recent structures (1) have also proved unsuccessful. We attribute the main problem in solving this structure is the low resolution of the data. This work is ongoing and new crystallization trials were performed with the same NarX construct in the presence of ADP and manganese ion. This is based on recent work revealing that NarX has a high affinity for ADP (11). Crystals were grown under two conditions from purchased screens. These crystals were confirmed to be composed of protein (Figure A-7) and are currently being optimized.

Acknowledgements

Michael Sawaya (for help with all aspects of structure determination), Maria Kaczor-Grzeskowiak, Woytek Bartowski, Mary Kopka, Duilio Cascio, Inna Pashkov, Todd Yeats, Robert P Gunsalus, and Richard E. Dickerson.

Figure A-1. Domain topology of the histidine kinase NarX. **(a)** The domain structure of NarX. NarX is divided into four modules: Transmembrane Signaling Conversion, Central, and Transmitter (or kinase core). Domains and conserved regions are represented by boxes, each colored differently: transmembrane regions (TM1 and TM2) in beige, conserved P box of the periplasmic domain in yellow, HAMP domain (green) and overlap (pink) with S helix (purple), cysteine-rich region in brown, conserved H and X boxes in red, with DxxQ motif as an orange bar, conserved N, D, and G boxes of the CA domain in blue. The scale is numbered according to amino-acid (AA) residue. This diagram was reproduced and slightly modified from a published source (17).

(a)

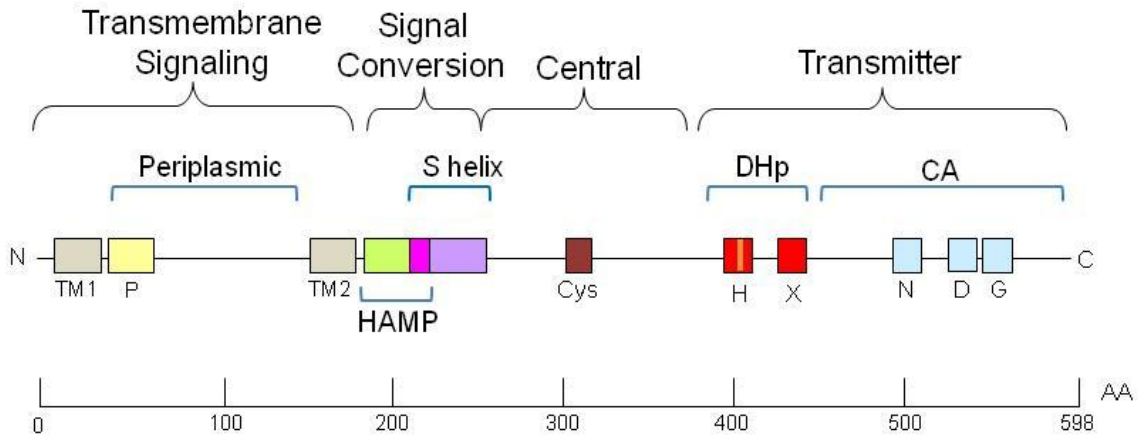
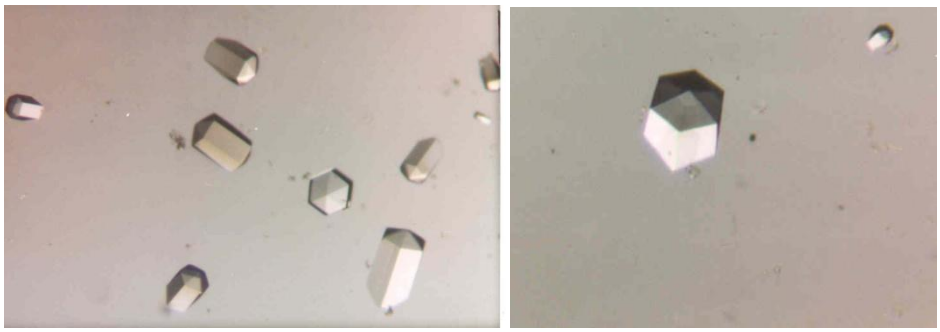


Figure A-2. NarX crystals and x-ray diffraction pattern. **(a)** Crystals grown in a drop containing an equal volume of 1.0M sodium formate 0.05M Mes 6.5 and 7mg/mL $NarX^{C}$. Those with a completely visual hexagon and at least 0.2mm in length have produced the best diffraction. **(b)** X-ray diffraction pattern of crystal NarX92b, a data set of which was collected to about 4Å. Crystals were cryoprotected with 1.2M Formate 0.05M Mes 6.5 20% glycerol, 5mM DTT.

(a)



(b)

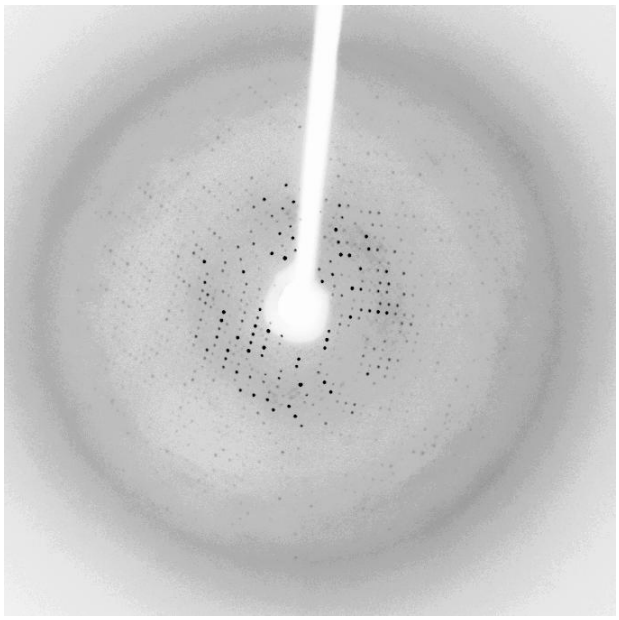


Figure A-3. 12.5% homogeneous native PHAST gel to test for heavy atom binding to $NarX^C$: lane 1, $NarX^C$; lane 2, $NarX^C$ with potassium gold cyanide; lane 3, $NarX^C$ with mersalyl; lane 4, $NarX^C$ with mercury acetate; lane 5, $NarX^C$ with mercury chloride; lane 6, NarXC with thimerosal (mercury compound); lane 7, $NarX^C$ with *p*-chloromercuriphenylsulfonic acid. Heavy atoms were mixed with $NarX^C$ in a 2:1 millimolar ratio. The gel was stained with coomassie blue. Arrows indicate the location of the unshifted $NarX^C$ band and the shifted $NarX^C$ bands in lanes 2 and 3.

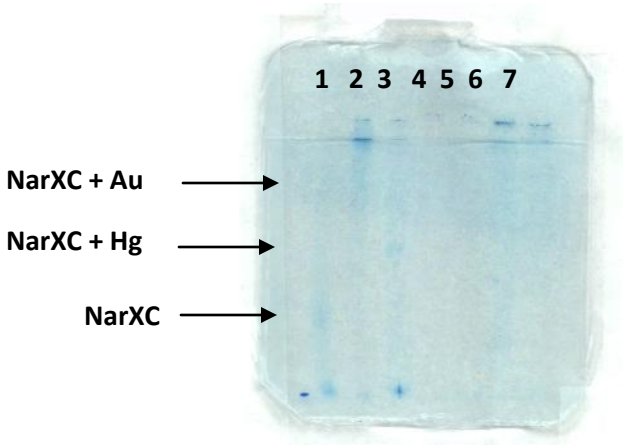


Figure A-4. Isomorphous difference Patterson maps for narx173b, at 4.8 Å at w= 0.500, 0.333, and 0.167 from left to right. The four Hg sites predicted by SHELXD are numbered as1 to 4 on the maps.

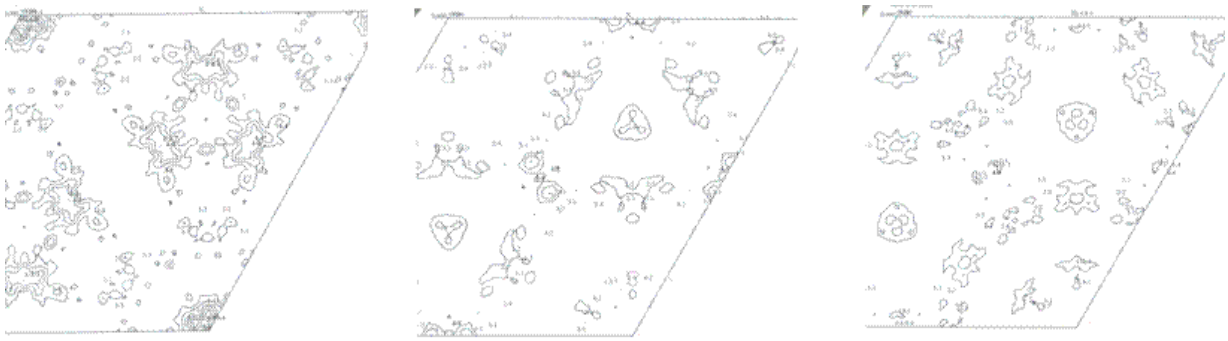


Figure A-5. Top panel shows isomorphous difference Patterson maps for narx183pb, w= 0.500, 0.333, and 0.167 from left to right, with predictions derived from anomalous coefficients used as input in SHELXD. Bottom panel shows anomalous difference Patterson maps at w= 0.500, 0.333, and 0.167 from left to right. The predictions are isomorphous coefficients used as input in SHELXD. Sites from one map correspond to peaks from the other map, and vice versa.

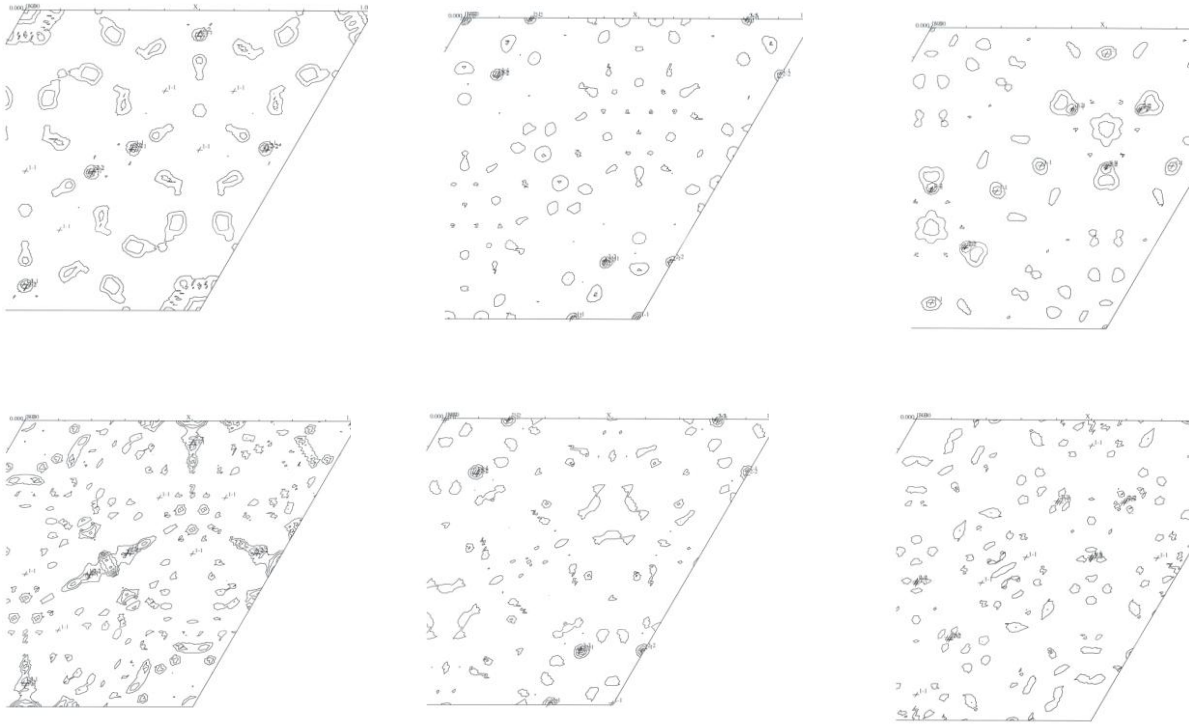


Figure A-6. Top and Bottom panels show a combination of isomorphous and anomalous difference Patterson maps for narx183pb at w= 0.500, 0.333, and 0.167 from left to right. The top panel shows predicted Pb sites found from SHELXD, while the bottom panel shows predicted Pb sites found by a self-anomalous difference Fourier. Sites 1 and 2 from the top panel correspond to sites 0 and 3 from the bottom panel; these sites are highlighted in yellow.

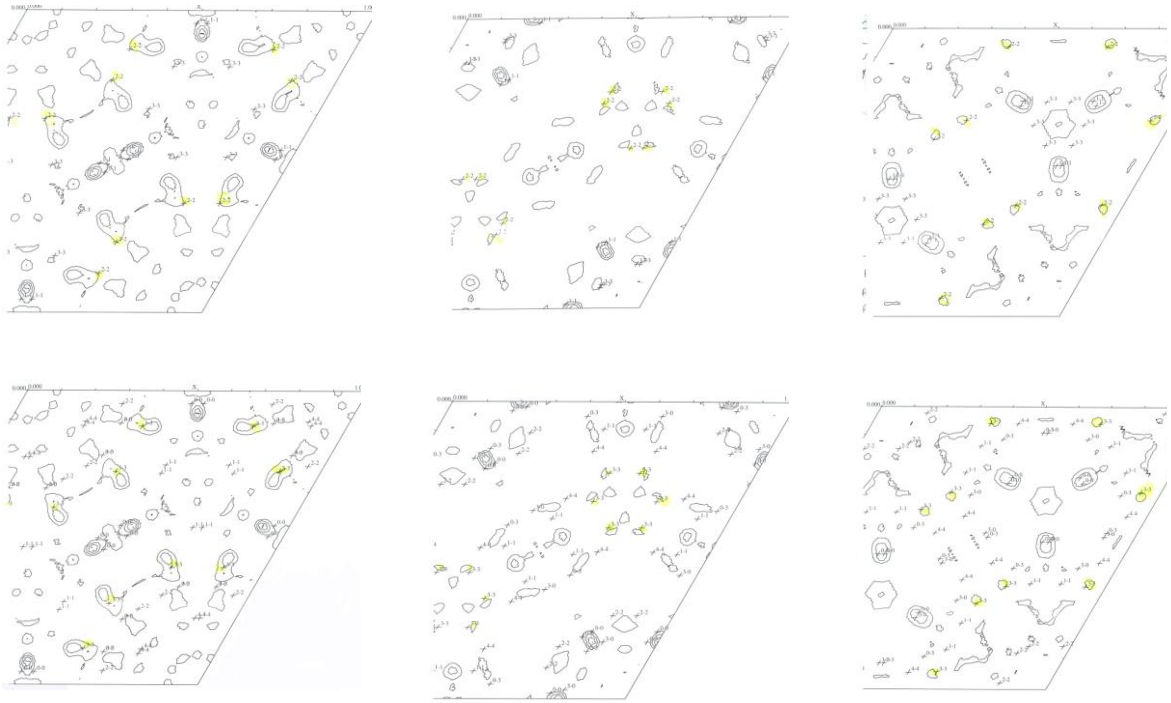


Figure A-7. New $NarX^{C}$ crystals. The left panel shows crystals grown in 0.1M Na/K phosphate pH6.2, 25% (v/v) 1,2 propanediol, 10% glycerol (Qiagen JCSG+ Screen #33). The right panel is the same well exposed to UV light, confirming that these crystals are protein.

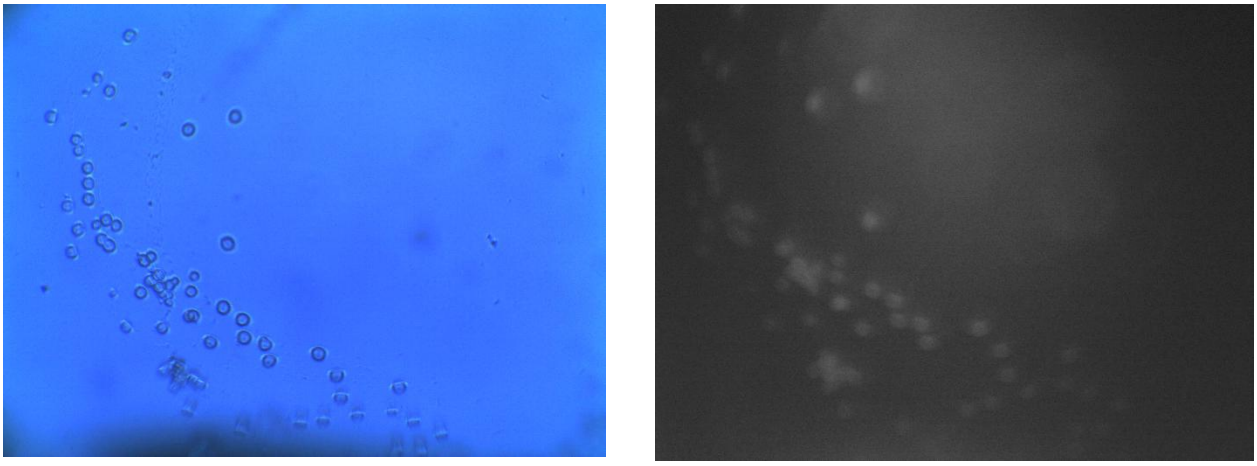


Table A-1. Summary of best $NarX^C$ native and heavy-atom soaked crystals.

Crystal	HA	Soaking Conditions	Res (Å)	Unit Cell dimensions (Å)/ Angles
narx92b	none	none	3.5*	158.641, 158.641, 160.798/ 90.0, 90.0, 120.0
narx150	none	none	4.3	
narx137	Au	3.8mM $KAu(CN)_2$ for 2.5hrs	7	
narx153b	Hg	0.5mM Mersalyl for 30min.	6.2	158.511, 158.511, 163.465/ 90.0, 90.0, 120.0
narx173b	Hg	1.0mM Mersalyl for 15min.	4.8	159.991, 159.991, 160.726/ 90.0, 90.0, 120.0
narx175	Pb	1mM $Pb(NO_3)_2$ for 45min.	5.4	159.006, 159.006, 157.321/ 90.0, 90.0, 120.0
narx179	Pb	0.5mM $Pb(NO_3)_2$ for 30min.	6.2	158.530, 158.529, 157.108/ 90.0, 90.0, 120.0
narx183pb	Pb	0.5mM $Pb(NO_3)_2$ for 15min.	4.0*	159.140, 159.140, 157.204/ 90.0, 90.0, 120.0

* Taken at the Advanced Light Source, Beamline 8.2.2

References

1. **Albanesi, D., M. Martín, F. Trajtenberg, M. C. Mansilla, A. Haouz, P. M. Alzari, D. de Mendoza, and A. Buschiazzo**. 2009. Structural plasticity and catalysis regulation of a thermosensor histidine kinase. Proceedings of the National Academy of Sciences of the United States of America **106**:16185–90.

2. **Bilwes, A. M., L. A. Alex, B. R. Crane, and M. I. Simon**. 1999. Structure of CheA, a signal-transducing histidine kinase. Cell **96**:131–41.

3. **Cheung, J., and W. a Hendrickson**. 2009. Structural analysis of ligand stimulation of the histidine kinase NarX. Structure (London, England : 1993). Elsevier Ltd **17**:190–201.

4. **Grebe, T. W., and J. B. Stock**. 1999. The histidine protein kinase superfamily. Advances in microbial physiology **41**:139–227.

5. **Gunsalus, R. P.** 1992. Control of electron flow in Escherichia coli: coordinated transcription of respiratory pathway genes. Journal of bacteriology **174**:7069–74.

6. **Huynh, T. N., C. E. Noriega, and V. Stewart**. 2010. Conserved mechanism for sensor phosphatase control of two-component signaling revealed in the nitrate sensor NarX. Proceedings of the National Academy of Sciences of the United States of America **107**:21140–5.

7. **Marina, A., C. Mott, A. Auyzenberg, W. a Hendrickson, and C. D. Waldburger**. 2001. Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. The Journal of biological chemistry **276**:41182–90.

8. **Marina, A., C. D. Waldburger, and W. a Hendrickson**. 2005. Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. The EMBO journal **24**:4247–59.

9. **McCoy, A., R. Grosse-Kunstleve, P. Adams, M. Winn, L. Storoni, and R. Read**. 2007. Phaser crystallographic software. Journal of Applied Crystallography **40**:658–74.

10. **McRee, D.** 1999. Practical Protein Crystallography, 2nd edition. Academic Press (London).

11. **Noriega, C. E., R. Schmidt, M. J. Gray, L.-L. Chen, and V. Stewart**. 2008. Autophosphorylation and dephosphorylation by soluble forms of the nitrate-responsive sensors NarX and NarQ from Escherichia coli K-12. Journal of bacteriology **190**:3869–76.

12. **Otwinowski, Z.** 1991. Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering. edited by W. Wolf, P. R. Evans & A. G. W. Leslie **Warrington**:80–6.

13. **Otwinowski, Z., and W. Minor**. 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology **276**:307–26, C.W. Carter, Jr. & R.M. Sweet, Eds., Acade.

14. **Sheldrick, G. M.** 2008. A short history of SHELX. Acta crystallographica. Section A, Foundations of crystallography. International Union of Crystallography **64**:112–22.

15. **Stewart, R. C.** 2010. Protein Histidine Kinases: Assembly of Active Sites and Their Regulation in Signaling Pathways. Curr Opin Microbiol. **13**:133–141.

16. **Stewart, V.** 2003. Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria. Biochemical Society Transactions **31**:1–10.

17. **Stewart, V., and L.-L. Chen**. 2010. The S helix mediates signal transmission as a HAMP domain coiled-coil extension in the NarX nitrate sensor from Escherichia coli K-12. Journal of bacteriology **192**:734–45.

18. **Stock, A. M., V. L. Robinson, and P. N. Goudreau**. 2000. Two-Component Signal Transduction. Annual review of biochemistry **69**:183–215.

19. **Tanaka, T., S. K. Saha, C. Tomomori, R. Ishima, D. Liu, K. I. Tong, H. Park, R. Dutta, L. Qin, M. B. Swindells, T. Yamazaki, A. M. Ono, M. Kainosho, M. Inouye, and M. Ikura**. 1998. NMR structure of the histidine kinase domain of the E. coli osmosensor EnvZ. Nature **396**:88–92.

20. **Terwilliger, T., and J. Berendzen**. 1999. Automated MAD and MIR structure solution. Acta crystallographica **D55**:849–61.

21. **Trajtenberg, F., M. Graña, N. Ruétalo, H. Botti, and A. Buschiazzo**. 2010. Structural and enzymatic insights into the ATP binding and autophosphorylation mechanism of a sensor histidine kinase. The Journal of biological chemistry **285**:24892–903.

22. **Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. a Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson**. 2011. Overview of the CCP4 suite and current developments. Acta crystallographica. Section D, Biological crystallography **67**:235–42.