BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

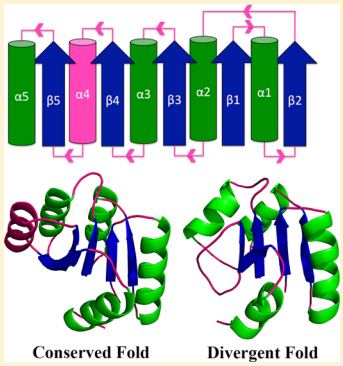
Nuclear Magnetic Resonance Structure and Dynamics of the Response Regulator Sma0114 from *Sinorhizobium meliloti*

Sarah R. Sheftic, Preston P. Garcia,[†] Emma White, Victoria L. Robinson, Daniel J. Gage, and Andrei T. Alexandrescu*

Department of Molecular and Cellular Biology, University of Connecticut, Storrs, Connecticut 06269, United States

Supporting Information

ABSTRACT: Receiver domains control intracellular responses triggered by signal transduction in bacterial two-component systems. Here, we report the solution nuclear magnetic resonance structure and dynamics of Sma0114 from the bacterium *Sinorhizobium meliloti*, the first such characterization of a receiver domain from the HWE-kinase family of two-component systems. The structure of Sma0114 adopts a prototypical α_5/β_5 Rossman fold but has features that set it apart from other receiver domains. The fourth β-strand of Sma0114 houses a PFxFATGY sequence motif, common to many HWE-kinase-associated receiver domains. This sequence motif in Sma0114 may substitute for the conserved Y-T coupling mechanism, which propagates conformational transitions in the 455 (α4−β5−α5) faces of receiver domains, to prime them for binding downstream effectors once they become activated by phosphorylation. In addition, the fourth α-helix of the consensus 455 face in Sma0114 is replaced with a segment that shows high flexibility on the pico- to nanosecond time scale by ^{15}N relaxation data. Secondary structure prediction analysis suggests that the absence of helix α4 may be a conserved property of the HWE-kinase-associated family of receiver domains to which Sma0114 belongs. In spite of these differences, Sma0114 has a conserved active site, binds divalent metal ions such as Mg^{2+} and Ca^{2+} that are required for phosphorylation, and exhibits micro- to millisecond active-site dynamics similar to those of other receiver domains. Taken together, our results suggest that Sma0114 has a conserved active site but differs from typical receiver domains in the structure of the 455 face that is used to effect signal transduction following activation.



Phosphorylation is a commonly used mechanism for achieving signal transduction. In bacteria, phosphorylation-mediated signal transduction is conducted by two-component systems. These signaling cascades utilize a sensor histidine kinase that autophosphorylates when triggered by a stimulus and a response regulator that mediates the downstream output response upon receiving a phosphate from its cognate histidine kinase.[1,2] The output response is maintained while the receiver domain component of the response regulator remains phosphorylated. The lifetime of the phosphorylated state is controlled primarily by the autophosphatase enzymatic activity of the receiver domain.

The two-component system encoded by the genes *sma0113* and *sma0114* from the bacterium *Sinorhizobium meliloti* was identified in a genetic screen for altered succinate-mediated catabolite repression.[3] *S. meliloti* is a Gram-negative soil bacterium in the α-proteobacteria. The bacterium can exist in the soil as a free-living organism, or it can live as a nitrogen-fixing symbiont in specialized organs called nodules, which it induces on the roots of certain legumes such as alfalfa. On the basis of previous work, it is expected that *sma0113* and *sma0114* play a role in *S. meliloti* catabolite repression and polyhydroxy butyrate synthesis.[3] The tandem genes encode a histidine-tryptophan-glutamate (HWE) histidine kinase (Sma0113), and a single-domain response regulator (Sma0114). HWE-kinases were first described in 2004[4] and constitute a subclass of the larger histidine kinase superfamily. Compared to the members of the well-characterized histidine kinase family, HWE-kinases have an altered ATP binding site, which lacks the F-box that is normally an integral component of the ATP lid.

The majority of response regulators have receiver and effector domains. The receiver domain acts as a "switch" that undergoes a conformational change in response to phosphorylation at a conserved aspartate.[1,5] The effector domain usually has a DNA binding function that regulates the transcription of its target genes.[1] In some cases, including response regulators CheY, Spo0F, DivK, and Sma0114, only the receiver domain is present.[6,7] The roles of single-domain response regulators are not as well understood as those of their two-domain counterparts, but there are three known functions. The receiver domain can act directly on a protein effector (e.g., CheY), can participate in a phospho-relay cascade (e.g., Spo0F), or can act as a histidine kinase inhibitor (e.g., DivK).[8−10] Structural studies of single-domain response regulators have shown that these are structurally similar in their inactive apo forms but show greater variability in their phosphorylated activated states.[11] This is probably because the function of the inactive

Received: July 11, 2012
Revised: August 9, 2012
Published: August 10, 2012

ACS Publications © 2012 American Chemical Society

receiver domain is to receive a phosphate from its cognate histidine kinase, whereas the active phosphorylated enzyme can bind to a wide range of downstream effector proteins.[12,13]

Receiver domains have a conserved α_5/β_5 Rossman fold, where five α-helices surround five parallel β-sheets. A conserved pair of acidic residues located in the loop between strand $\beta1$ and helix $\alpha1$ form part of the binding site for a divalent metal cation that is required for stabilizing the incoming phosphate group and forming the acyl−phosphate linkage.[5,14] The site of phosphorylation is a conserved Asp at the C-terminal end of strand $\beta3$.[1,5] The C-terminal end of strand $\beta5$ houses a conserved Lys that stabilizes the incoming phosphate group.[14] Phosphorylation induces a conformational switch, in which a conserved Thr at the C-terminal end of strand $\beta4$ hydrogen bonds with the phosphate group, and causes a rotameric change of an aromatic residue (Tyr or Phe) in strand $\beta5$ in a mechanism called "Y-T coupling".[14] Y-T coupling mediates the more global rearrangement of the "455 face" of the enzyme, comprised of secondary structure elements $\alpha4$, $\beta5$, and $\alpha5$. In Sma0114, the aromatic residue in the Y-T coupling mechanism is replaced with a leucine, and the conserved Thr is part of a PFxFATGY sequence motif that is common in the HWE-kinase-associated family of receiver domains.

The nuclear magnetic resonance (NMR) investigations described herein show that the unusual sequence features of Sma0114 lead to unique structural features, primarily affecting the 455 face of the enzyme. ^{15}N relaxation data show that there are also changes in the dynamics of the 455 face of Sma0114 compared to the corresponding regions of other receiver domains.[15,16] By contrast, the active site of the enzyme and the ability to bind metals are retained, suggesting that the unusual features of Sma0114 do not alter the activation mechanism but rather the conformational changes of the 455 face accompanying phosphorylation.

EXPERIMENTAL PROCEDURES

NMR Sample Preparation. Recombinant Sma0114 was ligated into a pET28(+) vector and transformed into *Escherichia coli* BL21(DE3) *pLysS* cells. Expression and purification of Sma0114 were performed as previously described.[17] [^{15}N]Sma0114 and [^{15}N,^{13}C]Sma0114 samples were dissolved in 50 mM Na_2HPO_4 buffered to pH 6.0. All samples for NMR had Sma0114 concentrations of 500 μM, with 0.02% NaN_3 to prevent bacterial growth and 1 mM DTT to prevent disulfide formation due to the sole cysteine at position 29 in the protein. For the metal titration with $CaCl_2$, we used a pH 6.0 MES buffer rather than phosphate to prevent precipitation of Ca^{2+}.

NMR Structure Determination. NMR spectra were collected on a Varian Inova 600 MHz instrument equipped with a cryogenic probe. Chemical shift assignments for Sma0114 have been published previously.[17] Three-dimensional (3D) ^{15}N- and ^{13}C-edited NOESY experiments[18] were used to obtain NOE-based distance restraints. Long-range HNCO[19] and deuterium isotope exchange experiments were used to identify hydrogen bonds. Restraints for backbone dihedral angles ϕ and ψ were calculated from the assigned HN, Hα, N, Cα, Cβ, and C′ chemical shifts using TALOS.[20] Additional 3D HNHA data were collected to check ϕ dihedral angles. Side chain χ_1 dihedral angles and stereospecific assignments for prochiral methylene protons were determined from 3D HNHB data[18] and short-mixing time two-dimensional NOESY spectra.[21] Stereospecific assignments for the prochiral methyl groups of Leu and Val residues were obtained from a sample fractionally labeled with 10% [^{13}C]glucose.[22] The NMR structure of Sma0114 was calculated using X-PLOR (version 3.851)[23] based on 1627 experimental restraints (Table 1). The 20 lowest-energy structures have been deposited in the Protein Data Bank (PDB) as entry 2LPM.

Table 1. Statistics for the 20 Lowest-Energy NMR Structures of Sma0114

total no. of NMR restraints	1627
total no. distance restraints	1431
no. of intraresidue NOEs	422
no. of sequential NOEs	486
no. of short-range NOEs ($1 < \|i - j\| < 5$)	109
no. of long-range NOEs ($5 < \|i - j\|$)	414
no. of hydrogen bonds (32×2)	64
no. of dihedrals (ϕ 61, ψ 54, χ_1 17)	132
residual restraint violations	
NOE (Å)[a]	0.039 ± 0.004
dihedral (deg)[b]	0.539 ± 0.018
rmsd from ideal geometry	
bonds (Å)	0.0040 ± 0.0003
angles (deg)	0.73 ± 0.02
improper torsions (deg)	0.54 ± 0.02
van der Waals energy E_{vdw} (kcal/mol)[c]	86.28 ± 2.37
Lennard-Jones energy $E_{L\text{-}J}$ (kcal/mol)[d]	−83.18 ± 7.40
ProcheckNMR Z score[e]	−2.52
Ramachandran plot (%)	
most favored	82.3
allowed	15.3
generously allowed	1.8
disallowed	0.6

Coordinate rmsd (Å)		
NMR ensemble to average	C_α, C, N	all heavy
entire domain (110 residues)[f]	0.94	1.46
excluding loops (72 residues)[g]	0.86	1.38

[a]Structure contains no NOE violations greater than 0.3 Å. [b]Structure contains no dihedral violations greater than 5°. [c]E_{vdw} was calculated using the X-PLOR *F*repel function[23] with van der Waals interactions and atomic radii set to 0.8 times their CHARMM[47] values. [d]$E_{L\text{-}J}$ was calculated using the CHARMM empirical energy function.[47] [e]ProcheckNMR used via Protein Structure Validation Suite (PSVS).[48] [f]Excluding the N-terminus (residues 1−9) and the C-terminus (residues 118−123). [g]Only residues in regular secondary structure (10−15, 19−31, 34−39, 42−51, 56−60, 70−78, 81−86, 102−105, and 109−117).

NMR Relaxation Measurements. Backbone dynamics of Sma0114 were investigated using ^{15}N R_1, ^{15}N R_2, and ^{1}H−^{15}N NOE experiments at a field strength of 600 MHz. Longitudinal relaxation rates (R_1) were characterized using relaxation delays of 0.02, 0.05, 0.13, 0.21, 0.31, 0.5, 0.71, and 1.0 s. Transverse relaxation rates (R_2) were measured using relaxation delays of 0.01, 0.03, 0.05, 0.07, 0.09, 0.15, 0.25, and 0.35 s. ^{1}H−^{15}N NOE values were determined from experiments in which the proton signals were saturated (s) for 4 s and control (c) experiments in which the saturation period was replaced with an equivalent preacquisition delay. Spectra were acquired in an interleaved manner on a Varian Inova 600 MHz instrument for all three relaxation data sets. Relaxation rates were calculated from least-squares fits of the data to an exponential decay model (eq 1),

dx.doi.org/10.1021/bi300922z | *Biochemistry* 2012, 51, 6932−6941

where I is the intensity for relaxation period τ, I_0 is the initial amplitude, and $R_{1,2}$ corresponds to relaxation rate R_1 or R_2.

$$I = I_0 \times \exp(-\tau/R_{1,2}) \quad (1)$$

Experimental uncertainties in relaxation parameters were taken as the standard errors of the fits. $^1H-^{15}N$ NOE values were calculated according to Equation 2, where $I(s)$ is the cross-peak intensity in the experiment with saturation (s) and $I(c)$ is the cross-peak intensity without saturation (c).

$$\text{NOE} = I(s)/I(c) \quad (2)$$

The errors for the $^1H-^{15}N$ NOE experiment were determined as described previously.[24,25] R_1, R_2, and $^1H-^{15}N$ NOE values were used as input for ModelFree calculations[26] using *Tensor2*.[27]

Metal Titration Studies. NMR titrations were performed using ^{15}N-labeled samples of Sma0114, over a range of concentrations from 0 to 500 mM $MgCl_2$. Because of the high $MgCl_2$ concentration of 150 mM needed to saturate the enzyme, we also looked at the binding of $CaCl_2$ over a concentration range from 0 to 10 mM. With $CaCl_2$, saturation of Sma0114 was achieved at a metal ion concentration of 1.5 mM.

Bioinformatics Analysis of Sma0114. A sequence alignment of Sma0114 was initially performed against 2036 receiver domains from the order Rhizobiales using the SMART database (version 6).[28] Sequences that were >97% homologous to that of Sma0114 were considered redundant and removed from further analysis. The 1792 remaining sequences were further aligned using MUSCLE (version 3.8.21).[29] From this alignment, 100 bootstrap replicate maximum-likelihood phylogenies were constructed using RAxML (version 7.3.0)[30] incorporating the WAG model of amino acid substitutions[31] and estimating the gamma model of rate heterogeneity with four discrete rate categories. The bootstrapped trees were mapped onto the best-scoring maximum-likelihood tree.

Analysis of 273 HWE-kinases collected from the MiST 1 database[32] showed that 89 of the kinases had associated receiver domains. Receiver domains were considered to be associated with an HWE-kinase if the genes were fused or immediately upstream or downstream of the kinase. The cognate receiver domains of the 88 HWE-kinases identified in this way plus Sma0114 were analyzed using the secondary structure prediction program PsiPred (version 2.6).[33] The final alignment excluded five of the 89 sequences because they were fewer than 100 amino acids in length and did not appear to encode full-length receiver domains. As a control, eight receiver domain sequences of known structure were employed in the alignment to verify that the α_5/β_5 motif could be detected accurately by PsiPred (Figure 6C).[33] The sequences used for analysis are provided in Table S1 of the Supporting Information.

RESULTS

Phylogenetic Distribution of Receiver Domains Associated with HWE-Kinases. To place the Sma0114 receiver domain in the context of a family of homologous proteins, we characterized the distributions of HWE two-component systems in prokaryotes. The SMART database (version 6) contains 1326 proteins with HWE histidine kinase domains, and the majority of these (70.8%) are found in the α-proteobacteria, with just more than half of the total in the order Rhizobiales (65.3%). The SMART database has some 2000 receiver domains from the Rhizobiales order of α-proteobacteria, of which ~1800 are nonredundant. Figure 1 shows a neighbor-joining tree of these receiver domains.

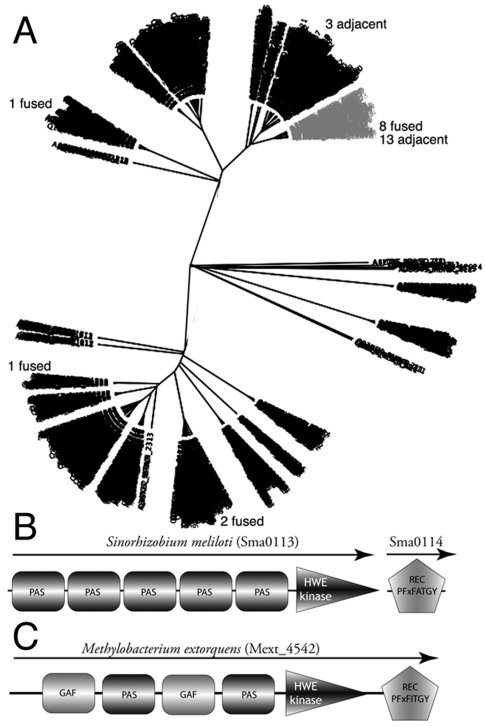


Figure 1. (A) Phylogenetic analysis of HWE-associated response regulators. Maximum-likelihood tree of 1792 receiver domains from the order Rhizobiales. The 28 receiver domains from the SMART database that are genetically adjacent (B) or fused (C) to HWE-kinases are labeled in the figure. Sma0114 is in the gray cluster. The bottom panels show examples of HWE-associated receiver domains. (B) The *Sma0113/Sma0114* pair is a case where the cognate kinase and receiver are separate proteins, encoded on adjacent genes. (C) The *Methylobacterium extorquens* Mext_4542 protein is a case in which the cognate kinase and receiver domain are fused in the same polypeptide chain.

There are 28 nonredundant HWE-kinases in the order Rhizobiales that are either fused to their cognate receiver domains or unlinked to their cognate receiver domains but are encoded in tandem genes (Figure 1). The genes encoding the unlinked kinase and receiver domain proteins are genetically adjacent, indicating that they form an operon and function together. The analysis summarized in Figure 1 shows a majority[21] of the HWE-associated receiver domains, including Sma0114, are in a single branch (shown in gray) indicating that they have a high degree of sequence similarity. Many of the other receiver domains in the gray branch of the neighbor-joining tree are likely to be response regulators that interact with HWE-kinases but were not considered in this analysis because they are encoded by genes distant from their cognate HWE-kinases.

The NMR Structure of Sma0114 Shows Differences from the Receiver Domain Superfamily. Figure 2 shows the NMR structure of Sma0114. The overall fold of the protein is similar to the canonical α_5/β_5 Rossman fold of other receiver domains, which is schematically illustrated in Figure 2A. Backbone (C^α, N, and C′) traces of the 20 lowest-energy NMR

dx.doi.org/10.1021/bi300922z | *Biochemistry* 2012, 51, 6932−6941

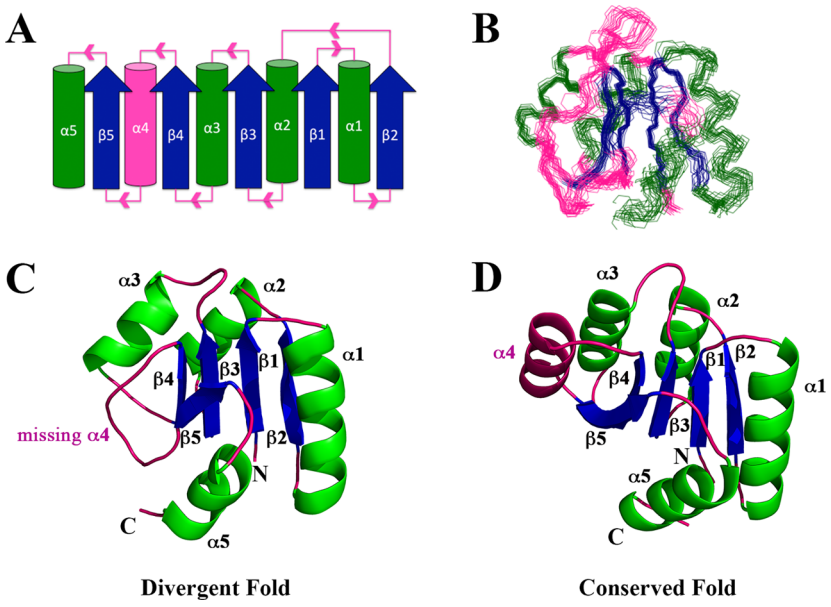


Figure 2. Fold and NMR structure of Sma0114. (A) Topology diagram for the canonical α_5/β_5 Rossman fold of receiver domains. Parallel β-sheets (blue) are surrounded by α-helices (green) with pink segments indicating loops. The fourth α-helix that is disordered in Sma0114 is also colored pink. (B) Backbone representation of the 20 lowest-energy NMR structures. Regular secondary structure elements 10−15 (β1), 19−31 (α1), 34−39 (β2), 42−51 (α2), 55−60 (β3), 70−78 (α3), 81−86 (β4), 102−105 (β5), and 109−117 (α5) were used to superpose the structures. (C) NMR structure of Sma0114 closest to the ensemble average. The fourth α-helix is replaced by a disordered segment. (D) NMR structure of Spo0F (PDB entry 1NAT).[49] The fourth α-helix is colored pink. N and C termini are labeled N and C, respectively.

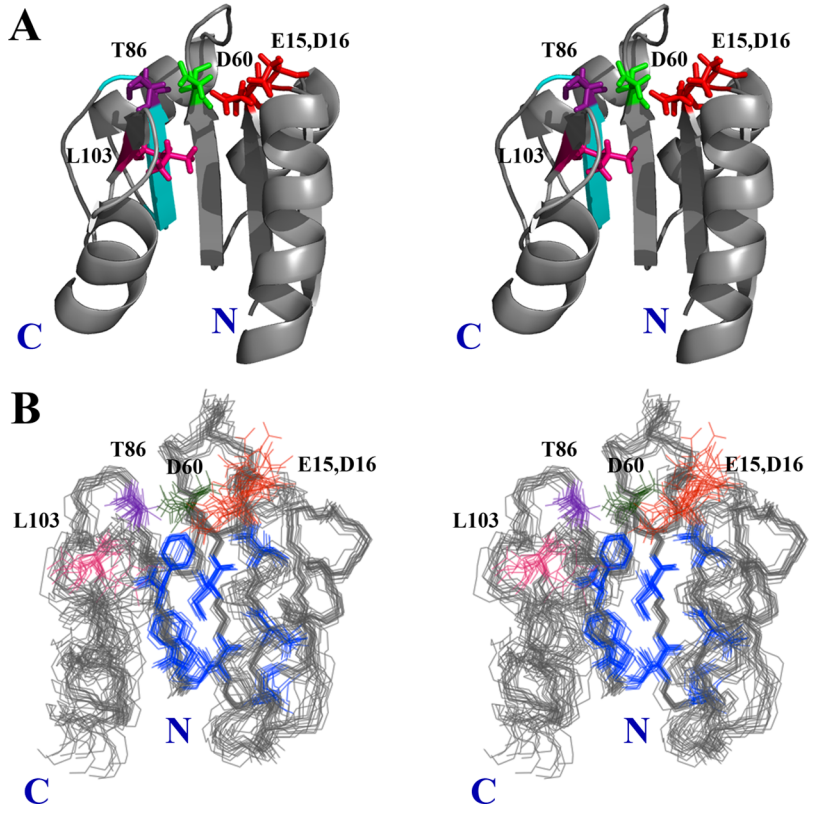


Figure 3. Active site of Sma0114. (A) Stereodiagram of the NMR structure of Sma0114 closest to the ensemble average, showing selected side chains of active-site residues: metal ligands Glu15 and Asp16 (red), phosphorylation-site Asp60 (green), and Thr86 and Leu103 of the Y-T coupling pair (purple and pink, respectively). The PFxFATGY motif is colored cyan. (B) Stereoview of the 20 lowest-energy NMR structures of Sma0114 illustrating the precision of side chains from different parts of the molecule. Side chains comprising the hydrophobic core (blue) are well defined, while those of active-site residues (color scheme described above) show poorer precision because of increased dynamics on the micro- to millisecond time scale.

dx.doi.org/10.1021/bi300922z | *Biochemistry* 2012, 51, 6932−6941

structures of Sma0114 are shown in Figure 2B. Structural statistics for the NMR models are listed in Table 1.

The NMR structure of Sma0114 closest to the ensemble average (Figure 2C) shows similarities to those of typical receiver domains like Spo0F (Figures 2D). The differences from the canonical receiver domain fold occur primarily along the 455 face of Sma0114 (Figure 2C). The most conspicuous difference is that helix α4 is replaced by a disordered segment (residues 89−93). Another difference is that strand β5, which is shorter than the other β-strands because of two flanking proline residues (P101 and P106), turns inward toward the core of the Sma0114 structure (Figure 2C).

To look more closely for similarities and differences with respect to other receiver domains, we used the DALI server[34] to find the 10 best structural matches to Sma0114. The 10 best matches were all structures of inactive states of receiver domains and had an average backbone rmsd of 3.17 Å from our Sma0114 structure over an average alignment length of 99 residues. The most prominent difference in Sma0114 is the absence of helix α4 (Figure 2C). Another important change is that the loop following strand β3 appears to cover and possibly restrict access to the phosphorylation-site residue, Asp60 (Figure 2C), whereas it extends away from the hydrophobic core in the homologous structures (Figure 2D).

The alignment with structurally homologous proteins using the DALI server[34] also indicates that the C-terminal portion of the PFxFATGY motif in Sma0114 differs from the corresponding region of the canonical receiver domain fold. The PFxFATGY motif is highlighted in cyan on the NMR structure of Sma0114 in Figure 3A. This motif encompasses all of strand β4 and forms the beginning of the loop between strands β4 and β5. The conserved Thr of the Y-T coupling pathway is part of this motif and is the last residue in strand β4 (colored purple in Figure 3A). The partner aromatic residue in the Y-T coupling mechanism is replaced with Leu103 in strand β5 (colored pink in Figure 3A). The absence of the conserved aromatic residue has also been seen in single-domain response regulator CheY2 from *S. meliloti*, an enzyme that has been reported to lack Y-T coupling.[35] These changes provide compelling evidence that the prototypical Y-T coupling mechanism is altered in Sma0114.

The two C-terminal residues of the PFxFATGY motif that form part of the loop immediately following strand β4 (G87 and Y88) show large structural differences in Sma0114 compared to the top 10 structural homologues. The two residues are located immediately before the start of the disordered segment that replaces helix α4 in Sma0114 and superpose with the N-terminal portion of helix α4 in the 10 closest structural matches. Together with the placement of the Y-T coupling threonine in the structure, this suggests that at least the C-terminal part of the PFxFATGY motif belongs to the 455 face of Sma0114.

While the 455 face of the enzyme shows deviations from the receiver domain superfamily, the active site of Sma0114 has features consistent with other receiver domains (Figure 3A). The metal binding residues, Glu15 and Asp16, are located at a conserved position in the loop between β1 and α1.[7,11,13] The phosphorylation site, Asp60, is at the C-terminal end of β3 like in other receiver domains. The hydroxylic residue, Thr86, that hydrogen bonds with the incoming phosphate in the Y-T coupling mechanism is located at the C-terminal end of strand β4. The basic residue, Lys105, which typically stabilizes the activated state by forming a salt bridge with the incoming phosphate, is in a conserved location in strand β5.

Strands β1−β4 and helices α1−α3 have the highest precision in the NMR structure (Figure 3B). Regions with lower precision occur in the 455 face and include the segments corresponding to helix α4, strand β5, and helix α5. The active site of Sma0114 also has lower precision in the NMR structure, as illustrated in Figure 3B, which compares side chains of residues in the active site and in the hydrophobic core of the protein. The equilibrium shift theory of activation hypothesizes that receiver domains exist in their inactive and active substates simultaneously in solution. Activation via phosphorylation shifts the population from the inactive to the active form.[14,16] It follows that the active site, which is experiencing exchange between inactive and active conformational substates on a micro- to millisecond time scale, would have a lower precision because of R_{2ex} line broadening. To verify that the lower precision of the active site and 455 face in the NMR structure of Sma0114 is due to genuine flexibility, as it is in other receiver domains,[7,36] we characterized the backbone dynamics of the protein using ^{15}N relaxation data.

NMR Relaxation Data Show Increased Flexibility for the 455 Face and Active Site. Relaxation data for Sma0114 are shown in Figure 4. The R_1 values are roughly constant over the length of the protein with a mean value of 2.00 ± 0.01 s^{-1}, except for increased rates in the region between residues 89 and 99, which also shows lowered ^{1}H−^{15}N NOEs (Figure 4A). The R_2 data reveal four distinct regions that have relaxation contributions greater than the mean value of 8.0 ± 0.3 s^{-1}, as indicated in Figure 4B. The mean ^{1}H−^{15}N NOE value is 0.75 ± 0.01, close to the theoretical maximum of 0.80.[37] Lowered ^{1}H−^{15}N NOEs are seen for the chain termini and the region between residues 88 and 95 that corresponds to the missing helix α4 (Figure 4C).

To interpret the relaxation data in terms of backbone dynamics, we performed a ModelFree analysis[26] using *Tensor2* (Figure 5). We obtained a correlation time of 5.1 ns for Sma0114, consistent with a monomeric protein of 13.5 kDa. The S^2 values that describe the amplitudes of internal motions on the pico- to nanosecond time scale in the ModelFree[26] analysis are shown in Figure 5A and are mapped on the NMR structure in Figure 5B. Except for the chain termini, the only region in the protein with low S^2 order parameters is the flexible segment that replaces helix α4 (Figure 5A). The R_{2ex} terms that describe exchange contributions to R_2 relaxation from dynamics on the micro- to millisecond time scale are shown in Figure 5C and are mapped on the NMR structure in Figure 5D. Residues that experience significant R_{2ex} contributions cluster to four regions of the protein: (I) the metal binding site, (II) the phosphorylation site, (III) Thr86, which is predicted to hydrogen bond with an incoming phosphate in the Y-T coupling mechanism, and (IV) Lys105, which stabilizes the incoming phosphate through a salt bridge. Thus, all four regions with R_{2ex} contributions are in the active site of the enzyme, which has been shown for other receiver domains to be subject to micro- to millisecond time scale interconversion between inactive and active substates as described in the equilibrium shift theory of activation.[15,16]

Despite their strong structural homology, receiver domains from various two-component systems exhibit different dynamics on the micro- to millisecond time scale. R_{2ex} terms for Spo0F, an intermediate in the sporulation relay of *Bacillus subtilis*, are larger for the metal binding site than for the 455

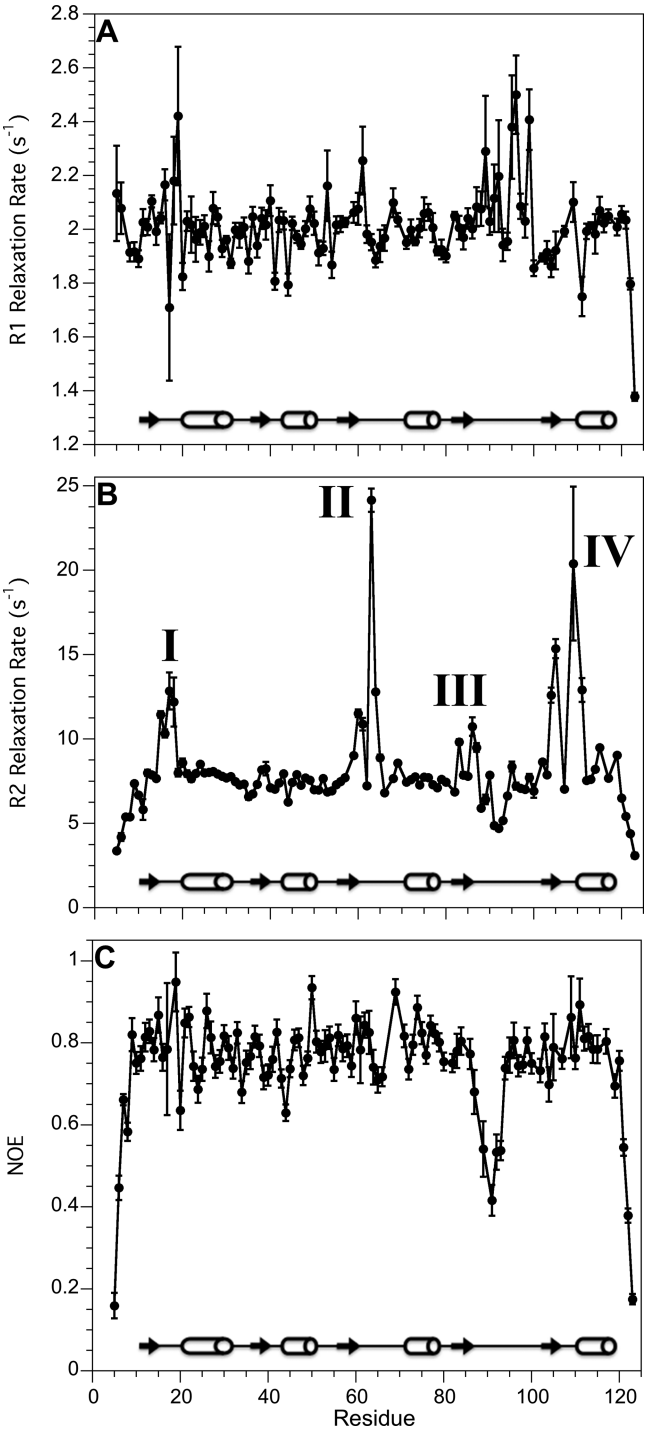


Figure 4. ^{15}N NMR relaxation data for Sma0114: (A) R_1 rates, (B) R_2 rates and (C) ^{1}H−^{15}N NOE values. Error bars are shown for all data points but in some cases are smaller than the symbols used to depict the data. The secondary structure of Sma0114 is indicated with arrows (β-sheets) and cylinders (α-helices) in each panel.

face of the enzyme.[15] Conversely, in NtrC, the nitrogen regulatory response regulator from *Salmonella typhimurium*, R_{2ex} contributions are large for the 455 face of the enzyme whereas the metal binding site shows no exchange broadening.[16,38] The R_{2ex} profile of Sma0114 differs from that of Spo0F and NtrC in that it shows exchange broadening in the metal-binding and phosphorylation sites but only in the β5 and α5 regions of the conserved 455 face.

The Family of Receiver Domains Associated with HWE-Kinases Is Predicted To Lack Helix α4. The NMR structure of Sma0114 is the first of an HWE-kinase associated response regulator. Because of the unusual features of this enzyme, we performed a bioinformatics analysis to determine if the absence of helix α4 is a conserved property of this subfamily. The structure prediction program PsiPred[33] was used to evaluate the secondary structure of 84 receiver domains with cognate HWE-kinase partners. Of the 84 proteins considered, 71 were predicted to lack helix α4 (Figure 6A). A subset of 13 HWE-associated receiver domains are predicted to have helix α4, but most of these appear to be missing the last helix α5 (Figure 6B). This suggests that the HWE-associated receiver domains in which helix α4 is present have compensatory perturbations that disrupt helix α5 of the 455 face. By contrast, all the α-helices and β-strands of the consensus Rossman fold are accurately predicted by PsiPred for a control group of eight receiver domains of known structure that are not associated with HWE-kinases (Figure 6C).

Sma0114 Binds Mg^{2+} and Ca^{2+}. We next wanted to see if divalent cations, which are required for the stabilization of the acyl−phosphate linkage,[14,39] are able to bind to Sma0114 in spite of the differences seen for the 455 face. Several groups have used Mg^{2+} to study the metal-bound states of receiver domains.[11,13,40] In the case of Sma0114, we found that Mg^{2+} binds very weakly, with a K_d near 75 mM and saturation at 150 mM Mg^{2+}. The $MgCl_2$ concentration required to stabilize the fully metal bound state leads to a decrease in sensitivity for NMR experiments because of the high solution ionic strength, especially for data collection with a cryogenic probe. An additional problem is that the high Mg^{2+} concentration needed to saturate Sma0114 causes nonspecific binding to clusters of acidic residues in the protein. Finally, Sma0114 is more susceptible to aggregation on the time scale of days to weeks needed for a detailed NMR characterization, in the presence of high (>100 mM) concentrations of $MgCl_2$.

Because of the problems associated with Mg^{2+}, we conducted an NMR titration using $CaCl_2$ as an alternative metal. Superimposed ^{1}H−^{15}N spectra of apo and Ca^{2+}-bound Sma0114 are shown in Figure 7A. In contrast to Mg^{2+}, which has a K_d of 75 mM (Figure 7B), Sma0114 binds Ca^{2+} with a K_d of ∼1 mM, and saturation is achieved at 1.5 mM $CaCl_2$ (Figure 7C). Residues that show the greatest chemical shift changes upon addition of Ca^{2+} are indicated with boxes in Figure 7A and for the most part cluster to the active site.

Figure 8 shows the chemical shift perturbations observed in ^{1}H−^{15}N HSQC spectra of Sma0114 due to addition of $CaCl_2$ (Figure 8A) and $MgCl_2$ (Figure 8B). The unusually high concentration of $MgCl_2$ necessary to achieve saturation induces nonspecific binding to acidic Asp and Glu residues as illustrated by the large number of chemical shift perturbations and line broadening effects observed (Figure 8B). By contrast, with $CaCl_2$ perturbations are mostly restricted to the active site of the enzyme (Figure 8A). Some residues become broadened rather than experiencing chemical shift changes upon addition of $CaCl_2$ or $MgCl_2$, and these are indicated with arrows in Figure 8. They include Glu17, which is one of the ligands for Ca^{2+}, and Lys105, which is part of the active site. Although most studies of metal-bound receiver domains have used Mg^{2+}, Ca^{2+} was used for structural studies of Spo0A and PhoP.[41,42]

dx.doi.org/10.1021/bi300922z | *Biochemistry* 2012, 51, 6932−6941

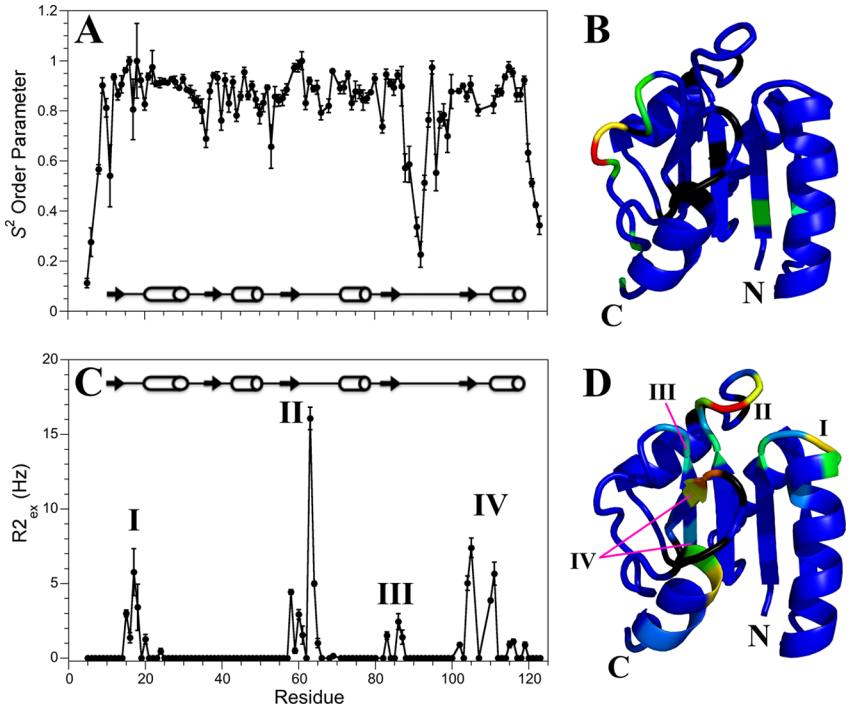


Figure 5. ModelFree analysis of Sma0114. (A) S^2 order parameters. (B) S^2 order parameters mapped onto the NMR structure of Sma0114. The color map goes from blue (rigid) to red (flexible). (C) R_{2ex} values. (D) R_{2ex} values mapped onto the NMR structure of Sma0114 with colors ranging from blue (negligible R_{2ex} contributions) to red (large R_{2ex} contributions). Residues for which no data were available are colored black.

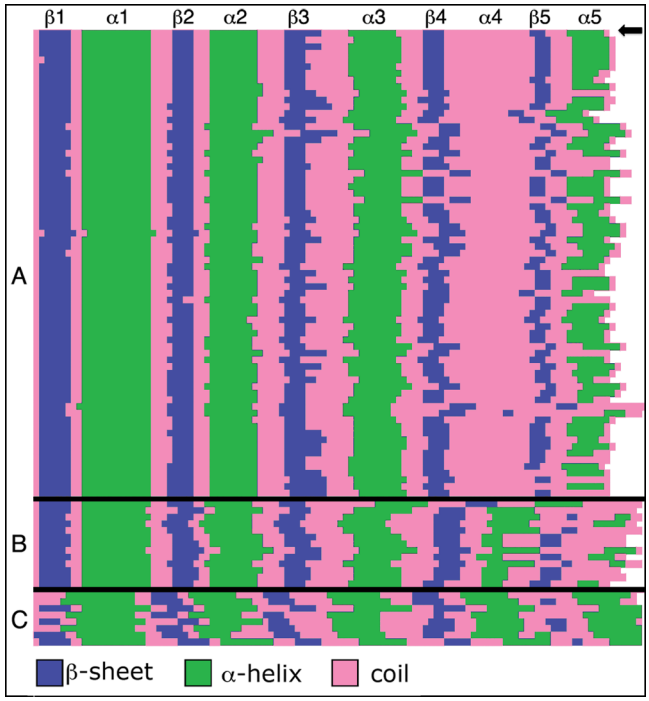


Figure 6. Secondary structure prediction of 84 HWE-associated receiver domains. Each row corresponds to one receiver domain. (A) Subset of 71 receiver domains with cognate HWE-kinases, all of which are predicted to be missing helix α4. Sma0114 is at the top of the figure labeled with an arrow. (B) Subset of 13 HWE-associated receiver domains predicted to have helix α4. Note that most of the members of this subset are predicted to be missing the alternate helix α5 of the 455 face. (C) Control group of eight receiver domains of known α_5/β_5 structures used to verify the secondary structure prediction algorithm.

DISCUSSION

Several groups have characterized the inactive and active conformations of receiver domains.[7,11,40] Understanding the structural features of these enzymes has provided a wealth of insight into their functions. Here, we extended the study of the receiver domain family to include an HWE-kinase-associated response regulator. The Sma0114 receiver domain, which may be prototypical of this class, has a conserved active site but shows substantial differences in the 455 face, which is used to propagate signal transduction intracellularly.

The structural and dynamic differences in the 455 face of Sma0114 suggest that this enzyme has an altered recognition interface for binding to downstream effectors. The PFxFATGY motif, which is unique to receiver domains associated with HWE-kinases, forms strand β4 in the Sma0114 structure (Pro81−Thr86) and the rigid part of the following loop (Gly87 and Tyr88). The last residue in the PFxFATGY motif demarcates the start of the Gly89−Leu93 segment, which replaces helix α4 in Sma0114, and shows low S^2 order parameters consistent with a dynamically disordered region. The only other receiver domain for which S^2 order parameters have been described is the inactive form of Spo0F.[15] In the case of Spo0F, it was shown that except for the chain termini, the protein is rigid on the pico- to nanosecond time scale. In typical receiver domains, helix α4 is a key element of the 455 face that undergoes concerted structural rearrangements following phosphorylation to form a new binding surface for downstream effector proteins.[11,36] Not only is helix α4 missing in Sma0114, but our secondary structure prediction analysis suggests that the absence of helix α4, or more rarely helix α5, is a conserved feature of HWE-associated receiver domains. A possible role for flexibility in the segment corresponding to helix α4 is to allow binding of downstream effectors through an induced fit

dx.doi.org/10.1021/bi300922z | *Biochemistry* 2012, 51, 6932−6941

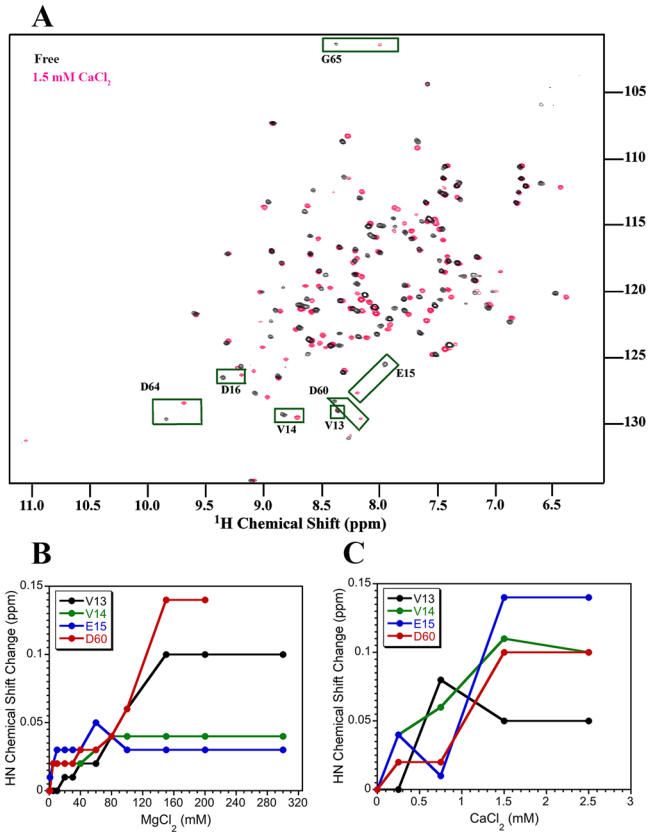


Figure 7. Metal binding of Sma0114 followed by NMR spectroscopy. (A) Superposition of $^{1}H-^{15}N$ HSQC spectra for the apo (black) and Ca^{2+}-bound forms of Sma0114 (pink). Active-site residues that show the largest chemical shift perturbations are indicated with green boxes. Titration curves for residues as a function of increasing $MgCl_2$ (B) or $CaCl_2$ (C) concentration. Note the different scale of the *x*-axis in panels B and C. Errors in chemical shift differences estimated from the digital resolution of the spectra were 0.02 ppm for panels B and C.

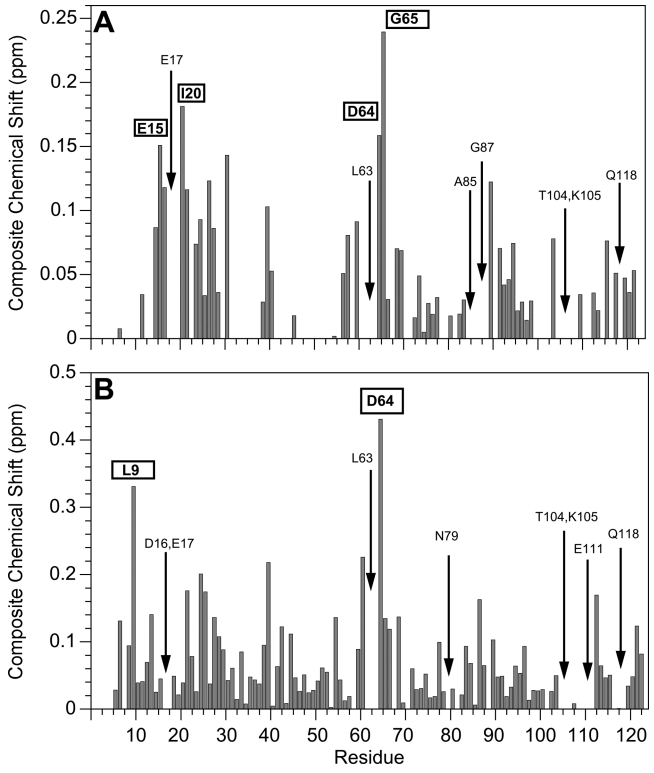


Figure 8. Perturbations of Sma0114 $^{1}H-^{15}N$ resonances due to metal binding. Bar graphs show composite HN and N chemical shift changes [$\Delta HN + 0.1(\Delta N)$] that occur in the presence of 1.5 mM $CaCl_2$ (A) or 150 mM $MgCl_2$ (B). Boxed residues indicate amino acids that experience large chemical shift perturbations upon metal binding. Arrows represent residues that broaden beyond detection in the presence of the metals. Errors in the composite chemical shift changes estimated from the digital resolution of the spectra were 0.03 ppm.

mechanism, thereby controlling the specificity of molecular association.

Residues Pro81−Gly87 of the PFxFATGY motif, which form strand β4 and the beginning of the following loop, while rigid on the pico- to nanosecond time scale have increased R_{2ex} contributions consistent with dynamics on the micro- to millisecond time scale. The exchange contributions to R_2 relaxation for these sites suggest that they experience the dynamic conformational equilibrium between the inactive and active substates of the enzyme, which is a hallmark of receiver domains.[7,14] The PFxFATGY sequence houses the conserved Thr86 of the Y-T coupling mechanism. The partnering aromatic residue for Y-T coupling in Sma0114 is replaced with Leu103. In a previous study of receiver domain CheY, it was shown that the β4−α4 loop and the N-terminal portion of helix α4 constitute the binding site for downstream effectors.[43] Because part of the PFxFATGY motif in Sma0114 occupies the same position in the structure as the β4−α4 loop region in CheY, it is likely that Sma0114 has a similarly reduced binding site for downstream effectors. The PFxFATGY sequence motif in Sma0114 may thus substitute or circumvent the Y-T coupling mechanism typically used in receiver domains. This change could be necessary in Sma0114 because of an altered 455 face, which is missing helix α4 and has additional structural differences in strand β5.

In spite of the differences in the 455 face, Sma0114 retains the ability to bind divalent metal ions needed for activation. The residues in Sma0114 that experience the largest chemical shift perturbations with increasing metal concentrations are consistent with the binding site predicted on the basis of sequence homology to other receiver domains (Figure 7). Like Spo0F, Sma0114 has a greater affinity for Ca^{2+} than Mg^{2+}.[44] The 1.5 mM concentration of Ca^{2+} required to saturate the metal binding site of Sma0114 is larger than the physiological concentration in bacteria.[45] A weak metal binding affinity could function to kinetically stabilize the activated state, by lowering the efficiency of the phosphatase enzymatic activity of the receiver domain, which in turn would effectively increase the lifetime of the phosphorylated active state.[46]

The conservation of active-site residues and the ability to bind metals suggest that the Sma0114 receiver domain can become activated by phosphorylation. The increased pliability of the 455 face may help Sma0114 achieve specificity in binding downstream effectors through an induced fit mechanism. We are now in a position to characterize the structure and dynamics of the metal-bound and activated states of Sma0114, which will allow us to obtain a more complete understanding of the structure−function relationships in the HWE-kinase-associated class of receiver domains.

dx.doi.org/10.1021/bi300922z | *Biochemistry* 2012, 51, 6932−6941

■ ASSOCIATED CONTENT

Supporting Information

One table listing the set of receiver domain sequences used for secondary structure prediction. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Department of Molecular and Cell Biology, University of Connecticut, 91 N. Eagleville Rd., Storrs, CT 06269-3125. Phone: (860) 486-4414. Fax: (860) 486-4331. E-mail: andrei@uconn.edu.

Present Address

†Castleton State College, Castleton, VT 05735.

Funding

This work was supported by a National Science Foundation Graduate Research Fellowship (GRFP) to S.R.S., a grant from the UConn Research Foundation (UCRF) to A.T.A. and D.J.G., and U.S. Department of Energy grants (DE-FG02-01ER15175 and DE-FG02-06ER15805) to D.J.G.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

DTT, dithiothreitol; HWE, histidine-tryptophan-glutamate; K_d, dissociation constant; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; R_{2ex}, chemical exchange contribution to transverse (R_2) relaxation; rmsd, root-mean-square deviation.

■ REFERENCES

(1) Robinson, V. L., Buckler, D. R., and Stock, A. M. (2000) A tale of two components: A novel kinase and a regulatory switch. *Nat. Struct. Biol. 7*, 626−633.

(2) Stock, J. B., Surette, M. G., McCleary, W. R., and Stock, A. M. (1992) Signal transduction in bacterial chemotaxis. *J. Biol. Chem. 267*, 19753−19756.

(3) Garcia, P. P., Bringhurst, R. M., Arango Pinedo, C., and Gage, D. J. (2009) Characterization of a two-component regulatory system that regulates succinate-mediated catabolite repression in *Sinorhizobium meliloti*. *J. Bacteriol. 192*, 5725−5735.

(4) Karniol, B., and Vierstra, R. D. (2004) The HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signaling. *J. Bacteriol. 186*, 445−453.

(5) Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) Two-component signal transduction. *Annu. Rev. Biochem. 69*, 183−215.

(6) Stock, A. M., Mottonen, J. M., Stock, J. B., and Schutt, C. E. (1989) Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature 337*, 745−749.

(7) Gardino, A. K., Volkman, B. F., Cho, H. S., Lee, S. Y., Wemmer, D. E., and Kern, D. (2003) The NMR solution structure of BeF_3^--activated Spo0F reveals the conformational switch in a phosphorelay system. *J. Mol. Biol. 331*, 245−254.

(8) Foreman, R., Fiebig, A., and Crosson, S. (2012) The LovK-LovR Two-Component System Is a Regulator of the General Stress Pathway in *Caulobacter crescentus*. *J. Bacteriol. 194*, 3038−3049.

(9) Jenal, U., and Galperin, M. Y. (2009) Single domain response regulators: Molecular switches with emerging roles in cell organization and dynamics. *Curr. Opin. Microbiol. 12*, 152−160.

(10) Paul, R., Jaeger, T., Abel, S., Wiederkehr, I., Folcher, M., Biondi, E. G., Laub, M. T., and Jenal, U. (2008) Allosteric regulation of histidine kinases by their cognate response regulator determines cell fate. *Cell 133*, 452−461.

(11) Hastings, C. A., Lee, S. Y., Cho, H. S., Yan, D., Kustu, S., and Wemmer, D. E. (2003) High-resolution solution structure of the beryllofluoride-activated NtrC receiver domain. *Biochemistry 42*, 9081−9090.

(12) Perez, E., Samper, S., Bordas, Y., Guilhot, C., Gicquel, B., and Martin, C. (2001) An essential role for phoP in *Mycobacterium tuberculosis* virulence. *Mol. Microbiol. 41*, 179−187.

(13) Cho, H. S., Lee, S. Y., Yan, D., Pan, X., Parkinson, J. S., Kustu, S., Wemmer, D. E., and Pelton, J. G. (2000) NMR structure of activated CheY. *J. Mol. Biol. 297*, 543−551.

(14) Bourret, R. B. (2010) Receiver domain structure and function in response regulator proteins. *Curr. Opin. Microbiol. 13*, 142−149.

(15) Feher, V. A., and Cavanagh, J. (1999) Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. *Nature 400*, 289−293.

(16) Gardino, A. K., and Kern, D. (2007) Functional dynamics of response regulators using NMR relaxation techniques. *Methods Enzymol. 423*, 149−165.

(17) Sheftic, S. R., Garcia, P. P., Robinson, V. L., Gage, D. J., and Alexandrescu, A. T. (2010) NMR assignments for the *Sinorhizobium meliloti* response regulator Sma0114. *Biomol. NMR Assignments 5*, 55−58.

(18) Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, Skelton, N. J., and Rance, M. (2006) *Protein NMR Spectroscopy Principles and Practice*, 2nd ed., Elsevier Inc., Amsterdam.

(19) Jaravine, V. A., Alexandrescu, A. T., and Grzesiek, S. (2001) Observation of the closing of individual hydrogen bonds during TFE-induced helix formation in a peptide. *Protein Sci. 10*, 943−950.

(20) Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR 13*, 289−302.

(21) Case, D. A., Dyson, H. J., and Wright, P. E. (1994) Use of chemical shifts and coupling constants in nuclear magnetic resonance structural studies on peptides and proteins. *Methods Enzymol. 239*, 392−416.

(22) Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuthrich, K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry 28*, 7510−7516.

(23) Brunger, A. T. (1996) *Xplor Version 3.8. A System for Crystallography and NMR*, Yale University Press, New Haven, CT.

(24) Nicholson, L. K., Kay, L. E., and Torchia, D. A. (1994) Protein Dynamics as studied by solution NMR techniques. In *NMR Spectroscopy and its Application to Biomedical Research* (Sarkar, S., Ed.) Elsevier Science Publishers, Amsterdam.

(25) Alexandrescu, A. T., and Shortle, D. (1994) Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *J. Mol. Biol. 242*, 527−546.

(26) Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Validity. *J. Am. Chem. Soc. 104*, 4546−4559.

(27) Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR 16*, 23−28.

(28) Letunic, I., Doerks, T., and Bork, P. (2009) SMART 6: Recent updates and new developments. *Nucleic Acids Res. 37*, D229−D232.

(29) Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res. 32*, 1792−1797.

(30) Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics 22*, 2688−2690.

(31) Whelan, S., and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol. 18*, 691−699.

(32) Ulrich, L. E., and Zhulin, I. B. (2007) MiST: A microbial signal transduction database. *Nucleic Acids Res. 35*, D386−D390.

(33) Buchan, D. W., Ward, S. M., Lobley, A. E., Nugent, T. C., Bryson, K., and Jones, D. T. (2010) Protein annotation and modelling servers at University College London. *Nucleic Acids Res. 38*, W563−W568.
(34) Holm, L., and Rosenstrom, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res. 38*, W545−W549.
(35) Riepl, H., Scharf, B., Schmitt, R., Kalbitzer, H. R., and Maurer, T. (2004) Solution structures of the inactive and BeF_3-activated response regulator CheY2. *J. Mol. Biol. 338*, 287−297.
(36) Peters, G. (2009) The effect of Asp54 phosphorylation on the energetics and dynamics in the response regulator protein Spo0F studied by molecular dynamics. *Proteins 75*, 648−658.
(37) Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry 28*, 8972−8979.
(38) Volkman, B. F., Lipson, D., Wemmer, D. E., and Kern, D. (2001) Two-state allosteric behavior in a single-domain signaling protein. *Science 291*, 2429−2433.
(39) Kojetin, D. J., Thompson, R. J., Benson, L. M., Naylor, S., Waterman, J., Davies, K. G., Opperman, C. H., Stephenson, K., Hoch, J. A., and Cavanagh, J. (2005) Structural analysis of divalent metals binding to the *Bacillus subtilis* response regulator Spo0F: The possibility for in vitro metalloregulation in the initiation of sporulation. *BioMetals 18*, 449−466.
(40) Stock, A. M., Martinez-Hackert, E., Rasmussen, B. F., West, A. H., Stock, J. B., Ringe, D., and Petsko, G. A. (1993) Structure of the Mg^{2+}-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry 32*, 13375−13380.
(41) Lewis, R. J., Muchova, K., Brannigan, J. A., Barak, I., Leonard, G., and Wilkinson, A. J. (2000) Domain swapping in the sporulation response regulator Spo0A. *J. Mol. Biol. 297*, 757−770.
(42) Menon, S., and Wang, S. (2011) Structure of the response regulator PhoP from *Mycobacterium tuberculosis* reveals a dimer through the receiver domain. *Biochemistry 50*, 5948−5957.
(43) Guhaniyogi, J., Robinson, V. L., and Stock, A. M. (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. *J. Mol. Biol. 359*, 624−645.
(44) Feher, V. A., Tzeng, Y. L., Hoch, J. A., and Cavanagh, J. (1998) Identification of communication networks in Spo0F: A model for phosphorylation-induced conformational change and implications for activation of multiple domain bacterial response regulators. *FEBS Lett. 425*, 1−6.
(45) Dominguez, D. C. (2004) Calcium signalling in bacteria. *Mol. Microbiol. 54*, 291−297.
(46) Feher, V. A., Zapf, J. W., Hoch, J. A., Whiteley, J. M., McIntosh, L. P., Rance, M., Skelton, N. J., Dahlquist, F. W., and Cavanagh, J. (1997) High-resolution NMR structure and backbone dynamics of the *Bacillus subtilis* response regulator, Spo0F: Implications for phosphorylation and molecular recognition. *Biochemistry 36*, 10015−10025.
(47) Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187−217.
(48) Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR 8*, 477−486.
(49) Madhusudan, M., Zapf, J., Hoch, J. A., Whiteley, J. M., Xuong, N. H., and Varughese, K. I. (1997) A response regulatory protein with the site of phosphorylation blocked by an arginine interaction: Crystal structure of Spo0F from *Bacillus subtilis*. *Biochemistry 36*, 12739−12745.