

# The *Escherichia coli* two-component signal sensor BarA binds protonated acetate via a conserved hydrophobic-binding pocket

Received for publication, October 3, 2021, and in revised form, October 27, 2021. Published, Papers in Press, November 4, 2021.  
<https://doi.org/10.1016/j.jbc.2021.101383>

Adrián F. Alvarez<sup>1</sup>, Claudia Rodríguez, Ricardo González-Chávez<sup>1</sup>, and Dimitris Georgellis<sup>\*1</sup>

From the Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, México City, México

Edited by Chris Whitfield

The BarA/UvrY two-component signal transduction system is widely conserved in  $\gamma$ -proteobacteria and provides a link between the metabolic state of the cells and the Csr post-transcriptional regulatory system. In *Escherichia coli*, the BarA/UvrY system responds to the presence of acetate and other short-chain carboxylic acids by activating transcription of the noncoding RNAs, CsrB and CsrC, which sequester the RNA-binding protein CsrA, a global regulator of gene expression. However, the state of the carboxyl group in the acetate molecule, which serves as the BarA stimulus, and the signal reception site of BarA remain unknown. In this study, we show that the deletion or replacement of the periplasmic domain of BarA and also the substitution of certain hydroxylated and hydrophobic amino acid residues in this region, result in a sensor kinase that remains unresponsive to its physiological stimulus, demonstrating that the periplasmic region of BarA constitutes a functional detector domain. Moreover, we provide evidence that the protonated state of acetate or formate serves as the physiological stimulus of BarA. In addition, modeling of the BarA sensor domain and prediction of the signal-binding site, by blind molecular docking, revealed a calcium channels and chemotaxis receptors domain with a conserved binding pocket, which comprised uncharged polar and hydrophobic amino acid residues. Based on the comparative sequence and phylogenetic analyses, we propose that, at least, two types of BarA orthologues diverged and evolved separately to acquire distinct signal-binding properties, illustrating the wide adaptability of the bacterial sensor kinase proteins.

The BarA/UvrY two-component system (TCS) is a central element for the modulation of the Csr posttranscriptional regulatory system that allows  $\gamma$ -proteobacteria, such as *Escherichia coli*, to coordinate numerous physiological processes, including carbon metabolism, motility, biofilm formation, peptide uptake, and virulence (1, 2). This TCS comprises the transmembrane sensor histidine kinase BarA, which belongs to the subclass of hybrid-tripartite histidine kinases (3, 4) and its cognate response regulator UvrY (5). Upon signal

perception, BarA autophosphorylates at the expense of ATP and transphosphorylates the response regulator UvrY (5, 6). The phosphorylated form of UvrY (UvrY-P), in turn, activates the transcription of the CsrB and CsrC noncoding RNAs (2, 7), which possess repeated sequence elements that allow them to interact with multiple molecules of the CsrA protein and prevent its interaction with mRNA targets (8, 9). CsrA is an RNA-binding protein that directly interacts with the 5' untranslated leaders of mRNA targets and controls gene expression by regulating their translation, stability, and/or elongation (10–14).

In an earlier study, we reported that the *E. coli* BarA protein senses the presence of acetate (15), which accumulates in the growth media as the cells metabolize acetogenic substrates, such as glucose (16), or other short-chain carboxylic acids, such as formate and propionate (15). However, the region of BarA that detects the physiological stimulus has not yet been identified. Nevertheless, the sensing domain of the BarA ortholog protein in *Pseudomonas aeruginosa*, GacS, was proposed to be in its periplasmic domain (PD). The structure of this domain was recently solved (17), revealing that it relies on a PhoQ, DcuS, CitA (PDC)/Per-ARNT-Sim (PAS)-like folding that includes an essential positively charged pocket, which was proposed to accommodate a negatively charged ligand. The PAS domains are widely distributed modules that are involved in sensing environmental and nutritional signals, such as metabolites, gases, light or redox potential (18–20) and are characterized by a conserved  $\beta$ -sheet core flanked by  $\alpha$ -helices and loops that provide ligand-binding specificity (18, 21). PDC/PAS-like domains are extracytoplasmic domains (22–24) and are often found as signal-recognition modules of sensor kinases or chemoreceptors (25). The PDC/PAS-like domains, which are closely related to the PAS superfamily, belong to the Cache (calcium channels and chemotaxis receptors) sensory domain superfamily (26). Various Cache sensors that detect carboxylate-harboring signals have been functionally and/or structurally characterized, and it has been demonstrated that all of them interact with the ligand in its anionic state through the formation of salt bridges with basic residues that constitute a positively charged pocket (17, 21, 22, 24, 27–30).

\* For correspondence: Dimitris Georgellis, [dimitris@ifc.unam.mx](mailto:dimitris@ifc.unam.mx).

## Identification of the sensor domain of BarA

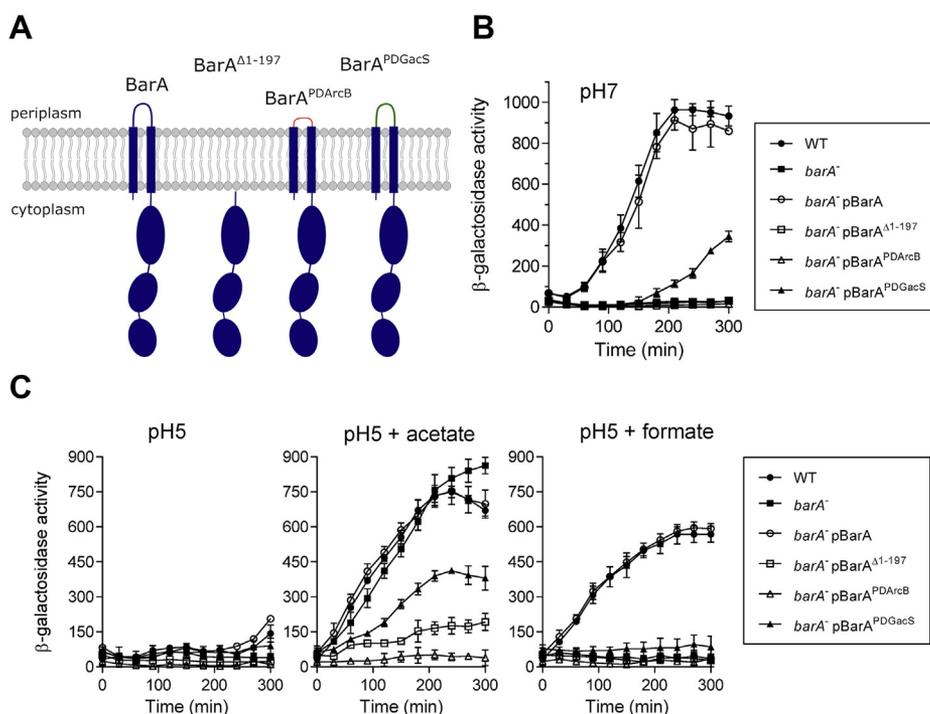
The present work aims to elucidate the role of the periplasmic region of the BarA HK in signal perception, by genetic, biochemical, and bioinformatic methods. We show that the PD of BarA is needed for proper switching between its phosphatase and kinase activities in response to the physiological stimulus and thereby for regulating the activity of the BarA/UvrY-signaling pathway. Moreover, we demonstrate that the putative signal-binding pocket does not rely on positively charged residues but instead on uncharged polar and hydrophobic residues. We, also, provide evidence that the protonated state of formic acid and acetic acid constitute the physiological stimuli for the BarA sensor kinase. Finally, our phylogenetic analysis of the PD of BarA ortholog proteins revealed that they could be grouped into two classes depending on the amino acid residue conservation pattern of the putative ligand-binding site, suggesting that the members of these two BarA groups have evolved to perceive different stimuli.

## Results and discussion

### The periplasmic region of BarA is required for activation of its kinase activity

It has been previously reported that short-chain carboxylic acids, such as formate and acetate, act as the physiological stimuli for BarA (15). However, the BarA signal reception site

remains elusive. Because the periplasmic region of most sensor kinases constitutes the sensor domain, we explored the functionality of the periplasmic segment of BarA. To this end, a low-copy number plasmid expressing the cytosolic part of the BarA protein ( $\text{BarA}^{\Delta 1-197}$ ), lacking the transmembrane (TM) and the periplasmic domains, and two plasmids carrying *barA* alleles in which the periplasmic segment of the protein was replaced by the corresponding section of either the *E. coli* ArcB HK ( $\text{BarA}^{\text{PD}_{\text{ArcB}}}$ ) or the *Azotobacter vinelandii* GacS HK ( $\text{BarA}^{\text{PD}_{\text{GacS}}}$ ) were constructed (Fig. 1A). ArcB was chosen because its periplasmic bridge is unusually short (16 aa residues), and it does not participate directly in the reception of any signal (31), whereas GacS is the BarA homolog of Pseudomonads, the activating signal of which remains unknown. The above constructed plasmids were transformed into the  $\Delta\text{barA}$  strain IFC5036, carrying a  $\lambda\Phi(\text{csrB-lacZ})$  fusion, and the activity of the BarA variants was analyzed by monitoring the *in vivo* levels of phosphorylated UvrY, as indicated by the expression of the *csrB-lacZ* reporter, which depends directly on the activity of the BarA/UvrY TCS (2). The transformants were grown in LB medium buffered at pH 7.0, and the *csrB-lacZ* expression was monitored. As to be expected, reporter expression was activated at the transition from exponential to stationary phase of growth in both the WT strain and the *barA* mutant harboring a plasmid-borne WT *barA* allele, whereas no activation of



**Figure 1. Testing the importance of the BarA periplasmic domain.** A, schematic representation of the BarA HK and the truncated or chimeric BarA proteins used in this study. B, overnight cultures of the isogenic strains KSB837 (WT) (filled circles), IFC5036 (*barA*<sup>-</sup>) (filled squares), and IFC5036 (*barA*<sup>-</sup>) carrying plasmid pEXT22-*barA* (pBarA) (open circles), pEXT22-*barA*cyt (pBarA <sup>$\Delta 1-197$</sup> ) (open squares), pEXT22-*barA*PDarcB (pBarA <sup>$\text{PD}_{\text{ArcB}}$</sup> ) (open triangles), or pEXT22-*barA*PDgacS (pBarA <sup>$\text{PD}_{\text{GacS}}$</sup> ) (filled triangles), were diluted to an  $A_{600}$  of 0.001 in LB medium. When the cultures reached an  $A_{600}$  of 0.15, the  $\beta$ -galactosidase activity was followed for 300 min. C, the cultures of the same strains described for (B) were grown in LB medium, the pH of which had been adjusted and buffered to 5.0. At an  $A_{600}$  of 0.15, a sample was withdrawn (time = 0 min), and the culture was divided into three parts: one of them continued its incubation without any supplement (left panel), whereas acetate (7 mM) (central panel) or formate (7 mM) (right panel) was added to the other cultures, and the  $\beta$ -galactosidase activity was followed. The averages from three independent experiments are presented, and the SDs (error bars) are indicated. PD, periplasmic domain.

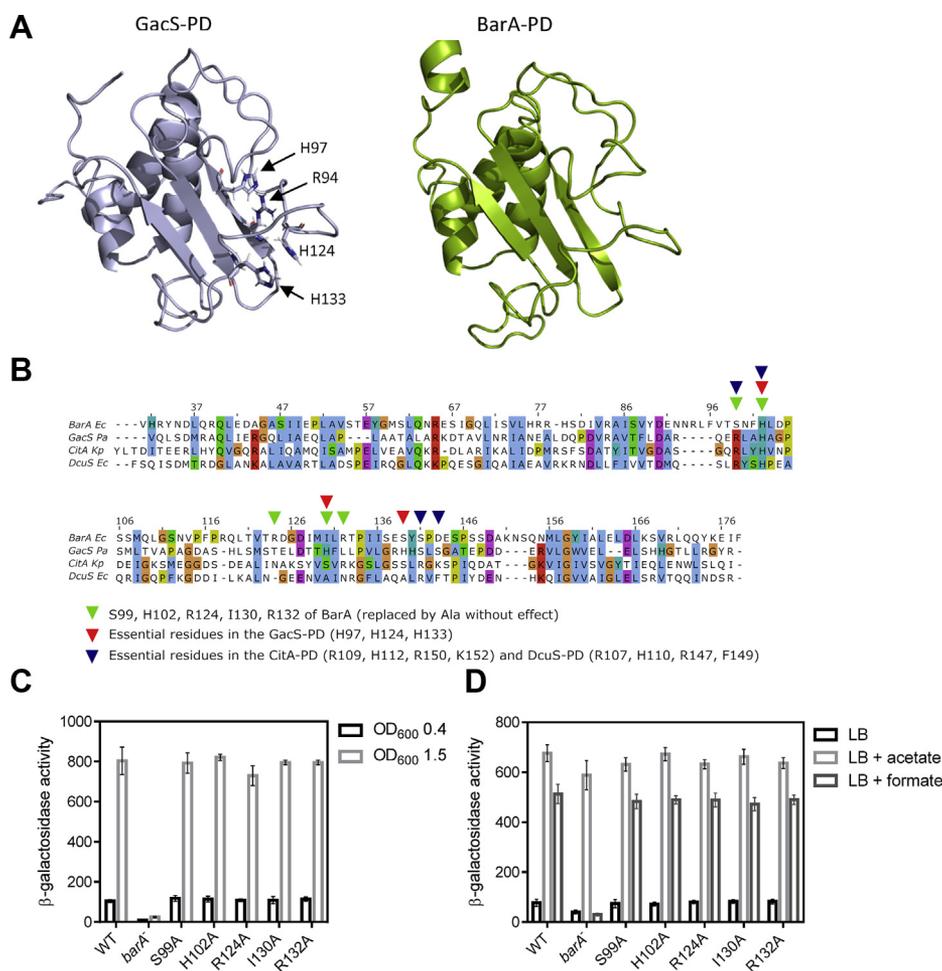
reporter expression was observed in the *barA*<sup>-</sup> strain (Fig. 1B). Interestingly, the construct where the BarA PD was replaced by the corresponding GacS region resulted in a delayed increase of reporter expression, reaching ~30% of the WT expression values (Fig. 1B). Thus, the PD of GacS may be able to respond to acetate, although less efficiently than the corresponding region of BarA, or it responds to another stimulus that is present at the transition from exponential to the stationary phase of growth. However, the possibility that structural distortions of the BarA-GacS hybrid do not allow its proper activation cannot be discarded. On the other hand, when the BarA TM and periplasmic domains were removed, or when the BarA periplasmic bridge was replaced by the ArcB counterpart, no activation of the *csrB-lacZ* expression was noted (Fig. 1B), suggesting that either these BarA variants are inactive proteins, or that they are unable to switch from their phosphatase to their kinase activity upon entry into the stationary phase of growth. To confirm that each BarA-hybrid variant remains functional, we evaluated the *csrB-lacZ* expression in the above strains grown in LB buffered at pH 5.0. Such a condition provides an environment that does not allow the activation of BarA (32) unless acetate or formate is added to the growth medium (15). It is relevant to mention that both acetate and formate act through BarA, whereas acetate can also activate *csrB* expression by promoting the acetyl-P-dependent phosphorylation of UvrY in the absence of BarA (15). This fact permits the evaluation of both the kinase and the phosphatase activities of BarA and its variants (33). As expected, no activation of the *csrB-lacZ* expression was observed in any of the tested strains at pH 5.0 (Fig. 1C), whereas addition of acetate or formate to the growth medium resulted in the immediate activation of the reporter expression in the WT strain and in the *barA* mutant harboring a plasmid-borne WT *barA* allele (Fig. 1C). Also, addition of acetate, but not formate, resulted in the activation of reporter expression in the  $\Delta barA$  mutant strain (Fig. 1C), due to acetyl-P-dependent UvrY phosphorylation, in agreement with previous reports (15, 34). On the other hand, no activation of the *csrB-lacZ* expression was observed in the strains expressing the BarA <sup>$\Delta$ 1-197</sup> or the BarA<sup>PDArcB</sup> proteins, indicating that these BarA variants fail to respond to acetate and formate but retain their phosphatase activity and are therefore functional proteins that remain locked in their phosphatase state. Moreover, activation of the *csrB-lacZ* expression was slower and reduced by about 2-fold in the BarA<sup>PDGacS</sup> expressing strain upon acetate addition, whereas no activation of reporter expression was observed after formate addition (Fig. 1C). Thus, the BarA<sup>PDGacS</sup> chimeric HK appears to comprise an impaired phosphatase activity. Alternatively, it may perceive a stimulus (other than acetate) whose production increases at pH 5.0 in the presence of acetate and is also produced at pH 7.0 in the stationary phase of growth. This is consistent with previous reports in which soluble metabolic molecules produced at the end of the exponential growth, such as tricarboxylic acid cycle intermediates, were proposed to activate the GacS sensor kinase

(35, 36). Nevertheless, the above results demonstrate that the PD of the *E. coli* BarA is required for proper switching between its phosphatase and kinase activities in response to the physiological stimulus.

### Structural modeling of the BarA periplasmic domain and identification of putative signal-binding residues

To explore the structural features of the putative BarA sensor domain, we generated a structural model of the BarA PD (residues 31–178 of BarA), based on the NMR structure of the *P. aeruginosa* GacS-PD (Protein Data Bank code: 5O7J) (17), using the Iterative Threading Assembly Refinement server (37) (Fig. 2A). The model shows the typical architecture of a group of signal-recognition domains that adopt a particular  $\alpha/\beta$  fold known as PDC/PAS-like domain (Fig. 2A), initially described for the sensor domains of the PhoQ, DcuS, and CitA histidine kinases (22–24). The PDC/PAS-like domain, currently regarded as Cache domain (26), is characterized by a central three-to-five antiparallel  $\beta$ -sheet flanked by an N-terminal  $\alpha$ -helix and additional loops and helices on each side (25). To identify residues implicated in signal sensing, we performed a comparative sequence analysis of the BarA-PD with the sensor domains of GacS, DcuS, and CitA, all of which contain well characterized positively charged binding pockets that allow the interaction with negatively charged ligands (Fig. 2B). The ligand binding pockets include residues Arg94, His97, His124, and His133 in GacS; Arg107, His110, and Arg147 in DcuS; and Arg109, His112, Arg150, and Lys152 in CitA. From these basic amino acid residues, the only one that appears to be conserved in the BarA-PD sequence is His102 (corresponding to His97 of GacS, His110 of DcuS, and His112 of CitA), whereas Ser99 seems to be at the position of the conserved Arg residue present in GacS (Arg94), DcuS (Arg107), and CitA (Arg109). It is important to mention that this pair of basic residues was shown to be essential for signal binding in DcuS and CitA (30, 38), whereas only His97 was needed for GacS signaling (17). Curiously, the other residues that constitute the positively charged pocket required for ligand binding in GacS (17), DcuS (21, 30) and CitA (38), were not found in the BarA sequence. Nevertheless, based on their positive charge and/or their relative position in the sequence alignment with known PDC/PAS-like sensor domains, the residues Ser99, His102, Arg124, Ile130, and Arg132 of BarA were selected to test whether they could be involved in signal reception. To this end, five strains carrying the *barA*<sup>S99A</sup>, *barA*<sup>H102A</sup>, *barA*<sup>R124A</sup>, *barA*<sup>I130A</sup>, or *barA*<sup>R132A</sup> mutant allele and the  $\lambda\Phi$ (*csrB-lacZ*) transcriptional fusion were generated. Alanine was chosen because it is uncharged, and it is not bulky enough to cause major structural distortions in the putative-binding pocket. The generated mutant strains, IFC5038 (*barA*<sup>S99A</sup>), IFC5039 (*barA*<sup>H102A</sup>), IFC5040 (*barA*<sup>R124A</sup>), IFC5041 (*barA*<sup>I130A</sup>), and IFC5042 (*barA*<sup>R132A</sup>) along the WT and  $\Delta barA$  strains were grown in LB buffered at pH 7.0 to an absorbance at 600 nm ( $A_{600}$ ) of 0.4 (nonstimulatory condition) or to an  $A_{600}$  of 1.5

## Identification of the sensor domain of BarA



**Figure 2. Analysis of conserved residues and structural model of the BarA periplasmic domain.** *A*, superposed solution structure of the GacS sensor domain (left) and predicted structure for the BarA-PD (right). The 3D structural model of BarA-PD was generated by using the program I-Tasser. The relative position of conserved residues involved in GacS signaling (17) is indicated. *B*, sequence alignment of the BarA-PD with related PDC/PAS-like containing sensor domains. The Clustal X color-scheme was used to visualize the residue conservation patterns. The numbers at the top of the alignment indicate the amino acid positions in full-length BarA. The position of essential residues in the GacS or in the Cita/DcuS sensor domains are indicated by red triangles or by blue triangles, respectively. The residues replaced by Ala in the BarA-PD are indicated by green triangles. *C*, the strains KSB837 (WT), IFC5035 (*barA*<sup>-</sup>), IFC5038 (*barA*<sup>S99A</sup>), IFC5039 (*barA*<sup>H102A</sup>), IFC5040 (*barA*<sup>R124A</sup>), IFC5041 (*barA*<sup>I130A</sup>), and IFC5042 (*barA*<sup>R132A</sup>) were grown in LB medium to an A<sub>600</sub> of 0.4 (nonstimulatory conditions, black bars) or 1.5 (stimulatory conditions, gray bars), and the β-galactosidase activity was measured. The average from three independent experiments is presented, and the SDs (error bars) are indicated. *D*, the strains KSB837 (WT), IFC5035 (*barA*<sup>-</sup>), IFC5038 (*barA*<sup>S99A</sup>), IFC5039 (*barA*<sup>H102A</sup>), IFC5040 (*barA*<sup>R124A</sup>), IFC5041 (*barA*<sup>I130A</sup>), and IFC5042 (*barA*<sup>R132A</sup>) were grown in LB medium buffered at pH 5.0. At an A<sub>600</sub> of 0.15, the culture was split in three, and 7 mM acetate (light light-gray bars) or formate (dark-gray bars) was added to two of them, whereas the third one was used as a control (black bars), and the cultures were incubated for 180 min before the samples were withdrawn for β-galactosidase quantification. The average and SDs (error bars) from three independent experiments are shown. Ec, *Escherichia coli*; I-Tasser, Iterative Threading Assembly Refinement; Kp, *Klebsiella pneumoniae*; Pa, *Pseudomonas aeruginosa*; PD, periplasmic domain.

(stimulatory condition), and the *csrB-lacZ* reporter expression was measured. It was found that substitution of Ser99, His102, Arg124, Ile130, or Arg132 to Ala did not affect reporter expression (Fig. 2C). Next, to test whether these substitutions affect the BarA phosphatase activity, the expression of the *csrB-lacZ* reporter was measured in the WT and mutant strains grown at pH 5.0 in the presence or the absence of either acetate or formate (Fig. 2D). No difference between the WT and the five BarA mutant variants was noted, indicating that none of the above amino acid residues is essential for BarA signal reception. In addition, these results suggest that the BarA sensor domain does not rely on a positively charged ligand-binding pocket, providing an important difference between the BarA sensor domain and other structural

related sensor domains, including the one of the *P. aeruginosa* GacS.

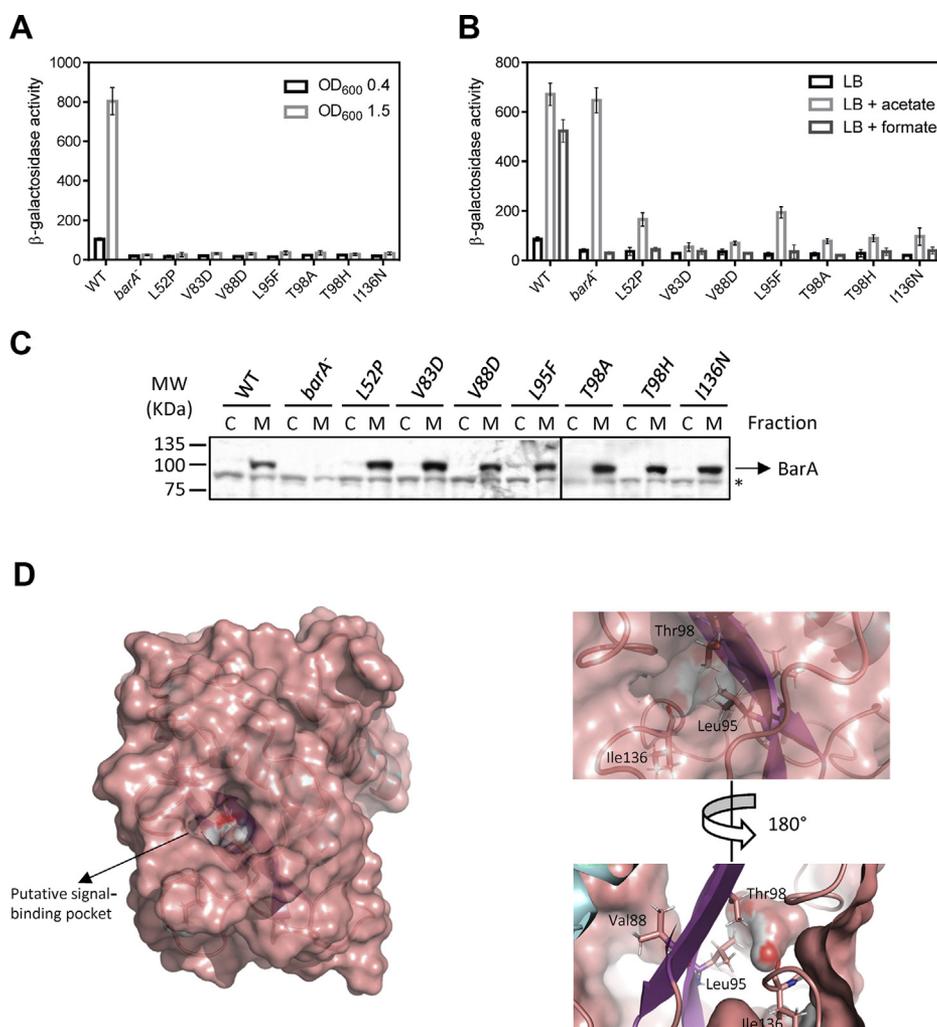
### Identification of amino acid residues required for proper BarA signaling

Attempts to identify specific amino acid residues within the BarA periplasmic portion that may participate in signal reception, by an unbiased approach, were then pursued. The DNA region coding the periplasmic portion of BarA was mutagenized by error-prone PCR, and the PCR products were used to replace the corresponding region of the *barA* gene that was previously cloned into the pACYCDuet-1 plasmid. In this construct, *barA* is under the control of a T7 promoter, whose

## Identification of the sensor domain of BarA

basal activity results in nearly WT BarA levels, as indicated by Western blot analysis (Fig. S1). The pool of plasmids was first screened for loss of kinase activity mutants, by streaking individual colonies onto LB agar at pH 7.0 supplemented with X-gal and selecting white or light-blue colonies. The selected clones were further screened for phosphatase active *barA* alleles by streaking them onto LB agar at pH 5.0 supplemented with X-gal and 7 mM acetate and selecting white or light-blue colonies. Sequencing of the selected clones identified seven alleles harboring single-point missense mutations, namely *barA*<sup>L52P</sup>, *barA*<sup>V83D</sup>, *barA*<sup>V88D</sup>, *barA*<sup>L95F</sup>, *barA*<sup>T98A</sup>, *barA*<sup>T98H</sup>, or *barA*<sup>I136N</sup>. To validate the kinase activity of these mutant *barA* alleles, the cells of WT,  $\Delta barA$ , and  $\Delta barA$  harboring either of the seven mutant plasmid-borne *barA* alleles were grown in LB buffered at pH 7.0 to nonstimulatory ( $A_{600}$  of 0.4)

or stimulatory ( $A_{600}$  of 1.5) conditions, and the *csrB-lacZ* reporter expression was measured. It was found that the *csrB-lacZ* reporter activation under stimulatory conditions occurred only in the WT but not in the  $\Delta barA$  or the  $\Delta barA$  harboring any of the seven mutant plasmid-borne *barA* alleles (Fig. 3A). To verify that the mutant BarA variants are functional proteins, the strains carrying the WT or mutant BarA alleles were grown at pH 5.0, and their *csrB-lacZ* expression was measured in the presence or the absence of acetate or formate (Fig. 3B). As anticipated, no activation of reporter expression was observed in any of the tested strains in the absence of acetate or formate, whereas the *csrB-lacZ* expression increased in the  $\Delta barA$  strain only in the presence of acetate, due to acetyl-P dependent UvrY phosphorylation (Fig. 3B). On the other hand, in contrast to the WT strain,



**Figure 3. Identification and spatial disposition of essential residues in the BarA sensor domain.** *A*, the cells of strains KSB837 (WT), IFC5035 (*barA*<sup>-</sup>), and IFC5035 carrying pACYCDuet-1 derivative plasmids that harbor the *barA* mutant variants were grown in LB buffered at pH 7.0 to midexponential growth phase (absorbance ~0.4) (nonstimulatory conditions, black bars) or to an  $A_{600}$  of 1.5 (stimulatory conditions, gray bars) and the  $\beta$ -galactosidase activity was measured. *B*, the strains KSB837 (WT), IFC5035 (*barA*<sup>-</sup>), and IFC5035 carrying pACYCDuet-1 derivative plasmids harboring the *barA* mutant variants were grown in LB medium buffered at pH 5.0; at an  $A_{600}$  of 0.15, the culture was split in three, and 7 mM acetate (light-gray bars) or formate (dark-gray bars) was added to two of them, whereas the third one was used as a control (black bars), and the samples for  $\beta$ -galactosidase quantification were withdrawn after 180 min. In (*A*) and (*B*), the average from three independent experiments is presented, and the SDs (error bars) are indicated. *C*, the BarA protein (102.5 kDa) levels in cytosolic (C) and membrane (M) fractions of strains KSB837 (WT), IFC5035 (*barA*<sup>-</sup>), and IFC5035 carrying pACYCDuet-1 derivative plasmids harboring the *barA* mutant variants, as determined by Western blot analyses using BarA polyclonal antibodies. \* indicates a nonspecific signal that was observed in all stains and fractions. *D*, a surface model depicting the predicted BarA signal-binding pocket. The whole view of the BarA-PD (left) and zoomed-in views of the putative binding pocket (right) are presented. The residues Leu95, Thr98, and I136 required for BarA signaling are indicated.

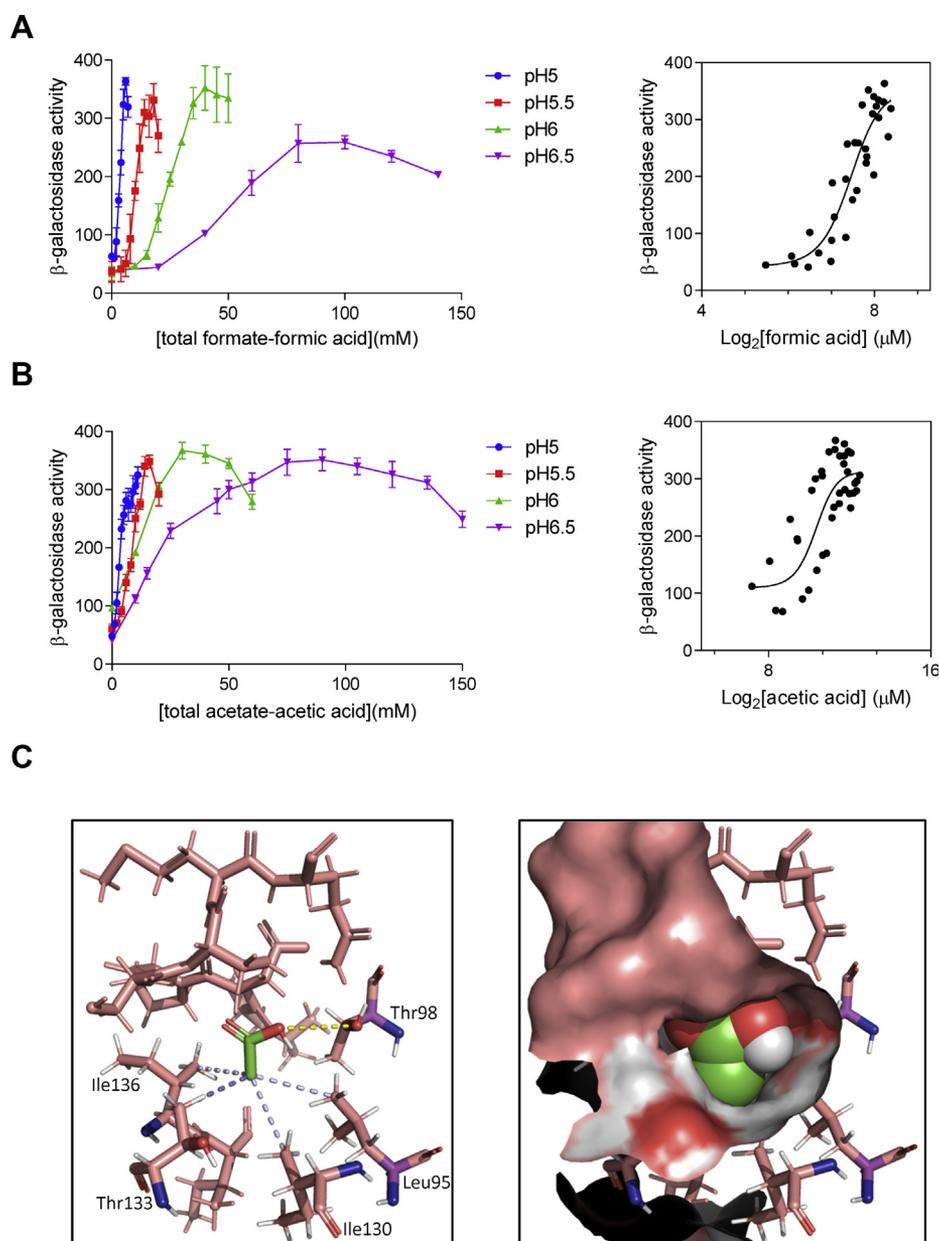
## Identification of the sensor domain of BarA

where the *csrB-lacZ* expression was activated in the presence of acetate or formate, the seven-BarA mutants were unable to respond to formate. In addition, in the presence of acetate, the V83D, V88D, T98F, T98H, and I136N BarA mutants showed a 7- to 10-fold lower *csrB-lacZ* expression than the WT or the  $\Delta$ barA strain, whereas the L52P or L95F BarA variants showed an almost 4-fold lower *csrB-lacZ* expression. To ascertain that each BarA mutant protein is properly expressed and remains membrane associated, we performed Western blot analyses on cytosolic and membrane fractions of the cells. In all cases, the mutant proteins were found to be expressed at nearly WT BarA levels and were associated with the cytoplasmic membrane (Fig. 3C). It can, therefore, be concluded that the selected BarA mutants are functional proteins that retain their UvrY-P phosphatase activity despite the presence of the stimulus. Next, we examined the relative position of the identified residues in the above-presented structural model of BarA-PD. It was noted that Leu95 and Thr98 are located in the second  $\beta$  strand, overlapping the position of Arg94 and His97 in the GacS-PD structure, whereas Ile136 is predicted to be located in the major loop between the second and the third  $\beta$  strand (Fig. 3D), and the side chains of these three amino acid residues appear to occupy the putative ligand-binding pocket (Fig. 3D), suggesting that they could be directly involved in signal binding. On the other hand, Val83 and Val88 are located in the first  $\beta$  strand but with their hydrophobic side chains facing to the opposite side with respect to the predicted signal binding site. The fact that the replacement of these two residues by Asp in the BarA protein resulted in a constitutive phosphatase phenotype could be explained by structural distortions caused by the negative-charged aspartate residue that may impact the conformation of the binding pocket. However, the possibility that the true orientation of these residues differs from the one predicted by the model cannot be excluded. In addition, the predicted position of Leu52 falls outside the putative ligand binding-pocket and would not be expected to be directly involved in acetate binding. Instead, the substitution of Leu52 by Pro may give rise to structural distortions, which affect the orientation of the whole periplasmic domain. Finally, we generated two additional BarA-PD structural models, based on the sensor domains of DcuS and CitA and explored the relative position of the above-mentioned BarA residues. As expected, Leu95, Thr98, and Ile130 of BarA overlap with residues Arg107, His110, and Arg147 of DcuS, and Arg109, His112, and Arg150 of CitA that are needed for ligand binding in these two proteins (Fig. S2). It is relevant to mention that the discrepancy in identified amino acid residues, important for ligand binding in BarA between the amino acid alignment (Fig. 2B) and the structure-based alignment (Fig. S2) highlights the limitation of approaches based on the sequence alignments for the analysis of poorly conserved domains. Indeed, the BarA-PD shares 19%, 22%, and 18% identity with the Cache domains of GacS, DcuS, and CitA at the amino acid sequence level, respectively. Taken together, our results indicate that the Cache folded periplasmic region of BarA is a functional detector domain and implicate residues Leu95, Thr98, and Ile136 to be directly involved in signal binding.

Curiously, these amino acid residues are not basic, as would be expected for a positively charged pocket able to bind a ligand bearing an anionic carboxylate, such as acetate or formate, but instead have polar hydroxyl group or hydrophobic side chains, leading us to reconsider the nature of the physiological stimulus of BarA.

### The protonated (neutral) state of formic and acetic acids provides the stimulus for the BarA/UvrY TCS

We previously demonstrated that the end products formate and acetate provide a physiological stimulus for BarA, and that an accessible carboxylate group in the signal molecule is essential for BarA activation (15). Now, we found that the putative ligand-binding pocket in the sensor domain of BarA seems to be comprised of hydrophobic and hydroxylated residues, rather than positively charged residues. This fact prompted us to speculate that the protonated state of acetate, acetic acid, or of formate, formic acid, may provide the physiological stimuli for the BarA-sensor kinase. To explore this possibility, the effect of increasing concentrations of formate to growth media buffered at different pHs on the activity of BarA/UvrY was examined. The following conditions were tested: LB medium buffered at (i) pH 5.0 in the presence of 0, 1, 2, 3, 4, 5, 6, or 7 mM formate, (ii) pH 5.5 in the presence of 0, 4, 6, 8, 10, 12, 14, 16, 18, or 20 mM formate, (iii) pH 6.0 in the presence of 0, 10, 15, 20, 25, 30, 35, 40, 45, or 50 mM formate, and (iv) pH 6.5 in the presence of 0, 20, 40, 60, 80, 100, 120, or 140 mM formate. In all cases, the cells were grown to an optical density at  $A_{600}$  of 0.15 before the indicated amount of formate was added to the culture medium, and the samples were withdrawn after 60 min for  $\beta$ -galactosidase activity determination. It was observed that the higher the pH, the greater amount of formate was needed to activate *csrB-lacZ* reporter expression (Fig. 4A). Moreover, when the obtained  $\beta$ -galactosidase activities were plotted against the  $\log_2$  of the formic acid concentration, calculated by using the Henderson–Hasselbalch equation and pKa values of formic acid/formate of 3.75, a correlation between the *csrB-lacZ* expression and the formic acid concentration was observed (Fig. 4A), revealing a threshold concentration of 6.5 to 7  $\mu$ M formic acid to be required for BarA activation, regardless of the pH. To provide further support to this conclusion, we carried out a similar experiment but now adding acetate to the growth media. In this case, to prevent acetyl-P-dependent UvrY phosphorylation, a strain in which the acetyl-P synthetic pathways were blocked by deletion of the acetate kinase (*ackA*) and phosphotransacetylase (*pta*) genes was used. The *ackA::Tet<sup>r</sup>::pta* mutant strain, carrying the *csrB-lacZ* reporter, was grown in LB medium buffered at pH 5.0, 5.5, 6.0, or 6.5, to an absorbance at  $A_{600}$  of 0.15. Thereafter, the indicated amount of acetate was added to the growth medium, and the cultures were incubated for 60 min before the samples were withdrawn for  $\beta$ -galactosidase quantification (Fig. 4B). It was observed that a higher concentration of acetate was needed to activate reporter expression at higher pH (Fig. 4B), and the  $\beta$ -galactosidase activity values correlated with the calculated acetic

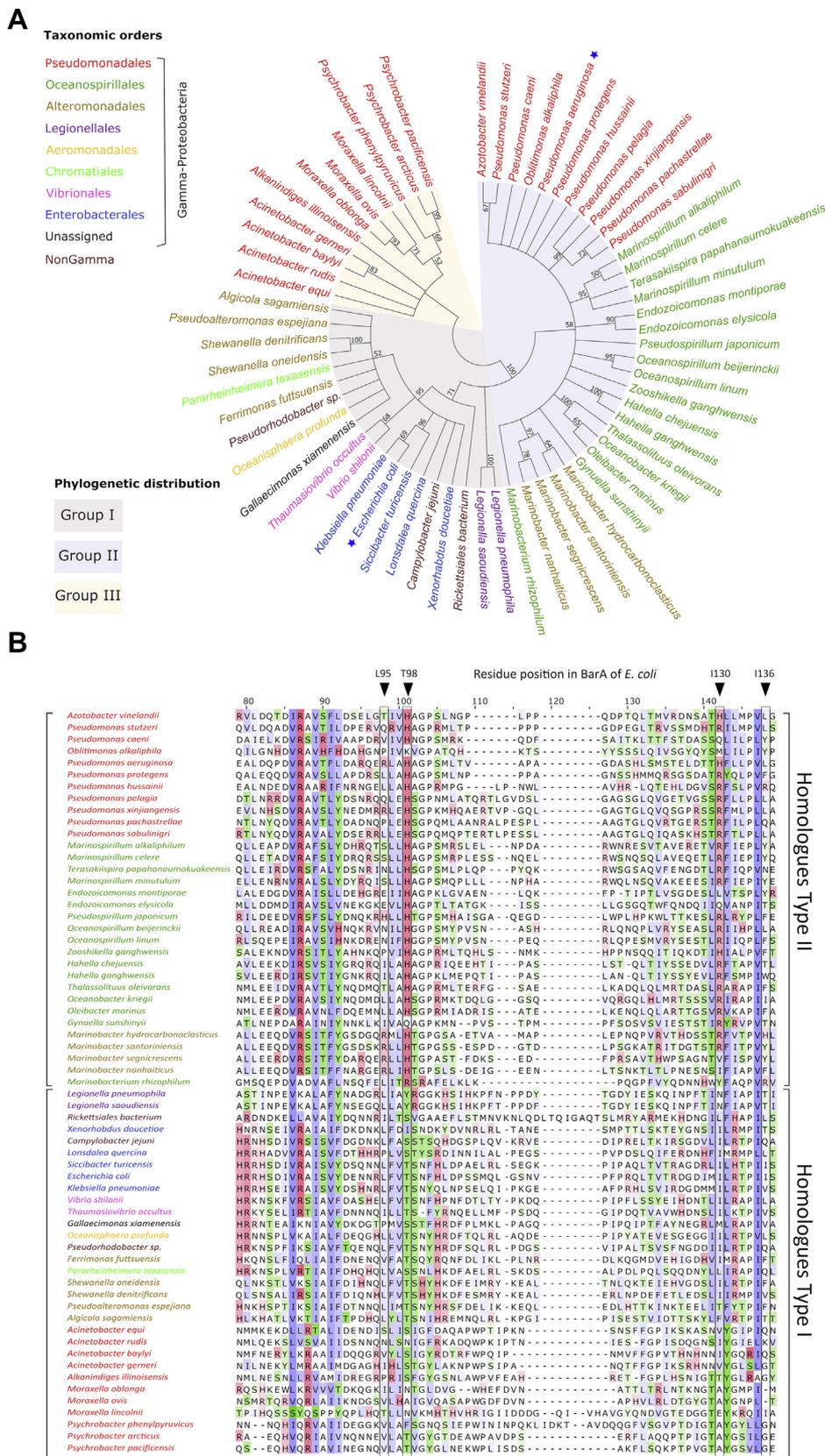


**Figure 4. The neutral state of formic and acetic acids provides the stimulus of BarA.** *A*, the strain KSB837 (WT) was grown in LB medium buffered at pH 5.0, 5.5, 6.0, or 6.5 to an  $A_{600}$  of 0.15, and a designated amount of formate was added to the culture medium (0, 1, 2, 3, 4, 5, 6, or 7 mM for the culture at pH 5.0; 0, 4, 6, 8, 10, 12, 14, 16, 18, or 20 mM for the culture at pH 5.5; 0, 10, 15, 20, 25, 30, 35, 40, 45, or 50 mM for the culture at pH 6.0; and 0, 20, 40, 60, 80, 100, 120, or 140 mM for the culture at pH 6.5). The cultures were incubated for 60 min before the samples were withdrawn for  $\beta$ -galactosidase quantification. *B*, the strain KSB-ackA-pta was grown in LB medium buffered at pH 5.0, 5.5, 6.0, or 6.5 to an  $A_{600}$  of 0.15, and a designated amount of acetate was added to the culture medium (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11 mM for the cultures at pH 5.0; 0, 2, 4, 6, 8, 10, 12, 14, 16, or 20 mM for the culture at pH 5.5; 0, 10, 20, 30, 40, 50, or 60 mM for the culture at pH 6.0; and 0, 10, 15, 25, 30, 45, 50, 60, 75, 90, 105, 120, 135, or 150 mM for the culture at pH 6.0). The cultures were incubated for 60 min before the samples were withdrawn for  $\beta$ -galactosidase quantification. *Left panels*, the  $\beta$ -galactosidase activity is plotted against the total concentration of formate – formic acid (*A*) or acetate – acetic acid (*B*) in the cultures grown at pH 5.0 (*blue*), pH 5.5 (*red*), pH 6.0 (*green*), and pH 6.5 (*purple*). The average from three independent experiments is presented, and the SDs (*error bars*) are indicated (*Right panels*). The nonlinear regression-fitting curve of the  $\beta$ -galactosidase activity plotted against the binary logarithm of the concentration of formic acid (*A*) or acetic acid (*B*). For clarity, only the average from the three independent experiments is presented in these panels. *C*, (*left panel*) position and atomic interaction of the acetic acid molecule into the binding site of the BarA sensor domain, as predicted by cavity detection and molecular docking. The acetic acid molecule is shown in a *stick representation* with carbon atoms colored *green* and oxygen atoms colored *red*. Only the BarA residues comprising the acetic acid-binding cavity are shown. The *dashed lines* represent the predicted hydrogen bond (*yellow*) and the hydrophobic interactions (*light blue*) between the acetic acid molecule and amino acid residues Leu95, Thr98, Ile130, Thr133, and Ile136 of BarA. *Right panel*, occupancy of the cavity by the ligand at the predicted BarA-binding pocket. The cavity of the BarA-PD is illustrated in a *surface style*, and the acetic acid molecule is shown in a *spheres representation*.

acid concentration ( $\text{pK}_{\text{a, acetic acid/acetate}}$ : 4.75), providing further support to the conclusion that the protonated state of acetate or formate provide the physiological stimulus of BarA. This is consistent with the occurrence of a noncharged-binding

pocket in the sensor domain of BarA. To strengthen our conclusion, we performed an acetic acid binding site prediction on the BarA-PD model, by molecular docking enhanced by cavity detection (39, 40). The acetic acid molecule was

# Identification of the sensor domain of BarA



**Figure 5. Sequence analysis and phylogenetic distribution of the putative sensor domain of BarA orthologous proteins.** A, phylogenetic tree based on the amino acid sequences of the periplasmic domains of BarA homologs. The tree was arbitrarily rooted and condensed so that only bootstrap support values higher than 50% are shown. The branches were grouped and shaded different colors according to the phylogenetic distribution. The taxon names of bacterial species are color coded according to the taxonomic order designation. The position of BarA of *Escherichia coli* and GacS of *Pseudomonas aeruginosa* in the tree is marked with a blue star. B, multiple sequence alignment of the 63 nonredundant periplasmic

docked into a cavity, which comprised the residues Leu95, Thr98, Leu103, Asp104, Pro105, Ser106, Ser107, Met108, Gln109, Ile130, Thr133, Ile135, and Ile136 (Fig. 4C), where the hydroxyl group of Thr98 associates with the protonated carboxylic group of acetic acid through hydrogen bonding, and the hydrophobic side chains of Leu95, Ile130, Thr133, and Ile136 are predicted to interact with the carbon chain of the organic acid (Fig. 4C). Taken together, our results suggest that a functional region within the BarA PD, which consists of a noncharged binding pocket, is involved in acetic acid or formic acid binding.

### Differential evolution of the BarA orthologs

The BarA/UvrY TCS of *E. coli* exerts global regulatory effects on gene expression by activating CsrB and CsrC transcription, thereby controlling CsrA activity (2, 9, 41). This signaling circuitry is highly conserved in  $\gamma$ -proteobacteria (42–44), preserving most of its general characteristics. However, the genes whose expression are modulated by the BarA/UvrY/CsrA regulatory cascade and hence the physiological response, may vary significantly among bacteria (10, 45–51). Here, we found that the input domain of the BarA/UvrY system of *E. coli* differs from that of the homologous GacS/GacA of *P. aeruginosa*, in that the former has evolved to recognize an uncharged signal. To gain further insight into the evolution and the diversity of the putative sensor domains of BarA orthologous proteins, we examined the phylogenetic relationships of the PDs in a group of species selected from fully sequenced bacteria encoding an identifiable BarA ortholog. About 500 deduced protein sequences were acquired from the GenBank database, using the Basic Local Alignment Search Tool (52), and the sequence of the BarA and GacS periplasmic domains as query. An unrooted maximum-likelihood phylogenetic tree was inferred from the nonredundant sequences (identity lower than 50%), which included members of nine gamma, two alpha, and one epsilon proteobacterial orders. The phylogenetic reconstruction showed that BarA orthologues could be classified into three main groups. Firstly, the bacteria belonging to the order Enterobacteriales cluster with the orders Vibrionales, Aeromonadales, Chromatiales, Legionellales, and various families from the order Alteromonadales in addition to the members of alpha and epsilon proteobacteria (group I) (Fig. 5A). Secondly, the BarA/GacS sensor domain seems to have early diverged within Pseudomonadales, because bacteria within the family Pseudomonadaceae were clustered together with members of the orders Oceanospirillales and Alteromonadales (group II) (Fig. 5A). Finally, genera belonging to the family Moraxellaceae was shown to be gathered in a separated branch (group III) (Fig. 5A). Moreover, a multiple-sequence alignment of the 63 nonredundant BarA-PD sequences revealed important differences in the putative signal-binding pocket of the three afore-mentioned phylogenetic groups. Particularly, most of the BarA homologs

clustered into group II, represented by GacS of *P. aeruginosa*, are characterized by a conserved residue pattern that contains the positively charged residues His and/or Arg at two specific positions (Fig. 5B). This pair of highly conserved residues (His97 and His133) has been shown to be essential for GacS/GacA signaling in *P. aeruginosa* (17). In contrast, members of group I, represented by the BarA of *E. coli*, contain highly conserved Ser/Thr/Tyr residues at the first position (Thr98 in BarA of *E. coli*) and Ile/Leu/Val residues at the second position (Ile130 in BarA of *E. coli*), (Fig. 5B). It is worth mentioning that although the replacement of Ile130 by Ala did not have an effect on the BarA function, this residue could be involved in signal detection (Figs. 4C and S2), because it is highly conserved in the BarA homologs Type I and because alanine is also a hydrophobic side-chain residue. Finally, the amino acid pattern of the putative-binding pocket of the BarA proteins clustered in group III is closer related to that of Type I because most of them contain conserved hydroxylated and hydrophobic residues in the two positions, suggesting that the detector domain of these HKs could also interact with noncharged signals. Thus, the differential residue conservation pattern of the putative signal-binding pocket, taken together with our genetic analysis for the BarA-PD and with the previously reported characterization of the GacS-PD (17), strongly suggest that Type I and Type II BarA proteins diverged and evolved separately to detect different kind of signal molecules.

### Conclusions

We previously showed that the end products formate and acetate constitute the physiological signal for the activation of the BarA/UvrY TCS of *E. coli*. In this work, we explored the role of the periplasmic region of the BarA HK in stimulus perception. Our results demonstrate that the PD of BarA constitutes a functional domain that is required for activation of the BarA–UvrY signaling pathway in response to the physiological stimulus. Our functional analyses combined with protein structure prediction and molecular docking indicate that the putative ligand-binding pocket of the BarA PD does not depend on positively charged residues, as would be expected for the stabilization of a negatively charged carboxyl group. Instead, it was found that the ligand-binding pocket of the BarA PD depends on polar hydroxyl groups or hydrophobic side chains, and that the neutral state of the carboxyl group moiety of acetic or formic acid provides the stimulus for the BarA sensor kinase, providing an explanation for the above observation. Nevertheless, the conclusion that acetic acid and formic acid directly activate BarA remains to be further supported by biochemical analysis with the purified sensor domain. Thus far, we have been unsuccessful in demonstrating the binding of acetic or formic acid to the purified periplasmic domain of BarA. One possible explanation could be the fact that the purified protein was shown to be insoluble at pHs below 7.0 (data not shown), frustrating the binding

domains of BarA orthologs. For clarity, only the central portion of the alignment is shown. The boxes indicate conserved residues that are predicted to be involved in signal binding, and their position in the sequence of BarA of *E. coli* is shown. The brackets enclose bacterial taxa and sequences grouped according to the nature of these residues (see text).

## Identification of the sensor domain of BarA

experiments at pHs at which the concentration of the protonated carboxylic acid were high enough to observe the interaction. However, the immediate response to the stimuli and the fact that acetic and formic acid are not readily converted to any common intermediary metabolite strongly suggest that their related chemical structures permit them to signal directly to BarA. Finally, our phylogenetic analysis prompted us to propose that the two families of BarA orthologous proteins have evolved separately to acquire differential signal-binding properties by their sensor domains, permitting these HKs to perceive different kinds of signal molecules, thereby highlighting the plasticity of this family of signaling proteins.

### Experimental procedures

#### Bacterial strains, plasmids, and growth conditions

The *E. coli* strains and plasmids used in this work are listed in Table S1. Strain IFC5035 ( $\Delta barA::Kan^r$  *csrB-lacZ*) was constructed by homologous recombination using the lambda-red recombinase system (53). Briefly, a PCR-amplified fragment, using primers barAdel-Fw and barAdel-Rv (Table S2) and plasmid pKD4 (53) as the template, was used to replace the *barA* gene with a kanamycin (Kan) cassette in strain KSB837 (WT, *csrB-lacZ*) (7). Then, the FRT-flanked  $Kan^r$  cassette was removed from the strain IFC5035 using the Flp recombinase encoded in the temperature-sensitive plasmid pCP20 (54), obtaining the strain IFC5036 ( $\Delta barA$  *csrB-lacZ*).

To construct plasmid pUC18-*barA*, the *barA* coding sequence was amplified by PCR using the primers barA-NdeI-Fw and BAfullR-Hind (Table S2), and the chromosomal DNA of strain KSB837 as a template. The PCR product was digested with NdeI-HindIII and cloned between the NdeI-HindIII sites of plasmid pUC18 (55), resulting in the plasmid pUC18-*barA*. To construct the low-copy number plasmid pEXT22-*barA*, which carries the *barA* native promoter and ribosomal binding site, an introduced NdeI site that includes the initiation codon of *barA* and the *barA* open reading frame, a 122-bp DNA fragment containing the transcriptional regulatory region of the *barA* gene was PCR-amplified using the primers BarA100up-Fw and BA5R (Table S2) and the chromosomal DNA of strain KSB837 as a template. The PCR product was digested with MluI-XbaI and cloned into the MluI-XbaI sites of a modified pEXT22 vector (56) (the NdeI site of pEXT22 was destroyed by cleaving, filling in, and ligation). Subsequently, the *barA* open reading frame was isolated as a 2.8 Kb NdeI-HindIII fragment from the plasmid pUC18-*barA* and inserted into the NdeI and HindIII sites of the above construct, resulting in the plasmid pEXT22-*barA*.

Bacteria were routinely cultured at 37 °C in LB medium. The LB agar medium was prepared by the addition of 1.5% (w/v) agar. When required, the LB medium was buffered at pH 7.0 with 100 mM 3-(N-morpholino)propanesulfonic acid; at pH 6.5, pH 6.0, and pH 5.5 with 100 mM 2-(N-morpholino)ethanesulfonic acid; and at pH 5.0 with 100 mM homopiperazine-1,4-bis(2-ethanesulfonic acid). The used concentration of the buffers was sufficient to maintain a constant pH

throughout all the experiments. When necessary, the growth medium was supplemented with chloramphenicol (20 µg/ml), kanamycin (50 µg/ml), ampicillin (100 µg/ml), or tetracycline (10 µg/ml). For semiquantitative detection of the *csrB-lacZ* reporter expression, LB agar was supplemented with 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (40 µg/ml). When indicated, acetate and formate were added to the growth media at a concentration of 7 mM, unless otherwise stated.

#### Generation of plasmids expressing cytosolic or chimeric BarA proteins

To generate a low-copy number plasmid expressing the cytosolic portion of the BarA protein (BarA<sup>Δ1-197</sup>), a PCR amplified DNA fragment, using the primers BarA5'(198-x) and BAfullR-Hind (Table S2) and chromosomal DNA of strain KSB837 as a template, was used to replace the NdeI-HindIII restriction fragment of pEXT22-*barA* to generate pEXT22-*barA*cyt. A *barA-gacS* fusion, encoding a hybrid protein in which the periplasmic segment of BarA was replaced by the corresponding section of GacS of *A. vinelandii*, was constructed by a three-step PCR procedure (57). Firstly, a 453-bp DNA fragment was PCR-amplified using the primers ChGS-perip-Fw and ChGS-perip-Rv (Table S2) and chromosomal DNA of *A. vinelandii* as a template. Secondly, the purified PCR product was used in combination with plasmid pEXT22-*barA* as templates in two PCR reactions, one of which using the primers GS-perip-Rv and BarA100up-Fw, and the other one containing primers GS-perip-Fw and BAfullR-Hind (Table S2), to amplify a 559-bp DNA and a 2664-bp DNA product, respectively. Finally, the full-length *barA-gacS* fusion was PCR-amplified by using the primers BarA100up-Fw and BAfullR-Hind and the purified PCR products from the previous step as templates. The resulting 2867-bp DNA product was digested with NdeI and HindIII, and the 2.7 Kb fragment was gel-purified and used to replace the NdeI-HindIII restriction fragment of pEXT22-*barA* to generate pEXT22-*barA*PDgacS. To construct the *barA-arcB* fusion, encoding a hybrid protein in which the periplasmic segment of BarA was replaced by the short periplasmic bridge of ArcB, a PCR reaction was performed using primers Arc-Bar-peripl-Fw and Arc-Bar-peripl-Rv (Table S2) and pEXT22-*barA* as a template. The 7508-bp length PCR product was purified and circularized by self-ligation to generate pEXT22-*barA*PDarcB.

#### Generation of chromosomal *barA* mutants by site-directed mutagenesis

To generate KSB837 derivative strains harboring chromosomal *barA* mutant alleles, a series of plasmids were created as follows. Plasmid pUC18-Cam was constructed by cloning a 1033-bp PCR-amplified fragment containing an FRT-flanked chloramphenicol resistance (*cat*) gene [using pKD-Kpn-Hind-Fw and pKDEco2-Rv (Table S2) as primers and plasmid pKD3 (53) as a template] into the HindIII-EcoRI sites of pUC18. BarA punctual mutant variants (*barA*<sup>\*</sup>) were created by site-directed mutagenesis according to a two-step PCR procedure (58). The first PCR amplifications were performed using

the plasmid pUC18-*barA* as a template and *barA*-NdeI-Fw and either BarA-S99A-Rv, BarA-H102A-Rv, BarA-R124A-Rv, BarA-I130A-Rv, or BarA-R132A-Rv as primers (Table S2). Each purified product was used as a primer with primer BAfullR-Hind for the second PCR, using the plasmid pUC18-*barA* as a template. The second PCR products were digested with NdeI and HindIII and cloned between the corresponding sites of vector pUC18-Cam, resulting in pBarAS99ACam, pBarAH102ACam, pBarAR124ACam, pBarAI130ACam, and pBarAR132ACam. To construct the strains IFC5038 (*barA*<sup>S99A</sup>::Cam<sup>r</sup>), IFC5039 (*barA*<sup>H102A</sup>::Cam<sup>r</sup>), IFC5040 (*barA*<sup>R124A</sup>-Cam<sup>r</sup>), IFC5041 (*barA*<sup>I130A</sup>-Cam<sup>r</sup>), and IFC5042 (*barA*<sup>R132A</sup>-Cam<sup>r</sup>), PCR-amplified fragments, using primers the *barA*mut-ins-Fw and *barA*del-Rv (Table S2), and plasmids pBarAS99ACam, pBarAH102ACam, pBarAR124ACam, pBarAI130ACam, or pBarAR132ACam as the template, were used to replace the  $\Delta$ *barA* allele in the strain IFC5036 with the corresponding mutant *barA*<sup>\*</sup>::Cam<sup>r</sup> allele by homologous recombination using the lambda-red recombinase system (53).

### In vitro random mutagenesis of *barA*

To construct the plasmid pDuetBarAXho, which was used as a template for error-prone PCR, the *barA* open reading frame was PCR-amplified, using the primers *barA*f1NcoI and BAfullR-Hind (Table S2) and chromosomal DNA of strain KSB837 as the template, and cloned between the NcoI-HindIII sites of a modified pACYCDuet-1 (Novagen) [pACYCDuet-1(-XhoI), in which the XhoI site was destroyed by cleaving, filling in, and ligation], resulting in pDuetBarA. A silent mutation that created an XhoI restriction site in *barA*, at a position corresponding to the second transmembrane domain of BarA, was introduced by site-directed mutagenesis of plasmid pDuetBarA using the QuikChange kit (Stratagene) and the mutagenic primers BarAXhoITM2-Fw and BarAXhoITM2-Rv (Table S2), resulting in the pDuetBarAXho plasmid. Random mutagenesis of the N-terminal region of BarA was performed by error-prone PCR according to the method of Cadwell and Joyce (59). The reaction mixtures contained 5 ng of plasmid pDuetBarAXho as a template, 2  $\mu$ M of primers *barA*f1NcoI and BarAXhoITM2-Rv, 50 mM KCl, 10 mM Tris-HCl buffer (pH 8.3), 7 mM MgCl<sub>2</sub>, 0.5 mM MnCl<sub>2</sub>, 0.2 mM each dATP and dGTP, 1 mM each dTTP and dCTP, and 5 U of Taq DNA polymerase (New England Biolabs) in a total volume of 100  $\mu$ l. The PCRs were performed with 20 cycles of the following steps: (i) 95 °C for 30 s, (ii) 60 °C for 30 s, and (iii) 72 °C for 1 min. To generate a mutant library, the purified PCR products were used to replace the NcoI-HindIII restriction fragment of pDuetBarAXho. The resulting library was introduced into the strain IFC5035 ( $\Delta$ *barA*::Kan<sup>r</sup> *csrB-lacZ*) by transformation, and the individual clones were screened by streaking on suitable agar plates, as described above.

### Determination of $\beta$ -galactosidase activity

The cells of KSB837 derivative strains, carrying the UvrY-P-activatable *csrB-lacZ* reporter, were grown aerobically in LB

adjusted to and buffered at the indicated pH, at 37 °C. When indicated, acetate or formate were added to the growth media. The samples were withdrawn at the indicated times or A<sub>600</sub>, and  $\beta$ -Galactosidase activity was assayed and expressed in Miller units, as described previously (60).

### Subcellular fractionation and Western blot analysis

The cultures grown aerobically in LB were harvested during midexponential growth, and the cells were washed with Tris/HCl buffer (50 mM Tris/HCl pH 7.6, 200 mM KCl, 10 mM MgCl<sub>2</sub>, and 0.1 mM EDTA). The cell pellet was resuspended in 5 ml of the same buffer and disrupted by sonication. Cell debris was removed by centrifugation for 10 min at 10,000g. The supernatant was centrifuged for 45 min at 30,000g to separate the cytosolic and the membrane fractions. The resultant supernatant fluid containing the soluble proteins was collected. The remaining pellet was resuspended in 0.5 ml of Tris/HCl buffer. The samples of cytosolic and membrane (containing 5  $\mu$ g of protein) fractions were separated by SDS-PAGE (10% polyacryl-amide gel), and the proteins were transferred to a Hybond-ECL filter (Amersham Biosciences). The filter was equilibrated in TTBS buffer (25 mM Tris, 150 mM NaCl, and 0.05% Tween 20) for 10 min and incubated in blocking buffer (1% milk in TTBS) for 1 h at room temperature. BarA polyclonal antibodies, raised against His6-BarA<sup>198–918</sup> (33), were added at a dilution of 1:10,000 and incubated for 1 h at room temperature. The bound antibody was detected by using anti-rabbit IgG antibody conjugated to horseradish peroxidase and the Immobilon Western detection system (Millipore). The protein bands were quantified using ImageJ software (61).

### Structural modeling and binding site prediction

The Iterative Threading Assembly Refinement server (37, 62) was used to model the BarA PD (residues 31–178), using the GacS periplasmic domain (5077) as a template, which shares 19% identity and 44% similarity at the protein level. The generated model had a C-score of 1.12 and a TM-score of 0.73  $\pm$  0.11. The model was improved by full-atomic simulations using ModRefiner (63). The final model has 76.7% of all residues residing in the most-favored region of the Ramachandran plot and 19.5% of residues in additionally allowed regions, as calculated by the PROCHECK module of the PDBSum server (64). The acetic acid (in its protonated form) binding site of BarA-PD was predicted by molecular docking simulations, using the CB-Dock server (39), which detects protein cavities to guide blind docking by the algorithm AutoDock Vina (40). The ribbon structures were viewed and aligned, and the figures were rendered using PyMOL (The PyMOL Molecular Graphics System, version 2.0, Schrödinger).

### Phylogenetic analysis

Approximately, 500 protein sequences were collected from the GenBank nonredundant protein sequences database by a BLASTP search using the BarA or the GacS

## Identification of the sensor domain of BarA

periplasmic domain sequence as a seed. The redundant sequences and those with identities higher than 50% were excluded, resulting in a dataset of 63 protein sequences (NCBI accessions are shown in Table S3). Multiple-sequence alignments were performed by MUSCLE (65). The phylogenetic and molecular evolutionary analyses were conducted by the Maximum Likelihood method using MEGA version X (66). The best-fit models of amino acid substitution were selected by the Bayesian information criterion. The bootstrap confidence levels were obtained using 500 replicates. Phylogenetic tree was visualized and arbitrarily rooted using the midpoint rooting method implemented in Mega software (66), and the tree was condensed so that only bootstrap support values higher than 50% are shown.

### Statistics

All the quantitative experiments were performed in triplicates, and the experimental results are expressed as mean  $\pm$  the SD value.

### Data availability

All data are contained in the article and in the Supporting information.

**Supporting information**—This article contains supporting information (7, 15, 53, 55, 56).

**Acknowledgments**—We are grateful to Adela Rodríguez Romero for critically reading the article.

**Author contributions**—A. F. A. and D. G. conceptualization; A. F. A. and D. G. funding acquisition; A. F. A., C. R., and R. G.-C. investigation; A. F. A. writing—original draft; A. F. A. and D. G. writing—review and editing; D. G. funding acquisition.

**Funding and additional information**—This work was supported by grants IN208721 (A. F. A.) and IN207921 (D. G.) from the Dirección General de Asuntos del Personal Académico (DGAPA), Universidad Nacional Autónoma de México (UNAM), and 140614 (A. F. A.) and 514856 (D. G.) from the Consejo Nacional de Ciencia y Tecnología (CONACyT). The funders had no role in study design, data collection, analysis and interpretation, decision to publish, or preparation of the manuscript.

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: Cache, calcium channels and chemotaxis receptors; PAS, Per-Arnt-Sim; PD, periplasmic domain; TCS, two-component system; TM, transmembrane.

### References

1. Pourciau, C., Lai, Y. J., Gorelik, M., Babitzke, P., and Romeo, T. (2020) Diverse mechanisms and circuitry for global regulation by the RNA-binding protein CsrA. *Front. Microbiol.* **11**, 601352
2. Suzuki, K., Wang, X., Weilbacher, T., Pernestig, A. K., Melefors, O., Georgellis, D., Babitzke, P., and Romeo, T. (2002) Regulatory circuitry of the CsrA/CsrB and BarA/UvrY systems of *Escherichia coli*. *J. Bacteriol.* **184**, 5130–5140
3. Ishige, K., Nagasawa, S., Tokishita, S., and Mizuno, T. (1994) A novel device of bacterial signal transducers. *EMBO J.* **13**, 5195–5202
4. Nagasawa, S., Tokishita, S., Aiba, H., and Mizuno, T. (1992) A novel sensor-regulator protein that belongs to the homologous family of signal-transduction proteins involved in adaptive responses in *Escherichia coli*. *Mol. Microbiol.* **6**, 799–807
5. Pernestig, A. K., Melefors, O., and Georgellis, D. (2001) Identification of UvrY as the cognate response regulator for the BarA sensor kinase in *Escherichia coli*. *J. Biol. Chem.* **276**, 225–231
6. Kinoshita-Kikuta, E., Kinoshita, E., Eguchi, Y., and Koike, T. (2016) Validation of *cis* and *trans* modes in multistep phosphotransfer signaling of bacterial tripartite sensor kinases by using Phos-Tag SDS-PAGE. *PLoS One* **11**, e0148294
7. Gudapaty, S., Suzuki, K., Wang, X., Babitzke, P., and Romeo, T. (2001) Regulatory interactions of Csr components: The RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*. *J. Bacteriol.* **183**, 6017–6027
8. Liu, M. Y., Gui, G., Wei, B., Preston, J. F., Oakford, L., Yüksel, U., Giedroc, D. P., and Romeo, T. (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J. Biol. Chem.* **272**, 17502–17510
9. Weilbacher, T., Suzuki, K., Dubey, A. K., Wang, X., Gudapaty, S., Morozov, I., Baker, C. S., Georgellis, D., Babitzke, P., and Romeo, T. (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol. Microbiol.* **48**, 657–670
10. Romeo, T., Vakulskas, C. A., and Babitzke, P. (2013) Post-transcriptional regulation on a global scale: Form and function of Csr/Rsm systems. *Environ. Microbiol.* **15**, 313–324
11. Figueroa-Bossi, N., Schwartz, A., Guillemardet, B., D’Heygère, F., Bossi, L., and Boudvillain, M. (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev.* **28**, 1239–1251
12. Dubey, A. K., Baker, C. S., Suzuki, K., Jones, A. D., Pandit, P., Romeo, T., and Babitzke, P. (2003) CsrA regulates translation of the *Escherichia coli* carbon starvation gene, *cstA*, by blocking ribosome access to the *cstA* transcript. *J. Bacteriol.* **185**, 4450–4460
13. Liu, M. Y., and Romeo, T. (1997) The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein. *J. Bacteriol.* **179**, 4639–4642
14. Schubert, M., Lapouge, K., Duss, O., Oberstrass, F. C., Jelesarov, I., Haas, D., and Allain, F. H.-T. (2007) Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat. Struct. Mol. Biol.* **14**, 807–813
15. Chavez, R. G., Alvarez, A. F., Romeo, T., and Georgellis, D. (2010) The physiological stimulus for the BarA sensor kinase. *J. Bacteriol.* **192**, 2009–2012
16. Wolfe, A. J. (2005) The acetate switch. *Microbiol. Mol. Biol. Rev.* **69**, 12–50
17. Ali-Ahmad, A., Fadel, F., Sebban-Kreuzer, C., Ba, M., Pélissier, G. D., Bornet, O., Guerlesquin, F., Bourne, Y., Bordi, C., and Vincent, F. (2017) Structural and functional insights into the periplasmic detector domain of the GacS histidine kinase controlling biofilm formation in *Pseudomonas aeruginosa*. *Sci. Rep.* **7**, 11262
18. Taylor, B. L., and Zhulin, I. B. (1999) PAS domains: Internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **63**, 479–506
19. Zhulin, I. B., Taylor, B. L., and Dixon, R. (1997) PAS domain S-boxes in Archaea, bacteria and sensors for oxygen and redox. *Trends Biochem. Sci.* **22**, 331–333
20. Ponting, C. P., and Aravind, L. (1997) PAS: A multifunctional domain family comes to light. *Curr. Biol.* **7**, R674–R677
21. Krämer, J., Fischer, J. D., Zientz, E., Vijayan, V., Griesinger, C., Lupas, A., and Uden, G. (2007) Citrate sensing by the C4-dicarboxylate/citrate sensor kinase DcuS of *Escherichia coli*: Binding site and conversion of DcuS to a C 4-dicarboxylate- or citrate-specific sensor. *J. Bacteriol.* **189**, 4290–4298

22. Pappalardo, L., Janausch, I. G., Vijayan, V., Zientz, E., Junker, J., Peti, W., Zweckstetter, M., Unden, G., and Griesinger, C. (2003) The NMR structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) DcuS of *Escherichia coli*. *J. Biol. Chem.* **278**, 39185–39188
23. Cheung, J., Bingman, C. A., Reyngold, M., Hendrickson, W. A., and Waldburger, C. D. (2008) Crystal structure of a functional dimer of the PhoQ sensor domain. *J. Biol. Chem.* **283**, 13762–13770
24. Reinelt, S., Hofmann, E., Gerharz, T., Bott, M., and Madden, D. R. (2003) The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. *J. Biol. Chem.* **278**, 39189–39196
25. Zhang, Z., and Hendrickson, W. A. (2010) Structural characterization of the predominant family of histidine kinase sensor domains. *J. Mol. Biol.* **400**, 335–353
26. Upadhyay, A. A., Fleetwood, A. D., Adebali, O., Finn, R. D., and Zhulin, I. B. (2016) Cache domains that are homologous to, but different from PAS domains comprise the largest superfamily of extracellular sensors in prokaryotes. *PLoS Comput. Biol.* **12**, e1004862
27. Wu, R., Gu, M., Wilton, R., Babnigg, G., Kim, Y., Pokkuluri, P. R., Szurmant, H., Joachimiak, A., and Schiffer, M. (2013) Insight into the sporulation phosphorelay: Crystal structure of the sensor domain of *Bacillus subtilis* histidine kinase, kind. *Protein Sci.* **22**, 564–576
28. Zhou, Y. F., Nan, B., Nan, J., Ma, Q., Panjkar, S., Liang, Y. H., Wang, Y., and Su, X. D. (2008) C4-Dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. *J. Mol. Biol.* **383**, 49–61
29. Pokkuluri, P. R., Dwulit-Smith, J., Duke, N. E., Wilton, R., Mack, J. C., Bearden, J., Rakowski, E., Babnigg, G., Szurmant, H., Joachimiak, A., and Schiffer, M. (2013) Analysis of periplasmic sensor domains from *Anaeromyxobacter dehalogenans* 2CP-C: Structure of one sensor domain from a histidine kinase and another from a chemotaxis protein. *Microbiologyopen* **2**, 766–777
30. Kneuper, H., Janausch, I. G., Vijayan, V., Zweckstetter, M., Bock, V., Griesinger, C., and Unden, G. (2005) The nature of the stimulus and of the fumarate binding site of the fumarate sensor DcuS of *Escherichia coli*. *J. Biol. Chem.* **280**, 20596–20603
31. Kwon, O., Georgellis, D., Lynch, A. S., Boyd, D., and Lin, E. C. (2000) The ArcB sensor kinase of *Escherichia coli*: Genetic exploration of the transmembrane region. *J. Bacteriol.* **182**, 2960–2966
32. Mondragón, V., Franco, B., Jonas, K., Suzuki, K., Romeo, T., Melefors, O., and Georgellis, D. (2006) pH-dependent activation of the BarA-UvrY two-component system in *Escherichia coli*. *J. Bacteriol.* **188**, 8303–8306
33. Camacho, M. I., Alvarez, A. F., Gonzalez Chavez, R., Romeo, T., Merino, E., and Georgellis, D. (2015) Effects of the global regulator CsrA on the BarA/UvrY two-component signaling system. *J. Bacteriol.* **197**, 983–991
34. Vázquez-Ciros, O. J., Alvarez, A. F., and Georgellis, D. (2020) Identification of Z nucleotides as an ancient signal for two-component system activation in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 33530–33539
35. Takeuchi, K., Kiefer, P., Reimann, C., Keel, C., Dubuis, C., Rolli, J., Vorholt, J. A., and Haas, D. (2009) Small RNA-dependent expression of secondary metabolism is controlled by Krebs cycle function in *Pseudomonas fluorescens*. *J. Biol. Chem.* **284**, 34976–34985
36. Heeb, S., Blumer, C., and Haas, D. (2002) Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0. *J. Bacteriol.* **184**, 1046–1056
37. Yang, J., and Zhang, Y. (2015) I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–W181
38. Gerharz, T., Reinelt, S., Kaspar, S., Scapozza, L., and Bott, M. (2003) Identification of basic amino acid residues important for citrate binding by the periplasmic receptor domain of the sensor kinase CitA. *Biochemistry* **42**, 5917–5924
39. Liu, Y., Grimm, M., Dai, W., Hou, M., Xiao, Z.-X., and Cao, Y. (2019) CB-dock: A web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacol. Sin.* **41**, 138–144
40. Trott, O., and Olson, A. J. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461
41. Babitzke, P., and Romeo, T. (2007) CsrB sRNA family: Sequestration of RNA-binding regulatory proteins. *Curr. Opin. Microbiol.* **10**, 156–163
42. Lapouge, K., Schubert, M., Allain, F. H. T., and Haas, D. (2008) Gac/Rsm signal transduction pathway of  $\gamma$ -proteobacteria: From RNA recognition to regulation of social behaviour. *Mol. Microbiol.* **67**, 241–253
43. Heroven, A. K., Bořme, K., and Dersch, P. (2012) The Csr/Rsm system of *Yersinia* and related pathogens: A post-transcriptional strategy for managing virulence. *RNA Biol.* **9**, 379–391
44. Zere, T. R., Vakulskas, C. A., Leng, Y., Pannuri, A., Potts, A. H., Dias, R., Tang, D., Kolaczowski, B., Georgellis, D., Ahmer, B. M. M., and Romeo, T. (2015) Genomic targets and features of BarA-UvrY (-SirA) signal transduction systems. *PLoS One* **10**, e0145035
45. Sobrero, P. M., and Valverde, C. (2020) Comparative genomics and evolutionary analysis of RNA-binding proteins of the CsrA family in the genus *Pseudomonas*. *Front. Mol. Biosci.* **7**, 127
46. Huertas-Rosales, Ó., Romero, M., Chan, K.-G., Hong, K.-W., Cámara, M., Heeb, S., Barrientos-Moreno, L., Molina-Henares, M. A., Travieso, M. L., Ramos-González, M. I., and Espinosa-Urgel, M. (2021) Genome-wide analysis of targets for post-transcriptional regulation by Rsm proteins in *Pseudomonas putida*. *Front. Mol. Biosci.* **8**, 624061
47. Butz, H. A., Mey, A. R., Ciosek, A. L., Crofts, A. A., Davies, B. W., and Payne, S. M. (2021) Regulatory effects of CsrA in *Vibrio cholerae*. *MBio* **12**, e03380-20
48. Ancona, V., Lee, J. H., and Zhao, Y. (2016) The RNA-binding protein CsrA plays a central role in positively regulating virulence factors in *Erwinia amylovora*. *Sci. Rep.* **6**, 1–12
49. Edwards, A. N., Patterson-Fortin, L. M., Vakulskas, C. A., Mercante, J. W., Potrykus, K., Vinella, D., Camacho, M. I., Fields, J. A., Thompson, S. A., Georgellis, D., Cashel, M., Babitzke, P., and Romeo, T. (2011) Circuitry linking the Csr and stringent response global regulatory systems. *Mol. Microbiol.* **80**, 1561–1580
50. Brencic, A., and Lory, S. (2009) Determination of the regulon and identification of novel mRNA targets of *Pseudomonas aeruginosa* RsmA. *Mol. Microbiol.* **72**, 612–632
51. Liu, J., Yu, M., Ge, Y., Tian, Y., Hu, B., and Zhao, Y. (2021) The RsmA RNA-binding proteins in *Pseudomonas syringae* exhibit distinct and overlapping roles in modulating virulence and survival under different nutritional conditions. *Front. Plant Sci.* **12**, 637595
52. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
53. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645
54. Cherepanov, P. P., and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of FLP-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9–14
55. Norrander, J., Kempe, T., and Messing, J. (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**, 101–106
56. Dykxhoorn, D. M., St Pierre, R., and Linn, T. (1996) A set of compatible *tac* promoter expression vectors. *Gene* **177**, 133–136
57. Grandori, R., Struck, K., Giovanielli, K., and Carey, J. (1997) A three-step PCR protocol for construction of chimeric proteins. *Protein Eng.* **10**, 1099–1100
58. Sarkar, G., and Sommer, S. S. (1990) The “megaprimer” method of site-directed mutagenesis. *Biotechniques* **8**, 404–407
59. Cadwell, R. C., and Joyce, G. F. (1992) Randomization of genes by PCR mutagenesis. *Genome Res.* **2**, 28–33
60. Miller, J. H. (1972)  $\beta$ -Galactosidase assay. In *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY: 352–355
61. Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675
62. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738

## Identification of the sensor domain of BarA

63. Xu, D., and Zhang, Y. (2011) Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.* **101**, 2525–2534
64. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., and Thornton, J. M. (2018) PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134
65. Edgar, R. C. (2004) Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797
66. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) Mega X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549