

Accepted Manuscript

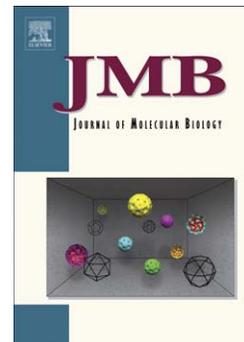
Structural and functional analysis of the signal transducing linker in the pH-responsive one component system CadC of *Escherichia coli*

Sophie Buchner, Andreas Schlundt, Jürgen Lassak, Michael Sattler, Kirsten Jung

PII: S0022-2836(15)00285-5
DOI: doi: [10.1016/j.jmb.2015.05.001](https://doi.org/10.1016/j.jmb.2015.05.001)
Reference: YJMBI 64746

To appear in: *Journal of Molecular Biology*

Received date: 27 January 2015
Revised date: 13 April 2015
Accepted date: 6 May 2015



Please cite this article as: Buchner, S., Schlundt, A., Lassak, J., Sattler, M. & Jung, K., Structural and functional analysis of the signal transducing linker in the pH-responsive one component system CadC of *Escherichia coli*, *Journal of Molecular Biology* (2015), doi: [10.1016/j.jmb.2015.05.001](https://doi.org/10.1016/j.jmb.2015.05.001)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Structural and functional analysis of the signal transducing linker in the pH-responsive one component system CadC of *Escherichia coli*

Sophie Buchner¹, Andreas Schlundt^{2,3}, Jürgen Lassak¹, Michael Sattler^{2,3} and Kirsten Jung^{1*}

¹Munich Center for Integrated Protein Science (CiPSM) at the Department of Microbiology, Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany

²Munich Center for Integrated Protein Science at Department of Chemistry, Technische Universität München, 85748 Garching, Germany

³Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Running title: The signal-transducing linker of *E. coli* CadC

*To whom correspondence should be addressed:

Prof. Dr. Kirsten Jung
Ludwig-Maximilians-Universität München
Department Biologie I, Bereich Mikrobiologie
Großhaderner Str. 2-4
82152 Martinsried
Germany
Phone: +49-89-2180-74500
Fax: +49-89-2180-74520
E-mail: jung@lmu.de

Key words: transcription activator, regulation, ToxR, acid stress, sensor

ABSTRACT

The pH-responsive one component signaling system CadC in *Escherichia coli* belongs to the family of ToxR-like proteins, whose members share a conserved modular structure, with an N-terminal cytoplasmic winged helix-turn-helix DNA-binding domain being followed by a single transmembrane helix and a C-terminal periplasmic pH-sensing domain. In *E. coli* CadC a cytoplasmic linker comprising approximately 50 amino acids is essential for transmission of the signal from the sensor to the DNA-binding domain. However, the mechanism of transduction is poorly understood. Using NMR spectroscopy, we demonstrate here that the linker region is intrinsically disordered in solution. Furthermore, mutational analyses showed that it tolerates a range of amino acid substitutions (altering polarity, rigidity, α -helix-forming propensity), is robust to extension, but is sensitive to truncation. Indeed, truncations either reversed the expression profile of the target operon *cadBA* or decoupled expression from external pH altogether. CadC dimerizes via its periplasmic domain, but light scattering analysis provided no evidence for dimerization of the isolated DNA-binding domain, with or without the linker region. However, bacterial two-hybrid analysis revealed that CadC forms stable dimers in a stimulus- and linker-dependent manner, interacting only at pH < 6.8. Strikingly, a variant with reversed *cadBA* expression profile, which lacks most of the linker, dimerizes preferentially at higher pH. Thus, we propose that the disordered CadC linker is required for transducing the pH-dependent response of the periplasmic sensor into a structural rearrangement which facilitates dimerization of the cytoplasmic CadC DNA-binding domain.

INTRODUCTION

Escherichia coli is commonly found in the intestine of warm-blooded mammals. In the human gastrointestinal tract *E. coli* is exposed to significant variations in pH [1]. *E. coli* can counteract low pH by employing a battery of pH-regulatory systems. Among them, the CadABC system responds to a combination of acidic pH and lysine availability [2, 3]. These two signals are sensed by an interaction between the pH-responsive one-component signaling system CadC and the lysine permease LysP [4]. LysP inhibits signal transduction by binding tightly to CadC when no lysine is available [4]. In the presence of lysine, CadC responds to low pH and activates the transcription of the *cadBA* operon leading to production of CadA and CadB. The lysine decarboxylase CadA converts lysine under consumption of a proton into the cadaverine and carbon dioxide, thus raising the pH of the cytoplasm [5-7]. As the polyamine cadaverine is more alkaline than the corresponding amino acid lysine, an antiport of both molecules mediated by CadB also leads to an increase of the extracellular pH [8]. Remarkably, *cadBA* expression is negatively feed-back regulated by the end products of the decarboxylase reaction, cadaverine and CO₂ [2, 9, 10]. For an appropriate stress response the expression of CadC is fine-tuned by elongation factor P (EF-P). EF-P alleviates translational stalling on polyproline stretches [11]. CadC is a membrane-integrated transcriptional activator of the ToxR family [12]. Thus, it combines sensory, signal transducing and DNA-binding functions within one polypeptide chain. Specifically, CadC is composed of a C-terminal periplasmic pH-sensing domain, a single transmembrane helix and an N-terminal cytoplasmic winged helix-turn-helix DNA-binding domain [13-16]. Signal transduction in ToxR-like transcriptional regulators is mediated without chemical modification [13] and is thus the simplest form of bacterial stimulus-response mechanism. Recently, it

was demonstrated that a patch of acidic amino acids at the interface between two CadC monomers in the periplasmic domain is responsible for pH sensing [10]. Furthermore the three-dimensional structure of the pH-sensing domain was solved at a resolution of 1.8 Å [17]. The DNA-binding domain (amino acids 1-107) and the transmembrane helix (amino acids 160-178) are interconnected by a linker (amino acids 108-159), which transduces the pH-dependent signal in an unknown fashion (Fig. 1). However, the mechanism of signal transduction from the transmembrane helix via the intervening linker to the DNA-binding domain has remained enigmatic. In the present study, a combined biochemical and genetic approach was used to investigate the structure and function of the linker in CadC. We demonstrate that the linker is intrinsically disordered. In addition, while the linker itself does not undergo dimerization, it determines the disposition of the two DNA-binding domains in the dimer in a stimulus-dependent manner and is thus intimately involved in signal transduction.

RESULTS

CadC includes a poorly conserved linker segment

CadC sequences from *Salmonella enterica*, *Shewanella piezotolerans*, *Aeromonas hydrophila*, *Vibrio fischeri* and *Yersinia ruckeri* were collected, and the degree of conservation was investigated separately for each domain (Fig 1.). This identified significant homologies within the DNA-binding and sensor domains. On the contrary the primary sequence of the linker is not conserved and thus is mainly responsible for the previously described divergence between CadC orthologs [18]. Beside the low degree of sequence conservation, the CadC linker also varies significantly in length (e.g., ~51 amino acids in *E. coli*, but 76 residues long in *V. fischeri*), suggesting that

signal transduction from the transmembrane domain to the DNA binding domain allows a high degree of variance of the amino acid sequence.

The CadC linker is disordered in solution

Secondary structure prediction of CadC₁₋₁₅₉ with PSIPRED [19] had suggested that the putative DNA binding domain ranges from amino acids 1 to approximately 107, while the segment beyond residue 107 contains no secondary structure (data not shown). To characterize the conformation of the CadC linker region (defined here as comprising residues 108-159), we carried out NMR spectroscopy on CadC fragments consisting of the DNA-binding domain alone (CadC₁₋₁₀₇) or the DNA-binding domain and the linker (CadC₁₋₁₅₉). ¹H, ¹⁵N NMR correlation experiments are ideally suited to study proteins at single-residue level, as the amide group for each residue (except prolines) gives rise to a cross peak in these spectra. ¹H, ¹⁵N correlation spectra of CadC₁₋₁₀₇ are consistent with a well-folded domain, as indicated by a large spectral dispersion of the NMR signals, typical for the presence of tertiary structure. The amide chemical shifts observed for residues 1-107 in the NMR spectra of CadC₁₋₁₅₉ and CadC₁₋₁₀₇ are basically identical, demonstrating that the globular DNA-binding domain fold is not significantly affected by the linker segment present in CadC₁₋₁₅₉ (Fig. 2). This becomes evident by superimposition of peaks from both samples (see color code) and is indicated with selected assignments for residue numbers smaller than 107. In contrast, all additional NMR signals that correspond to linker residues exhibit proton chemical shifts in the range of 7.5-8.5 ppm (additional signals seen in the spectrum of CadC₁₋₁₅₉, Fig. 2 with selected assignments of residue numbers >107 indicated). This limited chemical shift range is characteristic of the absence of secondary and tertiary structure and indicates that the linker region in CadC₁₋₁₅₉ is intrinsically disordered.

We were able to assign almost all protein backbone resonances in CadC₁₋₁₅₉ and first used ¹³C secondary chemical shifts as a an experimental measure of secondary structure [20] (Fig. 3 A, B). The differences of secondary chemical shifts ($\Delta\delta^{13}\text{C}_\alpha - \Delta\delta^{13}\text{C}_\beta$) are plotted (Fig. 3 B), where $\Delta\delta^{13}\text{C}$ is the difference between the observed chemical shift and the chemical shift found in a random coil conformation for a given amino acid. In the region spanned by amino acids 1-107, β -strand (consecutive values below -2) and α -helical (consecutive values above +2) secondary structure is observed, but the segment comprising residues 108-159 exhibits random-coil like values. This suggests that the C-terminal region is intrinsically unstructured and flexible. We then probed the conformational dynamics of the CadC₁₋₁₅₉ backbone based on ¹⁵N NMR relaxation measurements (Fig. 3 C-E). $\{^1\text{H}\}$ -¹⁵N heteronuclear Overhauser Effects values are sensitive to internal motions at rates faster than the overall tumbling rate of a molecule in solution (Fig. 3, panel C). Values below the theoretical maximum of ~ 0.8 for rigid amide groups indicate fast mobility of this residue. In fact, low $\{^1\text{H}\}$ -¹⁵N values are observed for CadC₁₀₈₋₁₅₉, indicating mobility on a pico- to nanosecond timescale, while the N-terminal DNA-binding domain comprises a rigid and compact fold based on heteronuclear NOE values around 0.8 (Fig. 3, panel C). Interestingly, the slightly less reduced values for residues 108-159 imply that this region is less mobile than the two termini of the polypeptide, which typically exhibit negative values (see residues 1-3 and 158-159 in Fig. 3 C). Thus, the linker region appears not to be completely unconstrained and a certain degree of order could be present transiently.

Local correlation times of the amides (τ_c) were derived from longitudinal (R_1) and transverse (R_2) ¹⁵N relaxation rates to further characterize overall tumbling and internal dynamics of the protein (Fig. 3, panels D-F). The CadC region beyond residue 107 shows consistently larger longitudinal and smaller transverse relaxation

rates (Fig. 3, panel D, E), implying a degree of conformational flexibility that is consistent with absence of secondary structure and with the $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE data. The ratio of R_1 and R_2 was used to derive the local, residue-specific values for the tumbling correlation times (τ_c) of the protein (Fig. 3 F). These data indicate two distinct average ranges with slower tumbling ($\tau_c \sim 8.2$ ns) for residues in the DNA-binding domain (residues 1-107) and higher mobility ($\tau_c \sim 5.5$ ns) for the linker residues 108-159. Thus, we conclude that the linker experiences higher internal mobility and tumbles independent from the CadC DNA-binding domain.

The isolated cytoplasmic segment of CadC exists exclusively in monomeric form

Disordered linkers have been shown to play crucial roles in effector domain oligomerization for signal transduction in bacteria [21]. Recently it was shown that CadC is active as a homodimer and that dimerization takes place in the periplasmic domain of CadC [17, 22]. To determine whether the CadC linker might be capable of oligomerizing the DNA-binding domain, we used static light scattering, a method which allows one to calculate a molecule's molecular weight based on its light scattering properties when freely tumbling in solution. Both CadC₁₋₁₀₇ and CadC₁₋₁₅₉ showed gel filtration profiles with a single UV-detected peak indicating sample homogeneity (Fig. 4). At a concentration of 5 mg/ml the corresponding refractive index data revealed both species to be monomeric, with virtually no deviation from the theoretical molecular weight of the proteins (see legend to Fig. 4). To exclude concentration-dependent effects, we also estimated the molecular weight of CadC₁₋₁₅₉ from the rotational correlation time based on the ^{15}N NMR relaxation data (see previous paragraph and Fig. 3, panel F), noting that the tumbling of a molecule in

solution is proportional to its molecular weight. From these data the molecular weight of the DNA-binding domain was found to be in the range of its theoretical value (13 ± 2 kDa) at a concentration of 12 mg/ml. We thus conclude that, even at concentrations well above the physiological range, the isolated cytoplasmic segment of CadC including the linker region does not form oligomers.

We then asked whether the linker is involved in the formation of CadC oligomers in the presence of or by interaction with membrane lipids. We prepared a DMPC:DMPG lipid mixture of 4:1 to mimic the presence of membranes with bicelles in vitro. As is shown in Supplementary Fig. 1, HSQC spectra of CadC₁₋₁₅₉ in the absence or presence of bicelles do not show differences in chemical shifts or site-specific line widths indicating that under our experimental conditions and at two different temperatures the cytoplasmic fragment of the protein does not interact with membranes. At the same time we conclude that in our lipid mixture the presence of bicelles does not induce the formation of secondary structure in the CadC flexible linker or promote protein dimerization.

The CadC linker tolerates a wide range of amino acid substitutions

These findings indicate that the CadC linker is disordered and is not involved in dimerization, leaving its role in signal transduction undefined. To probe this issue further, we tested the functional relevance of charged amino acids within the linker, since they might form salt bridges within the protein or interact with negatively charged phospholipids. Using site-directed mutagenesis, we generated a set of constructs in which negatively or positively charged amino acids were replaced by alanine, glutamate or arginine (Table 2). Western Blot analysis demonstrated that all the resulting CadC variants were produced and integrated into the membrane like wild-type CadC (data not shown). To test their effects on P_{cadBA} activation, we used

the *E. coli cadA::lacZ* reporter strain EP314, which lacks the chromosomal *cadC*, and complemented this strain with the constructs *in trans* [9]. β -Galactosidase activities were determined for cells grown in minimal medium [23] under *cadBA*-inducing (pH 5.8 in the presence of lysine) and non-inducing (pH 7.6 in the absence of lysine) conditions. First we focused on amino acids E125, D131, E147 and the two charged clusters ¹⁰⁹EEEGEE¹¹⁴ and ¹⁵²KSKR¹⁵⁵ in CadC. Variants E125A, D131A, E147A, ¹⁰⁹AAAGAA¹¹⁴ and ¹⁵²ASAA¹⁵⁵, as well as ¹⁵²ESEE¹⁵⁵, activated *cadBA* expression like wild-type CadC (Table 2). This result was unexpected, but supports the idea that the disordered structure of the linker in CadC, rather than its specific amino acid sequence, is important for its function.

Replacement of the negatively charged cluster of glutamate residues by positively charged arginines (CadC-¹⁰⁹RRRARR¹¹⁴) or inversion of the charges of the two clusters (CadC-¹⁰⁹RRRARR¹¹⁴ – ¹⁵²ESEE¹⁵⁵) resulted in non-responsive CadC variants (Table 2), which are presumably unable to bind P_{*cadBA*} *in vivo*.

As the secondary-structure prediction program PSIPRED [19] pinpointed a motif ¹⁴⁶PEQSPV¹⁵¹ within the CadC linker as a possible structural element, albeit with low confidence, chemical shifts of the linker (amino acids 108-159) were analyzed with the δ 2D program [24] that calculates populations of secondary structure in proteins, thereby including poly-proline-II (PPII)-helical content. Interestingly, the prediction supports PSIPRED in that residues 146-152 are the only motif that shows low (10-15 %) probability of PPII structure while the rest of the linker is predicted as random coil with more than 90% likelihood (data not shown). The motif ¹⁴⁶PEQSPV¹⁵¹ was analyzed with BLAST [25] and conservation with other proteins was found. In some proteins, e.g. neisserial surface protein A (NspA), this sequence is actually part of a larger structural element, which led to the hypothesis that conformational changes after stimulus perception might force this part of the

linker to adopt an ordered structure. Spontaneous disorder-to-order transitions have also been described for other proteins, such as the activator for thyroid hormone and retinoid receptors (ACTR). In its free state, the activation domain of ACTR is disordered but, after binding to the nuclear co-activator domain of the CREB binding protein, ACTR forms three α -helices in the complex [26]. We therefore performed alanine scanning mutagenesis of ¹⁴⁶PEQSPV¹⁵¹, since alanine has high helix-forming propensities [27]. However, all corresponding CadC variants (CadC-¹⁴⁶AEQSPV¹⁵¹, CadC-¹⁴⁶PEASPV¹⁵¹, CadC-¹⁴⁶PEQAPV¹⁵¹, CadC-¹⁴⁶PEQSAV¹⁵¹, CadC-¹⁴⁶PEQSPA¹⁵¹, CadC-¹⁴⁶AAAAAV¹⁵¹ and CadC-¹⁴⁶AAAAAA¹⁵¹) induced *cadBA* expression like wild-type CadC (Table 2). That alanine insertion does not disturb *cadBA* induction supports the notion that the linker acts as a spacer to maintain a minimal distance between DNA-binding domain and plasma membrane.

A cluster of consecutive prolines ¹²⁰PPPIP¹²⁴ is also present in the linker. We demonstrated recently that ribosomes stall at this polyproline stretch, and that translation elongation factor P (EF-P) alleviates stalling [11]. Further studies indicated the physiological relevance of this translational control, namely to fine-tune the copy number of CadC molecules [11].

Consecutive prolines are able to form rigid polyproline helices. Therefore, we also asked whether extension of the polyproline stretch in the linker of CadC had any effect on the activity of the protein. The following CadC variants CadC-¹¹⁹SPPPIE¹²⁵, CadC-¹¹⁹SPPPPPE¹²⁵ and CadC-¹¹⁹PPPPPPE¹²⁵ were tested for P_{cadBA} activation. However, all polyproline CadC variants behaved like wild-type CadC (Table 2). Again, increasing the rigidity of the linker had no influence on signal transduction.

Furthermore, a chimera was designed (designated as CadC-*Vibrio*), in which the *E. coli* CadC linker (51 amino acids) was replaced with the corresponding region

of *V. fischeri* (76 residues), lacking any sequential homology. Strikingly, CadC-*Vibrio* was able to induce *cadBA* expression in a pH dependent manner (Table 2). Thus, not amino acid composition but rather structural disorder is essential for proper signal transduction.

The CadC linker tolerates extensions but is sensitive to [large] truncations

The findings described above indicate that the linker in CadC is relatively robust against amino acid substitutions. We therefore generated CadC variants in which the linker was either extended or truncated (Fig. 5). All the resulting variants were produced in similar amounts to wild-type CadC and were stably integrated into the membrane (data not shown). Their functionality was tested in the reporter strain EP314 [9]. CadC derivatives containing a partial duplication of the linker (two copies of residues 120-151 or 145-151) activated *cadBA* expression like wild-type CadC (Fig. 5). However, deletion of short segments at or near the end of the linker (CadC Δ 145-151, CadC Δ 146-151 and CadC Δ 152-158) led to detectable activation at pH 7.6, while leaving *cadBA* expression under normal (acidic) inducing conditions largely unaffected (Fig. 5). Thus these amino acids are apparently dispensable. Moreover, CadC variants that lack larger portions of the linker region (CadC Δ 120-151, CadC Δ 128-151 and CadC Δ 133-151) induced *cadBA* expression at both acidic and slightly alkaline pH, albeit to different levels (Fig. 5). Therefore, these more extensive deletions result in constitutive, signal-independent activation of CadC, and the DNA-binding domain is apparently locked into an active conformation. Even more strikingly, the CadC variants CadC Δ 107-151, CadC Δ 108-151, CadC Δ 109-151 and CadC Δ 115-151 activated *cadBA* expression in cells exposed to pH 7.6, but not at pH 5.8, i.e., these truncated CadC variants exhibit the opposite response profile to the

wild-type (Fig. 5). This inverted pH response suggests that, in the absence of the central portion of the linker (CadC Δ 115-151), the external signal is still correctly perceived by the periplasmic domain, but is incorrectly interpreted on the cytoplasmic side.

Similar effects have been observed for regions flanking the PAS domain of NifL in *Azotobacter vinelandii*, where a helical linker connecting two PAS domains undergoes a conformational re-arrangement concomitantly with signal transduction [21]. The linker region in *E. coli* EnvZ, which has a helix 1-loop-helix 2 structure, also plays a crucial role in propagating a conformational change [28]. We therefore suggest that binding of CadC to its target DNA and/or interaction with the RNA polymerase is modulated by a conformational re-arrangement of the cytoplasmic linker that enables it to induce *cadBA* expression.

The linker controls the disposition of the DNA-binding domains in the CadC dimer formed in response to environmental signals

CadC is known to form dimers via its periplasmic region [17], but this apparently does not cause either its linker or its DNA-binding domain to interact (Fig. 4). How intrinsic disorder mediates biological function remains largely unclear. To uncover the role of the linker in controlling the conformation of the DNA-binding domains, we used the bacterial adenylate cyclase two-hybrid system (BACTH), which is based on the split adenylate cyclase (CyaA) from *Bordetella pertussis* [29]. We fused the complementary subunits of CyaA (T18-CadC and T25-CadC) to the N-terminus of full-length CadC [4] and screened for interaction, i.e., dimerization of the CadC segments signaled by β -galactosidase activation. The *Saccharomyces cerevisiae* yeast leucine-zipper fusion constructs zip-T18 and T25-zip [29] were used as positive controls (2,000 Miller Units) [4]. The effect of varying external pH values

on the interaction between the two CadC hybrid proteins was analyzed *in vivo*. As reported previously [4], we found a strong interaction between T18-CadC and T25-CadC when cells were grown under *cadBA*-inducing conditions ($\text{pH} \leq 6.8$), implying that CadC monomers are in close proximity (Fig. 6A). In contrast, no interaction between T18-CadC and T25-CadC was found when cells were grown under *cadBA* non-inducing conditions ($\text{pH} \geq 6.8$) (Fig. 6A). Cells producing the control fragments zip-T18 and T25-zip were tested under the same conditions and, as demonstrated earlier, the interaction of zip-T18 and T25-zip was not influenced by the external pH [4]. Previously, the relationship between induction of the Cad system and external pH had been monitored by determining cadaverine production as a measure of CadA activity [2]. At a lysine concentration of 10 mM, β -galactosidase activities of the BACTH system and cadaverine production follow exactly the same pattern: both activities begin to increase at a pH value of 6.8 (Fig. 6A). Hence the *in vivo* protein-protein analyses strongly argue that CadC homodimerizes under inducing conditions.

The CadC Δ 108-151 variant was also subjected to two-hybrid analysis, as truncation of the entire linker inverted the pattern of *cadBA* expression observed with wild-type CadC (Fig. 6B). And indeed, for T18-CadC Δ 108-151 and T25-CadC Δ 108-151, high β -galactosidase activities were measured at pH 7.6, and low activity was found at pH 5.8. Hence, this variant also exhibits an inverted interaction pattern in comparison to wild-type CadC.

DISCUSSION

Using a combination of bioinformatics and NMR spectroscopy we could show that CadC proteins encompass a disordered cytoplasmic linker region that connects the

DNA-binding domain to the transmembrane helix and hence is responsible for transduction of the pH-signal. Systematic mutagenesis of the linker sequence revealed relative robustness against amino acid substitutions. By contrast large truncation led to the identification of CadC variants being constitutively active or even producing an inverse activation pattern (e.g., activation at high rather than low pH).

Dimerization of CadC is necessary for proper signal transduction. However we found the linker only in a monomeric form. Thus dimer formation might be exclusively mediated by the periplasmic pH-sensing domain [10]. The linker fragment is also conformational independent of the compact DNA-binding domain and exhibits a degree of flexibility compatible with a lack of secondary structure. Altogether we conclude that structural disorder rather than the sequence *per se* govern signal transduction via the CadC linker. This idea is consistent with the observation that multi-domain proteins are often interconnected by disordered regions. The main role(s) of these linkers might be to act as a spacer between input and output domains and/or provide orientational freedom [30].

Our results suggest a model in which a drop in external pH causes the disordered linker to undergo a conformational transition that enables adjacent DNA-binding domains to interact productively, thus allowing them to activate their target genes (Fig. 7). From previous studies it was known that acidification of the environment leads to protonation of a cluster of negatively charged amino acids at the dimer interface in the periplasmic domain of CadC [10], which is thought to stabilize dimerization of the sensor domain [10]. We now propose that this also leads to structural re-arrangement of the cytoplasmic linker, which permits the DNA-binding domain to homodimerize and enables *cadBA* induction (Fig. 7). Truncated CadC linker variants cause either the converse (pH-dependent) or a pH-independent response at the level of *cadBA* expression, confirming the importance of the linker for

signal transduction and dimer formation of the DNA-binding domains (Fig. 7). The linker thus couples the conformation of the periplasmic domain directly to the DNA-binding domain, and virtually any truncation of the linker causes a misinterpretation of the incoming signal. It is important to note that information transfer is robust against amino acid replacements, secondary structures and extension of the linker. It is conceivable that natural (species-specific) variations in linker length are required to span various distances between the membrane and the DNA. The flexibility of the disordered linker would allow it to serve as a tether between the helix-turn-helix domain and the transmembrane helix. We assume that the linker in CadC transduces the signal and defines the orientation of the DNA-binding domain. Likewise a linker region in DesK, a multipass transmembrane histidine kinase of *Bacillus subtilis*, mediates signal transduction by adopting two conformational states [31]. The histidine kinase EnvZ also has a cytoplasmic linker of approximately 50 amino acids. EnvZ linker variants had no effect on dimerization, and a role of the linker for the correct positioning of two EnvZ molecules within a dimer was proposed [28].

MATERIALS AND METHODS

Molecular biological techniques. Plasmid DNA and genomic DNA were isolated using the HiYield Plasmid Mini-Kit (Suedlaborbedarf) and the DNeasy Blood and Tissue Kit (Qiagen), respectively. DNA fragments were purified from agarose gels using the HiYield PCR Clean-up/Gel Extraction Kit (Suedlaborbedarf). Oligonucleotides were synthesized by Sigma-Aldrich. Restriction enzymes, calf alkaline phosphatase (CIP), T4 DNA ligase, Phusion high-fidelity DNA polymerase and One Taq DNA polymerase were purchased from New England Biolabs and used

according to the manufacturer's directions. Monoclonal anti-His antibodies were obtained from Qiagen, and goat anti-(mouse IgG)-alkaline phosphatase conjugate was supplied by Biomol. ^{15}N -labelled ammonium chloride and $\text{U-}^{13}\text{C}$ glucose were purchased from Cambridge Isotope Laboratories.

Bacterial strains and growth conditions. *E. coli* strains (Table 1) were grown overnight in Lysogeny Broth [32] or KE minimal medium [23] with 0.4 % (w/v) glucose as C-source. Both media were buffered with 100 mM potassium phosphate to either physiologically induce (pH 5.8) or repress (pH 7.6) *cadBA* expression by CadC. We measured the pH of the medium after inoculation and before harvesting the cells and found no significant differences ensuring pH-stability over time. For selection purposes, media were supplemented with antibiotics (ampicillin sodium salt: 100 $\mu\text{g/ml}$; kanamycin sulphate: 50 $\mu\text{g/ml}$; chloramphenicol: 34 $\mu\text{g/ml}$). Solid media were prepared with 1.5% (w/v) agar. *E. coli* strain DH5 α [33] was used as cloning host for the plasmids listed in Table 1. To determine *cadBA* expression in vivo, *E. coli* EP314 [9], which carries a *cadA-lacZ* fusion and lacks the native *cadC* gene, was complemented with the plasmids described previously [34]. *E. coli* BL21(DE)pLysS [35] was used for expression of *cadC* from the T7 promoter. Competent *E. coli* cells were prepared by treatment with RbCl and transformed as described previously [36]. The adenylate cyclase-deficient *E. coli* strain BTH101 was used for the BACTH assay [29]. Genomic DNA extracted from *E. coli* MG1655 K-12 [37] served as template for routine PCRs. Isotope labeling of protein for NMR studies was performed in M9 minimal medium supplemented with ^{15}N ammonium chloride (0.5 g/l) and unlabeled or $\text{U-}^{13}\text{C}$ glucose (2 g/l).

Construction of *cadC* variants. All plasmids used in this study are listed in Table 1. The plasmids pET16b-*cadC* [38], pET16b-*cadC*-NheI (this work), pETM11 (Gunter Stier, Biochemiezentrum Heidelberg) and pETTrx1a (Gunter Stier, Biochemiezentrum Heidelberg) served as templates for mutagenesis. Mutagenesis was performed by two-step PCR using synthetic primers containing the desired mutations [39]. A list of mismatched oligonucleotide primers is available upon request. PCR fragments encoding the desired amino acid substitutions were cloned between the *Nco*I and *Bam*HI restriction sites in pET16b-*cadC* (CadC_D131A, Cad_E146A, CadC_E109-114A, CadC_E109-114K, CadC_K152-155A, CadC_K152-155E, CadCE-109-114K_K152-155E), the *Spe*I and *Bam*HI sites of pET16b-*cadC* (CadC-*vibrio*), the *Nco*I and *Nde*I sites of pET16b-*cadC*-NheI (CadC- Δ 107-151, CadC- Δ 108-151, CadC- Δ 109-151, CadC- Δ 115-151, CadC- Δ 120-151, CadC- Δ 128-151, CadC- Δ 133-151, CadC- Δ 145-151, CadC- Δ 146-151, CadC- Δ 152-158) or the *Nco*I and *Xho*I sites of pETM11 and pETTrx1a (CadC₁₋₁₀₇ and CadC₁₋₁₅₉). Vectors for expression and purification of cytoplasmic CadC contained TEV protease recognition sites for subsequent proteolytic removal of the tags. CadC variants carried an N-terminal 10xHis tag (pET16b-based variants) or a N-terminal 6xHis tag (pETM11- and pETTrx1a-based variants) for immunodetection.

Cloning, expression and purification of recombinant proteins. LB medium was inoculated with single fresh clones of *E. coli* BL21(DE3) or BL21(DE3)pLysS cells and cultures were grown overnight at 37°C. A 1-ml aliquot of an overnight culture was used to inoculate an expression culture. Cells were grown to an OD₆₀₀ of 0.9 at 37°C, induced with 0.5 mM isopropyl-D-thiogalactopyranoside (IPTG) and grown overnight at 20°C before harvesting. Pellets were resuspended in lysis buffer

(50 mM Tris, 300 mM NaCl, 4 mM TCEP, 15 mM imidazole, 1mg/ml lysozyme, 10 µg/ml Dnase I, and protease inhibitors, pH 8.0), and incubated on ice for 30 min prior to sonication (8x45 s, 70% amplitude). Cleared lysates were then exposed to Ni²⁺-agarose beads. After intensive washing, beads were incubated with 500 µg/l culture of TEV protease in lysis buffer for 3 h and gentle shaking at room temperature. Subsequently, the bead supernatant was collected, concentrated and passed over a gel-filtration column in 20 mM Bis-Tris (pH 6.5) containing 500 mM NaCl. The monomer peak was pooled and salt concentration was adjusted to 150 mM. Due to its susceptibility to proteolytic degradation, CadC₁₋₁₅₉ was additionally purified by anion exchange chromatography. The gel-filtration buffer was exchanged for 20 mM Tris-50 mM NaCl (pH 8.0) and the sample was loaded onto a 1-mL MonoQ column (GE Healthcare). A gradient from 50-500 mM NaCl was applied over a volume of 10 mL and fractions of 500 µL collected and analyzed by SDS-PAGE. Fractions devoid of degradation product were pooled and dialyzed against the buffer used for storage of CadC₁₋₁₀₇.

NMR spectroscopy. Heteronuclear single quantum coherence (HSQC) NMR analysis of CadC₁₋₁₀₇ and CadC₁₋₁₅₉ was performed in 20 mM BisTris (pH 6.5), 150 mM NaCl and 10% D₂O. Backbone chemical-shift assignments for both polypeptides were recorded at protein concentrations of 1.2 mM and 600 µM, respectively. HNCA, HNCACB, CBCAcoNH, HbHaNH, HNCO, HNcaCO and 3D ¹⁵N-edited NOESY spectra [40] were acquired at 298°K on BrukerAvance III spectrometers equipped with TCI cryogenic probe heads, at field strengths corresponding to proton Larmor frequencies of 600, 800 and 900 MHz. Amide ¹⁵N R₁ and R₂ relaxation data and steady-state heteronuclear {¹H}-¹⁵N NOE experiments were performed as described

[41]. Spectra were processed with Topspin 3.2 and analyzed with CCPNMR Analysis [42] and Sparky [43].

Static light scattering. Static Light Scattering (SLS) measurements on CadC length variants (CadC₁₋₁₀₇ and CadC₁₋₁₅₉) were performed by connecting a Viscotek TDA 305 triple array detector to an Äkta Purifier equipped with an analytical size-exclusion column at 4°C. CadC₁₋₁₀₇ and CadC₁₋₁₅₉ samples (5 mg/ml) in gel filtration buffer were passed through a GE Superdex200 10/300 column at a flow rate of 0.5 ml/min. The molecular masses of the samples were calculated from the experimentally determined refractive index and right-angle light scattering signals using the Omnisec software package (Malvern Instruments). The SLS detector was calibrated with a BSA solution (4 mg/ml), taking 66.4 kDa as the molecular mass of the BSA monomer and assuming a dn/dc value of 0.185 ml/g for all protein samples.

Determination of CadC-mediated signal transduction in vivo. The effects of all CadC variants on P_{cadBA} activation were determined using a β -galactosidase-based reporter gene assay. For this purpose the reporter strain *E. coli* EP314 was transformed with plasmids encoding CadC or a variant of CadC (pET16b-based). Cells from an overnight culture aerobically grown in KE minimal medium (pH 7.6) supplemented with 0.2% (w/v) glucose [23] were inoculated into fresh KE minimal medium containing 0.2% glucose (w/v) and buffered to pH 5.8 or pH 7.6 (100 mM phosphate buffer), and the cell density was adjusted to an OD₆₀₀ of 0.05. Where indicated, lysine was added to a final concentration of 10 mM. Bacteria were grown to mid-exponential phase (OD₆₀₀ of 0.3-0.5) under microaerobic conditions, and harvested by centrifugation. β -Galactosidase assays were performed as described

previously [34]. Enzyme activity was determined from at least three independent replicates, and is given in Miller Units [44].

In vivo protein-protein interaction studies using BACTH. Protein-protein interactions were assayed with the bacterial adenylate cyclase-based two-hybrid system (BACTH) [29]. *E. coli* BTH101 was co-transformed with different pUT18C and pKNT25 variants [29]. Leucine-zipper fusion constructs were used to test assay functionality [29]. The plasmids T25-zip and zip-T18 were purchased from Euromedex [29] and used as positive controls for interaction. The PCR fragments used to generate the bacterial two-hybrid variants T18-CadC Δ 108-151 and T25-CadC Δ 108-151 were obtained from plasmid pET16b-cadC- Δ 108-151 and cloned into pUT18C and pKT25 with enzymes *Xba*I and *Bam*HI [29]. Cells were grown aerobically overnight in LB supplemented with 0.5 mM IPTG at 30°C. These cultures were then used to inoculate fresh KE minimal medium buffered to pH 5.6-8.0 and supplemented with 0.5 mM IPTG, 0.2% glucose and 10 mM lysine at OD₆₀₀ of 0.05. Bacteria were grown to mid-exponential phase under microaerobic conditions at 30°C, and harvested by centrifugation. β -Galactosidase assays were performed as previously described [44].

Detection of CadC and its variants in the membrane fraction. *E. coli* BL21(DE)pLysS cells [35] containing the plasmid of interest were grown to an optical density of OD₆₀₀ = 0.5 in LB. Overproduction of CadC and its variants was induced by adding 0.5 mM IPTG, and bacteria were harvested 3 h post induction. Isolation of membrane vesicles was carried out as described in [45]. The proteins were fractionated by SDS-PAGE [46] on 12.5% acrylamide gels and transferred to a

nitrocellulose membrane. Tagged proteins were labelled with primary monoclonal anti-His antibodies, and immunodetection was performed colorimetrically utilizing a secondary goat anti-(mouse IgG)-alkaline phosphatase antibody.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (Exc114/1 and Exc114/2). S.B. granted financial support by the Women program of the Center for Integrated Protein Science Munich. We thank Ingrid Weigl for excellent technical assistance, and Cathia Rausch and Thomas Heydenreich for help with construction and analysis of some CadC variants. We are grateful to Arie Geerlof (Helmholtz Center Munich) for help with the SLS measurements.

DATA DEPOSITION

NMR chemical shifts of CadC amino acids 1-159 were deposited in the BMRB and, upon acceptance, the entry will be accessible under ID 25417.

REFERENCES

- [1] Foster J. *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol.* 2004;2:898-907.
- [2] Fritz G, Koller C, Burdack K, Tetsch L, Haneburger I, Jung K, et al. Induction kinetics of a conditional pH stress response system in *Escherichia coli*. *J Mol Biol.* 2009;393:272-86.

- [3] Meng SY, Bennett GN. Regulation of the *Escherichia coli* cad operon: location of a site required for acid induction. *J Bacteriol.* 1992;174:2670-8.
- [4] Rauschmeier M, Schüppel V, Tetsch L, Jung K. New insights into the interplay between the lysine transporter LysP and the pH sensor CadC in *Escherichia coli*. *J Mol Biol.* 2014;426:215-29.
- [5] Tabor H, Hafner EW, Tabor CW. Construction of an *Escherichia coli* strain unable to synthesize putrescine, spermidine, or cadaverine: characterization of two genes controlling lysine decarboxylase. *J Bacteriol.* 1980;144:952-6.
- [6] Sabo DL, Boeker EA, Byers B, Waron H, Fischer EH. Purification and physical properties of inducible *Escherichia coli* lysine decarboxylase. *Biochemistry.* 1974;13:662-70.
- [7] Meng SY, Bennett GN. Nucleotide sequence of the *Escherichia coli* cad operon: a system for neutralization of low extracellular pH. *J Bacteriol.* 1992;174:2659-69.
- [8] Soksawatmaekhin W, Kuraishi A, Sakata K, Kashiwagi K, Igarashi K. Excretion and uptake of cadaverine by CadB and its physiological functions in *Escherichia coli*. *Mol Microbiol.* 2004;51:1401-12.
- [9] Neely MN, Dell CL, Olson ER. Roles of LysP and CadC in mediating the lysine requirement for acid induction of the *Escherichia coli* cad operon. *J Bacteriol.* 1994;176:3278-85.
- [10] Haneburger I, Eichinger A, Skerra A, Jung K. New insights into the signaling mechanism of the pH-responsive, membrane-integrated transcriptional activator CadC of *Escherichia coli*. *J Biol Chem.* 2011;286:10681-9.
- [11] Ude S, Lassak J, Starosta A, Kraxenberger T, Wilson D, Jung K. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science.* 2013;339:82-5.
- [12] Watson N, Duniak DS, Rosey EL, Slonczewski JL, Olson ER. Identification of elements involved in transcriptional regulation of the *Escherichia coli* cad operon by external pH. *J Bacteriol.* 1992;174:530-40.
- [13] Miller VL, Taylor RK, Mekalanos JJ. Cholera toxin transcriptional activator toxR is a transmembrane DNA binding protein. *Cell.* 1987;48:271-9.
- [14] Merrell DS, Camilli A. Regulation of *Vibrio cholerae* genes required for acid tolerance by a member of the "ToxR-like" family of transcriptional regulators. *Journal of Bacteriology.* 2000;182:5342-50.
- [15] Hase CC, Mekalanos JJ. TcpP protein is a positive regulator of virulence gene expression in *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America.* 1998;95:730-4.
- [16] Egan S, James S, Kjelleberg S. Identification and characterization of a putative transcriptional regulator controlling the expression of fouling inhibitors in *Pseudoalteromonas tunicata*. *Applied and Environmental Microbiology.* 2002;68:372-8.
- [17] Eichinger A, Haneburger I, Koller C, Jung K, Skerra A. Crystal structure of the sensory domain of *Escherichia coli* CadC, a member of the ToxR-like protein family. *Protein Sci.* 2011;20:656-69.
- [18] Merrell DS, Camilli A. Regulation of *Vibrio cholerae* genes required for acid tolerance by a member of the "ToxR-like" family of transcriptional regulators. *J Bacteriol.* 2000;182:5342-50.
- [19] McGuffin L, Bryson K, Jones D. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000;16:404-5.
- [20] Berjanskii M, Wishart DS. A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc.* 2005;127:14970-1.
- [21] Little R, Slavny P, Dixon R. Influence of PAS domain flanking regions on oligomerisation and redox signalling by NifL. *PLOS one.* 2012;7:e46651.
- [22] Lindner E, White S. Topology, dimerization, and stability of the single-span membrane protein CadC. *J Mol Biol.* 2014.

- [23] Epstein W, Kim B. Potassium transport loci in *Escherichia coli* K-12. *J Bacteriol.* 1971;108:639-44.
- [24] Camilloni C, De Simone A, Vranken WF, Vendruscolo M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry.* 2012;51:2224-31.
- [25] McGinnis S, Madden T. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 2004;32:W20-W5.
- [26] Kjaergaard M, Iešmantavičius V, Poulsen F. The interplay between transient α -helix formation and side chain rotamer distributions in disordered proteins probed by methyl chemical shifts. *Prot Sci.* 2011;20:2023-34.
- [27] Padmanabhan S, Marqusee S, Ridgeway T, Laue TM, Baldwin RL. Relative helix-forming tendencies of nonpolar amino acids. *Nature.* 1990.
- [28] Park H, Inouye M. Mutational analysis of the linker region of EnvZ, an osmosensor in *Escherichia coli*. *J Bacteriol.* 1997;179:4382-90.
- [29] Karimova G, Pidoux J, Ullmann A, Ladant D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci USA.* 1998;95:5752-6.
- [30] Liu Z, Huang Y. Advantages of proteins being disordered. *Prot Sci.* 2014;23:539-50.
- [31] Inda ME, Vandenbranden M, Fernández A, de Mendoza D, Ruyschaert J, Cybulski L. A lipid-mediated conformational switch modulates the thermosensing activity of DesK. *Proc Natl Acad Sci USA.* 2014;111:3579-84.
- [32] Maniatis T, Fritsch E, Sambrook J. *Molecular cloning: a laboratory manual*: Cold Spring Harbor Laboratory Cold Spring Harbor, NY; 1982.
- [33] Meselson M, Yuan R. DNA restriction enzyme from *E. coli*. *Nature.* 1968;217:1110.
- [34] Tetsch L, Koller C, Haneburger I, Jung K. The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. *Mol Microbiol.* 2008;67:570-83.
- [35] Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol.* 1986;189:113-30.
- [36] Hanahan D. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol.* 1983;166:557-80.
- [37] Blattner F, Plunkett G, Bloch C, Perna N, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. *Science.* 1997;277:1453-62.
- [38] Küper C, Jung K. CadC-mediated activation of the *cadBA* promoter in *Escherichia coli*. *J Mol Microbiol Biotechnol.* 2006;10:26-39.
- [39] Ho S, Hunt H, Horton R, Pullen J, Pease L. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene.* 1989;77:51-9.
- [40] Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog NMR Spectrosc.* 1999;34:93-158.
- [41] Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, et al. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry.* 1994;33:5984-6003.
- [42] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* 2005;59:687-96.
- [43] Goddard TD, Kneller DG. SPARKY 3. University of California, San Francisco.
- [44] Miller J. *Experiments in molecular genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory 1972.
- [45] Jung K, Tjaden B, Altendorf K. Purification, reconstitution, and characterization of KdpD, the turgor sensor of *Escherichia coli*. *J Biol Chem.* 1997;272:10847-52.

[46] Laemmli U. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970;227:680-5.

ACCEPTED MANUSCRIPT

FIGURE LEGENDS

Fig. 1. Domain architecture of *E. coli* CadC

E. coli CadC consists of a DNA-binding domain with a helix-turn-helix motif [amino acids (aa) 1-107], a cytoplasmic linker region (aa 108-158), a single transmembrane helix (aa 159-178) and a C-terminal periplasmic sensor domain (aa 179-512).

Fig. 2. Comparison of NMR amide fingerprints of CadC₁₋₁₀₇ and CadC₁₋₁₅₉.

Overlay of ¹H,¹⁵N-HSQC NMR spectra of CadC₁₋₁₀₇ and CadC₁₋₁₅₉. Spectra are colored according to the respective protein, i.e. CadC₁₋₁₀₇ (red) or CadC₁₋₁₅₉ (black). The predicted secondary structural elements of the two polypeptides are schematically depicted at the top. Signals assigned to the backbone amides of residues Trp106 and Tyr107 show small differences between the two proteins, as they are located in the segment that connects the DNA-binding domain to the linker. Signals derived from the linker region beyond residue 107 in CadC₁₋₁₅₉ (black) cluster in the center of the plot, which is typical for disordered residues and indicated by the boxed region. Selected assignments are indicated by residue number.

Fig. 3. Structural analysis of the CadC linker region by NMR spectroscopy. (A)

Secondary structure of CadC₁₋₁₅₉ derived from prediction in PSIPRED [19] and analysis of secondary chemical shifts shown in B. Arrows indicate β -strands, cylinders represent α -helices. Unstructured regions are depicted as straight lines, links between secondary structural elements are indicated as loops. (B) Combined ¹³C secondary chemical shifts ($\Delta\delta^{13}\text{C}\alpha - \Delta\delta^{13}\text{C}\beta$) are plotted against CadC₁₋₁₅₉ residue numbers. The green dashed lines mark thresholds indicative of α -helical (+2) or β -

strand (-2) secondary structure at a residue's position. **(C)** $\{^1\text{H}\}$ - ^{15}N heteronuclear NOEs for backbone amides are plotted against CadC residue number. Errors derive from peak height uncertainties in the underlying spectra and their propagation. The green dashed line indicates the theoretical maximum value for rigid residues. **(D, E)** Longitudinal (R_1) and transversal (R_2) ^{15}N relaxation rates of CadC₁₋₁₅₉ plotted against residue number. Errors derive from relaxation curve fits. **(F)** Tumbling correlation time (τ_c) values derived from the R_2/R_1 in D and E ratio of CadC₁₋₁₅₉. The uncertainty in τ_c is propagated from individual R_1 and R_2 errors and does not exceed 2.6 ns. The dashed lines show the average values for residues that have two distinct average values for the tumbling correlation times within CadC₁₋₁₅₉. Gaps indicate prolines, unassigned residues or residues discarded from the analysis due to fitting errors.

Fig. 4. Oligomerization of CadC fragments probed by static light scattering.

Analytical gel-filtration chromatograms of CadC₁₋₁₀₇ (panel **A**, expected molecular weight (MW): 12.6 kDa; 5 mg/ml) and CadC₁₋₁₅₉ (panel **B**, expected MW: 18.2 kDa; 5 mg/ml) based on profiling of UV absorption (blue curves) and refractive indices. Measured molecular weights derived from the elution profiles by the Omnisec software are indicated. Insets show the main peak of interest (indicated by an arrow) with their refractive indices and elution curves plotted as binary logarithm ($\log_2\text{MW}$), respectively. Samples were run on a SD200 semi-analytical column (GE Healthcare) of 23 ml volume.

Fig. 5. Effects of linker truncation on *cadBA* expression. **A** Reporter gene assays were performed with *E. coli* EP314 ($\Delta cadC$, *cadA::lacZ* fusion), which was complemented with plasmid-encoded *cadC* or the indicated *cadC* variants. Cells

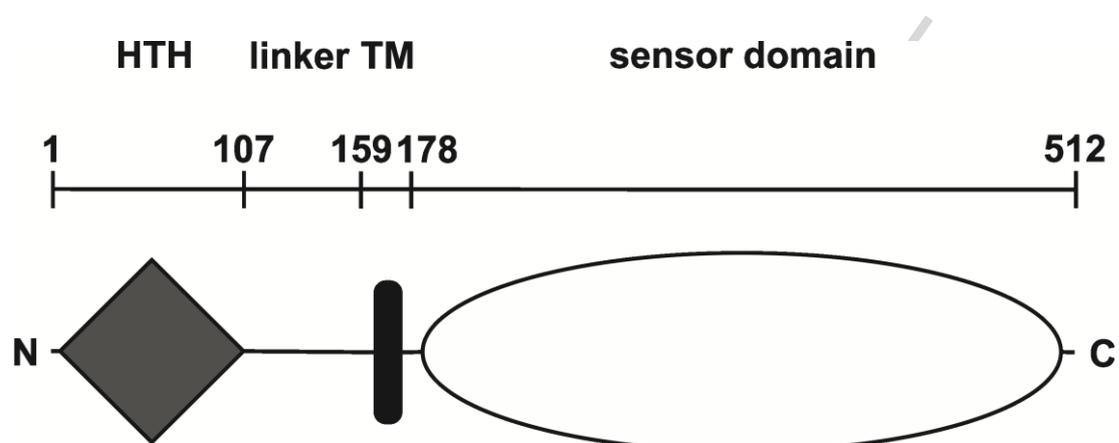
were grown to mid-exponential phase under microaerobic conditions at 37°C in minimal medium buffered at pH 5.8 (black bars) or pH 7.6 (white bars) in the presence of 10 mM lysine. The activity of the reporter enzyme β -galactosidase, which serves as a measure of *cadBA* expression, was determined according to Miller (1972) and expressed in Miller Units (MU). The experiments were performed in triplicate, and error bars indicate standard deviations of the mean. **B** Schematic representation of the cytoplasmic linker in CadC. Deleted segments of the linker region are indicated by residue numbers. In addition, two polyproline clusters (aa 120-124 and aa 145-147) are boxed.

Fig. 6. pH-dependent dimerization of CadC (A) and CadC- Δ 108-151 (B). A two-hybrid assay based on fragmented bacterial adenylate cyclase (CyaA) was used to detect interaction of the hybrid pairs T18-CadC and CadC-T25 in vivo. Cells were grown to mid-exponential phase at 30°C under microaerobic conditions in KE minimal medium buffered at pH 5.6-8.0 (**A**) or at pH 5.8 and pH 7.6 (**B**) and supplemented with 0.5 mM IPTG and 10 mM lysine. The activity of the reporter enzyme β -galactosidase was determined and served as a measure of the interaction strength [44]. The experiments were performed in triplicate, and error bars indicate standard deviations of the mean.

Fig. 7. Model for CadC-mediated signal transduction. The pH sensor and transcriptional activator CadC consists of a C-terminal periplasmic sensor domain (ON state in green, OFF state in red), a single transmembrane helix (shaded black) and an N-terminal cytoplasmic output domain comprising a linker and the DNA-binding domain (ON state in green, OFF state in red). **A** Signal perception by the wild-type periplasmic domain switches the sensor into the ON state by inducing a

structural rearrangement that is mediated by the cytoplasmic linker and thereby enables the DNA-binding domains to interact with each other and activate *cadBA* expression. **B** In contrast, the truncated CadC- Δ 108-151 linker variant induces *cadBA* expression in the absence of pH stress. This can be explained if truncation of the linker forces CadC to adopt the active conformation at physiological pH. At low pH the periplasmic domain perceives the signal, and undergoes conformational changes that are transduced to the DNA-binding domain in such a way as to switch the protein into the OFF state.

Figure 1



ACCEPTED MANUSCRIPT

Figure 2

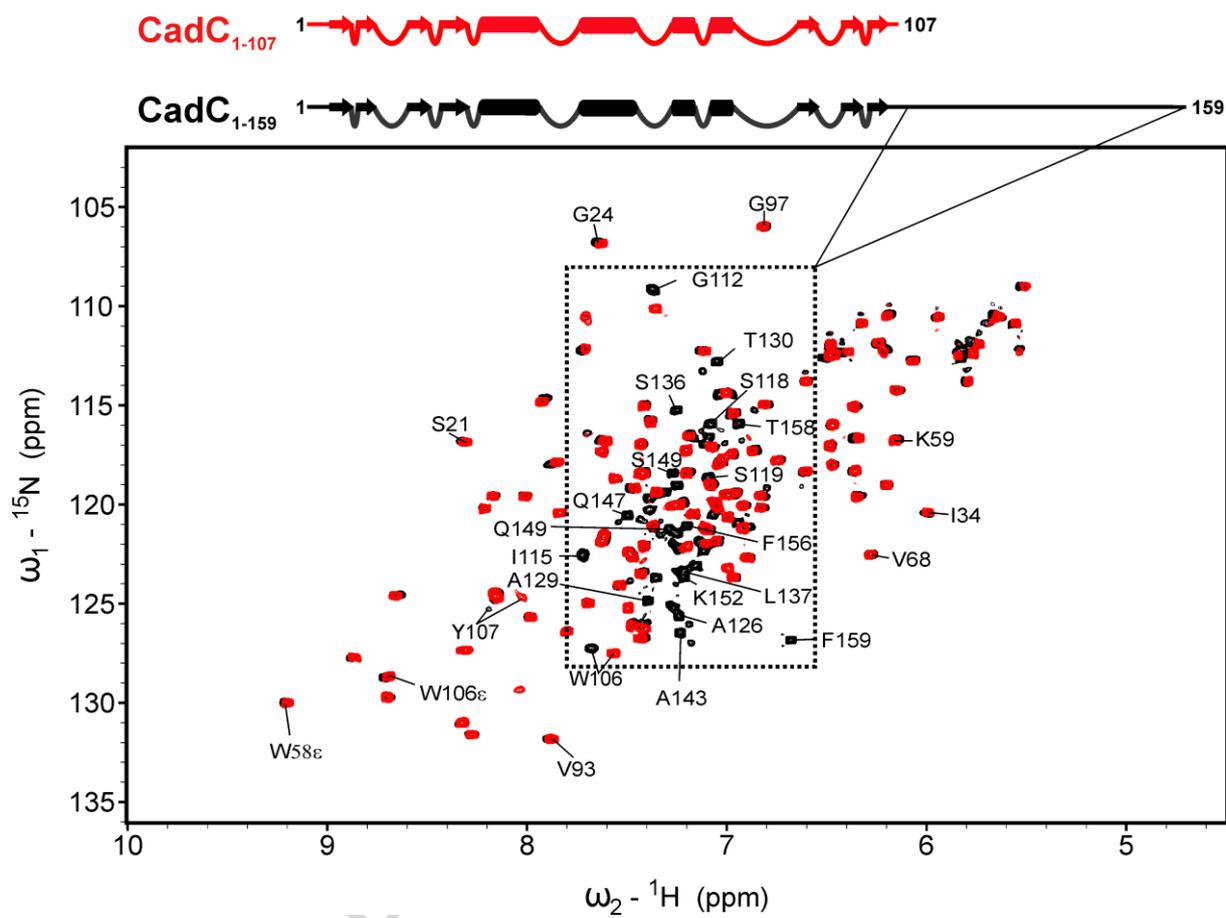


Figure 3

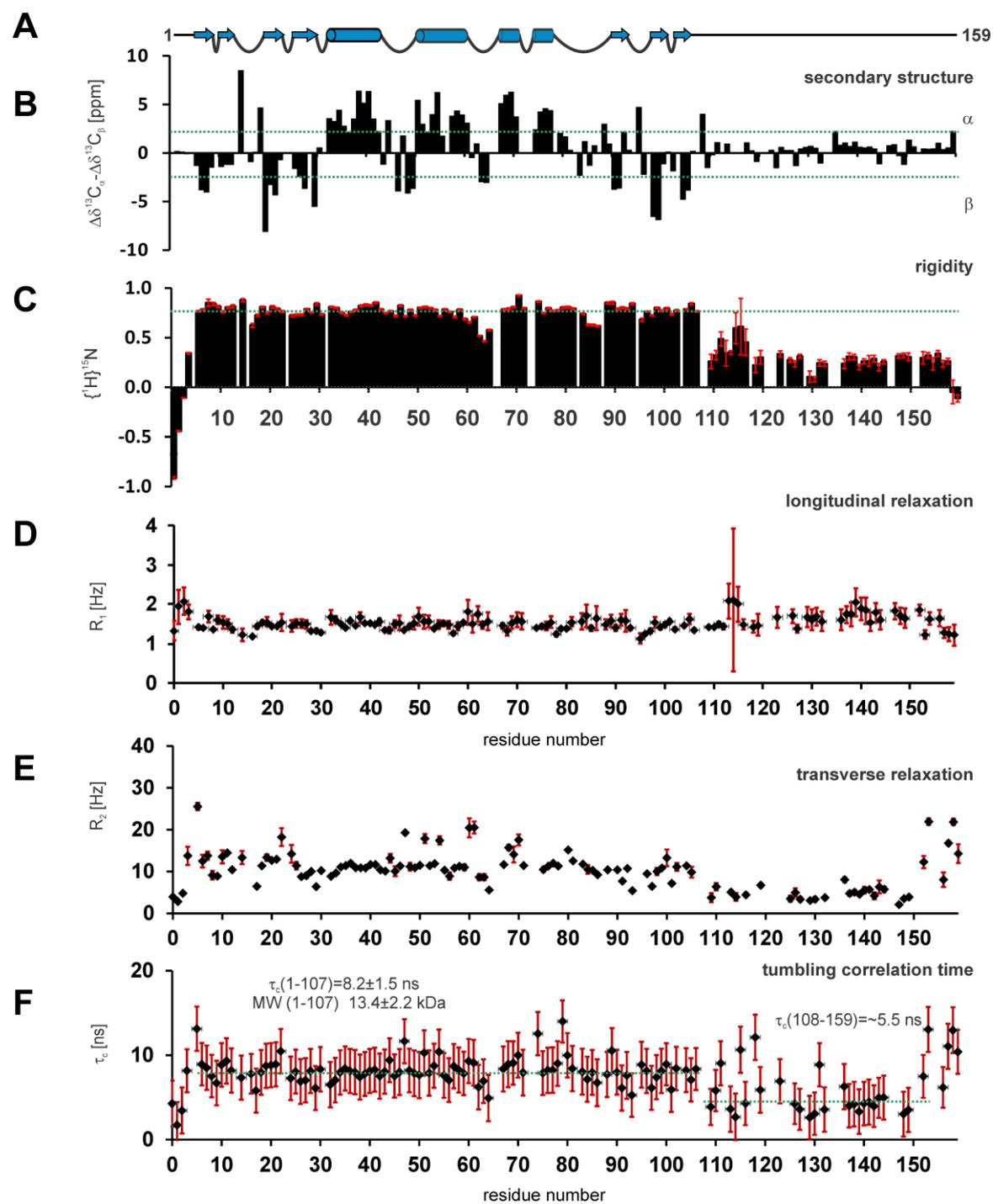


Figure 4

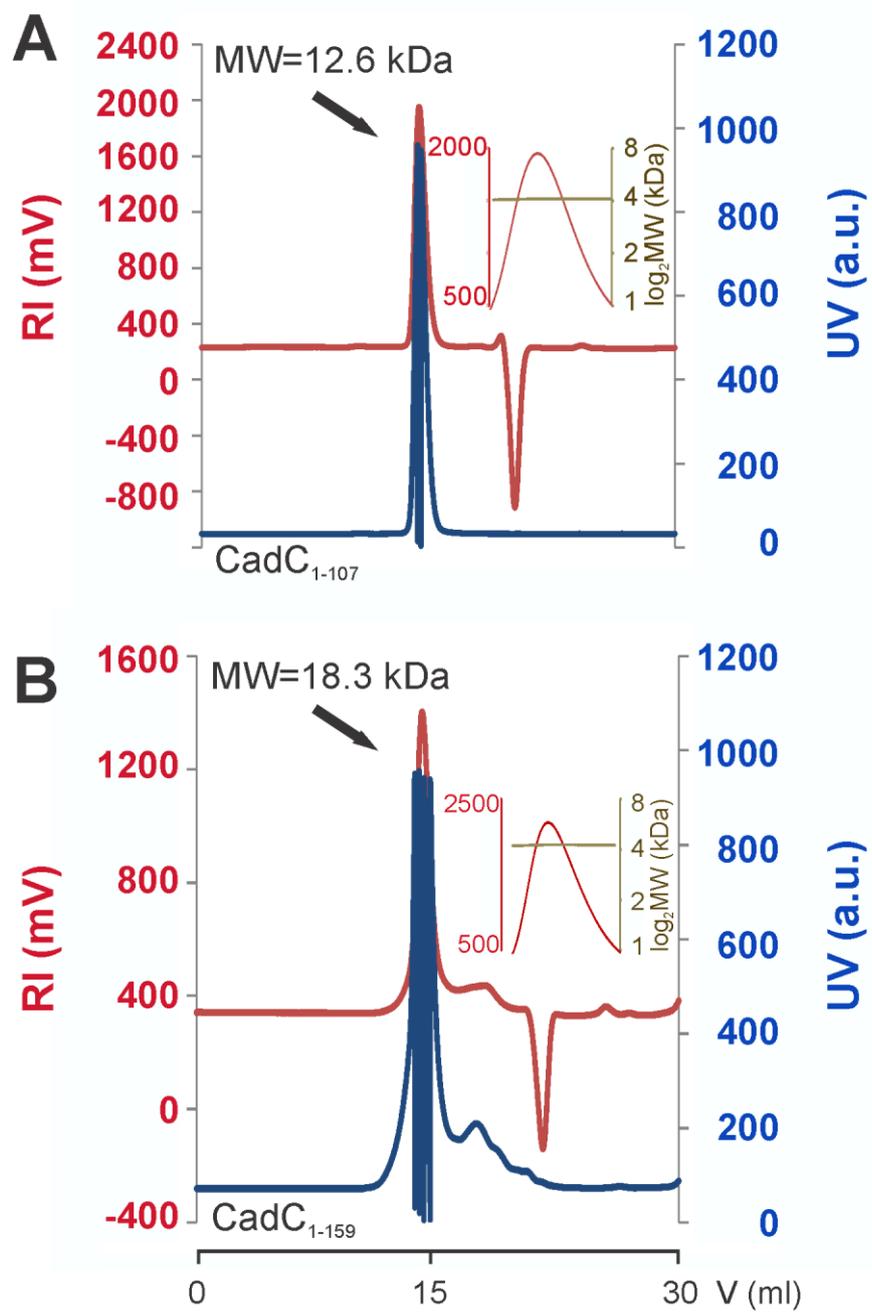


Figure 5

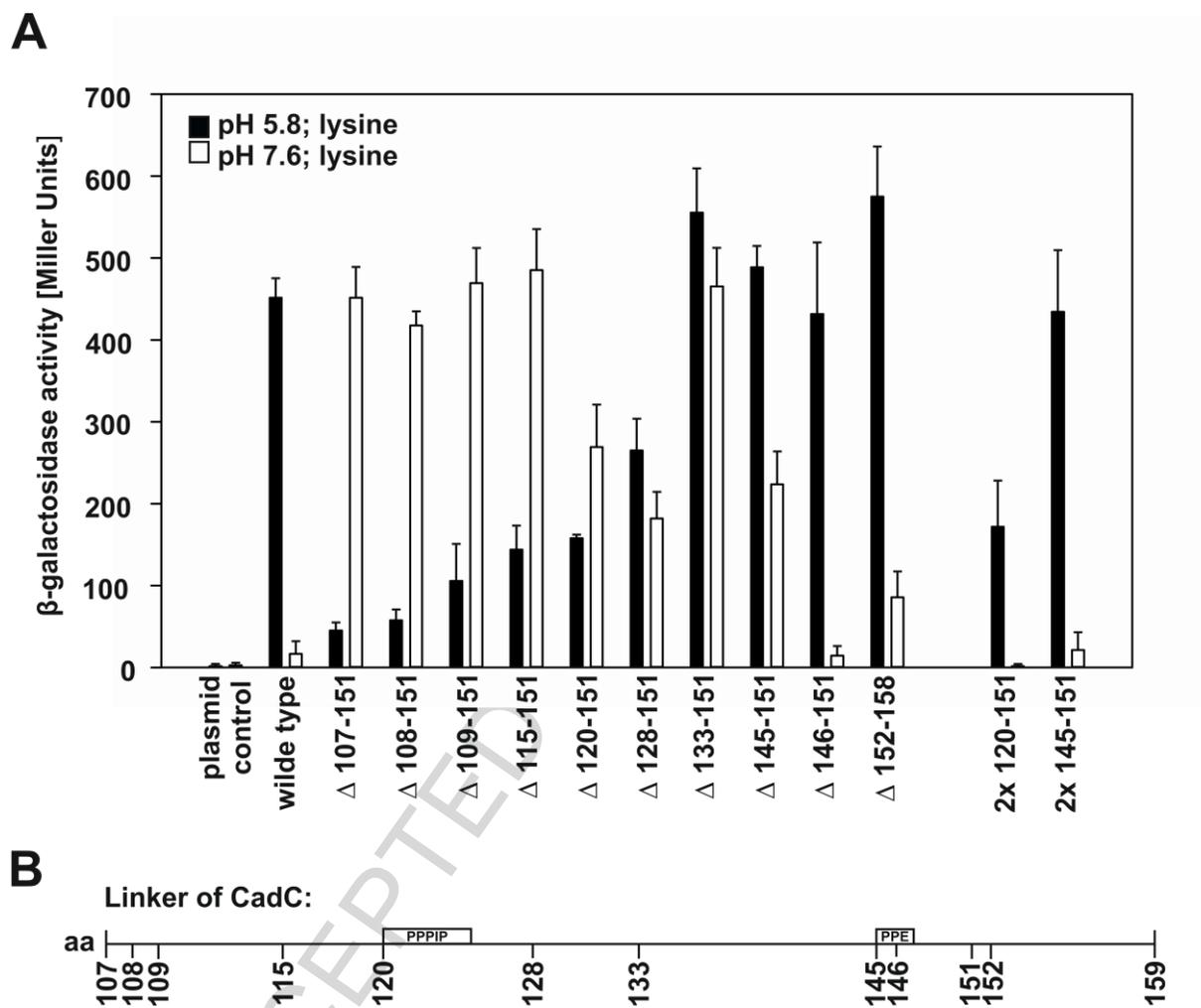
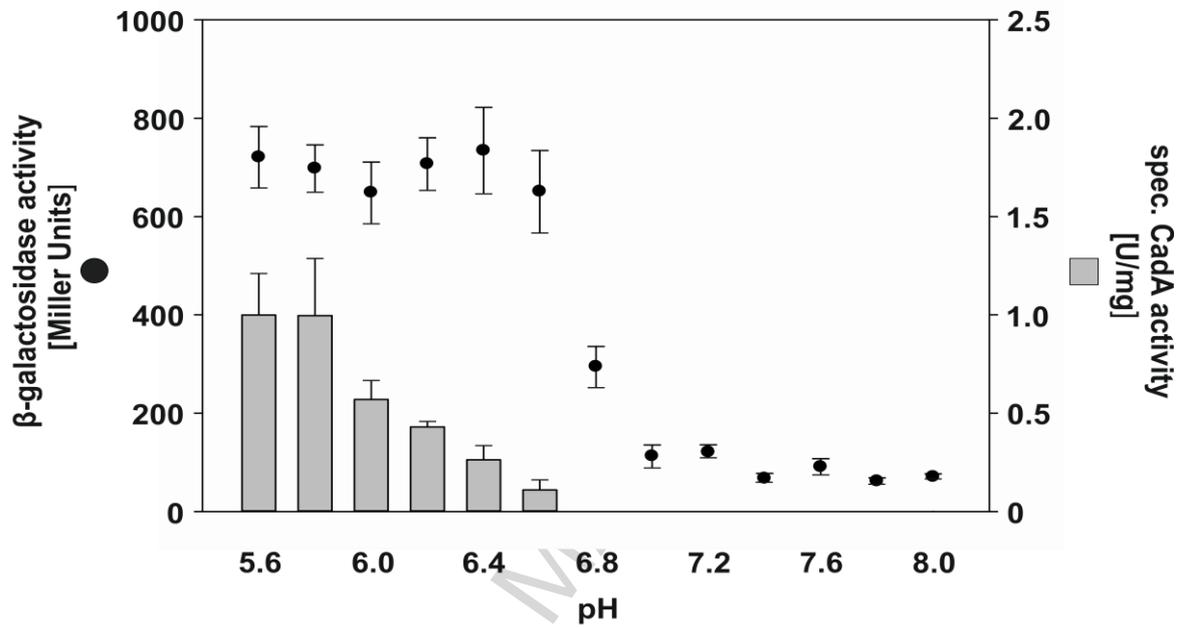


Figure 6

A



B

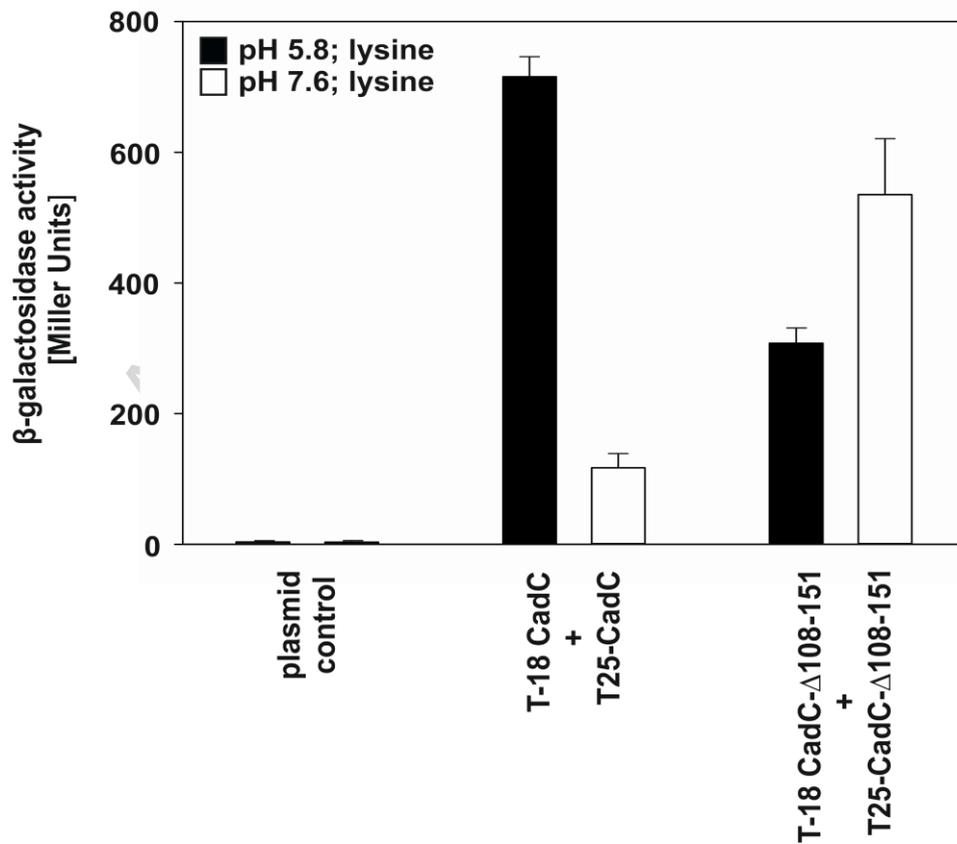
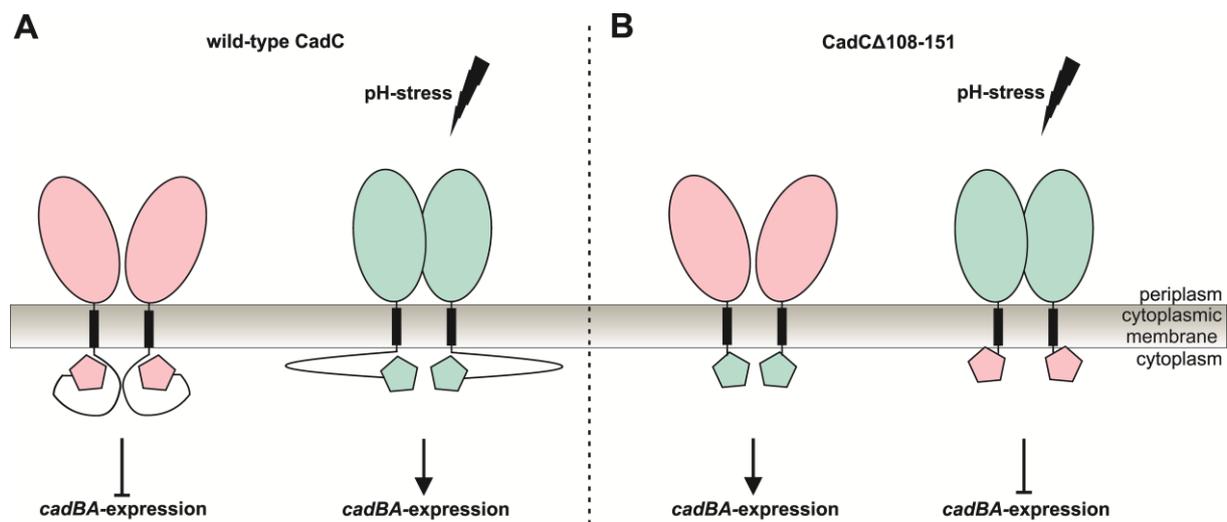


Figure 7



ACCEPTED MANUSCRIPT

TABLES

TABLE 1. Bacterial strains and plasmids used in this study

Strain	Relevant genotype or description	Reference or source
<i>E. coli</i> strains		
<i>E. coli</i> MG1655	F ⁻ λ ⁻ <i>ilvG rfb50 rph-1</i>	[36]
<i>E. coli</i> EP314	F ⁻ IN(<i>rmD-rmE</i>) Δ(<i>lacIOPZYA</i>) <i>exa-1::MuD11734 (Km lac)^a</i> <i>cadC1::Tn10</i>	[9]
<i>E. coli</i> DH5α	F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG</i> Φ80d <i>lacZΔM15</i> Δ(<i>lacZYA-argF</i>)U169, <i>hsdR17</i> (r _K ⁻ m _K ⁺), λ ⁻	[33]
<i>E. coli</i> BL21(DE3) pLysS	F ⁻ <i>ompT r_B m_B dcm gal</i> (DE3) pLysS (Cm ^R) <i>tonA</i>	[34]
<i>E. coli</i> BTH101	F ⁻ <i>cyaA-99 araD139 galE15 galK16 rpsL1 hsdR2 μrA1 μrB1</i>	[30]
Plasmids		
pET16b	Expression vector, Amp ^R	Novagen
pET16b- <i>cadC</i>	<i>cadC</i> in pET16b	[37]
pET16b- <i>cadC</i> -NheI	NheI restriction site in <i>cadC</i>	This work
pET16b- <i>cadC</i> -Δ107-151	<i>cadC</i> -Δ107-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ108-151	<i>cadC</i> -Δ108-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ109-151	<i>cadC</i> -Δ109-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ115-151	<i>cadC</i> -Δ115-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ120-151	<i>cadC</i> -Δ120-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ128-151	<i>cadC</i> -Δ128-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ133-151	<i>cadC</i> -Δ133-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ145-151	<i>cadC</i> -Δ145-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ146-151	<i>cadC</i> -Δ146-151 in pET16b	This work
pET16b- <i>cadC</i> -Δ152-158	<i>cadC</i> -Δ152-158 in pET16b	This work
pET16b- <i>cadC</i> -2x120-151	<i>cadC</i> -2x120-151 in pET16b	This work
pET16b- <i>cadC</i> -2x145-151	<i>cadC</i> -2x145-151 in pET16b	This work
pET16b- <i>cadC</i> -E125A	<i>cadC</i> -E125A in pET16b	This work
pET16b- <i>cadC</i> -D131A	<i>cadC</i> -D131A in pET16b	This work
pET16b- <i>cadC</i> -E147A	<i>cadC</i> -E147A in pET16b	This work
pET16b- <i>cadC</i> - ¹⁰⁹ AAAGAA ¹¹⁴	<i>cadC</i> - ¹⁰⁹ AAAGAA ¹¹⁴ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁰⁹ RRRGRR ¹¹⁴	<i>cadC</i> - ¹⁰⁹ RRRGRR ¹¹⁴ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁵² ASAA ¹⁵⁵	<i>cadC</i> - ¹⁵² ASAA ¹⁵⁵ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁵² ESEE ¹⁵⁵	<i>cadC</i> - ¹⁵² ESEE ¹⁵⁵ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁰⁹ RRRGRR ¹¹⁴ - ¹⁵² ESEE ¹⁵⁵	<i>cadC</i> - ¹⁰⁹ RRRGRR ¹¹⁴ - ¹⁵² ESEE ¹⁵⁵ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁴⁶ AEQSPV ¹⁵¹	<i>cadC</i> - ¹⁴⁶ AEQSPV ¹⁵¹ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁴⁶ PEASPV ¹⁵¹	<i>cadC</i> - ¹⁴⁶ PEASPV ¹⁵¹ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁴⁶ PEQAPV ¹⁵¹	<i>cadC</i> - ¹⁴⁶ PEQAPV ¹⁵¹ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁴⁶ PEQSAV ¹⁵¹	<i>cadC</i> - ¹⁴⁶ PEQSAV ¹⁵¹ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁴⁶ PEQSPA ¹⁵¹	<i>cadC</i> - ¹⁴⁶ PEQSPA ¹⁵¹ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁴⁶ AAAAAV ¹⁵¹	<i>cadC</i> - ¹⁴⁶ AAAAAV ¹⁵¹ in pET16b	This work
pET16b- <i>cadC</i> - ¹⁴⁶ AAAAAA ¹⁵¹	<i>cadC</i> - ¹⁴⁶ AAAAAA ¹⁵¹ in pET16b	This work
pET16b- <i>cadC</i> -vibrio	<i>cadC</i> -Δ101-159- <i>E.coli</i> _Ins101-185- <i>V. fischeri</i>	This work

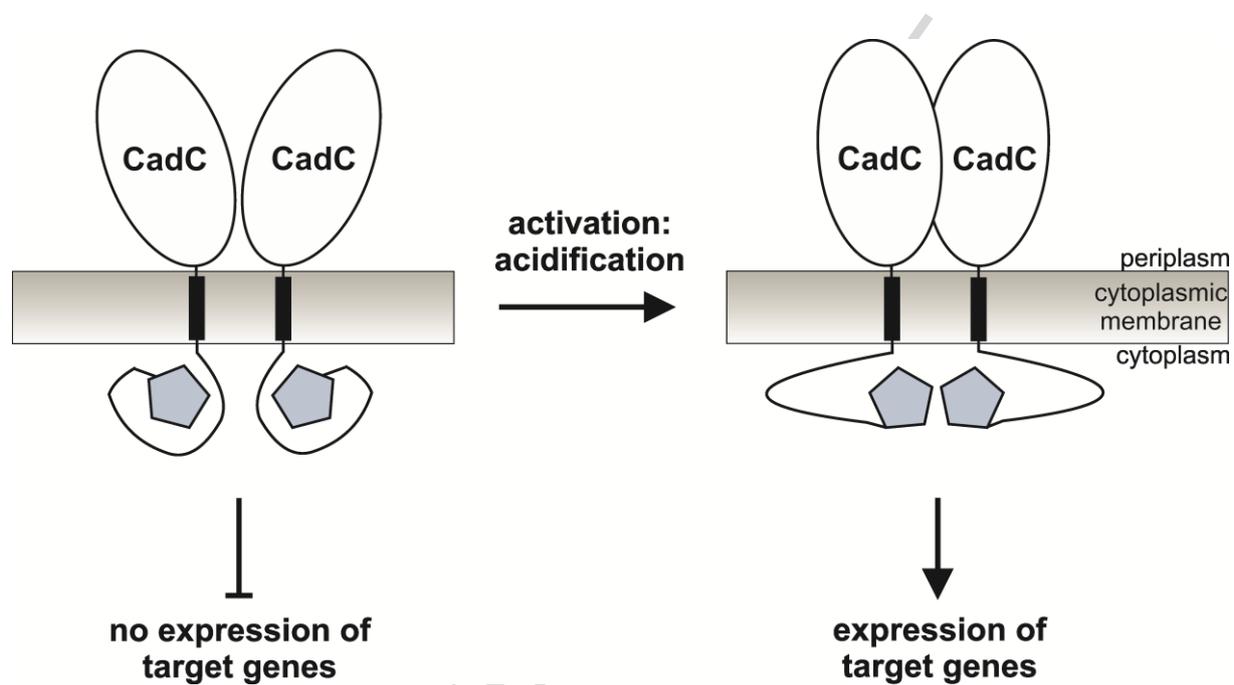
pET16b-cadC- ¹¹⁹ SPPPPPIE ¹²⁵	<i>cadC</i> - ¹¹⁹ SPPPPPIE ¹²⁵ in pET16b	This work
pET16b-cadC- ¹¹⁹ SPPPPPE ¹²⁵	<i>cadC</i> - ¹¹⁹ SPPPPPE ¹²⁵ in pET16b	This work
pET16b-cadC- ¹¹⁹ PPPPPE ¹²⁵	<i>cadC</i> - ¹¹⁹ PPPPPE ¹²⁵ in pET16b	This work
		This work
pUT18C	Expression vector, Amp ^r	[30]
pKT25	Expression vector, Kan ^r	[30]
pUT18C-zip	Control plasmid, N-terminal CyaA-T18-yeast leucine-zipper fusion, Amp ^r	[30]
pKT25-zip	Control plasmid, N-terminal CyaA-T25-yeast leucine-zipper fusion, Kan ^r	[30]
pUT18C-cadC	<i>cadC</i> in pUT18C (T18-CadC)	[4]
pKT25-cadC	<i>cadC</i> in pKT25 (CadC-T25)	[4]
pUT18C-cadC-Δ108-151	<i>cadC</i> -Δ108-151 in pUT18C	This work
pKT25- cadC-Δ108-151	<i>cadC</i> -Δ108-151 in pKT25	This work
pETM11	Expression vector, Kan ^R	Gunter Stier, BCZ Heidelberg
pETM11-cadC1-107	<i>cadC</i> -Δ108-151 in pETM11	This work
pETM11-cadC1-159	<i>cadC</i> -Δ160-512 in pETM11	This work
pETTrx1a	Expression vector, Kan ^R , Trx site	Gunter Stier, BCZ Heidelberg
pETTrx1a-cadC1-107	<i>cadC</i> -Δ108-512 in pETTrx1a	This work
pETTrx1a-cadC1-159	<i>cadC</i> -Δ160-512 in pETTrx1a	This work

TABLE 2. Characterization of CadC variants in *E. coli* EP314

Reporter gene assays were performed with *E. coli* EP314, which carries a chromosomal *cadA-lacZ* fusion and lacks the *cadC* gene. Cells were cultivated and β -galactosidase activities measured as described in the legend to Fig. 5, and are given in Miller Units [MU]. Production and integration of CadC variants into the cytoplasmic membrane was tested by Western immunoblotting (data not shown).

CadC variant	<i>cadBA</i> expression			
	(β -galactosidase activity [MU])			
	pH 5.8		pH 7.6	
	0 mM lysine	10 mM lysine	0 mM lysine	10 mM lysine
CadC wild-type	5 \pm 3	420 \pm 37	4 \pm 1	3 \pm 2
No CadC	3 \pm 2	4 \pm 1	2 \pm 1	3 \pm 1
charged residues				
CadC-E125A	8 \pm 7	590 \pm 62	2 \pm 1	18 \pm 2
CadC-D131A	3 \pm 1	504 \pm 42	4 \pm 3	15 \pm 3
CadC-E147A	8 \pm 7	518 \pm 42	3 \pm 2	9 \pm 2
CadC- ¹⁰⁹ AAAGAA ¹¹⁴	3 \pm 3	230 \pm 40	1 \pm 1	7 \pm 3
CadC- ¹⁰⁹ RRRGRR ¹¹⁴	2 \pm 1	22 \pm 5	3 \pm 1	11 \pm 2
CadC- ¹⁵² ASAA ¹⁵⁵	2 \pm 1	376 \pm 58	1 \pm 1	22 \pm 5
CadC- ¹⁵² ESEE ¹⁵⁵	4 \pm 1	541 \pm 73	3 \pm 2	13 \pm 4
CadC- ¹⁰⁹ RRRGRR ¹¹⁴ - ¹⁵² ESEE ¹⁵⁵	2 \pm 1	14 \pm 3	3 \pm 1	4 \pm 1
transient helix				
CadC- ¹⁴⁶ AEQSPV ¹⁵¹	3 \pm 2	369 \pm 22	2 \pm 1	3 \pm 1
CadC- ¹⁴⁶ PEASPV ¹⁵¹	1 \pm 1	380 \pm 40	1 \pm 1	11 \pm 7
CadC- ¹⁴⁶ PEQAPV ¹⁵¹	1 \pm 1	410 \pm 70	3 \pm 2	2 \pm 1
CadC- ¹⁴⁶ PEQSAV ¹⁵¹	4 \pm 3	422 \pm 45	1 \pm 1	13 \pm 4
CadC- ¹⁴⁶ PEQSPA ¹⁵¹	6 \pm 2	321 \pm 21	12 \pm 2	7 \pm 4
CadC- ¹⁴⁶ AAAAAV ¹⁵¹	6 \pm 4	435 \pm 51	5 \pm 3	4 \pm 1
CadC- ¹⁴⁶ AAAAAA ¹⁵¹	7 \pm 3	381 \pm 45	4 \pm 1	5 \pm 2
insertion of proline				
CadC- ¹¹⁹ SPPPPIE ¹²⁵	8 \pm 4	389 \pm 34	2 \pm 1	7 \pm 2
CadC- ¹¹⁹ SPPPPPE ¹²⁵	7 \pm 4	401 \pm 46	6 \pm 3	3 \pm 1
CadC- ¹¹⁹ PPPPPE ¹²⁵	4 \pm 1	433 \pm 56	4 \pm 2	3 \pm 1
CadC chimera				
CadC-vibrio	632 \pm 65	492 \pm 107	29 \pm 11	155 \pm 29

Graphical abstract



ACCEPTED

HIGHLIGHTS:

- Structure and function of the signaling linker in the pH-responsive one-component system CadC
- The cytoplasmic CadC linker is disordered in solution and robust against substitutions
- The CadC linker translates an environmental stimulus into a structural rearrangement

ACCEPTED MANUSCRIPT