Title
Topology, Dimerization, and Stability of the Single-Span Membrane Protein CadC

Journal
Journal of Molecular Biology, 426(16)

ISSN
0022-2836

Authors
Lindner, Eric
White, Stephen H

Publication Date
2014-08-01

DOI
10.1016/j.jmb.2014.06.006

Peer reviewed

Published in final edited form as:
J Mol Biol. 2014 August 12; 426(16): 2942–2957. doi:10.1016/j.jmb.2014.06.006.

Topology, Dimerization, and Stability of the Single-Span Membrane Protein CadC

Eric Lindner and **Stephen H. White***
Department of Physiology & Biophysics and the Center for Biomembrane Systems, University of California, Irvine, Irvine, CA 92697-4560

Abstract

Under acid stress, *Escherichia coli* induce expression of CadA (lysine decarboxylase) and CadB (lysine/cadaverine antiporter) in a lysine-rich environment. The ToxR-like transcriptional activator CadC controls expression of the *cadBA* operon. Using a novel Signal Peptidase I (SPase I) cleavage assay, we show that CadC is a Type II single-span membrane protein (MP) with a cytoplasmic DNA binding domain and a periplasmic sensor domain. We further show that, as long assumed, dimerization of the sensor domain is required for activating the *cadBA* operon. We prove this using a chimera in which the periplasmic domain of RodZ—a Type II MP involved in the maintenance of the rod shape of *E. coli*—replaces the CadC sensor domain. Because the RodZ periplasmic domain cannot dimerize, the chimera cannot activate the operon. However, replacement of the TM domain of the chimera with the glycophorin-A (GpA) TM domain causes intramembrane dimerization and consequently operon activation. Using a low-expression protocol that eliminates extraneous TM-helix dimerization signals arising from protein over-expression, we enhanced dramatically the dynamic range of the β-galactosidase assay for *cadBA* activation. Consequently, the strength of the intramembrane dimerization of the GpA domain could be compared quantitatively with the strength of the much stronger periplasmic dimerization of CadC. For the signal-peptidase assay, we inserted a SPase I cleavage site (AAA or AQA) at the periplasmic end of the TM helix. Cleavage occured with high efficiency for all TM and periplasmic domains tested, thus eliminating the need for the cumbersome spheroplast-proteinase K method for topology determinations.

Keywords

ToxR; RodZ; *cadBA* operon; signal peptidase; type II membrane protein; single-span membrane protein topology; transmembrane helix dimerization

*To whom correspondence should be addressed: stephen.white@uci.edu or Dept. of Physiology and Biophysics, Medical Sciences D346, University of California at Irvine, Irvine, CA 92697-4560.

Introduction

Crucial information about helix-helix interactions in membranes has come from studies of the single-span human erythrocyte sialoglycoprotein glycophorin A (GpA), which forms strong dimers in SDS [1] and lipid bilayers [2]. Engelman and co-workers took advantage of this observation to discover the location and properties of the dimerization domain, –L_{75}IxxG$_{79}$V$_{80}$xxG$_{83}$V$_{84}$xxT$_{87}$–, within the GpA transmembrane TM helix [3–4] and to determine the structure of the dimer in detergent micelles using NMR [5]. Langosch et al. [6] extended GpA dimerization measurements to *Escherichia coli* inner membranes by taking advantage of the properties of the single-span membrane protein (S-SMP) ToxR that regulates virulence-gene expression in *Vibrio cholerae* [7–8]. Dimeric ToxR binds to tandemly repeated DNA elements within the *ctx* promoter to initiate transcription of *ctx* genes. Langosch and his colleagues created ToxR chimeric proteins bearing GpA variants in the TM domain and maltose binding protein (MalE) as the periplasmic domain. placing the *lacZ* gene under the control of the *ctx* promotor, β-galactosidase (β-gal) activity could be used as an *in vivo* readout of TM dimerization. They showed that, as for dimerization in SDS micelles, G79A and G83A mutations disrupted formation of the GpA dimer interface and thereby significantly reduced β-gal activity.

ToxR is a member of the LysR-type transcription regulator (LTTR) family [9]. Members of the family are typically 300 residues long, have a helix-turn-helix (HTH) DNA-binding motif at the N-terminus, and a co-factor-binding domain at the C-terminus. An LTTR family member of particular interest to our lab is CadC (Fig. 1c) that regulates the expression of the *cadBA* operon (Fig. 1a). CadB, a lysine-cadaverine antiporter, and CadA, a lysine decarboxylase (Fig. 1b), are among several proteins that *E. coli* expresses at times of acid stress (see review by Kanjee & Houry [10]). A structurally similar protein of interest to us is RodZ (Figure 1c), which is involved in the maintenance of the rod shape of *E. coli* [11–12]. Oddly, it has an N-terminal HTH motif, but does not bind to DNA. Rather, it binds to MreB in a complex of proteins involved in shape control. CadC and RodZ attracted our attention, because both proteins have single TM segments that are more than 100 residues downstream of the N-terminus (Figure 1c). In contrast, most single-span Type II MPs have their signal-anchor sequences at, or very close to, the N-terminus. We have recently shown that membrane insertion of RodZ requires only the SecYEG translocon, the SecA ATPase motor, and the transmembrane proton motive force (PMF) [13]. The relative simplicity of RodZ insertion makes it an ideal model system for studying the biogenesis of S-SMPs. Studies in progress will reveal how similar the trafficking and assembly of CadC are to those of RodZ. Here, our goal is to report studies of CadC that show it to be useful for *in vivo* studies of TM helix stability and dimerization.

Recent studies of CadC, primarily in the laboratory of Kirsten Jung [14–18], have revealed the basic principles of CadC function. Because of its similarity to ToxR and related proteins, CadC is assumed to be a single-span Type II (N_{in}-C_{out}) membrane protein. But the topology has not been definitively established, probably, as we show here, because the periplasmic domain is completely resistant to proteinase K (protK) treatment of *E. coli* spheroplasts. To overcome this problem, we developed a new method of topology determination in which a

Signal Peptidase I (SPase I) cleavage site (-AXA-) is inserted on the periplasmic side of the TM helix. We show that this is a robust assay for testing the topology of S-SMPs.

Because dimerization is a common feature of gene activation by ToxR-like proteins, a reasonable assumption is that dimerization of the periplasmic pH-sensor domain of CadC is an essential feature of its activation of the *cadBA* operon (Fig. 1a). Although biochemical [18] and structural evidence [19] supports this assumption, the necessity for dimerization in CadC function has not been demonstrated directly. The strongest evidence is the recent structure of the presumed CadC periplasmic domain [19]. This structure and associated biochemical measurements show that CadC crystallizes as a biological dimer. Furthermore, the symmetry of the dimer places N- and C-termini close to one another in a manner consistent membrane anchoring of the dimer by the TM segments of the monomers (see the graphical abstract). Using the SPase I assay, we show that cleavage of the periplasmic domains results in loss of the ability of CadC to activate the *cadBA* operon. Also, a CadC chimera with the non-dimerizing RodZ periplasmic domain fails to activate the operon. Following the lead of Langosch et al. [6], we show that replacement of the CadC TM domain by the GpA TM helix in the CadC-RodZ chimera restores activation. Studies of other CadC-RodZ chimeras with TM segments of different amino acid compositions reveal that the application of the SPase I assay will likely be an effective tool for assessing TM helix stability.

A persistent problem with ToxR-like β-gal assays for TM helix dimerization as typically implemented (including the CadC β-gal assay) is high background β-gal activity. We show here that this background originates from high protein expression that can cause dimerization through simple crowding of highly hydrophobic TM segments lacking obvious dimerization motifs. We developed a low-expression *E. coli* system that eliminated overcrowding background signals and consequently increased dramatically the dynamic range of the CadC β-gal signal. This new approach allowed us to compare quantitatively the intramembrane dimerization strength of the GpA motif with the dimerization strength of the CadC periplasmic domain. Based on the β-gal assay, the CadC dimerization strength is twice that of the GpA motif.

Results

We used several tripartite CadC chimeras, shown in Fig. 1d. The CadC amino-terminal cytoplasmic domain (*N-CadC*) was common to all of the constructs. We placed the *lacZ* gene under the control of the *cadBA* promotor so that β-gal activity could be used as an *in vivo* readout of TM dimerization (Fig. 1a). The C-terminal domain of the chimeras generally consisted of the periplasmic domain of CadC (*CadC-*C_T) or RodZ (*RodZ-*C_T), but in some control experiments we used the secreted maltose-binding protein MalE (residues 27–396). We used a wide range of hydrophobic segments (H-segments) in place of the wild-type TM domain of CadC to test TM helix stability and dimerization. Because proteolytic cleavage of the periplasmic domain was an important tool for topology determination, all of the constructs carried a T7 immuno tag just before the H-segment and a His_6 immuno tag at the C-terminus (Fig. 1d and Methods). We used two versions of each construct: one version carried a SPase I cleavage (*clv*) site just after the H segment while the other (*null*) version

did not. The general nomenclature we adopt for describing the data from the various constructs is ***C-TM-X*** in which ***C*** = *N-CadC*, ***TM*** = the H-segment, and ***X*** = the chosen C-terminal domain (*CadC-*C_T, *RodZ-*C_T, or $MalE_{27–396}$). We assign a unique name, such as ***H*** or ***H****, to each H-segment used. This nomenclature provides compact shorthand for labeling the data presented in the figures and text. The relevant amino acid sequences of the H-segments are indicated in each figure.

The CadC periplasmic domain resists proteinase K cleavage

We first attempted to establish the topology of CadC *in vivo* using protK digestion of *E. coli* spheroplasts. We expressed wild-type CadC (*C-H-C*, Fig. 1d) carrying the T7 and His_6 tags (MW = 59.7 kDa) in BL21(DE3) cells after induction with 20 μM IPTG for 1 hr., and then prepared spheroplasts as described in Materials and Methods. Lanes 1 and 2 of Western blots prepared using the total fraction (less the periplasmic fraction) against the T7 tag (Fig. 2a) show that only full-length protein was present with or without protK treatment. As a control, we carried out the same protocol using a CadC-RodZ chimera (*C-H-R*, Fig. 1d) comprised of the CadC N-terminal domain, the wild-type (wt) CadC TM domain, and the RodZ C-terminal domain (MW = 45.8 kDa, but migrates anomalously on SDS-PAGE gels at about 60 kDa). In this case, full-length protein was seen only in the absence of protK (Fig. 2a, lane 5). After protK treatment, only T7-tagged *N-CadC* proteolytic fragments were detected (lane 6), consistent with type II S-SMP topology.

Signal Peptidase I cleavage sites allow facile determination of topology of CadC

To validate the topology of CadC, it occurred to us that SPase I, constitutively present in *E. coli* cytoplasmic membranes [20–22], might cleave the periplasmic domain if a SPase I cleavage site (*clv*) was introduced immediately after the TM helix (Fig. 1d). The consensus sequence for SPase I cleavage is AXA, where X is any amino acid [23]. We inserted *clv* = AAA after the CadC TM domain with immediate success. Fig. 2a, lanes 3 & 4, show that with or without protK treatment, a T7-*N-CadC* fragment was produced. Furthermore, because this fragment was not affected by protK treatment, we concluded that the fragment was in the cytoplasm, consistent with N_{in}-C_{out} (type II) topology. We tried the same approach using the *C-H-R* construct with *clv* = AAA with similar results (Fig. 2a, lanes 7 & 8). To verify the topology of both *C-H-C* and *C-H-R*, we examined the periplasmic fractions for the presence of the His_6-tagged periplasmic domains *CadC-*C_T and *RodZ-*C_T. Western blots using His_6 antibodies revealed these domains in the periplasmic fraction for the *clv* = AAA constructs but not for the *null* constructs lacking the cleavage site (Fig. 2b, compare lanes 1 & 2 and lanes 3 & 4).

As a final test of the method, we examined cell fractions to determine the location of the T7-tagged *N-CadC* fragment. In this case, rather than the wild-type CadC H-segment *H*, we used the artifical TM segment *H** consisting of sixteen leucines (L_{16}) sandwiched between GGPG and GPGG at the N- and C-terminals, respectively (Fig. 2). The rationale for this H-segment choice came from the constructs used by Hessa et al. [24] for determination of a biological hydrophobicity scale using dog pancreas microsomes. The purpose of the GGPG/GPGG segments is to isolate the hydrophobic TM domain from the surrounding sequence. But where should the *clv* site be inserted in these constructs, and importantly, would the

GGPG at the C-terminal for artifical TM segments interfere with cleavage? Jain et al. [25] determined the optimal placement of *clv* in mutant alkaline phosphatase signal sequences with ten leucines (Leu_{10}) immediately preceeding the cleavage site. They determined from sequential insertions of Gln following Leu_{10} that *clv* sites located 3 to 9 residues beyond Leu_{10} were completely processed. We duplicated Jain et al.'s experiment [25] experiments by sequential insertion of Cys residues following the GPGG segment. The use of Cys rather than Gln had the advantage that the formation of dimers provided a convenient marker of cleavage (data not shown). Dimers form only when there is cleavage, probably because of steric hindrance of intact periplasmic domains. We found optimal cleavage when *clv* was seven residues beyond the L_{16} segment.

The results shown in Fig. 2c reveal that *C-H*-R* (*null*) is found in the inner membrane (IM) fraction and, of course, the total (T) fraction (lanes 1 & 3). There are also traces of full-length protein in lane 2. For the *C-H*-R* (*clv*) construct (*clv* = AQA), the *N-CadC* fragment is found in the T and IM fractions only. These experiments show three things. First, *C-H*-R* behaves similarly to *C-H-R*; the *H** H-segment can replace the *H* H-segment. Second, SPase I cleavage does not depend on the amino acid composition of the H-segment as long as it is sufficiently hydrophobic to be inserted into membrane. Third, both the uncleaved protein and the cleaved protein are located in the inner membrane, consistent with stable insertion of CadC and cleavage of CadC at the periplasmic surface by SPase I.

Closer examination of Fig. 2c provides insights into membrane incorporation of expressed protein. Full-length *C-H*-R* (*clv*) is visible in the T fraction (lane 4), which should be absent if SPase I fully processed the protein. Two scenarios could lead to this result. Either the protein was membrane integrated, but was not processed by SPase I, or the protein was not integrated into the membrane at all. The amount full-length protein in the T fraction (lane 4) appears to be same as the amount in the C/P fraction (lane 5). Similarly, the amount of processed protein in the T fraction is similar to the amount in the IM fraction (lane 6), consistent with efficient and complete processing by SPase I of protein actually inserted into the membrane. If that is correct, then the unprocessed protein is protein that never entered the membrane and consequently could not be processed by SPase I. We show below that full-length protein carrying a *clv* site is never seen when protein expression is lowered sufficiently.

Although all of the results so far provide strong support for SPase I cleavage at the *clv* sites, the possibility existed that some other protease was responsible for cleavage. We therefore expressed the *C-H*-R* construct with *clv* = AQA in the SPase I depletion strain FTL85 in which the SPase I gene *lepB* is controlled by the arabinose operator. Lanes 1 & 2 of Fig. 3a show that T7-*N-CadC* fragments are observed only under non-depletion conditions, consistent with cleavage by SPase I. We conclude that the cleavage observed for all constructs carrying *clv* sites is due to SPase I. The results shown in Fig. 2 and Fig. 3 are consistent with lack of interference from the C-terminal GPGG sequence in the *H** constructs.

SecA translocase is required for CadC assembly

We noted earlier that RodZ and CadZ are unusual S-SMPs, because they lack a signal sequence and their TM domains occur more than 100 residues downstream from the amino terminus. We have established for RodZ that the only requirements for *in vivo* assembly are SecYEG, SecA, and the proton motive force (PMF) [13]. As a first step toward establishing the requirements for CadC assembly, we expressed *C-H**-*R* (*clv* = AQA) in the SecA depletion strain EO527 in which *secA* is regulated by the arabinose operator. Fig. 3b shows that the T7-tagged *N-CadC* fragment appears only under non-depletion conditions, consistent with SecA being essential for CadC biogenesis. Of course, we have here used a CadC-RodZ chimera. Future experiments will determine which components of the *E. coli* MP assembly system are required for insertion of wild-type CadC into the cytoplasmic membrane and whether the C-terminal periplasmic domain affects which components are necessary for insertion.

Dimerization of the CadC periplasmic domain is required for activation of the *cadBA* operon

The success of CadC topology determination by the SPase I method, opened the way to determining whether dimerization of the CadC periplasmic domain was necessary for activation of the *cadBA* operon. For this purpose, we measured transcription activation (*cadBA::lacZ*) by expressing the constructs (pET-21 derivatives) in BL21(DE3) cells harboring a single-copy plasmid (pETcoco-1 [NOVAGEN] containing *cadBA:lacZ*). BL21(DE3) cells containing the plasmids were grown in SOC media in the presence of chloramphenicol and ampicillin to the logarithmic phase (OD_{600} = ~ 0.5, 1 hr.). Protein expression was induced with 20 μM IPTG for an additional 1 hr. β-gal activities (Miller units) were determined as described in Materials and Methods. The Western blots in Fig. 2a,b show the expression levels and membrane topology of the CadC constructs used in the measurements.

We measured β-gal activities of *C-H-C* and *C-H-R* with or without *clv* = AAA sites. The β-gal activities are shown in Fig. 2d. The activity of the *C-H-C* (*null*) construct exceeded 2500 MU, whereas for the *C-H-C* (AAA) construct, activity was less than 500 MU, consistent with the belief that dimerization of the CadC periplasmic domain is required for activation of the operon. Cleavage of the dimeric periplasmic domain apparently destroys the geometry necessary for the cytoplasmic helix-turn-helix (H-T-H) domains to bind to the operon's promoter site.

The necessity for CadC dimerization via the periplasmic domains is further supported by the β-gal activities of the *C-H-R* constructs. The physiological function of the periplasmic H-T-H domain of RodZ is to bind to MreB in a complex of proteins involved in shape control; there is no evidence that RodZ dimerization is required for binding. Fig. 2d shows that *C-H-R* (*null*) does not activate the operon (β-gal activity ≈ 200 MU) and that cleavage of the RodZ-C_T domain has no significant effect on activity. We therefore conclude with considerable confidence that dimerization of the CadC periplasmic domains is a requirement for activation of the *cadBA* operon.

The CadC transmembrane domain is not involved in dimerization

Although CadC must dimerize to activate the *cadBA* operon, an important question is whether the TM domain might also be involved. We addressed this question by replacing the wild-type H-segment *H* with the polyleucine-based *H** segment. Zhou et al. [26] have shown that artificial TM segments composed only of leucine have relatively little tendency to dimerize *in vivo* in *E. coli* unless a polar residue, especially Asn, Asp, or Glu, is also present in the sequence to cause enhanced dimerization through hydrogen bond interactions. The *H** H-segments are thus unlikely to form dimers. Fig. 4 shows the results of experiments in which *H** replaced the wt H-segment, *H*. We first confirmed the expected topology of three constructs: *C-H*-C*, *C-H*-M*, and *C-H*-R*. Fig. 4a shows comparisons of the three constructs with and without *clv* = AQA. In all three cases, independent of C-terminal domain, T7-tagged *N-CadC* fragments are found in the periplasm-free total fractions when the cleavage site is present. The topologies implied by these results (panel a) are confirmed in panel b: The corresponding His_6-tagged C-terminal domains are found in the periplasmic fraction. These results not only confirm the topologies of the construct, but reveal as well the value of the SPase I cleavage site method for unequivocal topology determination. But the main conclusion is that the various constructs are assembled with the correct topology with the *H** H-segment replacing the wt segment.

What do the β-gal activities reveal about the involvement of the wt CadC TM helix in dimerization? The activities for *C-H-C (null*) and *C-H*-C* (*null*) are identical within experimental error (compare bar 1 with the control bar in Fig. 4c); the TM helix does not contribute significantly to dimerization. As expected, the activity of *C-H*-C* (*clv*) drops dramatically compared to *C-H*-C* (*null*). The activities for the other constructs shown by bars 3 – 6 in panel c are small, and confirm the conclusion that the CadC periplasmic domain is required for dimerization. Overall, the results shown in Fig. 4 demonstrate little, if any, contribution of the CadC TM domain to dimerization.

Glycophorin A TM helices cause dimerization of CadC

We were curious to know if CadC could be used as an indicator of TM helix dimerization in the spirit of the now widely used ToxR-based assays [27–28]. Using β-gal activity as a measure of dimerization, we examined the dimerization of *C-X-R* constructs using *X = H** as a neutral control. Two GpA constructs were used, *H1* and *H2: H1* was the wt GpA TM domain while *H2* was a hydrophobically augmented GpA TM domain [29] in which all residues not directly involved in dimerization are replaced by leucine (Fig. 5). To validate dimerization, we used as a control the insertion of a single Ala between GpA positions 81 and 82, which we call for convenience 82Ala constructs [30]. This strategy led to four additional constructs, *H1A* and *H2A* (Fig. 5) including *null* and *clv* = AQA. For the experiments summarized by the immuno blots (anti-T7 antibodies) in Fig. 5, we purposely over-expressed the proteins to see if oligomers of any of constructs were apparent. In addition, the proteins in the SDS sample buffer were boiled prior running on the gels. We show later that boiling is not a good idea. Lanes 5 & 7 show that oligomers of *H2* were formed, as might have been expected as a result of the hydrophobic augmentation of the GpA domain; full length protein and dimers are found in lane 5 and oligomers of the T7-tagged fragments in lane 7. Note that SPase I cleavage is incomplete under these over-

expression conditions, as observed to a lesser extent in Fig. 2c. This is consistent with the hypothesis that there are limits to the amount of protein *E. coli* can insert into the cytoplasmic membrane.

The β-gal activities for all of the constructs are shown in Fig. 5b, where the bars in the graph are numbered according to the lanes in panel a. The *C-H-C* control was again very high, as expected (over 3000 MU). Although smaller than the *C-H-C* control, the activities of some of the other constructs were also high, and some β-gal activities suggested that *clv* constructs had higher activities than the nulls. As we show next, the high and variable activities are an artifact of differences in protein expression levels and an apparent natural tendency of hydrophobic TM helices to associate, as is evident from greater dimerization of GpA helices with a leucine background (compare *H1* constructs with *H2* constructs in Fig. 5b).

Low expression levels increase the 'dynamic range' of β-gal activities

To examine the effect of expression levels on β-gal activities and the amount of uncleaved protein in the membrane, we reduced protein expression dramatically by using Top10 cells, which lack the T7 polymerase (See Materials and Methods). The expression of CadC in this system depends only on leakage transcription from the constitutive *E. coli* RNA polymerase. The results of this approach are shown in Fig. 6. Comparing lanes 1 & 2 with lanes 3 & 4 in Fig. 6a demonstrates that at low expression levels, cleavage of the *clv* constructs is complete; no full-length protein is observed. Notice, however, that no oligomers are observed in these lanes. This is not due to low expression, but rather because the samples were boiled in the SDS sample buffer prior to running SDS/PAGE gels. When the samples were not boiled, oligomers of the *H2* constructs were apparent (lanes 5 & 7). No oligomers were observed for the *H2A* constructs carrying the 82Ala insert (lanes 6 & 8). The difference between boiling and not boiling is likely due to complex interactions involving, among other things, the temperature dependence of SDS aggregate size and critical micelle concentration [31], protein denaturation, and protein aggregation. Fig. 6b shows that SPase I cleavage is also complete for *C-H*-C*, *C-H2-R*, and *C-H2A-R*, which supports the notion that SPase I processes all of the protein it encounters in the membrane at low expression levels.

The use of β-gal activities for judging dimerization provided much clearer answers when proteins were expressed at low levels. Fig. 6c shows that the activities were robust and statistically identical for *C-H2-R* (*null*) and *C-H2-R* (*clv*), which are the constructs carrying the GpA TM helix with Leu background. These results mean that the RodZ C-terminal domain in *C-H2-R* has no detectable effect: dimerization is driven entirely by GpA helix-helix interactions. This conclusion is supported fully by the very low β-gal activities of both GpA constructs carrying the 82Ala insert (*C-H2A-R*, *null* and *clv*). Together, the results of Fig. 6c,d demonstrate the high β-gal dynamic range that can be achieved using low protein expression levels. This low-expression approach allowed us to address the question of whether helix-driven dimerization or periplasmic-domain dimerization is more effective in activation of the *cadBA* operon. The answer is clear from the results shown in Fig. 6d: The β-gal activity for *C-H*-C* (about 3500 MU) is twice that of *C-H2-R* (*clv*) (about 1500 MU).

Membrane insertion of the H-segment depends upon hydrophobicity

Following the lead of earlier studies [24–32], which showed that polyalanine H-segments are inserted across membranes with a low probability, we used the Top10 low-expression system to examine the TM insertion of CadC constructs with polyleucine (*H**) or polyalanine ($H^{\#}$) H-segments with GGPG/GPGG flanks. Insertion can be judged by whether SPase I cleaves the periplasmic domain. Fig. 7a, lanes 1–4, shows that SPase I does not cleave the polyalanine construct *C-*$H^{\#}$*-R* (lanes 3 & 4) nor can the RodZ periplasmic domain be digested by protK treatment of spheroplasts. Therefore, *C-*$H^{\#}$*-R* is not inserted across the membrane; it must reside solely in the cytoplasm. *C-H*-R* carrying a polyleucine H-segment, on the other hand, is fully inserted judged both by protK digestion of spheroplasts (lanes 5 & 6) and SPase I cleavage (lanes 7 & 8). Note that because SPase I automatically cleaves *RodZ-*C_T domain leaving the *N-CadC* domain exposed only on the cytoplasmic surface of the membrane, the *N-CadC* domain is inaccessible to protK (lanes 7 & 8). Fig. 7b shows that the periplasmic domain does not affect insertion. Regardless of whether the C-terminal domain is that of RodZ or CadC, no insertion is observed for the polyalanine ($H^{\#}$) H-segment, as shown by the lack of SPase I cleavage of the peptides (lanes 1 – 4). As expected, however, when the H-segment is the polyleucine version (*H**), cleavage is complete, consistent with insertion. The β-gal assays for these various constructs provide a convenient measure of both insertion and dimerization. In Fig. 7c, significant β-gal activity is seen only for the C-H*-C (*null*) construct; activation of the *cadBA* operon requires both transmembrane insertion of CadC and dimerization of its periplasmic domain.

Discussion

Single-span MPs, abundant in all branches of life [33], generally have their TM domains near the N-terminus. There are several S-SMPs, however, without signal sequences whose TM segments occur a hundred or more residues downstream from the N-terminus. We recently identified six of these [13], including RodZ and CadC (Fig. 1c), which are of particular interest to our laboratory in the context of S-SMP assembly and stability. RodZ is important for the maintenance of the rod shape of *E. coli* [11–12] while CadC plays an important role in protecting *E. coli* from acid stress by activating the *cadBA* operon at low pH in the presence of lysine [16–34–35] (see Fig. 1a,b). We recently showed that *in vivo* assembly of RodZ requires only the SecA ATPase, the SecYEG translocon, and a transmembrane PMF [13]. A thorough analysis of the assembly requirements for CadC is presently lacking, but we showed that SecA is required for the assembly of a CadC-RodZ chimera (Fig. 3b).

The main goal of our CadC experiments was to explore the use of CadC for studies of the stability and dimerization of S-SMPs using CadC chimeras. CadC is appealing due its similarity to ToxR, which is widely used for examining helix-helix interactions *in vivo* [28–36]. A theoretical advantage of the CadC system is that the *cadBA* promoter has only two CadC binding sites and both must occupied for *cadBA* activation [35]. The *ctx* promoter sequence is simpler, consisting of three to eight tandemly repeated TTTTGAT DNA elements [8]. This raises the possibility that strict dimerization of ToxR is not enforced.

As expected, we found that by placing the *lacZ* gene under the control of the *cadBA* promotor that β-gal activity could be used as an *in vivo* readout of TM dimerization (Fig. 4). But we also found that β-gal activity could be used to ascertain whether a particular H-segment (*X*) allowed transmembrane insertion of *C-X-C* constructs (Fig. 7). Specifically, β-gal activity was absent when the H-segment was polyalanine-based whereas it was extremely high for a polyleucine-based H-segment (Fig. 7c). The lack of β-gal activity is consistent with physicochemical and *in vitro* measurements showing that polyalanine helices have a low probability of being stable across lipid membranes [24–32]. However, the situation is more complex than for cotranslational insertion studied in the experiments of Hessa et al. [24], because the SecA motor ATPase is involved. There are two issues. First, SecA must recognize the TM domain, and second, it must engage the SecYEG-based insertion apparatus. Failure to insert could thus be due to lack of SecA recognition, failure of the insertion mechanism, or both. We expect future experiments to clarify the SecA-driven recognition/insertion process.

Prior to our CadC experiments, several lines of evidence suggested that dimerization of the periplasmic domain of CadC is essential for activation of the *cadBA* operon (Fig. 1b) [18–19], but the necessity for CadC dimerization had never been demonstrated directly. Indeed, even the presumed N_{in}-C_{o}ut type II topology had not been validated experimentally. Before we could establish the usefulness of CadC for studying TM helix targeting, stability, and dimerization, we had first to confirm CadC's topology and to establish that dimerization of the periplasmic domain is necessary for activation of the *cadBA* operon. The standard method for testing S-SMP topology *in vivo* in *E. coli* is protK treatment of spheroplasts, but this approach failed for CadC, because the putative periplasmic domain was completely resistant to protK treatment (Fig. 2a, lanes 1 & 2). We therefore adopted a different strategy: we inserted a signal peptidase cleavage site (AAA or AQA) immediately after carboxy end of the helix. This approach worked uniformly well, as shown in Figs. 2 – 7. Whenever the cleavage site was present, appropriate tagged fragments were found that were consistent with N_{in}-C_{out} topology as shown, for example, in Figs. 2a,b. Furthermore, the CadC constructs and their N-CadC fragments were found in the membrane fraction of fractionated cells (Fig. 2c). To validate our assumption that cleavage was due to SPase I, we expressed *C-H*-R* in a SPase I depletion strain (FTL85) in which *lepB* is under the control the arabinose promoter/operator. The absence of arabinose led to SPase I depletion. When SPase I was depleted, no T7-tagged *N-CadC* fragment was found (Fig. 3a).

Our cleavage-site approach made it possible to prove unequivocally that dimerization of the CadC periplasmic domain is required for activation of the *cadBA* operon. The strategy was simple: measure β-gal activity of two CadC constructs, one with a cleavage site and one without, the idea being to sever the periplasmic domains, leaving only the TM and cytoplasmic domains. As shown in Fig. 4c, β-gal activity drops essentially to background when the cleavage site is present. Without its periplasmic domain, CadC cannot activate the *cadBA* operon. Furthermore, the H-segment had little effect on dimerization in the *null* constructs (compare *C-H-C* and *C-H*-C* in Fig. 4c). Further support for the necessity of periplasmic-domain dimerization was provided by constructs in which the RodZ periplasmic domain replaced the CadC domain (Fig. 4c, bars 5 & 6).

The question that arose during the CadC periplasmic-domain dimerization experiments was whether CadC could be used to explore TM helix dimerization. To answer that question, we turned to the glycophorin A (GpA) helix-dimerization motif – $L_{75}IxxG_{79}V_{80}xxG_{83}V_{84}xxT_{87}$- [3–4]. We began by replacing the CadC H-segment with the wild-type GpA TM domain (*H1*, Fig. 5) and replacing the CadC dimerization domain with the periplasmic RodZ domain. As a negative control, we inserted a single Ala at position 82 in the sequence to interrupt dimerization [30] to produce the H-segment *H1A* (Fig. 5). Because dimerization is strengthened by making the GpA TM segment more hydrophobic [29], we also made constructs in which all residues that are not part of the dimerization domain are replaced by leucine (*H2*, Fig. 5). A negative-control construct with the 82Ala insertion was also constructed (*H2A*, Fig. 5). Finally, for this family of constructs we created *clv* as well as *null* versions. Summarizing, we produced constructs of the form *C-X-R* (null) and *C-X-R (clv*) with *X = H1*, *H1A*, *H2*, and *H2A* (Fig. 5). The β-gal activities of the various constructs, summarized in Fig. 5b, were somewhat mixed with β-gal activities ranging from about 100 MU to about 2000 MU (the *C-H-C* control activity was about 3000 MU). Nevertheless, the activities of the *X = H1* and *H2* constructs were higher than for the respective *H1A* and *H2A* constructs, as expected. The effects of SPase I cleavage of the *clv* constructs tended to be ambiguous. Overall, the results were consistent with dimerization of the TM domains in the absence of the 82Ala insert, but we were dissatisfied with variations in activity arising from variations in expression levels. These variations, which limit sensitivity, have long been a problem with ToxR-based dimerization studies. We surmised that the problem was our inability to control closely protein expression, even by using very low IPTG concentrations.

To address the expression problem, our strategy was to reduce dramatically expression using Top10 rather than BL21(DE3) cells and to assure that constitutive β-gal activity was completely suppressed. Top10 cells lack the T7 polymerase and contain a defective *lacZ* gene, so that expression of the CadC constructs depended solely on ‘leakage’ due to weak activation by the constitutive Top10 promoter. The only disadvantage of this approach is that long expression times are required, but we did not find the overnight (16 hrs – 18 hrs) expression time to be burdensome. The results of the β-gal activities of the various constructs now became completely unambiguous (Fig. 6). The *C-H2-R* constructs yielded excellent β-gal activities and showed that the presence or absence of a cleavage site did not matter (Fig. 6c), while the β-gal activities of the *C-H2A-R* constructs dropped to essentially 0. The result of the low-expression strategy was thus a dramatic increase in the dynamic range of the β-gal assay. As shown in Fig. 6d, the strategy revealed that the β-gal activity for helix-helix driven dimerization was about one-half that of the activity observed using CadC periplasmic-domain dimerization. Finally, the low-expression results support the hypothesis that over-expression leads to the production of more protein than can be accommodated by the protein insertion processes, resulting in overcrowding that can cause false-positive dimerization.

CadC and CadC chimeras provide a robust system for studying single-span membrane protein topology, dimerization, and stability. The signal peptidase cleavage-site strategy seems to be insensitive to the composition of the TM and periplasmic domains used in the

constructs, suggesting that it may be broadly useful as tool for determining single-span MP topology. An inherent problem in studies of dimerization using β-gal and related assays is control of protein expression. We suggest that the accuracy and precision of such assays may be increased dramatically using the Top10 low-expression system.

Materials and Methods

Bacterial strains, plasmids, and materials

CadC and RodZ were amplified from chromosomal DNA (*E. coli* K12). We used the restriction sites *NdeI* and *XhoI* for gene insertion into the pET21 vector (*T7* promoter/*lac* operator, NOVAGEN). We inserted two additional unique restriction sites (*KpnI* and *BamHI*) to the *cadC* gene to exchange the H-segment using cassette cloning or overlap extension. All constructs were confirmed by sequencing. BL21(DE3) (F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) or Top10 (F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ^-) cells were used to express the various CadC constructs, which all carried an internal T7-tag and a C-terminal His_6-tag for Western blot detection. For SPase I depletion studies, we used *E. coli* strain FTL85 in which *lepB* is under the control of AraC [37]. For SecA depletion studies, we used *E. coli* strain EO527 in which *secA* is under the control of AraC. Both depletion strains were received from Ross E. Dalbey at the Ohio State University who obtained them from Tracy Palmer (FTL85) and Tom Rapoport (EO527), respectively.

Growth Conditions

Various CadC proteins were expressed from an IPTG-inducible and T7 polymerase-dependent system (pET-vector). We used two expression strategies: (a) 1h expression in BL21(DE3) cells (presence of T7 polymerase) or (b) overnight expression in Top10 cells (absence of T7 polymerase) which leads to very low protein yields. Protein expression in Figures 2 – 5 was done using BL21(DE3) cells containing the gene for T7-polymerase (CadC protein is regulated by the T7-promoter and the *lac*-operator). This leads to high protein expression levels even in the presence of small amounts IPTG inducer (20 μM) and short expression time (1 h). In Figures 6 and 7, Top10 cells were used to decrease the expression level dramatically, because this strain lacks the T7-polymerase, causing CadC to depend solely on leakage transcription from the constitutive *E. coli* RNA polymerase. To make sure that the *lac*-operator was not 'blocked', we used 1 mM IPTG, which binds to the *lac*-repressor LacI and makes it 'inactive'. The Top10 system obviously requires longer expression times than the BL21(DE3) system. We found that overnight expression (16 – 18 hrs) was sufficient to allow β-gal activity to be accumulated within the cells for accurate assays.

The experiments were done at pH 7 in Luria-Bertani (LB) medium or super optimal broth with catabolite repression (SOC) full media using glucose for repression [38]. We could detect activity under these conditions, because the media contained ~ 4 mM lysine. CadC activity depends on the presence of acidic conditions (pH < 5.7) as well as lysine. However, because we used a C-terminal His_6-tag, the pH dependence was absent, in agreement with

previous reports [39]. The C-terminal His_6-tag presumably interferes with the pH sensor domain [35].

SPase I and SecA Depletion Protocols

SPase-depletion experiments—*C-H*-R* constructs with *clv* = AQA (pET-vector, T7-RNA-polymerase dependent system) were transformed in FTL85 cells. Overnight cultures were grown in LB media in the presence of 0.02 % arabinose (non-depletion condition). A 400 µl inoculum from the culture was added to 10 ml of fresh LB media with or without 0.02 % arabinose. After 1 h (OD_{600} ~ 0.6) protein expression was induced by adding 100 µM IPTG (note FTL85 strain does not contain the gene for T7-RNA-polymerase). After 3 h of protein expression cells were pelleted and analyzed.

SecA-depletion experiments—*C-H*-R* constructs with *clv* = AQA (modified pET-vector, T7-RNA-polymerase independent system using a T5 promoter sequence which is recognized by the wt *E. coli* RNA-polymerase.) were transformed in EO527 cells. Overnight cultures were grown in SOC media in the presence of 0.02 % arabinose (non-depletion condition). A 400 µl inoculum from the culture was added to 10 ml fresh SOC media with or without 0.02 % arabinose. After 2 h (OD_{600} ~ 0.6) protein expression was induced by adding 10 µM IPTG. After 0.5 h of protein expression, cells were pelleted and analyzed.

Cell Fractionation

Cell fraction was performed by cell lysis using freeze-thaw and DNaseI treatment [38]. The bacterial cells were harvested, centrifuged, and the pellet resuspended in Lysis-Equilibration-Wash buffer (LEW buffer: 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) containing DNaseI enzyme, DNaseI buffer, lysozyme, and phenylmethanesulfonyl fluoride (PMSF). Thereafter, the cell pellet was subjected to 10 cycles of freeze (liquid Nitrogen) and thaw (at 37°C water bath) followed by incubation at 37°C for 10 minutes. The cell suspension was centrifuged at 13,000×g for 10 minutes at 4°C, and the supernatant containing the soluble and periplasmic proteins (called the *C/P* fraction) was transferred to a new tube. The pellet was resuspended in LEW+1.5% CHAPS to solubilize membrane proteins. The suspension was centrifuged at 13,000×g at 4°C for 15 minutes. The supernatant contains the inner membrane (IM) fraction.

Periplasmic fraction

Cells were grown to mid-logarithmic phase and harvested by centrifugation. Osmotic shock was performed by a method adapted from Neu and Heppel [40] and Thorstenson et al. [41] as follows: Cell pellets were resuspended in 100µl osmotic shock buffer (0.5 M sucrose, 0.2 M Tris, 0.5 mM EDTA) and incubated on ice for 15 min, followed by the addition of 400 µl of 5 mM $MgSO_4$. The cells were incubated on ice for an additional 30 min, followed by pelleting at 13,000×g at 4°C for 15 minutes. The supernatant (periplasmic fraction) and the pellet were mixed separately with SDS sample buffer and analyzed by SDS-PAGE [42].

Protease treatment studies

Cells were grown to mid-logarithmic phase and harvested by centrifugation. Cell pellets were resuspended in 100μl osmotic shock buffer (0.5 M sucrose, 0.2 M Tris, 0.5 mM EDTA) and incubated on ice for 15 min. Then, 400 μl of 5 mM $MgSO_4$ containing prot K (80 ng) was added and the cells incubated on ice for an additional 30 min, followed by pelleting at 13,000×g at 4°C for 15 minutes. The supernatant was discarded, the pellet resuspended in SDS sample buffer, and analyzed by SDS-PAGE [42].

Western Blotting

The pellet was resuspended in SDS sample buffer and analyzed by SDS-PAGE (4–20 %) [42] and then Western blotted using iBLOT from Invitrogen® (Invitrogen Corp., Carlsbad, CA), which guarantees complete protein transfer that is necessary under low-expression conditions. The protein was detected by a T7-tag alkaline phosphatase-conjugated antibody from Novagen® (Novagen (EMD) Biosciences, Madison, WI) or by a His_6-tag antibody from Roche® (Hoffman La Roche, Basel, Switzerland).

CadC activity assays

Transcription activation (*cad*BA::*lac*Z) was mediated by expressing the constructs on pET-21 derivatives in BL21(DE3) or Top10 cells (Invitrogen) harboring a single-copy plasmid (pETcoco-1, Novagen®) containing *cad*BA:*lac*Z). BL21(DE3) cells containing the plasmids were grown in SOC media in the presence of chloramphenicol and ampicillin to the logarithmic phase (OD_{600} = ~ 0.5) and protein expression was induced with 20 μM IPTG for an additional 1 h. Top10 cells containing the plasmids were grown in LB media for 16 h in the presence of 1 mM IPTG. β-gal activities were quantitated in crude cell lysates after addition of o-nitrophenylgalactoside and monitoring the reaction at 405 nm for a period of 5 min at intervals of 1 sec with a UV-Visible spectrophotometer. Specific β-gal activities (Miller units) were calculated from the slope of the reaction curves, and the OD_{600} previously measured for the cell suspension [27–43]. All β-gal activities were determined in duplicate for each experiment. The “error bars” shown represent standard deviations of the duplicates. We choose to use duplicates because of the close agreement of completely independent measurements of the activity of the *C-H-C* (Figs. 2d, 4c, and 5b) and the *C-H*-C* constructs (Figs. 6d and 7c). By independent, we mean each experiment began with a new transformation, cell culture, etc. The average activity (MUs) was 2960 ± 323 (SD) for the *C-H-C* constructs 3973 ± 413 (SD) for the *C-H*-C* constructs. These results indicate an experimental uncertainty of about 10%, which is about the uncertainty found for the duplicates.

Acknowledgments

This work was supported in part by grant GM074637 from the National Institute of General Medical Sciences. We thank Dr. Gargi Dasgupta for her careful and thoughtful reading of the manuscript.

Abbreviations

MP membrane protein

S-SMP	single-span membrane protein
TM	transmembrane
PMF	proton motive force
protK	proteinase K
β-gal	β-gal
SPase I	Signal Peptidase I
H-T-H	helix-turn-helix
wt	wild-type
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani
SOC	super optimal broth with catabolite repression
LEW	lysis equilibration wash
SDS	sodium dodecyl sulfate
PMSF	phenylmethanesulfonyl fluoride

References

1. Furthmayr H, Marchesi VT. Subunit structure of human erythrocyte glycophorin A. Biochemistry. 1976; 15:1137–1144. [PubMed: 175830]
2. Bormann B-J, Knowles WJ, Marchesi VT. Synthetic peptides mimic the assembly of transmembrane glycoproteins. J Biol Chem. 1989; 264:4033–4037. [PubMed: 2783929]
3. Lemmon MA, Flanagan JM, Treutlein HR, Zhang J, Engelman DM. Sequence specificity in the dimerization of transmembrane alpha- helices. Biochemistry. 1992; 31:12719–12725. [PubMed: 1463743]
4. Lemmon MA, Flanagan JM, Hunt JF, Adair BD, Bormann BJ, Dempsey CE, Engelman DM. Glycophorin-A dimerization is driven by specific interactions between transmembrane a-helices. J Biol Chem. 1992; 267:7683–7689. [PubMed: 1560003]
5. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: Structure and implications. Science. 1997; 276:131–133. [PubMed: 9082985]
6. Langosch D, Brosig B, Kolmar H, Fritz H-J. Dimerisation of the glycophorin A transmembrane segment probed with the ToxR transcription activator. J Mol Biol. 1996; 263:525–530. [PubMed: 8918935]
7. Miller VL, Mekalanos JJ. Synthesis of cholera toxin is positively regulated at the transcriptional level by *toxR* . Proc Natl Acad Sci USA. 1984; 81:3471–3475. [PubMed: 6374658]
8. Miller VL, Taylor RK, Mekalanos JJ. Cholera toxin transcriptional activator *toxR* is a transmembrane DNA binding protein. Cell. 1987; 48:271–279. [PubMed: 3802195]
9. Maddocks S, Oyston PCF. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiology. 2008; 154:3609–3623. [PubMed: 19047729]
10. Kanjee U, Houry WA. Mechanisms of acid resistance in *Eschericia coli* . Annu Rev Microbiol. 2013; 67:65–81. [PubMed: 23701194]
11. van den Ent F, Johnson CM, Persons L, de Boer P, Löwe J. Bacterial actin MreB assembles in complex with cell shape protein RodZ. EMBO J. 2010; 29:1081–1090. [PubMed: 20168300]

12. Alyahya SA, Alexander R, Costa T, Henriques AO, Emonet T, Jacob-Wagner C. RodZ, a component of the bacterial core morphogenic apparatus. Proc Natl Acad Sci USA. 2009; 106:1239–1244. [PubMed: 19164570]
13. Rawat S, Zhu L, Lindner E, Dalbey R, White SH. SecA drives transmembrane insertion of RodZ, an unusual single-span membrane protein. J Mol Biol. 2014
14. 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; 67:570–583. [PubMed: 18086202]
15. Tetsch L, Jung K. The regulatory interplay between membrane-integrated sensors and transport proteins in bacteria. Mol Microbiol. 2009; 73:982–999. [PubMed: 19708919]
16. 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; 286:10681–10689. [PubMed: 21216950]
17. Tetsch L, Koller C, Dönhöfer A, Jung K. Detection and function of an intramolecular disulfide bond in the pH-responsive CadC of *Escherichia coli* . BMC Microbiology. 2011; 11:1–12. [PubMed: 21194490]
18. Haneburger I, Fritz G, Jurkschat N, Tetsch L, Eichinger A, Skerra A, Gerland U, Jung K. Deactivation of the *E.coli* pH stress sensor CadC by cadaverine. J Mol Biol. 2012; 424:15–27. [PubMed: 22999955]
19. 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; 20:656–669. [PubMed: 21308846]
20. Zwizinski C, Wickner W. Purification and characterization of leader (signal) peptidase from *Escherichia coli* . J Biol Chem. 1980; 255:7973–7977. [PubMed: 6995457]
21. Date T, Wickner W. Isolation of the *Escherichia coli* leader peptidase gene and effects of leader peptidase overproduction *in vivo* . Proc Natl Acad Sci USA. 1981; 78:6106–6110. [PubMed: 6273848]
22. Paetzel M, Dalbey RE, Strynadka NCJ. The structure and mechanism of bacterial type 1 signal peptidases: a novel antibiotic target. Pharmacol Ther. 2000; 87:27. [PubMed: 10924740]
23. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 1997; 10:1–6. [PubMed: 9051728]
24. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson IM, White SH, von Heijne G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature. 2005; 433:377–381. [PubMed: 15674282]
25. Jain RG, Rusch SL, Kendall DA. Signal peptide cleavage regions: Functional limits on length and topological implications. J Biol Chem. 1994; 269:16305–16310. [PubMed: 8206936]
26. Zhou FX, Merianos HJ, Brunger AT, Engelman DM. Polar residues drive association of polyleucine transmembrane helices. Proc Natl Acad Sci USA. 2001; 98:2250–2255. [PubMed: 11226225]
27. Langosch D, Brosig B, Kolmar H, Fritz H-J. Dimerisation of the Glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. J Mol Biol. 1996; 263:525–530. [PubMed: 8918935]
28. Russ WP, Engelman DM. TOXCAT: A measure of transmembrane helix association in a biological membrane. Proc Natl Acad Sci USA. 1999; 96:863–868. [PubMed: 9927659]
29. Lemmon MA, Treutlein HR, Adams PD, Brünger AT, Engelman DM. A dimerization motif for transmembrane alpha-helices. Nat Struct Biol. 1994; 1:157–163. [PubMed: 7656033]
30. Mingarro I, Whitley P, Lemmon MA, von Heijne G. Ala-insertion scanning mutagenesis of the glycophorin A transmembrane helix: A rapid way to map helix-helix interactions in integral membrane proteins. Protein Sci. 1996; 5:1339–1341. [PubMed: 8819166]
31. Hammouda B. Temperature effect on the nanostructure of SDS micelles in water. J Res Natl Inst Stand Technol. 2013; 118:151–167.
32. Jayasinghe S, Hristova K, White SH. Energetics, stability, and prediction of transmembrane helices. J Mol Biol. 2001; 312:927–934. [PubMed: 11580239]

33. Worch R, Bökel C, Höfinger S, Schwille P, Weidemann T. Focus on composition and interaction potential of single-pass transmembrane domains. Proteomics. 2010; 10:4196–4208. [PubMed: 21058338]
34. Watson N, Dunyak 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; 174:530–540. [PubMed: 1370290]
35. Küper C, Jung K. CadC-mediated activation of the cadBA promoter in *Escherichia coli* . J Mol Microbiol Biotechnol. 2005; 10:26–39. [PubMed: 16491024]
36. Laage R, Langosch D. Strategies for prokaryotic expression of eukaryotic membrane proteins. Traffic. 2001; 2:99–104. [PubMed: 11247308]
37. Lüke I, Handford JI, Palmer T, Sargent F. Proteolytic processing of *Escherichia coli* twin-argine signal peptides by LepB. Arch Microbiol. 2009; 191:919–925. [PubMed: 19809807]
38. Green, MR.; Sambrook, J. Molecular Cloning. A Laboratory Manual. 4 edit. Vol. 3. Cold Spring Harbor: Cold Spring Harbor Press; 2012.
39. Dell CL, Neely MN, Olson ER. Altered pH and lysine signalling mutants of *cadC*, a gene encoding a membrane-bound transcriptional activator of the *Escherichia coli cadBA* operon. Mol Microbiol. 1994; 14:7–16. [PubMed: 7830562]
40. Neu HC, Heppel LA. The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. J Biol Chem. 1965; 240:3685–3692. [PubMed: 4284300]
41. Thorstenson YR, Zhang Y, Olson PS, Mascarenhas D. Lederless polypeptides efficiently extracted from whole cells by osmotic shock. J Bacteriol. 1997; 179:5333–5339. [PubMed: 9286985]
42. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970; 227:680–685. [PubMed: 5432063]
43. Miller, JH. Experiments in molecular genetics. Cold Spring Harbor Laboratory; 1972.
44. Fritz G, Koller C, Burdack K, Tetsch L, Haneburger I, Jung K, Gerland U. Induction kinetics of a conditional pH stress response system in *Escherichia coli* . J Mol Biol. 2009; 393:272–286. [PubMed: 19703467]
45. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G. The molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature. 2007; 450:1026–1030. [PubMed: 18075582]

Highlights

- SPase I cleavage sites can be used for determining the topology of single-span MPs.
- CadC, which regulates the *cadBA* operon, is a type II single-span membrane protein.
- Dimerization of CadC is required for activation of the *cadBA* operon.
- CadC-RodZ chimeras can be used to examine TM helix stability and dimerization.
- CadC-RodZ chimeras can dimerize and activate *cadBA* via TM helix dimerization.

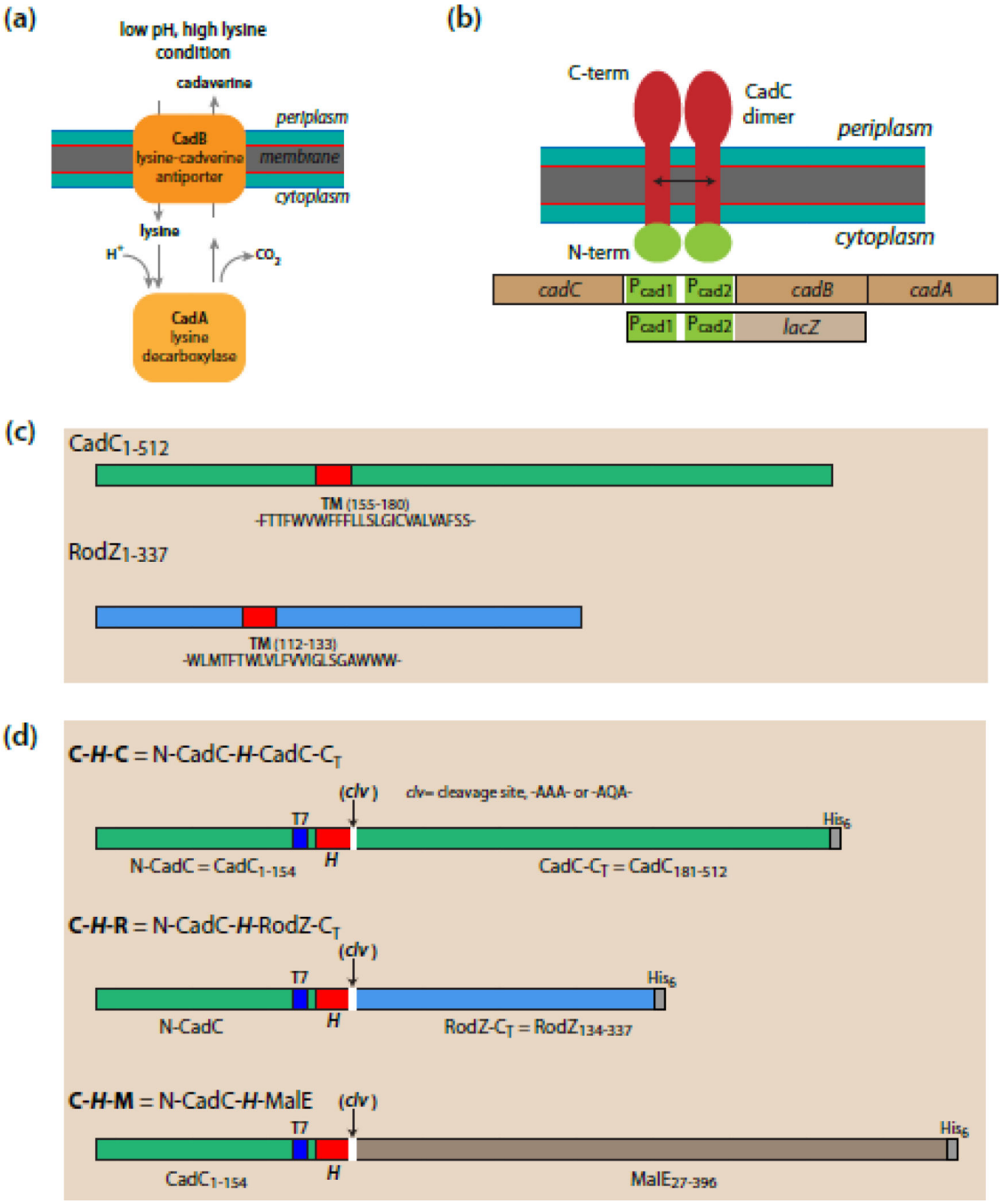

Figure 1.

Overview of the function and organization of the Cad system for protecting *Enterobacteriaceae* against acid stress, and the CadC-based constructs used in this study. (**a**) Expression of the *cadBA* operon is dependent on the transcriptional activator CadC, which belongs to the OmpR/PhoB response-regulator subfamily, defined primarily by the winged helix-turn-helix DNA binding motif within the effector domain. CadC shares sequence homology with the ToxR protein, which is a widely used method for *in vivo* transmembrane helix interaction analysis [6]. (**b**) Low-pH stress in the presence of lysine causes the

regulatory protein CadC to induce expression of CadA (lysine decarboxylase) and CadB (lysine/cadaverine antiporter). Internal protons are consumed by conversion of lysine into cadaverine by CadA and removed from the cell by CadB [44]. In addition to the lysine-based acid resistance offered by the Cad system, three additional decarboxylase systems are present in *E. coli*, based upon glutamate, arginine, and ornithine [10]. (**c**) Schematic overview of wild-type CadC and RodZ. Both are members of a small class of single-span membrane proteins lacking an N-terminal signal sequence and whose TM helix occurs more than 100 residues from the amino terminus [13]. The locations and amino acid compositions of the TM helices (red) are indicated. (**d**) Schematic overview of the CadC constructs used in our study. Color code: Green, CadC components; blue, RodZ components; red, hydrophobic segment (*H*-segment); purple, T7 tag (MASMTGGQQMG) inserted at aa position 140; grey, His_6-tag (HHHHHH); brown, $MalE_{27–396}$; white, optional Signal Peptidase I (SPase I) cleavage sites (*clv*) based on the consensus sequence A-X-A [23] inserted at the C-terminal end of the *H*-segment. In this study, *clv* = AAA or AQA.

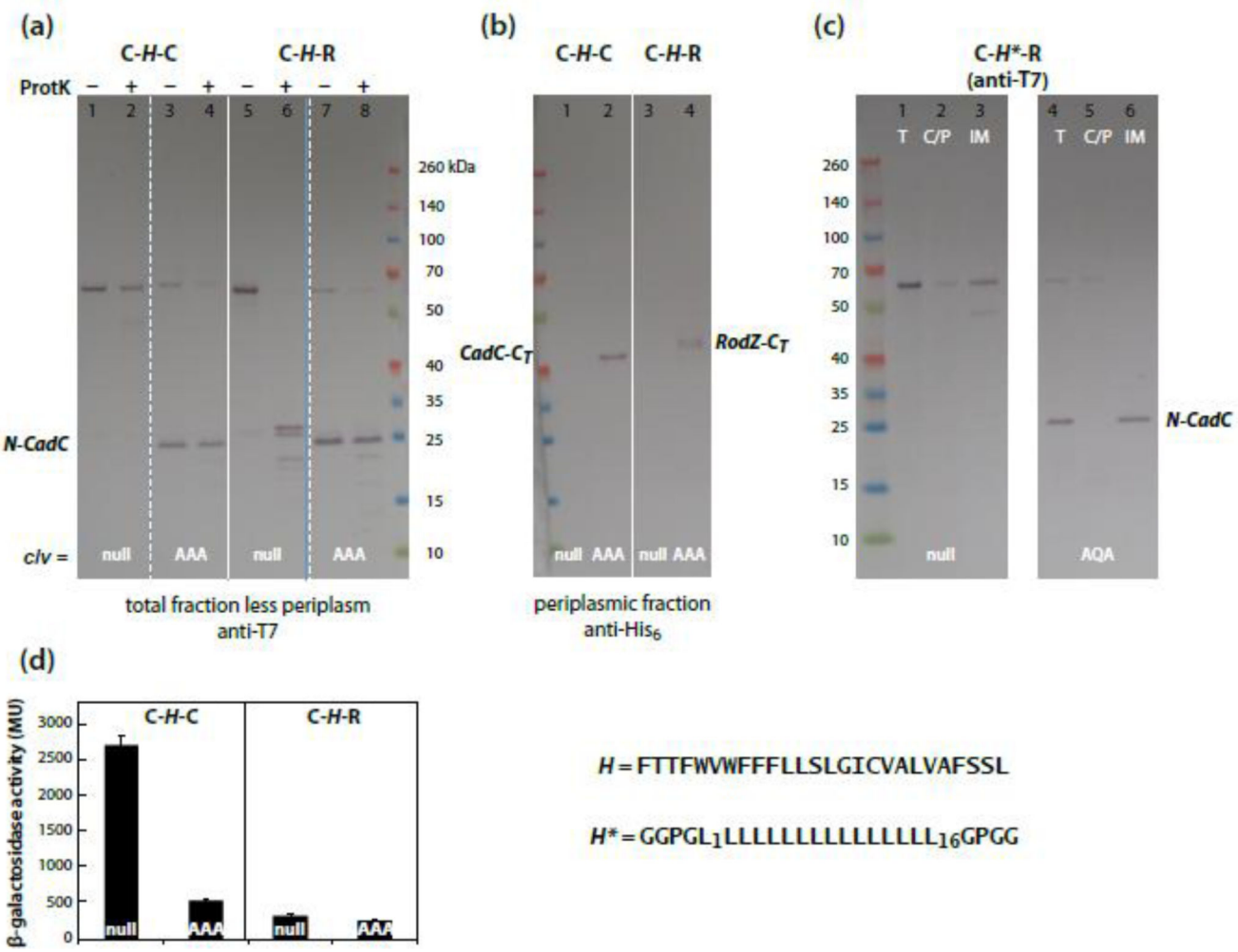

Figure 2.

CadC has an N_{in}-C_{out} topology and must dimerize to activate the *cadBA* operon. The nomenclature of the constructs used is shown in Fig. 1c,d. The various CadC constructs (Fig. 1d) were expressed in BL21(DE3) cells after induction with 20 μM IPTG for 1 h. In all figures, *null*, indicates the absence of a SPase I cleavage site following the H-segment. (**a**) CadC constructs carrying -AAA- sites are cleaved with or without protK treatment of spheroplasts. Wild-type CadC (*C-H-C*) was impervious to protK digestion, because only the full-length protein was seen when spheroplasts are treated with protK (lanes 1 & 2). On the other hand, the CadC-RodZ chimera (*C-H-R*) readily yielded to protK digestion, as indicated by the presence of the T7-tagged *N-CadC*- fragments (lanes 5 & 6). To test the topology of wild-type CadC (*C-H-C*), an -AAA- cleavage site was introduced immediately after the H-segment. Lanes 3 & 4 suggest that SPase I in the cytoplasmic membranes cleaved off the periplasmic CadC C-terminal domain to yield T7-tagged N-terminal fragments (*N-CadC*), regardless of protK treatment. The periplasmic domain of a *C-H-R* chimera carrying the -AAA- cleavage site was also cleaved, leaving behind T7-tagged *N-CadC* (lanes 7 & 8; compare with lanes 5 & 6). Although the molecular weight of T7-tagged *N-CadC* is 20.5 kDa, it migrates anomalously on SDS-PAGE gels at 25 kDa. (**b**) CadC is a type II (N_{in}-C_{out})

single-span membrane protein. When an -AAA- site is present, His_6-tagged CadC and RodZ C-terminal domains (*CadC-*C_T and *RodZ-*C_T) are found in the periplasm, consistent with type II topology. In the absence of an -AAA- site, no fragment is observed. We have previously shown that RodZ has type II topology [13]. The *RodZ-*C_T domain has a MW of 29.1 kDa, but it migrates anomalously at about 45 kDa. (**c**) The *C-H**-R* construct and its cleaved *N-CadC* fragment are located in the inner membrane (IM) fraction of fractionated cells. These results show that (1) *C-H**-R* behaves similarly to *C-H-R* the *H** H-segment can replace the *H* H-segment, that SPase I cleavage does not depend on the amino acid composition of the H-segment as long as it is sufficiently hydrophobic to be inserted into membrane, and that both the uncleaved and cleaved proteins are located in the inner membrane, consistent with stable insertion of CadC and cleavage of CadC at the periplasmic surface by SPase I. (**d**) Dimerization of the periplasmic domain of CadC is required for activation of the *cadBA* operon, as determined by reporter gene (*lacZ*) experiments with various CadC constructs in *vivo*. The Western blot in panel b shows the expression levels and membrane topology of the respective CadC proteins. Wild-type CadC (*C-H-C*) without an AAA cleavage site induced very strong β-gal activity (greater than 2500 MU). When an AAA cleavage site was introduced to allow cleavage of the periplasmic domain of CadC, β-gal activity dropped dramatically (~500 MU). The CadC-RodZ chimera (*C-H-R*) yielded very low β-gal activity (~200 MU), because the periplasmic domain does not, apparently, dimerize. Consistent with that conclusion, insertion of the AAA cleavage site had no effect on β-gal activity.

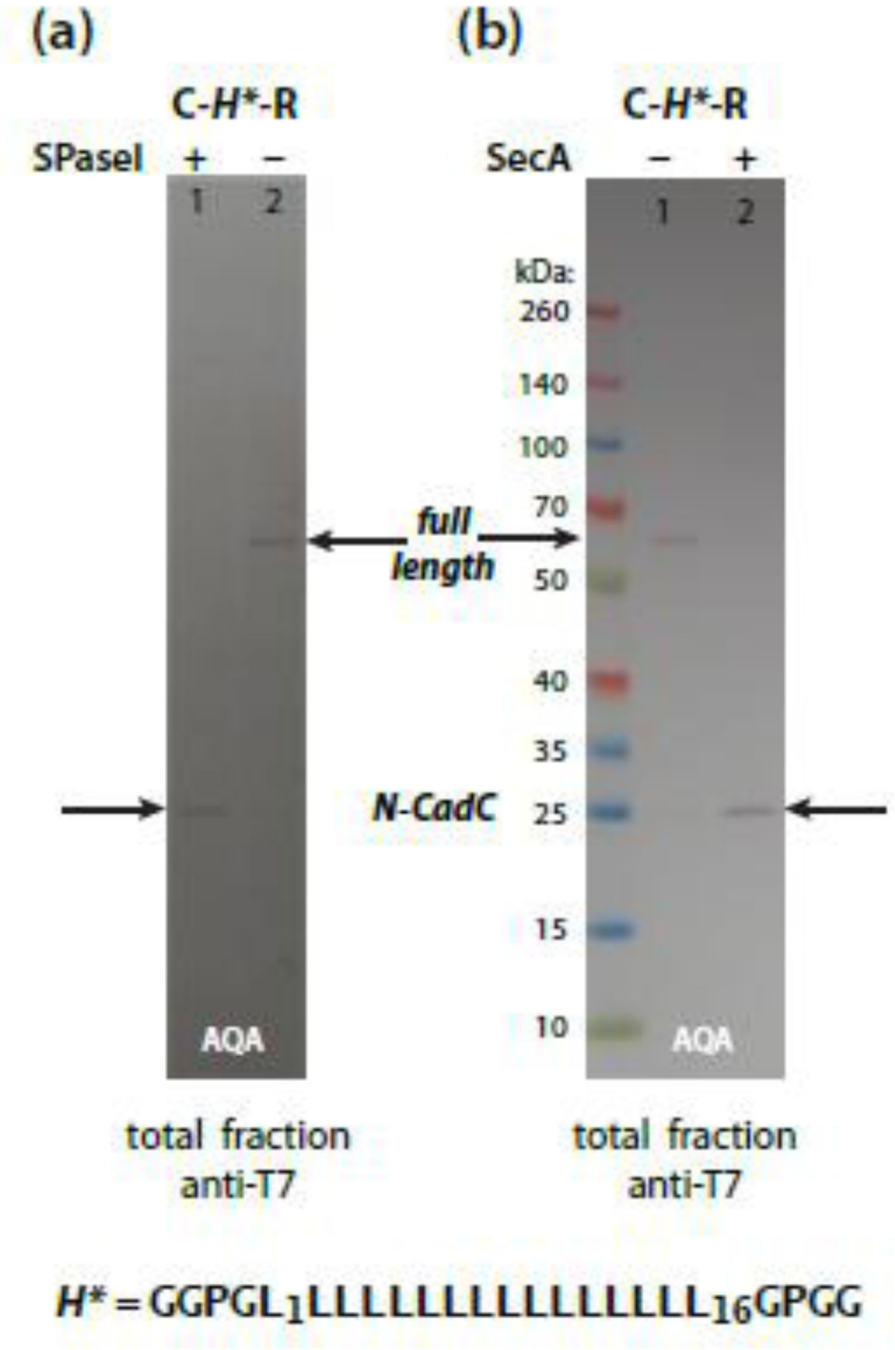

Figure 3.

Depletion experiments demonstrate that cleavage is due to Signal Peptidase I and that membrane insertion of CadC requires SecA. (**a**) Signal peptidase I is responsible for cleavage. The CadC construct *C-H**-R* carrying an AQA site was expressed in a SPase I depletion strain (FTL85) in which *lepB* is under the control the arabinose promoter/operator. The absence of arabinose leads to SPase I depletion. When SPase I is depleted, no T7-tagged *N-CadC* fragment is found (compare lanes 1 & 2). (**b**) SecA translocase is essential for membrane insertion of CadC constructs. The *C-H**-R* construct with an AQA cleavage site

was expressed in the SecA depletion strain EO527 in which *secA* is regulated by the arabinose promoter/operator. The absence of arabinose leads to SecA depletion. Under depletion conditions, only full-length *C-H*-R* is found, consistent with failure to insert the construct into the cytoplasmic membrane.

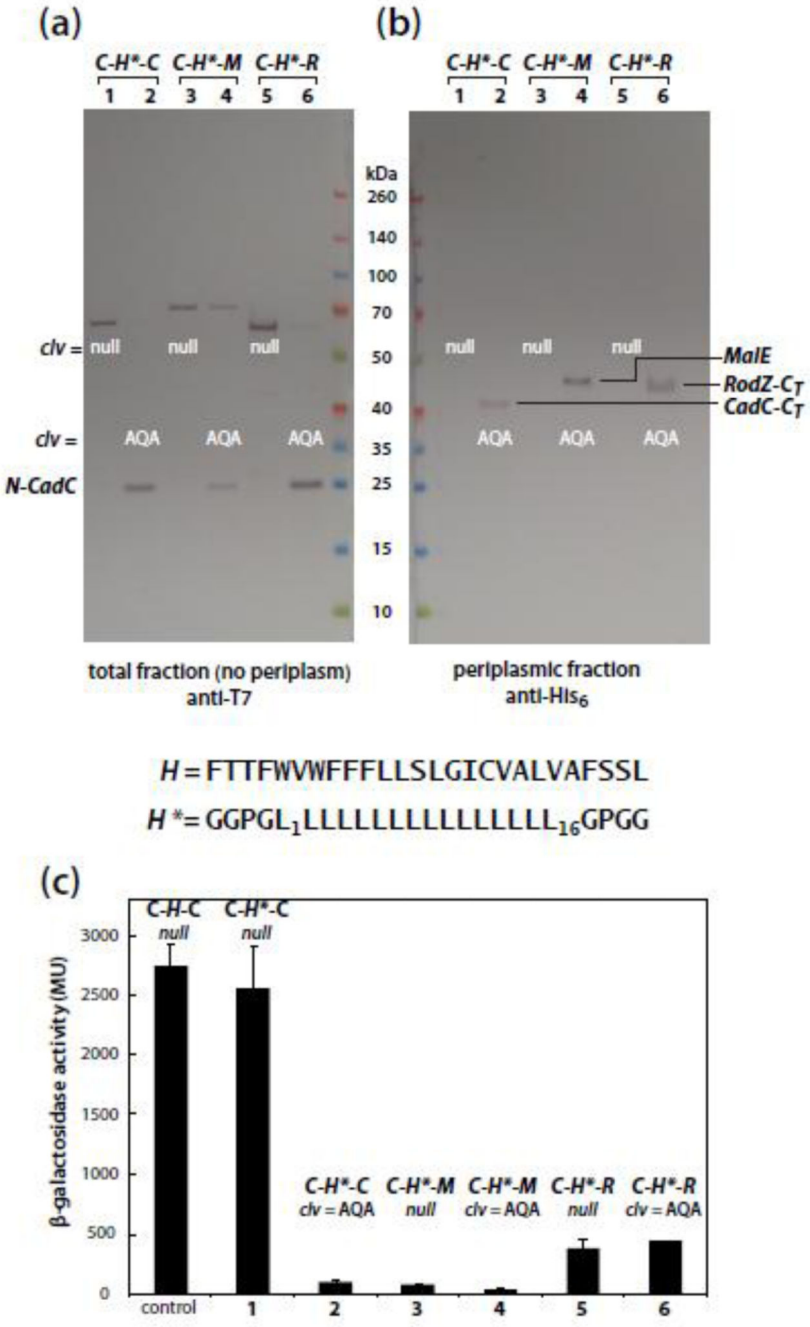

Figure 4.

The CadC TM helix can be changed to a polyleucine variant without effect. (**a,b**) Changing the H-segment from wild-type *H* to *H** comprised of sixteen leucines bounded by GGPG- and -GPGG [24–45] has no effect on topology or dimerization. Insertion of an -AQA- cleavage site results in a periplasmic His_6-tagged fragment (panel b, lanes 1 & 2) and a non-periplasmic T7-tagged fragment (panel a, lanes 1 & 2). Changing the periplasmic domain of the *H** CadC variant to $MalE_{27–396}$ or RodZ-C_T has no effect on cleavage or topology. Compare lanes 3 – 6 of panel a with lanes 3 – 6 of panel b. (**c**) Regardless of H-segment

composition, only the wild-type CadC periplasmic domain can activate the *cadBA* operon. The β-gal activity measured for each construct is shown. The far-left bar labeled 'control' is the activity of wild-type CadC measured independently of the values shown in Fig. 2. The numbers below the other bars in the graph correspond to the numbered lanes in panels a and b. Notice the activity of *C-H**-C* (bar 1) is statistically equivalent to *C-H-C*, indicating that the change in the H-segment from *H* to *H** had little effect on activation. But when an -AQA- cleavage site is present in *C-H**-C*, the β-gal activity collapses to a negligible value (bar 2). The other constructs, *C-H**-M* and *C-H**-R*, with or without a cleavage site have very low β-gal activities (bars 3 – 6). Bars 5 & 6 are higher than the others, because the expression level of *C-H**-R* was higher, as evidenced by the relative intensities of the bands on the blots.

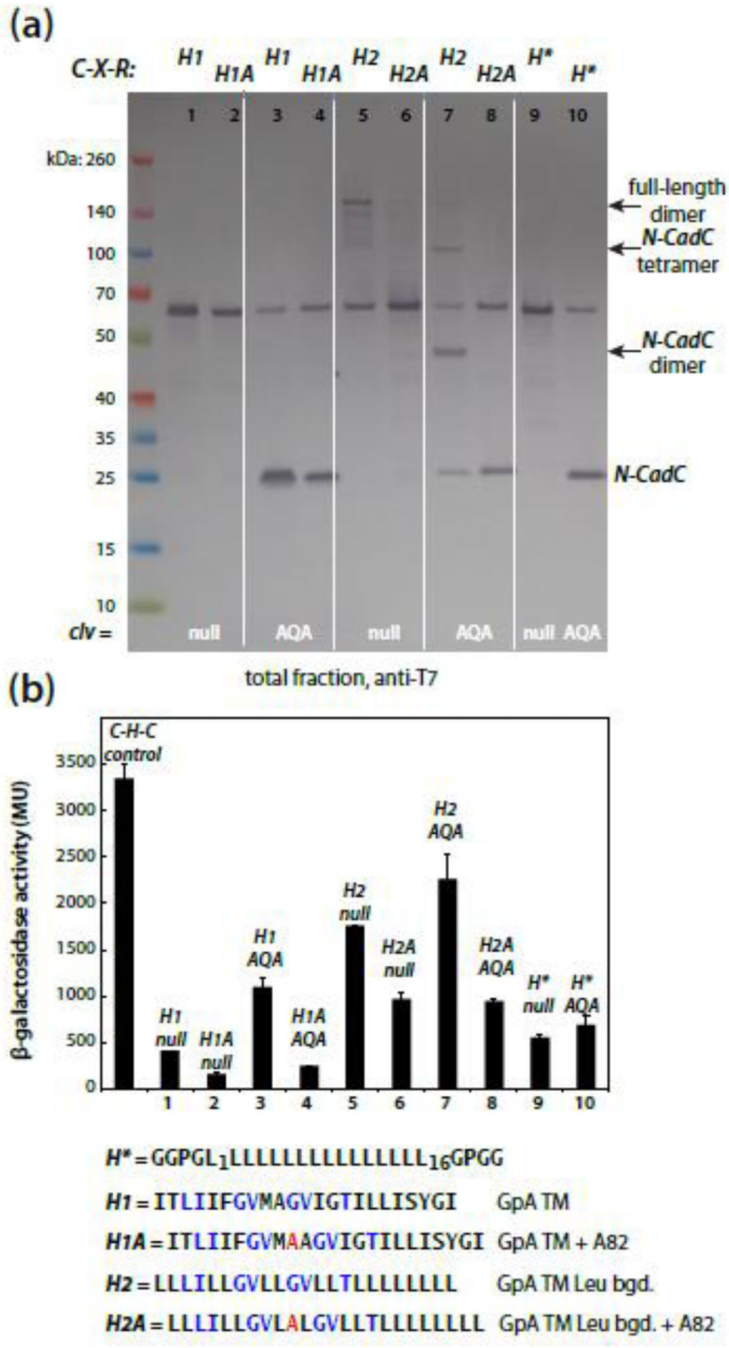

Figure 5.

Glycophorin A (GpA) TM domains can drive dimerization in the absence of the CadC periplasmic domain. We used variants of GpA TM helices to test for detection of the helix-helix dimerization. We used a CadC construct in which CadC-C_T was replaced by RodZ-C_T so that dimerization would depend only on the H-segment. The polyleucine construct *C-H*-R* served as a neutral control. We began with the wild-type GpA H-segment (labeled *H1*), and then added a single Ala residue at the GpA position 82 (H-segment *H1A*), which is known to disrupt GpA dimerization in SDS micelles [30]. We next increased the overall

hydrophobicity of the GpA-based H-segment by replacing all residues not involved in dimerization with leucines (*H2* and *H2A*). (In the sequences shown, the residues involved directly in GpA dimerization are colored purple.) (**a**) SPase I processed all constructs when an AQA cleavage site was present. However, because the expression levels were very high in these experiments, SPase I processing was incomplete. High expression levels were used in the hope of seeing dimers on the blots, which we did. Oligomers are clearly visible in *H2* constructs, but not the *H2A* constructs, consistent with Ala82 disrupting dimerization. The oligomerization occurred despite boiling the samples in SDS sample buffer. (**b**) The various GpA constructs have different β-gal activities. The bars are numbered according to the lanes in panel a. The far-left bar is the wild-type CadC activity, measured independently of wild-type activities in Figs. 2 & 3. Notice that the *H2* constructs have higher activities than the *H1* constructs, indicating that hydrophobicity as well as the GxxxG motif contribute to dimerization. Ignoring the wild-type control, the highest activity was observed for *H2* with -AQA- cleavage site (bar 7). The *H2 (null*) level was lower, perhaps because the RodZ-C_T domain interfered with helix dimerization. Indeed, ignoring expression levels, there seems to be a trend toward higher activities for the AQA constructs: $H1_{AQA} > H1_{null}$ and $H2_{AQA} > H2_{null}$, and $H1A_{AQA} > H1A_{null}$, $H^*_{AQA} > H1^*_{null}$ although $H2A_{AQA} \approx H2A_{null}$. The large variations in the β-gal activities make interpretation of the data problematic. As we show in Fig. 6, the variations are apparently a result of variations in protein expression levels.

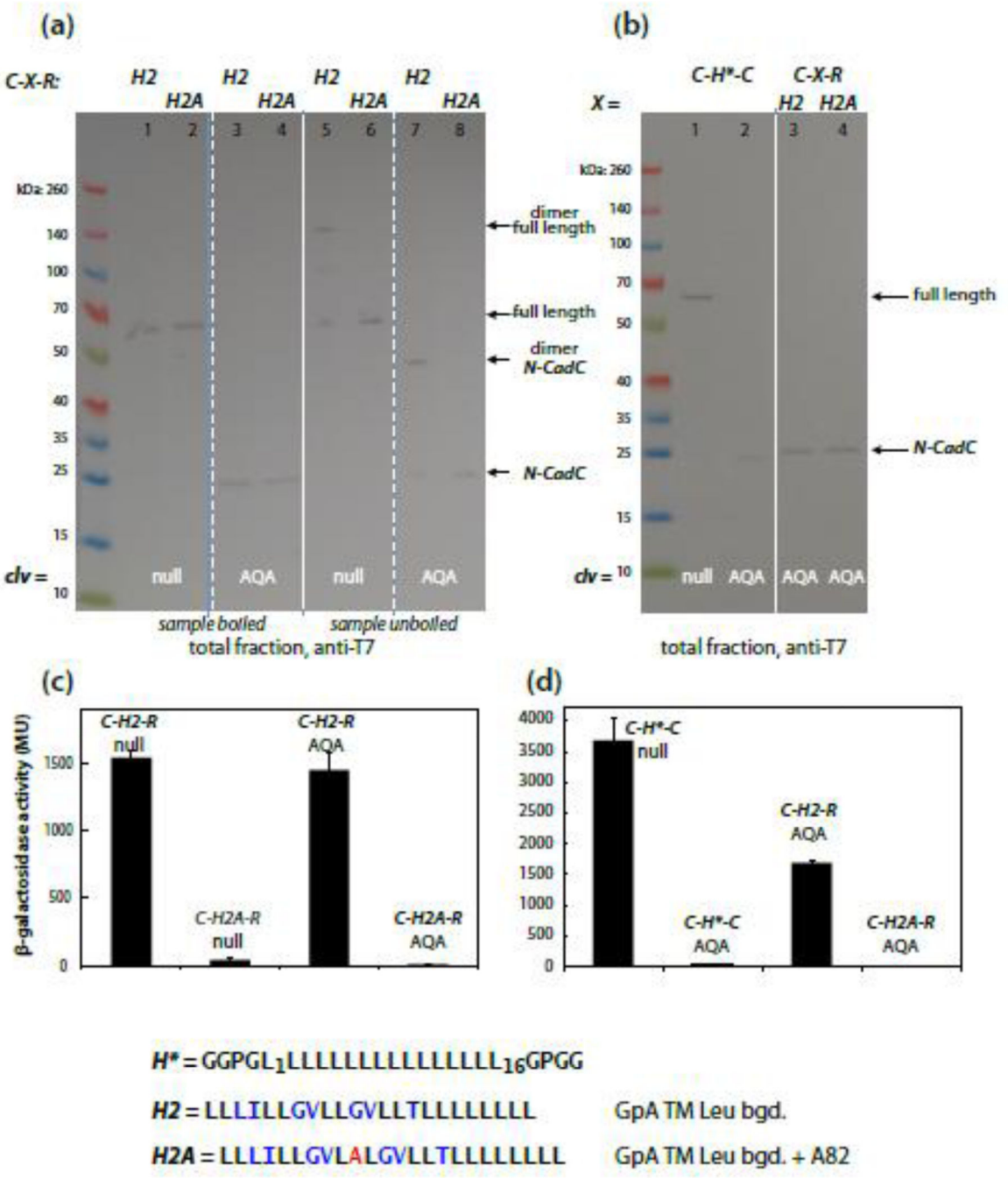

Figure 6.

Lower expression levels increase the 'dynamic range' between interacting and non-interacting H-segments, and leads to full processing by SPase I. As discussed in the main text and Methods & Materials, we reduced protein expression dramatically by using Top10 cells. (**a**) Dimerization of GpA-based CadC constructs is obvious at low expression levels if the samples in SDS samples are not boiled (compare lanes 1 – 4 with lanes 5 – 8). Dimers are seen for full-length *C-H2-R (null*) but not for *C-H2A–R (null*). For *H2* constructs carrying an AQA cleavage site, dimers of N-CadC are observed for *C-H2-R* but not for *C-*

H2A-R (compare lanes 7 & 8). (**b**) CadC with an *H** H-segment is cleaved fully at low expression levels (compare lanes 1 & 2) as are *C-H2-R* and *C-H2A-R* (lanes 3 & 4). (**c**) Activation of the *cadBA* operon by dimerizing species is strongly apparent in the β-gal assays at low protein expression levels. *C-H2-R* constructs with and without -AQA- have equally high β-gal activities, consistent with dimerization of only the H-segments irrespective of the absence or presence of the periplasmic RodZ-C_T domain. The full extent of the blockage of dimerization GpA by 82Ala is now apparent; both the *null* and -AQA- have virtually no β-gal activity. (**d**) Activation of the *cadBA* operon is stronger for periplasmic domain dimerization than for H-segment dimerization. Under the reduced-expression conditions, β-gal activity for *C-H*-C* is very high without the AQA cleavage, but drops practically to 0 when the cleavage site is present. The data for *C-H2-R* from panel c have been re-plotted in panel d for comparison (note the difference in vertical scales between the two panels).

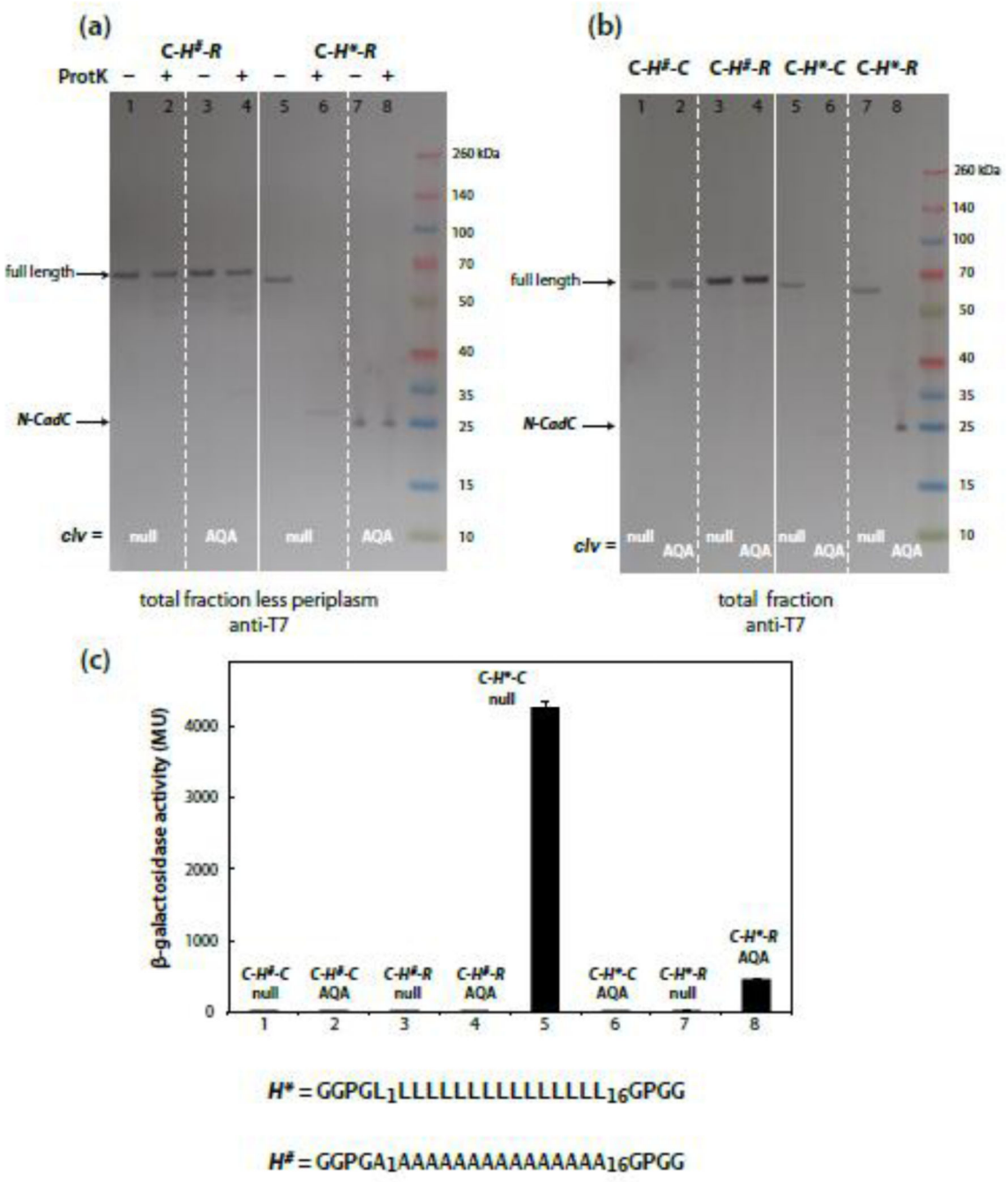

Figure 7.

H-segments with low hydrophobicity are not inserted as transmembrane segments. Following the lead of earlier studies [24–32] that showed polyalanine H-segments are inserted across membranes with a low probability, we examined the TM insertion of CadC constructs with polyleucine (*H**) or polyalanine ($H^{\#}$) H-segments using the Top10 expression system. (**a**) CadC-RodZ constructs (*C-X-R*) with polyalanine H-segments are not inserted across membranes whereas those carrying polyleucine segments are inserted. The RodZ-C_T domain of *C-H#-R* cannot be digested by protK treatment of spheroplasts (lanes 1

& 2) nor can it be cleaved by SPase I (lanes 3 & 4), indicating that *C-H*$^{\#}$*-R* is not inserted across the cytoplasmic membrane. It must reside only in the cytoplasm. *C-H*-R* carrying a polyleucine H-segment, on the other hand, is fully inserted as judged both by protK digestion of spheroplasts (lanes 5 & 6) and SPase I cleavage (lanes 7 & 8). Note that because SPase I automatically cleaves *RodZ-C*$_T$domain leaving the *N-CadC* domain exposed only on the cytoplasmic surface of the membrane, the *N-CadC* domain is inaccessible to protK (lanes 7 & 8). (**b**) Failure to insert polyalanine H-segments across the cytoplasmic membrane is independent of the periplasmic domain. No *N-CadC* fragment is observed for *C-H*$^{\#}$*-C or* for *C-H*$^{\#}$*-R* with or without the -AQA- cleavage site (lanes 1 – 4). When the H-segment is polyleucine, however, *C-H*-C* or for *C-H*-R* are fully inserted (lanes 5 – 8). (**c**) The CadC periplasmic domain (CadC-C$_T$) must be in the periplasm to promote binding of the CadC to the *cadBA* operon. The bars in the β-gal activity chart are numbered according the lane numbers of panel b. High β-gal activity is seen only for *C-H*-C* lacking a cleavage site (bar 5). The activity seen for *C-H*-R* (AQA) is a result of much higher expression levels.