University of Alberta

Biochemical and structural characterization of CpxP and CpxA, key components of an envelope stress response in *Escherichia coli*

by

Gina Lea Thede

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

©Gina Lea Thede

Fall 2012

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Abstract

The Cpx two-component signal transduction pathway of *Escherichia coli* consists of an inner membrane-localized sensor histidine kinase, CpxA, the response regulator, CpxR, and the novel periplasmic accessory protein, CpxP. These proteins mediate a bacterial stress response, sensing envelope perturbations including damage caused by misfolded periplasmic or inner membrane proteins, and regulating the expression of a host of factors that contribute to preservation of the envelope.

While structural information exists for a number of bacterial two-component systems, there is limited data to describe the constituents of the Cpx system. Since CpxP has no homologues of known function, we initially focused on its biophysical and structural characterization. Biochemical studies demonstrated that CpxP maintains a dimeric state, but may undergo a slight structural adjustment in response to the Cpx pathway inducing cue, alkaline pH. The crystal structure of CpxP, determined to 2.85 Å resolution, revealed an antiparallel dimer of intertwined α-helices with a highly basic concave surface. We identified that the core fold adopted by CpxP is also found in a number of other periplasmic stress response proteins. Additionally, we proposed that the conserved LTXXQ motifs that define a family of proteins have a structural role in the formation of diverging turns. Finally, we identified several sites, including two solvent-exposed residues and the charged surfaces of CpxP, which are likely involved in intermolecular interactions.

In an effort to understand the molecular mechanism(s) by which CpxA detects specific inducing signals, we set out to describe the uncharacterized periplasmic sensor domain of CpxA biochemically and structurally. We demonstrated that the isolated sensor domain likely exists in a unique tetramer-dimer equilibrium. Further, we suggested that the periplasmic domain of CpxA forms a PAS-like PDC fold, and identified regions that could possibly be involved in a structural rearrangement upon stimulus perception. Lastly, we established preliminary crystallization conditions that will be optimized for high-resolution structure determination. Ultimately, these studies will provide insight into the molecular characteristics of CpxA and CpxP that contribute to the regulation of the response to envelope stress in pathogenic bacteria such as *E. coli*.

Acknowledgements

First and foremost, I need to thank my supervisor, Dr. Mark Glover, for his support and encouragement over the years, through my many projects big and small. His immeasurable patience, and his cultivation of an environment which promotes independence have been instrumental in shaping my time in the lab, and I hope to continue to develop the skills I have learned far into the future.

Several members of the lab, in particular, have helped to instigate or guide the key projects that played a part in my graduate studies. I would like to thank Megan Lee for introducing me to my first project on MDC1, Steve Chaulk for opening my eyes to the world of microRNAs, and Dave Arthur for initiating the Cpx project in our lab. I am indebted to Ross Edwards, who I can't thank enough for his help and expertise, and for being there through it all. Thanks also to my other lab members, past and present, for their daily input and suggestions: Ruth, Jody, Jason, Nina, Jun, Charles, Stephen, Joyce, Danny, Nico, Zahra, Curtis, Sheraz, Inbal, Yun, Mike, and Lucy.

I owe a huge thank-you to Dr. Howard Young for taking me under his wing to co-supervise my microRNA EM project, and to the members of his lab for an enjoyable two and a half years. It was an exciting experience to go from "EM for Dummies" to writing part of a paper on the topic. A special thanks to Dr. Tracy Raivio, for initiating the collaboration on the Cpx pathway with our lab. She has been an invaluable wealth of knowledge and help in supervising this project. Thanks also to her lab members, especially Daelynn Buelow, Roxana Malpica, and Julia Wong for their contributions to our work.

I would also like to thank the other members of my committee, my external examiner Dr. Mirek Cygler (University of Saskatchewan) for providing his expertise and thoughtful comments, and Dr. Andrew MacMillan for his supportive ideas, particularly with regards to the microRNA project.

Importantly, I would like to thank the staff at beamline CMCF 1 at the Canadian Light Source (Saskatoon) for their support during X-ray crystallographic data collection, and to the Canadian Institutes of Health Research and Alberta Heritage Foundation for Medical Research for supporting me and my work.

Thank you to friends and the members of our 70+ intramural/sports teams (Biochemical Disasters, Kiss My Fatty Acid, Muffins, and Ligers) for the fun and laughs, and for helping me to get out of the lab at night. A special thank you to everyone who contributed to the success of the winter charity drive for the past eight years for making this time in the department memorable.

To my parents, Gary and Theresa, my sister, Whitney, and to Jason – I could not have done this without you.

Table of Contents

List of Tables

Page

List of Figures

Page

List of Abbreviations

ATP	adenosine triphosphate
BSA	bovine serum albumin
CD	circular dichroism
Cpx	conjugative plasmid expression
DLS	dynamic light scattering
DTT	dithiothreitol
ECF	extracytoplasmic function
EM	electron microscopy
FSC	Fourier shell correlation
GndCl	guanidine hydrochloride
GST	glutathione-S-transferase
HAMP	histidine kinase, adenylyl cyclase, methyl-accepting chemotaxis protein, and phosphatase
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HK	histidine kinase
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani
MAD	multi-wavelength anomalous dispersion
MALLS	multi-angle laser light scattering
MBP	maltose-binding protein
MES	2-(N-morpholino)ethanesulfonic acid
miRNA	microRNA
mRNA	messenger RNA
M_z	molar mass
NCS	non-crystallographic symmetry

NlpE	new lipoprotein E
NMR	nuclear magnetic resonance
$O.D._{600}$	optical density at 600 nm
PAS	Per, ARNT, Sim
PCR	polymerase chain reaction
PDB	Protein Data Bank
PDC	PhoQ-DcuS-CitA
PMSF	phenylmethylsulfonyl fluoride
pri-miRNA	primary-microRNA
R_h	radius of hydration
RMSD	root mean square deviation
RR	response regulator
SAD	single-wavelength anomalous dispersion
SAXS	small-angle X-ray scattering
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
Se-Met	selenomethionine
TCS	two-component system
TLS	translation libration screw
T_m	melting temperature
Tris	tris(hydroxymethyl)aminomethane

CHAPTER 1

Introduction

Bacterial extracytoplasmic stress responses

Bacteria have evolved mechanisms to sense and respond to continually changing conditions, allowing them to survive in a wide range of environments. Five major stress responses that are necessary for maintenance of the bacterial cell envelope, which consists of the outer membrane, the periplasm, and the inner membrane, have been identified in Gram-negative bacteria such as *Escherichia coli*. These consist of the σ^E, and Psp responses, and the Bae, Rcs, and Cpx two-component systems (TCSs). These responses are involved in a variety of protective tasks that encompass diverse aspects of the envelope. The σ^E pathway is mainly involved in the upkeep of outer membrane proteins (1, 77), while maintenance of inner membrane integrity and restoration of the proton motive force is controlled by the Psp response (39, 98). The Bae signal transduction system has been implicated in expulsion of toxic compounds from the cell (46), and the Rcs phosphorelay is involved in monitoring envelope proteins and capsule synthesis (50). The Cpx pathway appears to mediate a broad response to envelope stress including remediation of damage caused by misfolded periplasmic proteins (91). These stress responses are activated by distinct perturbations, and regulate discrete groups of genes, although there is evidence of some overlap (47). Recently, a sixth envelope stress response was identified in *E. coli*. The vesicle release response regulates the removal of both insoluble and soluble molecules from the envelope in the form of secreted outer membrane vesicles (44). In contrast to the aforementioned five pathways, the vesicle release response appears to be a general mechanism to combat extracytoplasmic stress.

While the σ^E and Psp stress responses utilize an extracytoplasmic function (ECF) sigma factor pathway or an alternate signal transduction mechanism, respectively, the Bae, Rcs, and Cpx stress responses utilize the widespread two-component system (TCS) signalling strategy to enhance survival

under adverse environmental conditions. Two-component systems sense particular signals *via* a histidine kinase (HK) that transmits the information through a phosphotransfer event to a cognate response regulator (RR) in the cytoplasm. The RR in most cases subsequently controls the expression of a limited set of target genes that mediate adaptation to the specific cue. Here, the pathway components, activating signals, and functional role of the Cpx TCS will be examined.

The Cpx stress response: a two- (or three- ?) component system

The Cpx system was first identified in 1980 in an investigation of *E. coli* K-12 genes that were necessary for conjugative plasmid expression (Cpx), providing the first indication that the function of the pathway would be associated with the bacterial envelope (55). The *cpx* locus was found to code for two proteins: an inner membrane-localized sensor histidine kinase CpxA, and the response regulator CpxR (2, 23). In addition to these two classic components, a third constituent, CpxP, was discovered in a screen for genes regulated by the Cpx system (18). A schematic of these three proteins, including the domains identified by the PROSITE, UniProt, and Pfam databases is shown in Figure 1-1 (37, 49, 72).

A role for the novel accessory protein CpxP

While CpxA and CpxR were recognized as typical two-component signal transduction proteins, the function of CpxP was unclear. After closer examination, the CpxP protein was established to be a resident of the periplasm where it was capable of counteracting the toxicity derived from the aggregation of extracytoplasmic proteins (18). Further, CpxP overexpression resulted in the negative regulation of genes controlled by the Cpx response, and this repression

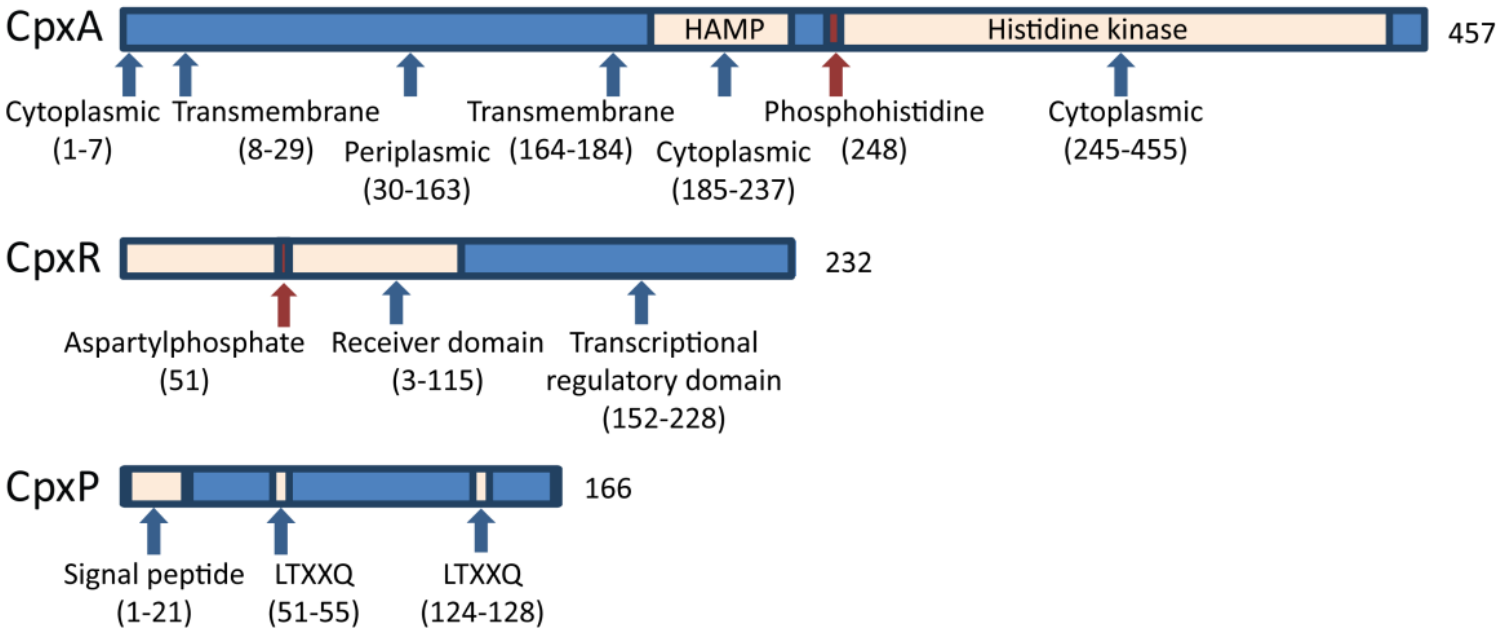


Figure 1-1. Domain architecture of the Cpx proteins. Schematic diagram of CpxA, CpxR, and CpxP depicting the predicted domains, phosphorylation sites, and motifs identified by the PROSITE (37), UniProtKB (49), and Pfam (72) databases.

required the periplasmic sensing domain of CpxA (75). Additionally, the Cpx system could not be fully activated by envelope stress caused by spheroplasting (elimination of the cell wall) when CpxP was tethered to the inner membrane (74). Together, these results suggest that CpxP functions as an inhibitor that likely interacts with the CpxA sensing domain to repress the Cpx response, and that CpxP must be removed for activation of the pathway (74, 75).

The mechanism by which CpxP is released from its inhibitory role is not fully understood. Previous studies have shown that CpxP is degraded by the major periplasmic protease, DegP, upon pathway activation, and CpxP is required for DegP-mediated proteolysis of misfolded pilus proteins (8, 38). Therefore it has been speculated that CpxP may have a dual function in the Cpx response as both a negative regulator of CpxA in the absence of inducing cues, and upon signal sensing as a chaperone to assist in the delivery of certain misfolded envelope proteins to DegP, where it is itself ultimately degraded (38).

Induction of the Cpx stress response

The Cpx response is activated by a wide variety of signals, however the nature of the majority of the inducing cues are poorly understood. Most stimuli are sensed by CpxA, while some have been reported to bypass CpxA and activate the Cpx response *via* CpxR (reviewed in (91)). Cpx inducing signals include alkaline pH (18, 60), salts (26), copper (43, 99, 100), and modified membrane lipid composition resulting from phosphatidylethanolamine deficiency (57), the accretion of the enterobacterial common antigen lipid II precursor in the inner membrane (16), or mutation of phospholipase A in the outer membrane (45). Other triggers arise from the overproduction of membrane-associated proteins that aggregate at the periplasmic face of the inner membrane. These include the overexpression of pilus subunits such as the specific P-pilus subunits, PapE and PapG, in the absence of their chaperone PapD (40), and the major subunit, BfpA,

of the type IV bundle-forming pilus (63). Mutation of the inner membrane protease FtsH also induced the Cpx pathway, possibly due to accumulation of proteins that would normally be degraded in the inner membrane (82). The Cpx response is also activated by overexpression of the outer membrane new lipoprotein E (NlpE), which leads to its mislocalization at the inner membrane (84), and NlpE-dependent adhesion of bacteria to abiotic (hydrophobic) surfaces (67). Many of these inducing cues are bacterial stresses that lead to modified membrane composition or the accumulation of misfolded proteins at the inner membrane, suggesting that inner and outer membrane perturbations are a major signal for activation of the Cpx stress response by CpxA, although the molecular mechanism by which CpxA senses these signals is unknown. In addition, CpxR appears to integrate stimuli independently from CpxA. Growth signals upon entering stationary phase downstream of CpxA (22), and metabolic signals from surplus glucose or pyruvate (94) have both been shown to induce the Cpx response through CpxR. Although many activating signals have been identified, how they are recognized by CpxA (or CpxR) to induce the Cpx response is not yet clear.

The Cpx regulon

Activated CpxR generally binds to consensus sequences within 100 bp of the transcription start site of Cpx-regulated genes to mediate adaptation (70). Some of the most strongly controlled regulon members have key associations with envelope protein quality control (70), yet with each new study the range of regulated genes is rapidly expanding. A number of genes encoding periplasmic protein folding and degrading factors are upregulated by the Cpx response including *degP*, which codes for the major periplasmic chaperone/serine protease (19), *dsbA*, a disulphide bond oxidoreductase (17, 69), *ppiA*, specifying a peptidyl-prolyl isomerase (69), and *spy*, which is the closest homolog of *cpxP*,

and is now known to function as a chaperone (73, 74). Several inner membrane proteins that are upregulated by the Cpx response include HtpX, a protease that degrades misfolded inner membrane proteins (82), and YccA, which inhibits degradation of blocked Sec translocation machinery (88). Genes encoding peptidoglycan remodelling enzymes, such as the *N*-acetylmuramoyl-L-alanine amidases AmiA and AmiC (93) are also upregulated by CpxR. In contrast, protein complexes such as pili and flagella, and inner membrane transporters are downregulated by the Cpx system, perhaps to curtail nonessential processes or reduce unnecessary protein traffic in the periplasm during periods of stress (20, 48, 70, 90). The Cpx pathway upregulates its own constituents, CpxRA and CpxP, which provides a mechanism for feedback and autoregulation. Interestingly, the CpxP promoter is the most highly induced of those tested, and is often used as a reporter of Cpx pathway activity (22, 70). The Cpx regulon forms links to other pathways as well through the upregulation of *mzrA*, which connects to the osmolarity detecting EnvZ/OmpR TCS (30) and *ydeH*, encoding a diguanylate cyclase which synthesizes the signalling molecule cyclic di-GMP (100), and by downregulating *rpoErseABC*, specifying members of the σ^E envelope stress response (21, 70). While many members of the Cpx regulon remain uncharacterized, the wide range of target genes known to be under the control of CpxR reflect the broad spectrum of processes and signalling networks necessary to maintain the integrity of the envelope in the presence of stress.

Cellular functions of the Cpx two-component system

Following the first experiments implicating the Cpx response in suppression of the toxicity resulting from protein aggregates in the periplasm (84), many additional studies have provided support for a role for the Cpx pathway in adaptation to envelope stress caused by misfolded secreted or periplasmic proteins. Beyond envelope protein maintenance, the Cpx TCS affects

physiological functions including biofilm formation, virulence, and antibiotic resistance. An examination of differential gene expression between cells in growth phase and those in a mature biofilm environment revealed that the Cpx pathway is activated in the formation of biofilms (4, 24). Additionally, CpxR was found to repress the expression of genes encoding curli, extracellular structures that mediate bacterial adhesion (71). The Cpx system also influences virulence in numerous bacterial pathogens through regulation of the expression of extracellular surface structures. The Cpx response is necessary for elaboration of pili including the type IV bundle-forming pilus in enteropathogenic *E. coli*, and the Pap pilus in uropathogenic *E. coli*, and regulates factors involved in type III and type IV secretion. In contrast, the Cpx response negatively affects virulence through repression of adhesins and adherence to host cells and decreased expression of protein secretion systems including the LspB-LspA2 secretion system (reviewed in (91)). Lastly, the Cpx response induces changes in the bacterial envelope that affect antibiotic resistance. The Cpx regulon members, AmiA, and AmiC, which modify the cell wall by cleaving peptidoglycan, were revealed to be involved in resistance to cationic antimicrobial peptides in *Salmonella enterica* serovar Typhimurium (93). In addition, activation of the Cpx response in *Haemophilus ducreyi* resulted in reduced resistance to cationic antimicrobial peptides (80). Furthermore, the Cpx TCS induces the expression of multi-drug efflux pumps involved in antibiotic resistance in *E. coli* (65). Given that the Cpx response functions to maintain the integrity of the bacterial envelope, and that many of the structures involved in biofilm formation, virulence, and sensitivity to antimicrobials are localized to the envelope, it is unknown whether these represent true functions that can attributed to the Cpx TCS or if they are secondary effects in a broad response to envelope stress.

While the inducing cues and downstream effects of the Cpx envelope stress response have been extensively studied, little is known about the structure of Cpx pathway components. An examination of existing structures

from different TCSs may give us insight into the mechanisms of stimulus detection, signal transduction, and gene regulation used by the CpxAR TCS.

Two-component system constituents: versatile signalling molecules with a modular design

Two-component systems are abundant in bacteria, regulating a variety of fundamental processes from nutrient sensing to nitrogen fixation. There is extensive variation in the prevalence of these proteins in bacteria, with the number of TCSs ranging from 0 to over 100. The TCS offers an effective strategy for transduction of environmental signals *via* phosphotransfer events from a histidine protein kinase to a response regulator protein in order to mediate a cellular adaptive response. Diversity is achieved through different domain interactions and alternative modes of regulation that arise from the modular organization of individual HK and RR domains, which are generally structurally conserved (27, 96). TCS pathway activity is dictated by the levels of HK autophosphorylation, kinase, and phosphatase activities. Additionally, accessory molecules may impart their own modes of regulation at different points in the system to modify the functions of the HK or RR to influence the cellular response. With regards to the Cpx response, the HK CpxA and RR CpxR are typical modular TCS proteins, while the structural composition of the novel auxiliary regulator, CpxP, is largely unknown.

Histidine kinases: the signal sensors and transmitters

HKs in TCSs consist of variable sensor domains that may be localized to assorted cellular compartments, coupled to conserved cytoplasmic catalytic core domains. They sense a diverse set of stimuli, ranging from small molecules to light, and function in a dimeric state. The HK must recognize a specific signal,

transmit this information in many cases through several domains to the histidine kinase domain, which autophosphorlyates at a histidine residue and communicates with a cognate RR through its often bifunctional kinase/phosphatase activities.

The recognized domain topology for CpxA consists of a short N-terminal cytoplasmic segment, two transmembrane regions that flank an uncharacterized periplasmic domain, a cytoplasmic HAMP linker domain, and a cytosolic histidine kinase catalytic core that contains the conserved histidine residue involved in phosphorylation (Figure 1-1). As observed in CpxA, histidine kinase catalytic modules are expected to be paired with a HAMP linker in 24% of HKs (85). Additionally, due to the enormous sequence variation observed in HK sensor domains, including CpxA, it is common for the constitution of the sensor fold to go unrecognized from sequence alone (Figure 1-1).

The sensor domain

HK sensor domains can be organized into three classes: those that are cytoplasmic, localized within the inner membrane, or periplasmic, depending on the signal that is detected (54). While the inner membrane-localized sensor domains consist of 2 to 20 membrane-spanning helices, cytoplasmic and periplasmic sensor domain structures fit into a few domain fold families (27).

The periplasmic segment of CpxA was first identified as the domain responsible for sensing envelope stress after an examination of mutations in this region revealed that CpxA was insensitive to normal signals (76). Though there is a substantial quantity of high-resolution structural information available for HK cytoplasmic domains, the first two structures of true sensor kinase periplasmic domains were revealed in 2003. A number of structures have now been elucidated, including the periplasmic ligand-binding domains of the chemotaxis

aspartate receptor Tar, which is not a true sensor HK (56, 101), the trimethylamine-N-oxide-binding TorS (59), the nitrate/nitrite sensor NarX (13), the citrate sensor CitA (78, 81), fumarate- and malate-binding DcuS (12, 68), the divalent cation and antimicrobial peptide sensor PhoQ (11, 15), the quorum-sensing LuxQ (61, 62), succinate-binding domain of DctB (12, 102), the complex carbohydrate sensor AbfS (25), and phosphate limitation sensor PhoR (9). Instead of an array of diverse folds, the structures of these sensor domains reveal two main types: an all-α-helical fold, and a mixed α/β fold (Figure 1-2).

The all-α-helical structures of Tar, TorS, and NarX share a similar four-helix bundle structure, although the nature of their ligand binding differs. These structures are distinct from the rest, which reveal a PAS- (Per, ARNT, Sim)-like fold composed of a central five-stranded anti-parallel β-sheet flanked on both sides by α-helices (and sometimes other modifications), to create an overall concave surface (Figure 1-3). This structural arrangement has been coined an α/β PDC fold, for PhoQ-DcuS-CitA, due to the differences between the periplasmic sensor kinase folds and cytoplasmic PAS domains (11). While most of the PDC sensor domains exhibit a single PDC fold, LuxQ and DctB reveal a distinct tandem PDC domain repeat. Given that the PDC domains share little primary sequence identity, it is surprising that they are related by such high structural conservation. The sequence variability in these sensing domains may suggest a means for their versatility in signal perception. Recently, a third extracytoplasmic HK sensor domain fold was described for $HK29_s$ (14). The lone example of this new class presents a periplasmic binding protein-like fold for which the stimulus is unknown. In all, computational analysis suggests that the majority of extracytoplasmic HK sensor domains are predicted to adopt the PAS-like PDC fold (9, 85).

Although HKs function as dimers in the cell, the isolated PDC domain subunits appear to possess a low intrinsic affinity for dimerization (12, 78, 92).

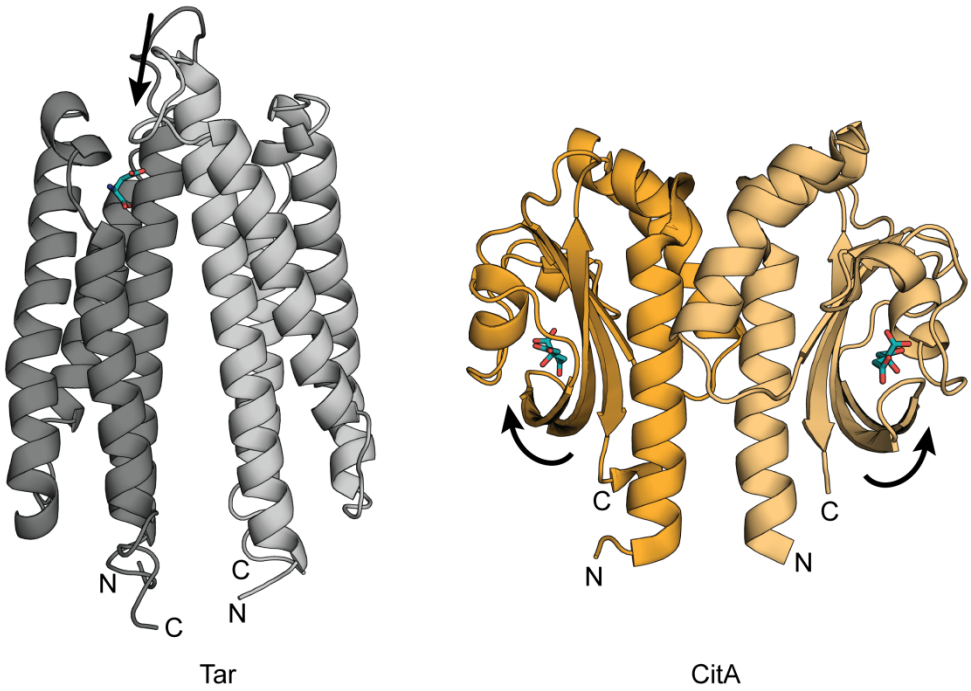


Figure 1-2. Dimeric structures of Tar and CitA periplasmic sensor domains. Cartoon representation of the all-α-helical fold of the aspartate-bound Tar periplasmic domain from *Salmonella enterica* serovar Typhimurium (PDB ID 1VLT), and the PAS-like PDC fold of the citrate-bound CitA periplasmic domain from *Klebsiella pneumoniae* (2J80). The N- and C-termini are labelled, and point towards the inner membrane. The individual protomers are shown in light and dark shades of grey and orange for Tar and CitA, respectively. Ligands are displayed in teal. The arrows illustrate the piston-like displacement of one subunit relative to the other in Tar, and the contraction of the β-sheet upon ligand binding to CitA.

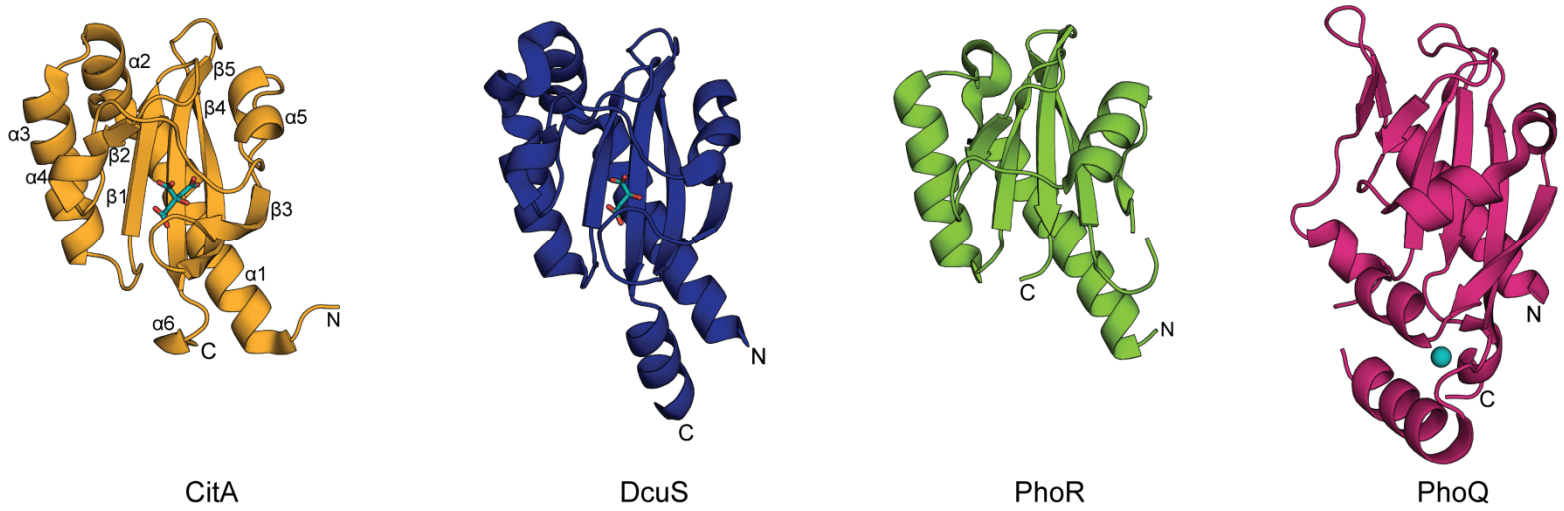


Figure 1-3. Structural conservation of the predominant periplasmic histidine kinase sensor domain fold. Structural alignment of several single PDC domains including the citrate-bound sensor CitA from *Klebsiella pneumoniae* (PDB ID 2J80), malate-bound sensor DcuS from *Escherichia coli* (3BY8), phosphate limitation sensor PhoR from *Bacillus subtilis* (3CWF; no ligand present), and divalent cation- (nickel)-bound PhoQ sensor from *E. coli* (3BQ8). Structures are illustrated as cartoons with the N- and C-termini indicated and ligands shown in teal. The secondary structure elements of CitA are also labelled. Structure-based alignments were generated using Dali with CitA as the query molecule (34, 35). DcuS shares 28% sequence identity with CitA over 126 aligned residues, with a root mean square deviation (RMSD) of 1.6 Å and a Z-score of 20.1. PhoR shares 11% sequence identity with CitA over 106 aligned residues, with an RMSD of 2.4 Å and a Z-score of 11.6. PhoQ shares 10% sequence identity with CitA over 99 aligned residues, with an RMSD of 3.3 Å and a Z-score of 7.1.

However, many of the PDC domains form dimers in the crystal lattice, and reveal different modes of dimerization. Given that the isolated PDC domains exhibit weak dimerization, it has been suggested that crystal lattice packing interactions and minor electrostatic forces might compete and even prevail over the physiologically relevant dimeric orientation in some cases (12). Mutagenesis studies, *in vivo* activity assays, cross-linking, and computational analyses were performed to identify the physiologically relevant dimer of PhoQ (11, 32). In addition to demonstrating that particular features of the dimer contacts were critical to for function, one of the criteria rested on the relative positions of the periplasmic domain N- and C-termini. The transmembrane helices are presumed to pack into a dimeric four-helix bundle similar to the phototaxis transducer (33). Therefore, the N- and C-terminal ends of dimeric PDC domains must be proximal and run parallel to each other in order to connect with the transmembrane four-helix bundle. This was determined to be true for one of the PhoQ dimeric orientations (11), and similar putative physiologically relevant subunit arrangements are observed in structures of NarX (13), DcuS (12), LuxQ (62), DctB (102), Tar (101), and citrate-bound CitA (81), depicted in Figure 1-2.

The HK PDC sensing domains have been implicated in binding a wide range of ligands, from ions to small molecules and even proteins, and in most cases this interaction is inhibitory. Consequentially, the PDC domains exhibit diverse modes of interaction with their ligands. Small molecules such as citrate, fumarate or malate, succinate and malonate bind in a concave pocket formed by the β-sheet and connecting loops of CitA, DcuS, and DctB, respectively (12, 68, 78, 81, 102). Uniquely, PhoQ binds divalent cations through a helix-turn-helix insertion (11, 15). However, none of these contacts describe protein-protein interactions that might occur in the potential recognition of CpxA by CpxP. Perhaps the most relevant example comes from the protein-PDC interaction described in the structures of the periplasmic ligand-binding protein LuxP in complex with LuxQ (61, 62). However, LuxP contacts a groove between the

double PDC repeat of LuxQ, a motif that is not observed in CpxA. If the CpxA sensing domain adopts a PDC fold, it is probable that CpxP interacts with this domain in a novel manner.

The mechanism by which stimulus perception is translated into a signal that is transmitted through the membrane in periplasmic PDC sensor domains is poorly understood. A conformational change in the PDC domain has been speculated to be involved in signalling for CitA (81), DcuS (12), PhoQ (28, 89), and DctB (12, 102). Structures of CitA in the presence and absence of citrate reveal a contraction of the β-sheet around the ligand in the citrate-bound state, and this structural change was proposed to represent the regulatory switch that activates signalling through the transmembrane region (Figure 1-2) (81). In contrast, evidence suggests that LuxQ does not undergo a conformational change, and instead it is the subunits that shift relative to one another to induce a response (62). Taken together, it is likely that these diverse signal sensors utilize different mechanisms to relay a signal to the cytoplasmic kinase core.

Signal transmission through the transmembrane domain

For HKs with periplasmic sensor domains, the signal originating from an extracytoplasmic compartment must be transduced through the membrane to regulate the histidine kinase catalytic domain. As mentioned previously, the transmembrane helices are thought to be arranged in a four-helix bundle (33), with the N- and C- termini of the periplasmic domains connecting at the inner membrane. Aspartate binding to Tar was observed to induce an asymmetric displacement of one helix towards the transmembrane helices, suggesting a piston-like movement for signal transmission (Figure 1-2) (10, 56, 66, 101). The contraction observed in CitA upon citrate binding also supports a piston-like movement of the transmembrane helices (81). Although ligand binding to NarX resulted in a symmetrical displacement of two N-terminal helices towards the

membrane region, this motion was consistent with a piston-like movement as well (13). These results agree with a common mechanism for transmembrane signalling through a piston-like motion. However, the LuxQP system may employ an alternative method to affect kinase activity that exploits conformational changes to alter interactions at the dimer interface (62). Although a piston-like movement might be one mechanism to transmit the signal through the transmembrane helices, it does not appear to be a universal strategy.

The HAMP linker

Cytoplasmic HAMP domains, found in HKs, adenylyl cyclases, methyl-accepting chemotaxis proteins, and phosphatases, are central to signal transmission. Each protomer consists of two helices connected by a loop, which dimerize to form a parallel four-helical coiled-coil (36) (Figure 1-4A). The nuclear magnetic resonance (NMR) structural ensemble illustrates an unusual coiled-coil knobs-to-knobs arrangement; however the characteristic coiled-coil knobs-into-holes packing can be satisfied by a 26° rotation of adjacent helices in opposite directions (36). This helical rotation in the HAMP linker is proposed to be involved in the mechanism for signal transmission to the kinase core, likely through the central coiled-coil that is predicted to connect directly with the dimerization domain. How this rotation would communicate with the piston-like motion described as the signalling mechanism through the transmembrane region is unknown.

The histidine kinase domain

The histidine kinase core module, consisting of dimerization and catalytic domains, exhibits a highly conserved structure and signalling mechanism. For

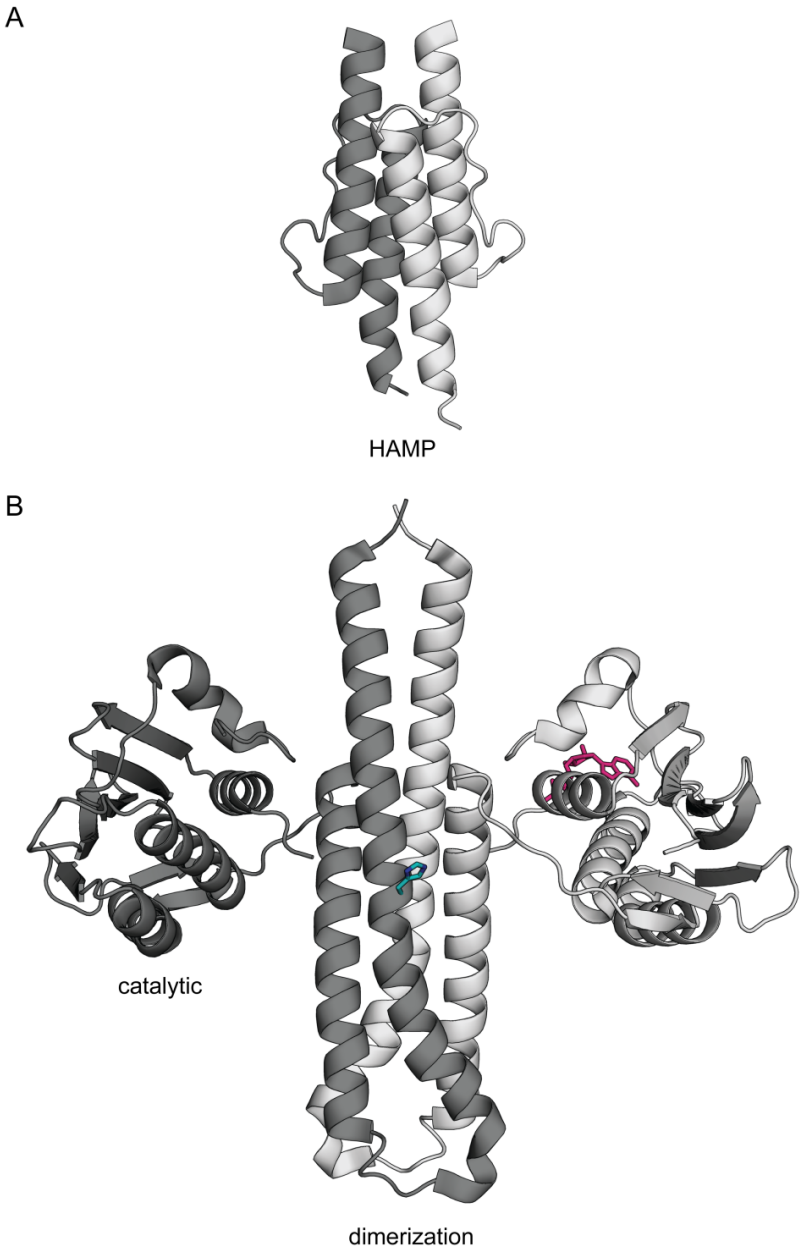


Figure 1-4. Structures of histidine kinase cytoplasmic domains. A) Four-helix bundle NMR solution structure of the Af1503 cytoplasmic HAMP domain dimer from *Archaeoglobus fulgidus* (PDB ID 2L7H). The protomers are shown in light and dark grey. B) Histidine kinase core including dimerization and catalytic domains from *Thermotoga maritima* HK TM0853 (PDB ID 2C2A). The solvent-exposed conserved histidine (teal) in the dimerization domain of one subunit (dark grey) is positioned for *trans*-phosphorylation by ATP (here, the analogue ADPβN in pink) in the catalytic domain of the second subunit (light grey) of the dimer. The structures are represented as cartoons.

most HKs, the histidine kinase core domain controls the functional level of the pathway through bifunctional kinase/RR phosphatase activities. The dimerization domain mediates the association of two subunits to form the functional HK, and contains the conserved histidine residue involved in autophosphorylation and phosphotransfer. The catalytic domain contains highly conserved ATP-binding and enzymatic sites to catalyze the phosphotransfer reaction.

A structure of the full kinase core of HK TM0853 from *Thermotoga maritima* reveals that two long antiparallel helices from each subunit mediate dimerization, forming a four-helix bundle through hydrophobic interactions between the helices (51) (Figure 1-4B). The conserved histidine residue is exposed on the surface of the dimerization domain in a manner that is accessible to the catalytic domain or RR for phosphotransfer. NMR studies of the osmosensor EnvZ demonstrated that the region surrounding the conserved histidine residue is flexible (87), which might imply this segment is accommodating in catalytic reactions.

The catalytic domain possesses an α/β sandwich fold in which the helices pack against a five-stranded β-sheet. The catalytic domain binds a divalent cation and adenosine triphosphate (ATP; in TM0853, the analogue ADPβN) in a highly conserved cavity that is covered by a flexible loop. This loop is proposed to undergo a conformational change upon nucleotide binding to close the cavity and ultimately induce catalysis (5, 27, 51, 96).

The conserved histidine residue in the dimerization domain accepts a phosphoryl group from ATP in the catalytic domain in a *trans*-autophosphorylation reaction, where the catalytic domain of one subunit phosphorylates the conserved histidine residue in the other subunit (64). Autophosphorylation generates a high-energy phosphoryl group on a histidine residue that is subsequently transferred to an aspartate residue in the RR. In the

structure of HK TM0853, the ATP (ADPβN) binding site in the catalytic domain is approximately 20 Å away from the conserved histidine residue in the dimerization domain (51, 96). In order for the *trans*-autophosphorylation reaction to occur, it was estimated that the catalytic domain was required to rotate approximately 70° around the dimerization domain. The structure revealed a substantial interface between the dimerization and catalytic domains that, in this conformation, is predicted to promote the phosphotransferase and phosphatase activities of the HK. Results from mutagenesis studies suggest that perturbation of the interface upon signal sensing results in HK autophosphorylation (51). Together, these observations outline a mechanism for *trans*-autophosphorylation that involves dynamic movement of the catalytic domain. The catalytic subunits are proposed to return to their original position after the autophosphorylation reaction to facilitate binding of the RR.

An analysis of covarying residues in HK and RR pairs was performed to elucidate the elements that confer specificity to the phosphotransfer reaction and prevent cross-talk between highly conserved HK dimerization and RR receiver domains from different TCSs. Two groups of residues were identified in HKs, which through mutational studies could convert the specificity of recognition to that of a different RR. One group resides close to the conserved histidine residue in the dimerization domain, and the other is localized to the membrane-distal terminus of the dimerization domain (83). Structural studies of a HK-RR complex demonstrated that the RR does bind to the conserved histidine-containing helix of the dimerization domain in an orientation that is poised for phosphotransfer (95, 97).

Response regulators: the output mediators

RRs generally contain a highly conserved receiver domain coupled to an effector domain that varies with the TCS output response. Out of 32 RRs from

TCSs in *E. coli*, 29 are transcription factors that mediate altered gene expression (96). RRs are activated upon phosphorylation, which often induces a conformational change that releases the RR to carry out its function. CpxR is composed of a receiver domain that accepts a phosphoryl group from CpxA, and a winged-helix DNA binding effector domain (Figure 1-1).

The receiver domain

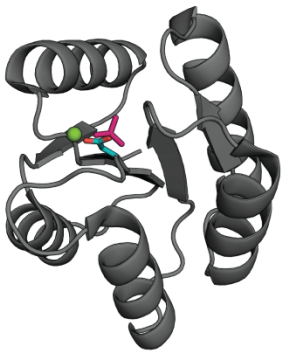
RR receiver domains contain a conserved aspartate residue that accepts a phosphoryl group from a conserved histidine in the HK. The receiver domain is involved in phosphoryl transfer and regulates the activity of the effector domain. Numerous structures of receiver domains reveal a similar α/β fold, consisting of a central parallel five-stranded β-sheet flanked on one side by two α-helices, and on the other side by three α-helices (3, 27, 96) (Figure 1-5A). Upon phosphorylation, the receiver domain active site undergoes a conformational change, the extent of which varies with each RR (27).

NMR studies of the RR, NtrC, establish that the unphosphorylated RR, while principally in an inactive state, is continuously undergoing conformational adjustments to sample both the inactive and active states. Upon phosphorylation, the equilibrium shifts towards the active conformation (29). The conformational switch induces changes in regulatory interactions that are particular to each RR (27).

The effector domain

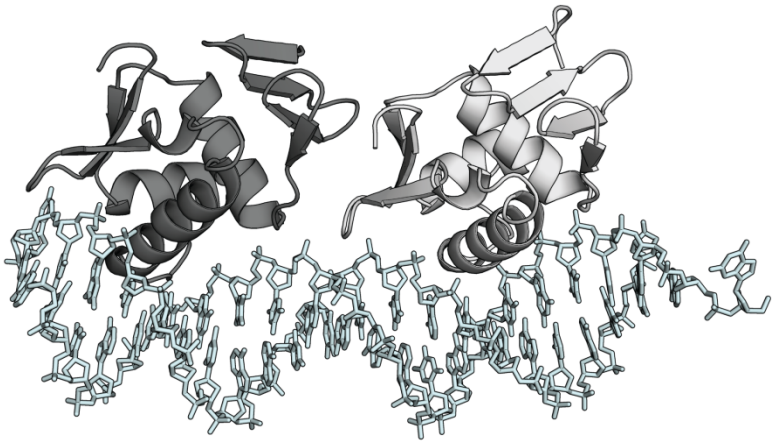
The extensive array of input signals that enter into TCSs also yield a vast number of output responses to mediate adaptation. While the majority of RR effector domains bind to DNA as transcription factors, others participate in RNA-

A



receiver

B



winged-helix effector

Figure 1-5. Structures of response regulator receiver and effector domains. A) Cartoon representation of the structure of the *E. coli* PhoP receiver domain (PDB ID 2PL1). The conserved aspartate residue is shown in teal with the bound phosphoryl group mimic beryllofluoride (BeF_3^-) in pink. A Mg^{2+} ion is represented in green. B) Structure of the DNA-binding effector domains of *E. coli* PhoB bound to pho box DNA (PDB ID 1GXP). The winged-helix domains reveal a head-to-tail tandem arrangement on the DNA. The PhoB domains are represented as cartoons in light and dark grey. The DNA is depicted in cyan sticks.

binding or binding to other proteins, exhibit enzymatic activities, or exist as stand-alone receiver domains that might interact with motor proteins in chemotaxis pathways or participate in phosphorelay systems as intermediates (27).

The effector domain of CpxR is a member of the OmpR subfamily of winged-helix transcription factors that are proposed to bind DNA after a conformational change induced by phosphorylation (42, 52, 53, 79). The OmpR subfamily is the predominant fold adopted by DNA-binding RRs and constitutes 30% of all effector domains (27). Even within the OmpR subfamily, RRs are subjected to diverse regulatory mechanisms. In general, the inactive forms are varied and illustrate different modes of inhibition by the receiver domain. However, upon phosphorylation the receiver domains of OmpR subfamily members form symmetric dimers, bringing the DNA-binding domains within proximity to bind specific direct repeat DNA recognition sequences and regulate transcription (reviewed in (27)). A structure illustrating the recognition of tandem DNA sequences by OmpR subfamily member PhoB is shown in Figure 1-5B (6). The recognition helix of the winged-helix domain contacts the major groove of DNA, while the flanking "wings" make minor groove contacts. Although the regulatory mechanisms that govern the inactive state of the RR differ for each protein, and the exact nature of the CpxR binding sites remain to be determined (70), it is likely that CpxR will bind to DNA in a manner similar to that of other winged-helix DNA-binding proteins.

Accessory molecules: auxiliary regulation in two-component systems

Whereas HKs and their cognate RRs constitute the fundamental elements of TCSs, there are an increasing number of examples of noncanonical external regulators that modulate the output of TCSs by directly or indirectly affecting the levels of phosphorylated RR. These auxiliary factors can act by targeting the

sensing, transmembrane, and catalytic domains of HKs to activate or inhibit the processes of autophosphorylation or RR dephosphorylation, and can affect function of the RR by influencing its phosphorylation state, sequestration or degradation, DNA-binding ability, or the recruitment of RNA polymerase (reviewed in (7, 58)). Accessory proteins may also act as a scaffold to aid in the physical assembly of a pathway (41). Auxiliary regulators can often influence signalling by increasing the range of stimuli that activate a particular response, or interconnecting TCSs in regulatory networks to promote a greater response to a stimulus.

In the Cpx system, there are several examples of external regulatory factors that are mediated through the pathway (7). The first is the TCS connector protein, MzrA. The expression of this small inner membrane-bound protein is upregulated by the CpxAR TCS (100). MzrA specifically interacts with the HK EnvZ, an osmolarity sensor, to either activate its kinase activity or repress its phosphatase activity (30, 31). In this way, MzrA broadens the scope of the response by establishing a regulatory link between the Cpx pathway and the EnvZ/OmpR TCS, which controls the expression of outer membrane porins.

As previously described, CpxP is a periplasmic auxiliary regulator of the Cpx TCS that acts as an inhibitor of the kinase activity of CpxA (75) (see "A role for the novel accessory protein CpxP"). Since the expression of CpxP is upregulated by the Cpx response, CpxP has been suggested to interact with the HK CpxA to facilitate rapid "shut-off" of the pathway (75). Additionally, CpxP may serve to maintain CpxA in an inactive state to reduce signalling noise in the absence of inducing signals (8, 75). CpxP has also been speculated to act as a co-sensor to interact with misfolded proteins (38). While evidence suggests that CpxP interacts with the CpxA sensing domain to repress the Cpx response, a direct interaction has yet to be established. The combination of functions that has been proposed for CpxP is entirely novel in TCS proteins. CpxP possesses no

recognized domains, although two LTXXQ motifs were identified by Pfam (72) (Figure 1-1). No function has yet been attributed to the LTXXQ motif, however in CpxP these regions are known to be important as six *cpxP* loss-of-function mutations were identified in five highly conserved residues in and surrounding the two conserved motifs (8). Elucidation of the structure of CpxP and investigation into its mode of interaction with CpxA will provide molecular insight into the interaction of a unique inhibitory protein with a HK. With reports of new accessory proteins rapidly increasing, it is highly probable that further study of the interactions of these auxiliary factors with TCSs, including the Cpx pathway, will reveal that TCSs and their networks are much more complex than previously imagined.

Model for Cpx pathway activation and signal transmission

Although high-resolution structures are not available for the components of the Cpx system, information extracted from structural studies of highly conserved elements of TCS domains, and data from genetic experiments investigating Cpx pathway induction and output responses can be extrapolated to construct a model for the Cpx stress response in *E. coli* (Figure 1-6).

In the absence of inducing cues, the inner membrane-localized sensor HK, CpxA, acts primarily as a phosphatase to maintain its cognate cytoplasmic RR, CpxR, in an inactive, unphosphorylated state. The autokinase activity of CpxA is inhibited partly due to the presumed interaction of the CpxA sensor domain with the periplasmic inhibitory protein, CpxP (26, 74, 75). The Cpx response is activated by envelope stresses such as alkaline pH, metals, modified membrane lipid composition, bacterial adhesion, and the overproduction of membrane-associated proteins including NlpE and various pilus components (see "Induction of the Cpx stress response"). Many of these inducing cues are predicted to result in misfolded periplasmic proteins or aggregates that localize to the periplasmic

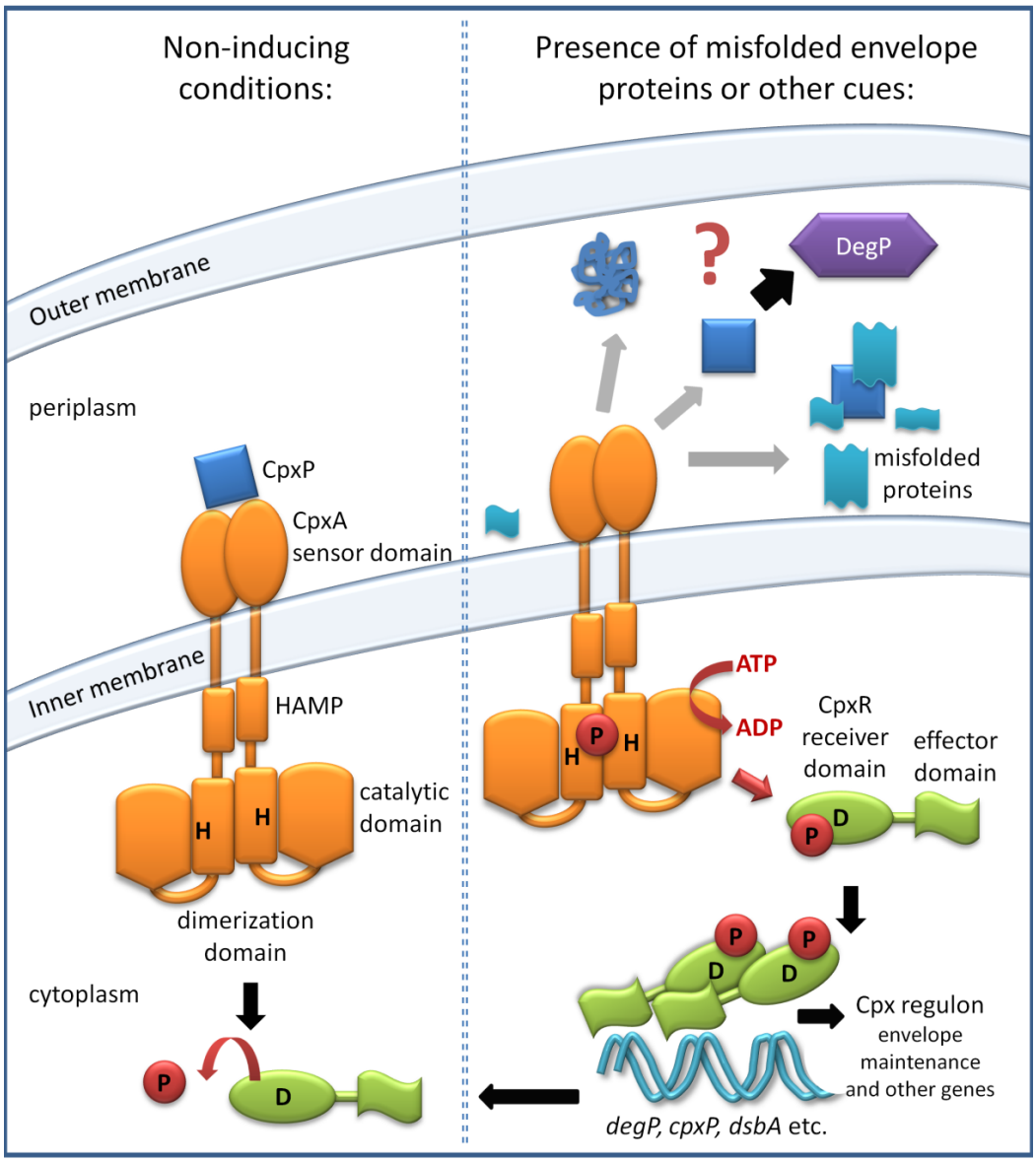


Figure 1-6. Model of the Cpx envelope stress response pathway. Model for activation and signal transduction through the Cpx two-component system. See text for details. "H" and "D" represent the conserved histidine and aspartate residues involved in phosphotransfer, respectively. "P" designates a phosphoryl group.

face of the inner membrane, therefore misfolded proteins and membrane perturbations may represent a unifying signal for Cpx pathway activation.

Upon sensing the stimulus, CpxA is relieved from inhibition by CpxP. The mechanism by which CpxP dissociates from the sensor domain of CpxA is not yet clear. There is evidence to demonstrate that CpxP is degraded by the periplasmic protease, DegP (8, 38), and that CpxP is necessary for DegP-mediated proteolysis of misfolded pilus proteins (38). From these findings it has been hypothesized that denatured periplasmic proteins might bind to CpxP in a way that would result in the dissociation of CpxP from CpxA, thus freeing CpxA from inhibition. CpxP could then act as an adaptor to transport the misfolded proteins to DegP where they would be degraded together (8, 38). Alternatively, it was suggested that the structure of CpxP might be compromised by the same stress that caused the denaturation of other periplasmic proteins, and consequently CpxP would be targeted to DegP for proteolysis (47). In the simplest case, CpxP and misfolded proteins might be degraded independently by DegP in the presence of inducing cues.

Once CpxA is relieved from inhibition, the signal must travel from the periplasmic sensing domain, possibly through a piston-like or other movement of the transmembrane helices (10, 56, 66, 101) and a rotational conformational change in the cytoplasmic HAMP linker (36), to the kinase core domain where its autokinase activity is triggered. The Cpx system is believed to employ the traditional two-component signal transduction mechanism where the catalytic domain utilizes ATP to catalyze the *trans*-phosphorylation of a conserved histidine residue at the interface of the dimerization domains (26, 64, 76, 96). Transduction of the signal continues through recruitment of the receiver domain of CpxR to CpxA and phosphotransfer from the histidine to the conserved aspartate residue in the RR (26, 76, 96). Growth and metabolic signals have been

suggested to be integrated by CpxR in a manner that is independent of CpxA, and these signals may enter the pathway at this point.

The effector domain of CpxR is a member of the OmpR subfamily of RR winged helix-turn-helix transcription factors that are proposed to dimerize and bind to DNA after a conformational change upon phosphorylation (42, 52, 53, 79, 86). Phosphorylated CpxR binds to consensus sequences and possibly to other sites upstream of Cpx-regulated genes to activate or repress their expression (21, 70, 76). Several of the strongly upregulated gene targets in the Cpx regulon encode proteins involved in envelope protein folding and degradation, such as DegP, DsbA, and PpiA, as well as genes controlling pathway amplification and autoregulation, such as *cpxRA* and *cpxP* (see "The Cpx regulon") (47, 70). Other positively- and negatively-controlled regulon members specify inner membrane proteins, peptidoglycan remodelling enzymes, protein complexes such as pili and flagella, inner membrane transporters, and factors that link to other pathways. Ultimately, the Cpx response affects not only the integrity of the envelope, but also influences physiological functions including biofilm formation, virulence, and antibiotic resistance.

Upon removal or lessening of the envelope stress, CpxA functions as a phosphatase to dephosphorylate CpxR, which diminishes the activity of the Cpx pathway (26). CpxP interacts again with the periplasmic sensing domain of CpxA to facilitate "shut-down" of the pathway and maintain the inactive state in the absence of inducing stimuli.

Research objectives

Genetic studies have generated a wealth of information regarding the activating signals and downstream effects of the Cpx envelope stress response. Beginning in the 1980's, the CpxA and CpxR proteins were first identified and

established to be part of a two-component system. Their role in envelope stress was characterized through the 1990's and investigations were further augmented by the discovery of the novel inhibitory protein CpxP. With the detection of numerous TCS interconnecting partners and the elucidation of signalling networks over the last few years, it has become increasingly recognized that these accessory factors are an integral part of TCS signalling, and that the extent and impact of these pathways are broader than previously expected.

Whereas there is a large quantity of molecular data outlining the activities of the Cpx pathway components, there is little information to describe their structure and how this relates to the mechanisms of stimulus detection and signal transduction used specifically by the CpxAR TCS. This work will focus on the periplasmic compartment of *E. coli*, with biochemical and structural studies of the novel accessory protein CpxP and the uncharacterized sensor domain of the HK CpxA.

In Chapter 2, the crystal structure of CpxP is reported and the overall structure is verified using small-angle X-ray scattering studies of CpxP in solution. Biochemical experiments are performed to identify any inherent structural rearrangements or changes in the oligomeric state of CpxP that may occur in response to the pathway inducing cue of alkaline pH. An analysis of previously identified *cpxP* loss-of-function mutations is performed within the context of the crystal structure to give insight into any sites that may be involved in intermolecular interactions. Finally, through evaluation of the CpxP structure and those of newly ascertained structural homologues with potentially related periplasmic roles, the functional significance of the high degree of sequence conservation in the LTXXQ motif is explored.

Biochemical and structural characterization of the periplasmic domain of CpxA is initiated in Chapter 3. Size-exclusion chromatography, multi-angle laser

light scattering, and far-UV circular dichroism experiments are used to investigate the oligomeric state, secondary structure composition, and structural stability of this domain. The evidence suggesting that the CpxA extracytoplasmic domain assumes a PAS-like PDC fold, and exists in a unique tetramer-dimer equilibrium is examined. In addition, the possibility of a conformational change in this domain upon signal perception is considered. Preliminary crystal hits are obtained in the first step towards solving the high-resolution structure of CpxA. The elucidation of a structure will provide insight into the core fold and distinct features of the CpxA periplasmic domain that are specific for its signalling function.

In the final chapter (Chapter 4), recent advances in the field are noted, and the new functions attributed to CpxP and its structural homologues are discussed. Further, the nature of the signal detected by CpxA, and the potential for interactions between CpxA and CpxP, and CpxP and DegP are addressed. Finally, the possible application of the Cpx stress response as a target for antibacterial drugs is considered.

With this work, we attempt to describe, structurally, the molecular details that provide a foundation for signal sensing by the periplasmic domain of the HK CpxA, and its regulation by the novel auxiliary protein CpxP, to offer broader insight into the mechanism by which the Cpx TCS combats stress in the Gram-negative bacterial envelope.

Lastly, the low-resolution single-particle electron microscopy structure of the microRNA cluster, miR-17~92, is described in Appendix A.

References

1. Ades SE. 2008. Regulation by destruction: design of the σ^E envelope stress response. *Curr Opin Microbiol* 11: 535-40

2. Albin R, Weber R, Silverman PM. 1986. The Cpx proteins of *Escherichia coli* K12. Immunologic detection of the chromosomal *cpxA* gene product. *Journal of Biological Chemistry* 261: 4698-705

3. Bachhawat P, Stock AM. 2007. Crystal structures of the receiver domain of the response regulator PhoP from *Escherichia coli* in the absence and presence of the phosphoryl analog beryllofluoride. *J Bacteriol* 189: 5987-95

4. Beloin C, Valle J, Latour-Lambert P, Faure P, Kzreminski M, et al. 2004. Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. *Mol Microbiol* 51: 659-74

5. Bilwes AM, Quezada CM, Croal LR, Crane BR, Simon MI. 2001. Nucleotide binding by the histidine kinase CheA. *Nat Struct Biol* 8: 353-60

6. Blanco AG, Sola M, Gomis-Ruth FX, Coll M. 2002. Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* 10: 701-13

7. Buelow DR, Raivio TL. 2010. Three (and more) component regulatory systems - auxiliary regulators of bacterial histidine kinases. *Mol Microbiol* 75: 547-66

8. Buelow DR, Raivio TL. 2005. Cpx signal transduction is influenced by a conserved N-terminal domain in the novel inhibitor CpxP and the periplasmic protease DegP. *J Bacteriol* 187: 6622-30

9. Chang C, Tesar C, Gu M, Babnigg G, Joachimiak A, et al. 2010. Extracytoplasmic PAS-like domains are common in signal transduction proteins. *J Bacteriol* 192: 1156-9

10. Chervitz SA, Falke JJ. 1996. Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. *Proc Natl Acad Sci U S A* 93: 2545-50

11. Cheung J, Bingman CA, Reyngold M, Hendrickson WA, Waldburger CD. 2008. Crystal structure of a functional dimer of the PhoQ sensor domain. *J Biol Chem* 283: 13762-70

12. Cheung J, Hendrickson WA. 2008. Crystal structures of C4-dicarboxylate ligand complexes with sensor domains of histidine kinases DcuS and DctB. *J Biol Chem* 283: 30256-65

13. Cheung J, Hendrickson WA. 2009. Structural analysis of ligand stimulation of the histidine kinase NarX. *Structure* 17: 190-201

14. Cheung J, Le-Khac M, Hendrickson WA. 2009. Crystal structure of a histidine kinase sensor domain with similarity to periplasmic binding proteins. *Proteins* 77: 235-41

15. Cho US, Bader MW, Amaya MF, Daley ME, Klevit RE, et al. 2006. Metal bridges between the PhoQ sensor domain and the membrane regulate transmembrane signaling. *J Mol Biol* 356: 1193-206

16. Danese PN, Oliver GR, Barr K, Bowman GD, Rick PD, Silhavy TJ. 1998. Accumulation of the enterobacterial common antigen lipid II biosynthetic intermediate stimulates *degP* transcription in *Escherichia coli*. *Jounal of Bacteriology* 180: 5875-84

17. Danese PN, Silhavy TJ. 1997. The σ^E and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. *Genes and Development* 11: 1183-93

18. Danese PN, Silhavy TJ. 1998. CpxP, a stress-combative member of the Cpx regulon. *Journal of Bacteriology* 180: 831-9

19. Danese PN, Snyder WB, Cosma CL, Davis LJ, Silhavy TJ. 1995. The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. *Genes and Development* 9: 387-98

20. De Wulf P, Kwon O, Lin ECC. 1999. The CpxRA signal transduction system of *Escherichia coli*: growth-related autoactivation and control of unanticipated target operons. *Journal of Bacteriology* 181: 6552-778

21. De Wulf P, McGuire AM, Liu X, Lin EC. 2002. Genome-wide profiling of promoter recognition by the two-component response regulator CpxR-P in *Escherichia coli*. *J Biol Chem* 277: 26652-61

22. DiGiuseppe PA, Silhavy TJ. 2003. Signal detection and target gene induction by the CpxRA two-component system. *J Bacteriol* 185: 2432-40

23. Dong JS, Iuchi S, Kwan S, Lu Z, Lin ECC. 1993. The deduced amino-acid sequence of the cloned *cpxR* gene suggests the protein is the cognate regulator for the membrane sensor, CpxA, in a two-component signal transduction system of *Escherichia coli*. *Gene* 136: 227-30

24. Dorel C, Lejeune P, Rodrigue A. 2006. The Cpx system of *Escherichia coli*, a strategic signaling pathway for confronting adverse conditions and for settling biofilm communities? *Res Microbiol* 157: 306-14

25. Emami K, Topakas E, Nagy T, Henshaw J, Jackson KA, et al. 2009. Regulation of the xylan-degrading apparatus of *Cellvibrio japonicus* by a novel two-component system. *J Biol Chem* 284: 1086-96

26. Fleischer R, Heermann R, Jung K, Hunke S. 2007. Purification, reconstitution, and characterization of the CpxRAP envelope stress system of *Escherichia coli*. *J Biol Chem* 282: 8583-93

27. Gao R, Stock AM. 2009. Biological insights from structures of two-component proteins. *Annu Rev Microbiol* 63: 133-54

28. Garcia Vescovi E, Soncini FC, Groisman EA. 1996. Mg^{2+} as an extracellular signal: environmental regulation of Salmonella virulence. *Cell* 84: 165-74

29. Gardino AK, Kern D. 2007. Functional dynamics of response regulators using NMR relaxation techniques. *Methods Enzymol* 423: 149-65

30. Gerken H, Charlson ES, Cicirelli EM, Kenney LJ, Misra R. 2009. MzrA: a novel modulator of the EnvZ/OmpR two-component regulon. *Mol Microbiol* 72: 1408-22

31. Gerken H, Misra R. 2010. MzrA-EnvZ interactions in the periplasm influence the EnvZ/OmpR two-component regulon. *J Bacteriol* 192: 6271-8

32. Goldberg SD, Soto CS, Waldburger CD, Degrado WF. 2008. Determination of the physiological dimer interface of the PhoQ sensor domain. *J Mol Biol* 379: 656-65

33. Gordeliy VI, Labahn J, Moukhametzianov R, Efremov R, Granzin J, et al. 2002. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* 419: 484-7

34. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38 Suppl: W545-9

35. Holm L, Sander C. 1996. Mapping the protein universe. *Science* 273: 595-603

36. Hulko M, Berndt F, Gruber M, Linder JU, Truffault V, et al. 2006. The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell* 126: 929-40

37. Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, et al. 2008. The 20 years of PROSITE. *Nucleic Acids Res* 36: D245-9

38. Isaac DD, Pinkner JS, Hultgren SJ, Silhavy TJ. 2005. The extracytoplasmic adaptor protein CpxP is degraded with substrate by DegP. *Proc Natl Acad Sci U S A*

39. Joly N, Engl C, Jovanovic G, Huvet M, Toni T, et al. 2010. Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. *FEMS Microbiol Rev* 34: 797-827

40. Jones CH, Danese PN, Pinkner JS, Silhavy TJ, Hultgren SJ. 1997. The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems. *EMBO Journal* 21: 6394-406

41. Jung K, Fried L, Behr S, Heermann R. 2012. Histidine kinases and response regulators in networks. *Curr Opin Microbiol* 15: 118-24

42. Kenney LJ. 2002. Structure/function relationships in OmpR and other winged-helix transcription factors. *Curr Opin Microbiol* 5: 135-41

43. Kershaw CJ, Brown NL, Constantinidou C, Patel MD, Hobman JL. 2005. The expression profile of *Escherichia coli* K-12 in response to minimal, optimal and excess copper concentrations. *Microbiology* 151: 1187-98

44. Kulp A, Kuehn MJ. 2010. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu Rev Microbiol* 64: 163-84

45. Langen GR, Harper JR, Silhavy TJ, Howard SP. 2001. Absence of the outer membrane phospholipase A suppresses the temperature-sensitive phenotype of *Escherichia coli degP* mutants and induces the Cpx and σ^E extracytoplasmic stress responses. *J Bacteriol* 183: 5230-8

46. Leblanc SK, Oates CW, Raivio TL. 2011. Characterization of the induction and cellular role of the BaeSR two-component envelope stress response of *Escherichia coli*. *J Bacteriol* 193: 3367-75

47. MacRitchie DM, Buelow, D.R., Price, N.L., Raivio, T.L. 2007. Two-component signaling and Gram negative envelope stress response systems. In *Bacterial signal transduction: network and drug targets*, ed. R Utsumi. Austin, TX: Landes Bioscience

48. Macritchie DM, Ward JD, Nevesinjac AZ, Raivio TL. 2008. Activation of the Cpx Envelope Stress Response Down-regulates Expression of Several LEE-encoded genes in Enteropathogenic *Escherichia coli*. *Infect Immun* 76: 1465-75

49. Magrane M, Consortium U. 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011: bar009

50. Majdalani N, Gottesman S. 2005. The Rcs phosphorelay: a complex signal transduction system. *Annu Rev Microbiol* 59: 379-405

51. Marina A, Waldburger CD, Hendrickson WA. 2005. Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* 24: 4247-59

52. Martinez-Hackert E, Stock AM. 1997. The DNA-binding domain of OmpR: crystal structure of a winged helix transcription factor. *Structure* 5: 109-24.

53. Martinez-Hackert E, Stock AM. 1997. Structural relationships in the OmpR family of winged-helix transcription factors. *J Mol Biol* 269: 301-12

54. Mascher T, Helmann JD, Unden G. 2006. Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev* 70: 910-38

55. McEwen J, Silverman PM. 1980. Chromosomal mutations of *Escherichia coli* that alter expression of conjugative plasmid functions. *Proceedings of the National Academy of Sciences USA* 77: 513-7

56. Milburn MV, Prive GG, Milligan DL, Scott WG, Yeh J, et al. 1991. Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. *Science* 254: 1342-7

57. Mileykovskaya E, Dowhan W. 1997. The Cpx two-component signal transduction pathway is activated in *Escherichia coli* mutant strains lacking phosphatidylethanolamine. *Journal of Bacteriology* 179: 1029-34

58. Mitrophanov AY, Groisman EA. 2008. Signal integration in bacterial two-component regulatory systems. *Genes Dev* 22: 2601-11

59. Moore JO, Hendrickson WA. 2009. Structural analysis of sensor domains from the TMAO-responsive histidine kinase receptor TorS. *Structure* 17: 1195-204

60. Nakayama S-I, Watanabe H. 1995. Involvement of *cpxA*, a sensor of a two-component regulatory system, in the pH-dependent regulation of expression of *Shigella sonnei virF* gene. *Journal of Bacteriology* 177: 5062-9

61. Neiditch MB, Federle MJ, Miller ST, Bassler BL, Hughson FM. 2005. Regulation of LuxPQ receptor activity by the quorum-sensing signal autoinducer-2. *Mol Cell* 18: 507-18

62. Neiditch MB, Federle MJ, Pompeani AJ, Kelly RC, Swem DL, et al. 2006. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell* 126: 1095-108

63. Nevesinjac AZ, Raivio TL. 2005. The Cpx Envelope Stress Response Affects Expression of the Type IV Bundle-Forming Pili of Enteropathogenic *Escherichia coli*. *J Bacteriol* 187: 672-86

64. Ninfa EG, Atkinson MR, Kamberov ES, Ninfa AJ. 1993. Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): *trans*-phosphorylation between subunits. *J Bacteriol* 175: 7024-32

65. Nishino K, Yamasaki S, Hayashi-Nishino M, Yamaguchi A. 2010. Effect of NlpE overproduction on multidrug resistance in *Escherichia coli*. *Antimicrob Agents Chemother* 54: 2239-43

66. Ottemann KM, Xiao W, Shin YK, Koshland DE, Jr. 1999. A piston model for transmembrane signaling of the aspartate receptor. *Science* 285: 1751-4

67. Otto K, Silhavy TJ. 2002. Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. *Proc Natl Acad Sci U S A* 99: 2287-92

68. Pappalardo L, Janausch IG, Vijayan V, Zientz E, Junker J, et al. 2003. The NMR structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) DcuS of *Escherichia coli*. *J Biol Chem* 278: 39185-8

69. Pogliano JA, Lynch S, Belin D, Lin ECC, Beckwith J. 1997. Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. *Genes and Development* 11: 1169-82

70. Price NL, Raivio TL. 2009. Characterization of the Cpx regulon in *Escherichia coli* strain MC4100. *J Bacteriol* 191: 1798-815

71. Prigent-Combaret C, Brombacher E, Vidal O, Ambert A, Lejeune P, et al. 2001. Complex regulatory network controls initial adhesion and biofilm

formation in *Escherichia coli via* regulation of the *csgD* gene. *J Bacteriol* 183: 7213-23

72. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. 2012. The Pfam protein families database. *Nucleic Acids Res* 40: D290-301

73. Quan S, Koldewey P, Tapley T, Kirsch N, Ruane KM, et al. 2011. Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. *Nat Struct Mol Biol* 18: 262-9

74. Raivio TL, Laird MW, Joly JC, Silhavy TJ. 2000. Tethering of CpxP to the inner membrane prevents spheroplast induction of the Cpx envelope stress response. *Molecular Microbiology* 37: 1186-97

75. Raivio TL, Popkin DL, Silhavy TJ. 1999. The Cpx envelope stress response is controlled by amplification and feedback inhibition. *Journal of Bacteriology* 181: 5263-72

76. Raivio TL, Silhavy TJ. 1997. Transduction of envelope stress in *Escherichia coli* by the Cpx two-component system. *Journal of Bacteriology* 179: 7724-33

77. Raivio TL, Silhavy TJ. 1999. The σ^E and Cpx regulatory pathways: overlapping but distinct envelope stress responses. *Current Opinion in Microbiology* 2: 159-65

78. Reinelt S, Hofmann E, Gerharz T, Bott M, Madden DR. 2003. The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. *J Biol Chem* 278: 39189-96

79. Rhee JE, Sheng W, Morgan LK, Nolet R, Liao X, Kenney LJ. 2008. Amino acids important for DNA recognition by the response regulator OmpR. *J Biol Chem* 283: 8664-77

80. Rinker SD, Trombley MP, Gu X, Fortney KR, Bauer ME. 2011. Deletion of *mtrC* in *Haemophilus ducreyi* increases sensitivity to human antimicrobial peptides and activates the CpxRA regulon. *Infect Immun* 79: 2324-34

81. Sevvana M, Vijayan V, Zweckstetter M, Reinelt S, Madden DR, et al. 2008. A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA. *J Mol Biol* 377: 512-23

82. Shimohata N, Chiba S, Saikawa N, Ito K, Akiyama Y. 2002. The Cpx stress response system of *Escherichia coli* senses plasma membrane proteins and controls HtpX, a membrane protease with a cytosolic active site. *Genes Cells* 7: 653-62

83. Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, et al. 2008. Rewiring the specificity of two-component signal transduction systems. *Cell* 133: 1043-54

84. Snyder WB, Davis LJB, Danese PN, Cosma CL, Silhavy TJ. 1995. Overproduction of NlpE, a new outer membrane lipoprotein, suppresses the toxicity of periplasmic LacZ by activation of the Cpx signal transduction pathway. *Journal of Bacteriology* 177: 4216-23

85. Szurmant H, White RA, Hoch JA. 2007. Sensor complexes regulating two-component signal transduction. *Curr Opin Struct Biol* 17: 706-15

86. Tapparel C, Monod A, Kelley WL. 2006. The DNA-binding domain of the *Escherichia coli* CpxR two-component response regulator is constitutively active and cannot be fully attenuated by fused adjacent heterologous regulatory domains. *Microbiology* 152: 431-41

87. Tomomori C, Tanaka T, Dutta R, Park H, Saha SK, et al. 1999. Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. *Nat Struct Biol* 6: 729-34

88. van Stelten J, Silva F, Belin D, Silhavy TJ. 2009. Effects of antibiotics and a proto-oncogene homolog on destruction of protein translocator SecY. *Science* 325: 753-6

89. Vescovi EG, Ayala YM, Di Cera E, Groisman EA. 1997. Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg^{2+} and Ca^{2+}. *J Biol Chem* 272: 1440-3

90. Vogt SL, Nevesinjac AZ, Humphries RM, Donnenberg MS, Armstrong GD, Raivio TL. 2010. The Cpx envelope stress response both facilitates and inhibits elaboration of the enteropathogenic *Escherichia coli* bundle-forming pilus. *Mol Microbiol* 76: 1095-110

91. Vogt SL, Raivio TL. 2012. Just scratching the surface: an expanding view of the Cpx envelope stress response. *FEMS Microbiol Lett* 326: 2-11

92. Waldburger CD, Sauer RT. 1996. Signal detection by the PhoQ sensor-transmitter. Characterization of the sensor domain and a response-impaired mutant that identifies ligand-binding determinants. *J Biol Chem* 271: 26630-6

93. Weatherspoon-Griffin N, Zhao G, Kong W, Kong Y, Morigen, et al. 2011. The CpxR/CpxA two-component system up-regulates two Tat-dependent peptidoglycan amidases to confer bacterial resistance to antimicrobial peptide. *J Biol Chem* 286: 5529-39

94. Wolfe AJ, Parikh N, Lima BP, Zemaitaitis B. 2008. Signal integration by the two-component signal transduction response regulator CpxR. *J Bacteriol* 190: 2314-22

95. Yamada S, Akiyama S, Sugimoto H, Kumita H, Ito K, et al. 2006. The signaling pathway in histidine kinase and the response regulator complex

revealed by X-ray crystallography and solution scattering. *J Mol Biol* 362: 123-39

96. Yamada S, Shiro Y. 2008. Structural basis of the signal transduction in the two-component system. *Adv Exp Med Biol* 631: 22-39

97. Yamada S, Sugimoto H, Kobayashi M, Ohno A, Nakamura H, Shiro Y. 2009. Structure of PAS-linked histidine kinase and the response regulator complex. *Structure* 17: 1333-44

98. Yamaguchi S, Darwin AJ. 2012. Recent findings about the *Yersinia enterocolitica* phage shock protein response. *J Microbiol* 50: 1-7

99. Yamamoto K, Ishihama A. 2005. Transcriptional response of *Escherichia coli* to external copper. *Mol Microbiol* 56: 215-27

100. Yamamoto K, Ishihama A. 2006. Characterization of copper-inducible promoters regulated by CpxA/CpxR in *Escherichia coli*. *Biosci Biotechnol Biochem* 70: 1688-95

101. Yeh JI, Biemann HP, Prive GG, Pandit J, Koshland DE, Jr., Kim SH. 1996. High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. *J Mol Biol* 262: 186-201

102. Zhou YF, Nan B, Nan J, Ma Q, Panjikar S, et al. 2008. C4-dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. *J Mol Biol* 383: 49-61

CHAPTER 2

Structure of the periplasmic regulatory protein, CpxP

A version of this chapter is published in:

Thede, G. L., D. C. Arthur, R. A. Edwards, D. R. Buelow, J. L. Wong, T. L. Raivio, and J. N. M. Glover. 2011. Structure of the periplasmic stress response protein CpxP. J Bacteriol **193**:2149-57.

Introduction

CpxP is a small, 147 amino acid, periplasmic protein with no homologues of known function. The *cpxP* gene was originally identified as an upregulated product of the Cpx envelope stress response pathway in *Escherichia coli* (9). The Cpx response senses and mediates adaptation to a variety of stresses that are thought to result in periplasmic protein misfolding (reviewed in (21, 35, 59)). These include the overexpression of pilus subunits in the absence of their cognate chaperones or assembly machineries (23, 40), as well as alkaline pH (9). A two-component sensor histidine kinase (HK), CpxA, detects inducing signals and conveys this information by means of phosphorylation to the cognate response regulator (RR), CpxR, which directs altered transcription of the Cpx regulon (15, 49). A number of Cpx-regulated genes are directly involved in protein folding or degradation in the bacterial envelope (inner membrane, periplasm, and outer membrane), including the chaperone/protease DegP (8, 10, 35, 43, 46, 59).

CpxP functions as both a signalling protein to control the activity of CpxA and as a proteolytic adaptor to facilitate degradation of some misfolded proteins (11, 22, 47, 48). Overexpression of CpxP leads to downregulation of the Cpx response by inhibition of the autokinase activity of CpxA (11, 15, 47, 48), possibly through direct interaction with the periplasmic sensing domain of CpxA. Inhibition of CpxA activity is relieved in the presence of inducing cues through DegP-mediated degradation of CpxP (5, 22). This step is likely preceded by titration of CpxP from the CpxA sensing domain by either or both misfolded proteins and DegP, as degradation of some misfolded pilus subunits by DegP is dependent upon CpxP (22).

In an effort to understand the molecular mechanism(s) by which CpxP carries out its signalling function, we performed a genetic screen for mutations that negate the ability of CpxP to inhibit CpxA (5). Mutations identified from the

screen affect residues in or near two conserved LTXXQ motifs, a sequence for which no function has been associated that defines a large family of proteins (Pfam protein family database accession number PF07813 [http://pfam.sanger.ac.uk/family?entry=PF07813&type=Family]). Many of these proteins share similarity with CpxP and Spy, the only known homologue of CpxP. In this study, we determined the crystal structure of CpxP to 2.85 Å resolution, performed additional biochemical studies using purified protein, and further characterized the previously identified mutants using *in vivo* techniques. Our results define a role for the conserved LTXXQ motif in the large PF07813 family of proteins and identify several sites on the surface of CpxP that are likely involved in intermolecular interactions.

Materials and Methods

Strains and plasmids

Bacterial strains and plasmids used in this study are described in Table 2-1.

Media, antibiotics, and growth conditions

Strains were grown on Luria-Bertani (LB) agar (50) at 30°C or 37°C. All liquid cultures were grown in LB broth with aeration at either 30°C or 37°C. Strains were maintained with the appropriate antibiotic selection, 100 μg/mL ampicillin (Amp), 25 μg/mL chloramphenicol (Cam), 25 μg/mL Kanamycin (Kn) or 25 μg/mL tetracycline (Tet) (Sigma).

Expression and purification of wild-type CpxP for biochemical assays

Wild-type CpxP was overexpressed as a maltose-binding protein (MBP) fusion from pMCP in TR757, a *degP* null strain (SS1) (Table 2-1). Cells containing

Table 2-1. Bacterial strains and plasmids used in this study

Strain or plasmid	Description	Reference
Strains		
MC4100	F⁻ *araD139* Δ(*argF-lac*)U169 *rpsL150* (Str^R) *relA1 flbB5301 decC1 ptsF25 rbsR*	(6)
TR50	MC4100 λRS88 (*cpxP'-lacZ*$^+$)	(49)
TR61	MC4100 λRS88 (*cpxP'-lacZ*$^+$) *ara*^{R}	This study
TR757	MC4100 λRS88 (*cpxP'-lacZ*$^+$) *degP*::Tn10	(2)
TR930	TR50 (pBBR1MCS)	(5)
TR932	TR50 (pPB)	(5)
DB4	TR757 (pPB)	(5)
DB5	TR757 (pBBR1MCS)	This study
DB31	TR50 (pPB2)	(5)
DB40	TR50 (pPB6)	(5)
DB71	TR757 (pPB2)	(5)
DB75	TR757 (pPB6)	(5)
SS1	TR757 (pMCP)	This study
Plasmids		
pCpxP	*cpxP* overexpression vector (Amp^R)	(48)
pCpxPQ$_{55}$P	*cpxPQ$_{55}$P* overexpression vector (Amp^R)	(5)
PCpxPQ$_{128}$H	*CpxPQ$_{128}$H* overexpression vector (Amp^R)	(5)
pMCP	Over-expresses a functional MBP-CpxP fusion protein (Amp^R)	(48)
pPB	CpxP'-'Bla translational fusion overexpression vector (Amp^R, Cam^R)	(5)
pPB2	pPB encoding a CpxPQ$_{55}$P'-'Bla mutation (Amp^R, Cam^R)	(5)
pPB6	pPB encoding a CpxPQ$_{128}$H'-'Bla mutation (Amp^R, Cam^R)	(5)
p*trc*99A	High expression vector with a multiple cloning site following an IPTG inducible *trc* promoter (Amp^R)	Pharmacia
pCpxP(40-151)	CpxP(40-151) was amplified from pMCP and subcloned into pGEX-6P-1 using BamHI and EcoRI restriction sites.	This study

the pMCP plasmid were grown at 30°C in LB medium with 2 g/L glucose and ampicillin until they reached an $O.D._{600}$ of 0.7, at which point they were induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown overnight at 22°C. Cells were then harvested, resuspended, and osmotically shocked to release the periplasmic fusion protein as per the pMAL™ protein fusion and purification system manual (New England Biolabs). The shock fluid was then applied to amylose resin (New England Biolabs) which was pre-equilibrated with wash buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT)). The MBP-CpxP fusion protein was eluted with elution buffer (wash buffer + 10 mM final concentration maltose), buffer exchanged into 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 1 mM $CaCl_2$, and concentrated to approximately 2 mL using a 10,000 MWCO spin concentrator (Millipore, Fisher Scientific). The concentrated sample was quantitated using the theoretical extinction coefficient at 280 nm of 78 840 M^{-1} cm^{-1} calculated using ProtParam (16) from the primary sequence of MBP-CpxP. Factor Xa protease (GE Healthcare) was then added to 2.5 % w/w ratio Factor Xa:MBP-CpxP and incubated for 16 h at 4°C. The cleavage reaction was monitored by SDS-PAGE and the reaction was stopped with phenylmethylsulfonyl fluoride (PMSF) (Sigma). The cleavage reaction was diluted in wash buffer and re-flowed over the amylose resin to separate CpxP from MBP and residual fusion protein. The CpxP-containing flow through was concentrated using a 5,000 MWCO spin concentrator (Millipore, Fisher Scientific) and applied to a Superdex™ 75 26/60 gel filtration column (GE Healthcare) which was equilibrated in buffer containing 250 mM NaCl, 1 mM EDTA and 50 mM sodium phosphate (pH 5.8 or 8.0). The purified CpxP was subjected to MALDI-TOF mass spectrometry (University of Alberta Chemistry Mass Spectrometry Facility), where it was found to have an average m/z of 17 336 ± 5 Da, which corresponded to the protein having its last two C-terminal amino acids removed. Amino acid analysis (Alberta Peptide

Institute) produced an experimental extinction coefficient at 280 nm of 12 214 $M^{-1}cm^{-1}$ which was subsequently used for protein quantitations.

Size-exclusion chromatography with multi-angle laser light scattering (MALLS)

Fifty microlitres of a 2 mg/mL CpxP sample was injected at 0.5 mL/minute onto a Superose™ 6 HR 10/300 gel filtration column (GE Healthcare) pre-equilibrated with 50 mM sodium phosphate (pH 5.8 or 8.0), 250 mM NaCl, and 1 mM EDTA. After flowing over the column, the effluent was directly passed over an in-line DAWN EOS™ MALLS, an Optilab rEX™ differential refractive index detector (Wyatt Technologies, Santa Barbara, CA). Light scattering data was processed using the ASTRA version 4.90 software. Average molecular weights were calculated from the elution peak. A minimum of two runs were collected at each pH, from which the average and standard deviation were determined.

Formaldehyde-mediated in vivo cross-linking and Western blot analysis

Cross-linking experiments were performed as described in Larsen *et al*. (29), with minimal changes. Cells were subcultured 1:50 in 5 mL fresh media with the corresponding antibiotics and grown until mid log phase ($O.D._{600}$ = 0.6). Cells were harvested and resuspended in the original volume of 0.1 M sodium phosphate buffer (pH 6.8). Fifty microlitres of formaldehyde (37 % w/w) (Fisher) was added to 1 mL sample aliquots. Samples were mixed at room temperature using a Labquake mixer (Fisher) for either 10 or 15 minutes, centrifuged and resuspended in 50 μL of 2X SDS-PAGE loading buffer. Samples were analyzed *via* western blot analysis. Strains were grown overnight at 30°C with aeration. The next day, overnight cultures were diluted 1:50 into 5 mL of LB with the appropriate antibiotic. Cultures were then grown at 37°C for 4 h, and the $O.D._{600}$'s of the cultures were measured. Whole cell extracts were prepared by

pelleting volumes that were adjusted to correspond to 1 mL of the culture that reached the lowest $O.D._{600}$ and resuspended in 50 μL of 2X SDS-PAGE loading dye. The membrane was labelled with polyclonal CpxP antibody (1:25,000), maltose-binding protein antibody (1:50,000), or β-lactamase monoclonal antibody (1:1000 dilution) (QED BioScience Inc.). A secondary anti-mouse antibody conjugated to alkaline phosphatase (Sigma) was used at a 1:25,000 dilution to detect the CpxP-Bla fusion and a secondary anti-rabbit antibody conjugated to alkaline phosphatase (Sigma) was used at a 1:50,000 dilution to detect the CpxP-MBP fusion and native CpxP protein. Proteins were visualized using the Immun-Star™ AP substrate pack chemiluminescent kit (BioRad).

Small-angle X-ray scattering analysis (SAXS)

SAXS curves were collected on Beamline 12.3.1 at the Advanced Light Source, Lawrence Berkeley National Laboratory (Berkeley, California U.S.A.). Data were collected on wild-type CpxP at 7.5 mg/mL dialyzed against 250 mM NaCl, 50 mM sodium phosphate (at either pH 5.8 or 8.0), 1 mM EDTA, 10 % glycerol. CpxP exhibited no radiation-induced damage as determined by comparison of a sequential series of exposure times, two 6 s, a 60 s and a final 6 s exposure (data not shown). Guinier plots (data not shown) plotted in PRIMUS (26) were linear in the range $s.Rg < 1.3$, an approximation that holds true for non-aggregated samples. The radii of gyration (*Rg*), calculated from the scattering data using the Guinier plots, were 23.3 ± 0.1 and 22.3 ± 0.1 Å for the protein at pH 5.8, and 8.0, respectively. The distance distribution function, a transformation from reciprocal to real-space, calculated using the program GNOM (51) gave mean estimates of the maximum particle dimension (D_{max}) of 66 and 60 Å for all four exposures in the series for the pH 5.8 and 8.0 samples, with corresponding *Rg* of 22.8 and 21.9 Å, respectively. The program GASBOR (52) was used to generate an *ab initio* low-resolution envelope of wild-type CpxP

at pH 5.8. The envelope was constructed from the 60 s exposure using 15 models from independent GASBOR simulations having P2 symmetry imposed. The resulting models were superimposed and averaged with DAMAVER and filtered with DAMFILT (60). The X-ray model was docked into the SAXS envelope at pH 5.8 using SITUS (62).

Circular dichroism (CD)

CD experiments were performed on a Jasco J720 spectropolarimeter. Near-UV (320-255 nm) experiments used 1 mg/mL samples of CpxP in a 1 cm path length thermostated fused silica cell at 24°C. CpxP samples were in 50 mM sodium phosphate pH (5.8 or 8.0), 250 mM NaCl, 1 mM EDTA. Raw ellipticities were subtracted from the buffer and converted into mean residue ellipticities using the average amino acid weight of 117.1 Da, calculated from the CpxP primary sequence. Near-UV spectra were an average of 12 scans. For melting temperature (T_m) experiments, ellipticity was recorded at 220 nm for five minutes at each temperature, measuring a reading every second. From the data, an average ellipticity at each temperature was calculated. CpxP samples were at a concentration of 0.25 mg/mL in 250 mM NaCl, 1 mM EDTA, and 50 mM sodium phosphate (pH 5.8 or 8.0) in a 0.05 cm silica cell. The samples were equilibrated for five minutes at each temperature before data collection. Temperatures measured ranged from 24 to 62°C, employing a Lauda water bath (Brinkmann Instruments) to control the temperature of the cell. The loss of ellipticity at 220 nm indicated unfolding of CpxP, and the T_m is defined as the temperature at which the ellipticity at 220 nm decreases by half. Melting temperatures were calculated using the SigmaPlot 2001 software (SPSS Inc.). T_m values represent an average of two experiments.

Limited proteolysis of wild-type CpxP

Five micrograms of purified wild-type CpxP in 50 mM sodium phosphate pH 7.0, 250 mM NaCl, and 1 mM EDTA was digested with 0, 5 x10^{-4}, 0.001, 0.005, 0.01, 0.05, and 0.1 µg of trypsin (Sigma-Aldrich) for 60 minutes at 37℃. Reactions were stopped with PMSF and the reaction products were separated on an 18 % SDS-PAGE gel. MALDI-TOF mass spectrometry (University of Alberta Chemistry Mass Spectrometry Facility) was used to identify the molecular mass of the two digestion products from a parallel trypsin digest of CpxP at a 1/100 w/w ratio of trypsin:CpxP. The two fragments had average m/z of 15 990 and 13 336 ± 5 Da. Analysis using the program MS-Digest (ProteinProspector suite; University of California, San Francisco) showed that the fragments corresponded to residues 20-151 and 40-151, respectively.

Cloning and purification of recombinant CpxP for structure determination

The sequence corresponding to the CpxP(40-151) proteolytic fragment was amplified from full-length *cpxP* in the pMCP vector (Table 2-1) using the PCR primers 5'(5'-GGCGGCGGATCCAGTACGCAGAGCCATATGTTCGACGGCATAAGTTTA) and 3' (5'-GGCGGCGAATTCCTATTTTTGCCATTGCGTCACGTCACG). The fragment was ligated into the BamHI and EcoRI restriction sites of the pGEX-6P-1 expression vector (GE Healthcare), which contains N-terminal GST tag and PreScission™ Protease cleavage site sequences. The subclones were confirmed by DNA sequencing (The Applied Genomics Centre, University of Alberta). The CpxP(40-151)-containing plasmid was transformed into *E. coli* BL21(DE3) cells. Cells were cultured at 37°C in Luria-Bertani medium with 100 µg/mL ampicillin. Upon reaching an optical density at 600 nm of 0.8-0.9, CpxP overexpression was induced with 0.05 mM IPTG, and the cells were grown for a further 20 h at 30°C. CpxP(40-151) was purified using a protocol modified from the GST Gene Fusion System Handbook (GE Healthcare). Cells were harvested by centrifugation and

the cell pellet was resuspended in buffer composed of 50 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 5.8 and 150 mM NaCl, plus protease inhibitors. Cells were lysed by chicken egg white lysozyme and sonication, and the lysate was fractionated by centrifugation. The clarified lysate was applied to Glutathione Sepharose™ 4B resin (GE Healthcare) and the captured GST fusion protein was eluted using resuspension buffer containing 20 mM reduced glutathione. The elutions were pooled, buffer exchanged into resuspension buffer, and concentrated to a small volume (~2 mL) using a 10,000 MWCO centrifugal filter device (Millipore, Fisher Scientific). The GST tag was removed by PreScission™ Protease (GE Healthcare) over 16 h at 4°C, and the cleavage reaction was applied to a second GST affinity column to remove proteolyzed GST and any remaining fusion protein. The flow-through was concentrated using a 5,000 MWCO spin concentrator (Millipore, Fisher Scientific). Size-exclusion chromatography was performed as a final purification step using a HiLoad Superdex™ 75 26/60 gel filtration column equilibrated in resuspension buffer on an ÄKTA™purifier (GE Healthcare). Fractions were pooled and the protein was dialyzed into storage buffer containing 50 mM MES pH 5.8 and 50 mM NaCl, and concentrated by ultrafiltration using a 5,000 MWCO spin concentrator. The protein concentration was estimated from the absorbance at 280 nm using the theoretical extinction coefficient of 6990 $M^{-1}cm^{-1}$, which was derived from the amino acid sequence of CpxP(40-151) using ProtParam (16). Selenomethionine (Se-Met)-substituted CpxP(40-151) was produced in cells grown in M9 minimal medium supplemented with amino acids and seleno-L-methionine (Sigma-Aldrich). The Se-Met-labeled protein was purified using the same methods as above, with resuspension buffer composed of 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, and 0.1 % β-mercaptoethanol, and storage buffer consisting of 50 mM HEPES pH 7.0, 50 mM NaCl, and 1 mM DTT.

Crystallization and data collection

Initial crystallization conditions for CpxP(40-151) were identified from the Emerald Wizard II suite, N11 (Emerald Biosystems) and were further optimized. Se-Met-substituted CpxP(40-151) was crystallized by hanging drop vapour diffusion at room temperature (~21°C) by mixing 1 µL of CpxP at 24 mg/mL in the final storage buffer with 1 µL of reservoir solution containing 9 % (v/v) 2-propanol, 0.1 M sodium cacodylate pH 5.4, 0.7 M $Zn(OAc)_2$, and 1 mM DTT. Native protein crystallized in similar conditions by mixing 2 µL of CpxP at 37 mg/mL in the final storage buffer with 1.5 µL of reservoir solution containing 9 % (v/v) 2-propanol, 0.1 M sodium cacodylate pH 5.6, and 0.7 M $Zn(OAc)_2$). No additional cryoprotectant was required and the crystals were flash frozen in liquid nitrogen. All diffraction data were collected at the Canadian Light Source, beamline CMCF 08ID-1 (Saskatoon, Canada). A multi-wavelength anomalous dispersion (MAD) data set was collected from the Se-Met-substituted crystal at three wavelengths (peak, 0.97945 Å; inflection, 0.97961 Å; and remote, 0.97754 Å). Zinc single-wavelength anomalous dispersion (SAD), and native data sets were collected from the native crystal at the zinc peak, 1.28197 Å, and remote, 0.97949 Å, wavelengths, respectively.

Structure solution and refinement

All diffraction data sets were processed and integrated with DENZO, and scaled in the space group $P6_122$ using SCALEPACK (HKL2000 suite) (42). The structure of CpxP was solved by the MAD method using the selenomethionine derivative data sets processed to 2.9 Å resolution. The heavy atom positions for 15 of the 16 expected selenium atoms (given two molecules in the asymmetric unit) were determined and used to obtain preliminary phases in SOLVE (56). Density modification of phases and automatic model building into the calculated electron density map were performed by the program RESOLVE (54, 55).

However, the partial models obtained from automatic building needed to be rebuilt manually using both 2.9 Å- and 2.6 Å-resolution remote $2F_o$-F_c and F_o-F_c electron density maps, and the known locations of the selenium atoms to place the CpxP sequence into the maps in COOT (14). The starting model was completed by iterative rounds of model building and refinement using REFMAC5 in the CCP4 suite (7, 39, 45, 61). Although it was speculated from large positive difference densities remaining in the maps and peaks in the difference Fouriers that zinc ions could be present, placing these in the model did not aid the refinement. Therefore, additional data sets were collected from a single native crystal at both the zinc peak, and a remote wavelength (Table 2-2). SAD phases were determined for the zinc peak data using SOLVE (56), and zinc sites were located. Further model building and refinement was carried out using COOT (14) and PHASER (37) against the 2.85 Å data collected on the native CpxP crystal at the remote wavelength. Nine zinc sites identified from the zinc SAD solution were added to the model. Geometry restraints were applied to zinc sites where possible using ideal values obtained from Tamames *et al.* (53). One translation libration screw (TLS) group was refined for each of the two chains in the later stages of refinement using PHENIX (1). Although CpxP is a dimer, the application of non-crystallographic symmetry (NCS) was not beneficial at any stage of refinement. The final model consisted of amino acid residues 44-151 of protomer A, and 40-151 of protomer B. The four N-terminal residues of protomer A were not observed in the electron density map as a result of disorder. The R_{work} and R_{free} of the final model were 24.7 %, and 29.8 %, respectively. Analysis of the final model with PROCHECK (30) revealed 89.4 % of residues in the most favoured regions of the Ramachandran plot, and 9.6 % in additional allowed regions with no residues in the generously allowed region. Two residues, corresponding to Leu51 from each protomer, were found to be in disallowed regions. However, these residues form a critical part of the conserved LTXXQ motif and pack in a hydrophobic region under the turn for

Table 2-2. Data collection and refinement statistics

	CpxP(40-151) Native (Remote)	CpxP(40-151) Native (Zinc Peak)
Data collection[1]		
Space group	$P6_122$	$P6_122$
Unit cell parameters		
a=*b*, *c* (Å)	88.03, 134.93	88.10, 135.05
α=β, γ (°)	90.0, 120.0	90.0, 120.0
Wavelength (Å)	0.97949	1.28197
Resolution (Å)	41.85-2.85 (2.95-2.85)	44.05-3.20 (3.31-3.20)
R_{sym}[2]	0.042 (0.489)	0.045 (0.510)
I / σ(I)	34.2 (3.1)	27.7 (3.2)
Completeness (%)	99.8 (100.0)	100.0 (100.0)
Redundancy	9.9 (5.9)	8.9 (9.0)
Refinement		
Resolution (Å)	44.0-2.85	
No. of reflections[3]	13 092 (588)	
R_{work} / R_{free}[4]	0.247 / 0.298	
No. of atoms		
Protein	1834	
Zinc ions	9	
Water	15	
B-factors		
Protein	90	
Zinc ions	104	
Water	62	
RMSD		
Bond lengths (Å)	0.015	
Bond angles (°)	1.183	
Ramachandran statistics (%)		
Most favoured regions	89.4	
Additional allowed regions	9.6	
Generously allowed regions	0.0	
Disallowed regions	1.0	

[1] Data collection statistics for the highest resolution shell are shown in parentheses.

[2] $R_{\text{sym}} = \sum \left| I - \langle I \rangle \right|^2 / \sum \left| I^2 \right|$

[3] Value in parentheses represents the number of reflections used in the calculation of R_{free}.

[4] $R = \sum \left\| F_o \right| - \left| F_c \right\| / \sum \left| F_o \right|$

which there is good electron density. The data collection and refinement statistics are summarized in Table 2-2.

The dimeric interface of CpxP was analyzed using PISA (27). The electrostatic surface potential was computed and visualized using the PDB2PQR server (12, 13) in conjunction with the APBS plugin in The PyMOL Molecular Graphics System, Version 1.2 (Schrodinger, LLC) (4, 33). The two CpxP protomers were superpositioned, and the root mean square deviation (RMSD) between them was calculated using the maximum-likelihood alignment method in THESEUS (57). The secondary structure in the atomic model was delineated by DSSP (3, 24). The structural coordinates of CpxP were compared to those of all structures in the Protein Data Bank (PDB) in a search performed by the Dali server (19, 20). Figures were generated using PyMOL.

β-*Galactosidase assays*

β-Galactosidase assays were performed in triplicate as described previously by Buelow & Raivio (5).

Protein structure accession number

The atomic coordinates and structure factors for CpxP(40-151) have been deposited in the RCSB Protein Data Bank under accession code 3QZC.

Results

CpxP forms a dimer that undergoes a subtle structural rearrangement in response to alkaline pH

In order to determine the oligomeric state of wild-type *Escherichia coli* CpxP, size-exclusion chromatography was performed in conjunction with multi-angle laser light scattering (MALLS) at pH 8.0, a condition where the Cpx pathway is activated, and pH 5.8, where the pathway is repressed (9). The elution volume of CpxP from the gel filtration column was found to be independent of pH (Figure 2-1A). Additionally, light scattering data established that at both pHs CpxP formed a dimer (pH 5.8, MW of 34 400 ± 500; pH 8.0, MW of 33 800 ± 1000), which indicates that CpxP does not undergo a change in oligomeric state in response to the inducing cue of alkaline pH *in vitro*. To establish the oligomeric state of CpxP *in vivo*, wild-type CpxP or various CpxP fusion proteins were overexpressed and cross-linked with formaldehyde. In the absence of cross-linker, monomers of the expected molecular weights for MBP'-'CpxP, CpxP'-'Bla and CpxP were observed (Figure 2-1B, lanes 1). Within 10 min of incubation with formaldehyde, major cross-reactive protein species were present at 114 kDa, 98 kDa, and 34 kDa, which are consistent with double the monomeric weight of MBP'-'CpxP, CpxP'-'Bla, and CpxP, respectively (Figure 2-1B, lanes 2). This experiment suggests that *in vivo* CpxP exists as a dimer.

Since size-exclusion chromatography and MALLS experiments did not reveal a change in the oligomeric state of full-length CpxP in response to the pathway-inducing cue of alkaline pH (Figure 2-1A), small-angle X-ray scattering (SAXS) and circular dichroism (CD) spectroscopy methods were used to investigate whether the structure of CpxP was altered under the same conditions (Figure 2-2A-D). The values for the radius of gyration (*Rg*) and maximum particle dimension (D_{max}) from SAXS revealed a small but reproducible compression of the structure at pH 8.0 compared to pH 5.8 (Figure 2-2B). A small increase in the protein fold stability was also detected through a slight rise in the melting temperature (T_m) of CpxP at the higher pH as measured by CD (Figure 2-2C). Near-UV CD experiments (320-255 nm) identified variation in the chemical environment of one or both tryptophan residues in CpxP (W30 and W149) with a

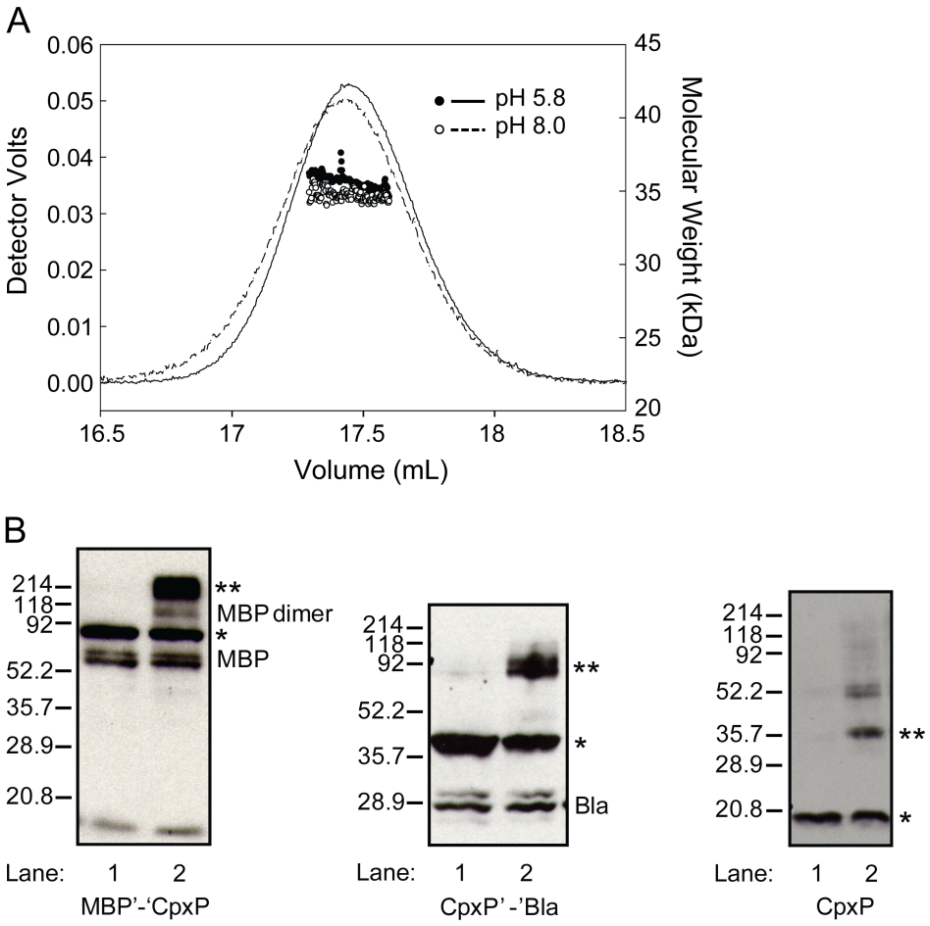
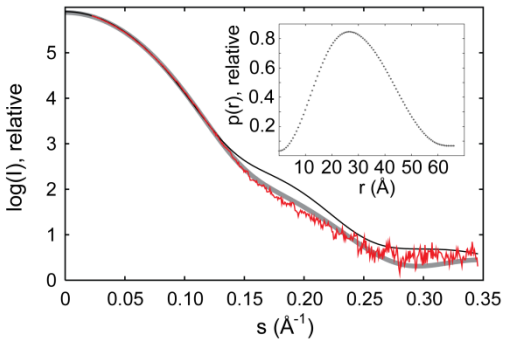


Figure 2-1. CpxP forms a dimer *in vitro* at both pH 5.8 and 8.0, and *in vivo*. A) Size-exclusion chromatography and multi-angle laser light scattering determination of the molecular weight of CpxP at pH 5.8 and 8.0. Light scattering over the elution profile of CpxP shown in solid (pH 5.8) and dashed (pH 8.0) lines. Superimposed across the peaks, filled (pH 5.8) and empty (pH 8.0) circles indicate molecular weights as a function of elution volume. B) Western blots performed on whole-cell lysates of wild-type strains expressing different CpxP fusion proteins that had been cross-linked with 1 % v/v formaldehyde for 10 minutes. From left to right: cross-linking of MBP'-'CpxP (DB345), CpxP'-'Bla (DB359), and native CpxP (DB239) expressed from the ptrc promoter. Lane 1 represents uncross-linked sample, while lane 2 represents cross-linked sample after 10 minutes in the presence of formaldehyde. Single asterisks indicate the position of the monomeric protein, double asterisks indicate the position of the dimer. Each blot was performed three times and a representative blot is shown. A fainter band at approximately 51 kDa was also detected in the cross-linked sample derived from cells over-expressing the native CpxP (panel 3, lane 2), which may correspond to a trimer.

A



B

pH	*Rg* (Guinier)	*Rg* (GNOM)	D_{max}
5.8	23.3 ± 0.1 Å	22.8 Å	66 Å
8.0	22.3 ± 0.1 Å	21.9 Å	60 Å

C

pH	Melting Temperature
5.8	48.3 ± 0.2 °C
8.0	51.6 ± 0.2 °C

D

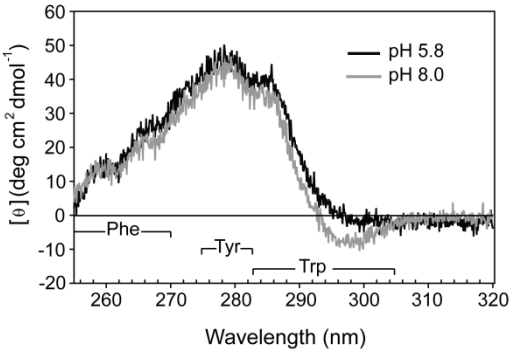


Figure 2-2. A small structural rearrangement occurs in CpxP in response to the inducing cue of alkaline pH. A) Small-angle X-ray scattering curve of wild-type CpxP at pH 5.8, 60 s exposure (red). Overlaid are the back-calculated theoretical scattering curves of the CpxP X-ray crystal structure (thin black line) and a model CpxP dimer constructed from chain B of the X-ray structure (thick grey line). The chain B-chain B dimer model provides a better fit (χ^2 = 1.8) to the solution SAXS data than the X-ray crystal structure (χ^2 = 4.2) by effectively removing the crystal-pack constrained orientation of α4 from chain A. Inset, distance distribution function, p(r), calculated using GNOM. The symmetrical bell-shaped curve is typical of a folded, globular particle. B) Summary of values for the radius of gyration (*Rg*) calculated using both the Guinier approximation and GNOM, and the maximum dimension of the particle (D_{max}) for CpxP at pH 5.8 and pH 8.0. C) Summary of the melting temperatures (T_m) for CpxP at pH 5.8 and 8.0, calculated from the temperature at which the average ellipticity at 220 nm is decreased by half. D) Near-UV CD spectra of CpxP at pH 5.8 (black) and 8.0 (grey). The Phe, Tyr and Trp signal wavelength ranges are indicated. Each of the CD experiments was repeated in duplicate and representative spectra are shown.

change in pH from 5.8 to 8.0 (Figure 2-2D). Cumulatively, these results suggest that CpxP undergoes slight conformational adjustment to a more compact and stable form upon a shift from pH 5.8 to 8.0.

CpxP is a curved α-helical dimer with a highly basic concave face

Wild-type CpxP was subjected to limited proteolysis to identify an ordered protein domain suitable for crystallization (Figure 2-3). Mass spectrometry revealed that the smallest stable domain corresponded to residues 40-151, which was sufficient to inhibit the Cpx response (not shown). CpxP(40-151) was purified and crystallized, and the structure was determined to 2.85 Å resolution using selenomethionine multi-wavelength anomalous dispersion and zinc single-wavelength anomalous dispersion methods (Table 2-2).

The crystal structure revealed that CpxP is a largely helical dimer (Figure 2-4A). The two protomers are intertwined in an antiparallel arrangement with an extensive, buried interface of ~1600 $Å^2$. The dimer adopts an elongated, bowl-shaped structure, with a highly basic interior concave surface and a rather acidic convex outer surface (Figure 2-4B). Each protomer can be described as a long, bent, and hooked hairpin made up of four α-helices (Figure 2-4A). Helices α1 and α2 combine to span the length of the protein, and are separated by a short region of random coil that permits a kink between the two helices. After a hairpin turn, helix α3 returns along the length of the protein, approximately parallel to α1 and α2 and with a bend opposite the coil region between α1 and α2. Together, these three helices form the bent hairpin fold. After another short turn, α4 runs alongside the N-terminus and completes the protein, with the N-terminus and α4 comprising the hooks on the end of the hairpin.

The two protomers are similar to each other in overall structure yet are distinct, with an overall root mean square deviation (RMSD) of 3.1 Å between

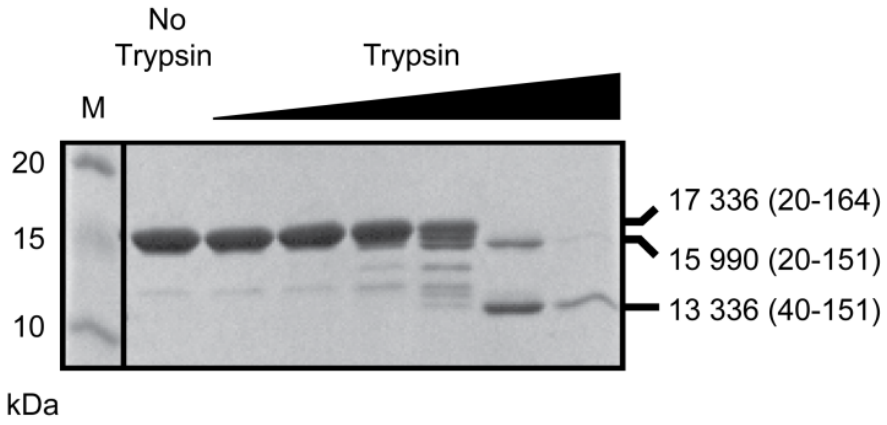


Figure 2-3. Limited proteolysis of wild-type CpxP. Reactions contained 5 µg of purified CpxP with 0, $5x10^{-4}$, 0.001, 0.005, 0.01, 0.05, and 0.1 µg of trypsin. The molecular masses and corresponding residues of full-length CpxP minus the two C-terminal residues that are cleaved during protein purification (17 336 Da; residues 20-164) and its digested fragments (15 990 Da; residues 20-151 and 13 336 Da; residues 40-151) as determined by MALDI-TOF mass spectrometry are shown on the right. Molecular weight markers are indicated in the left-most lane.

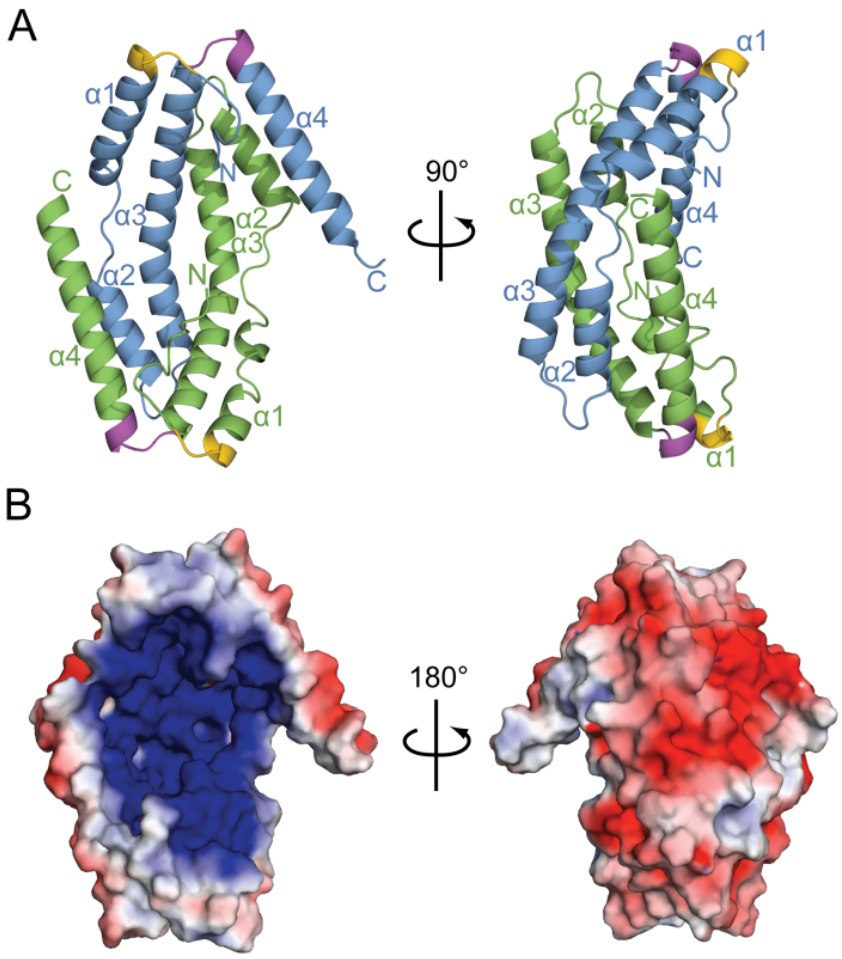


Figure 2-4. The structure of CpxP reveals an α-helical dimer with a highly basic surface. A) Cartoon representation of the structure of CpxP(40-151), shown in two orientations that are rotated by 90° about the vertical axis. Chain A is depicted in blue, while chain B is green. The conserved N- and C-terminal LTXXQ motifs are displayed in gold and purple, respectively. Helices α1-4 and the N- and C-termini are labeled. B) Electrostatic potential distribution on the molecular surface of the CpxP dimer. The concave surface is highly basic (dark blue), while the convex surface (180° rotation around the vertical axis) is mainly acidic (red). Uncharged regions are shown in white.

the C_α atoms of the two chains. Most of the variation occurs in the coil region between helices α1 and α2, the angle of offset of α4, and the N- and C-termini of the protein (Figure 2-5A). The divergence between the protomers in all of these regions is likely due to distortion caused by the close packing of a symmetry-related molecule to the coil region and N-terminus of chain B, and to helix α4 and the C-terminus of chain A. The deformation of these segments is correlated with high *B*-factors for the affected atoms, which are indicative of high thermal motion or flexibility. The crystal structure was found to be in good agreement with data and an *ab initio* molecular envelope generated from SAXS experiments of CpxP in solution, taking into account the distortion due to the crystal pack (Figure 2-2A and Figure 2-5B and C). The consistency between the X-ray and SAXS structures demonstrates that the crystal structure is representative of the CpxP dimer in solution.

Mutations in and near the conserved LTXXQ motifs disrupt CpxP regulatory function

The crystal structure of CpxP revealed that the two highly conserved LTXXQ motifs are situated proximal to each other at one end of the monomer, with the beginning of the first motif (residues 51-55) positioned before α1 and the second motif (residues 124-128) between α3 and α4 (Figure 2-4A and Figure 2-6A). Each LTXXQ motif forms a diverging turn, stabilized by the conserved leucine (L51 or L124) that packs in a hydrophobic region under the turn, and the conserved threonine (T52 or T125), that caps the N-terminus of the following α-helix (Figure 2-6B and C). Hydrogen bond interactions between the conserved glutamine residues of the N- and C-terminal LTXXQ motifs (Q55 and Q128) link the two motifs to one another and likely stabilize the overall structure of the protomer (Figure 2-6B and D).

All six of the *cpxP* loss-of-function mutations previously identified (5)

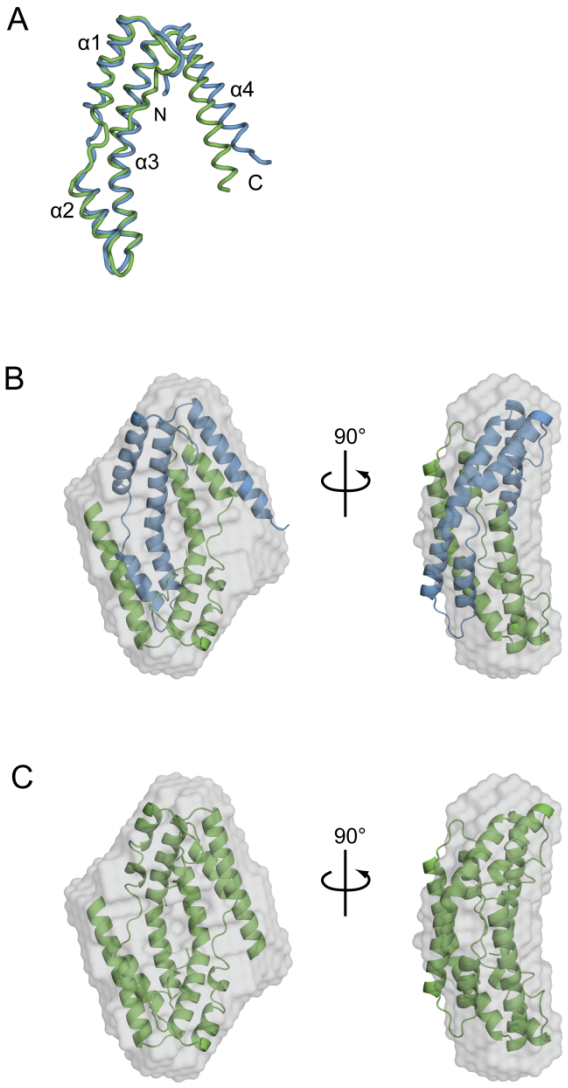


Figure 2-5. Comparison of the two CpxP protomers to the solution structure. A) Ribbon representation of the two superpositioned CpxP protomers, aligned with an overall root mean square deviation (RMSD) of 3.1 Å between C_α atoms of the two chains. Helices α1-4 and the N- and C-termini are labeled. B) Overlay of the CpxP(40-151) X-ray structure in cartoon representation with the SAXS molecular envelope from wild-type CpxP at pH 5.8. The low-resolution solution structure envelope was calculated *ab initio*, from an average of 15 independent GASBOR models. C) Overlay of a model CpxP dimer generated from two protomers of chain B from the X-ray structure with the SAXS molecular envelope from wild-type CpxP at pH 5.8. Panels B) and C) illustrate the better fit of the modeled chain B-chain B dimer ($\chi 2$ = 1.8) to the SAXS data than the experimental chain A-chain B dimer ($\chi 2$ = 4.2), due to distortion caused by the crystal pack. B) and C) are shown in two orientations that are rotated by 90° about the vertical axis. Chain A is portrayed in blue, while chain B is green, and the molecular envelopes from SAXS are shown in grey.

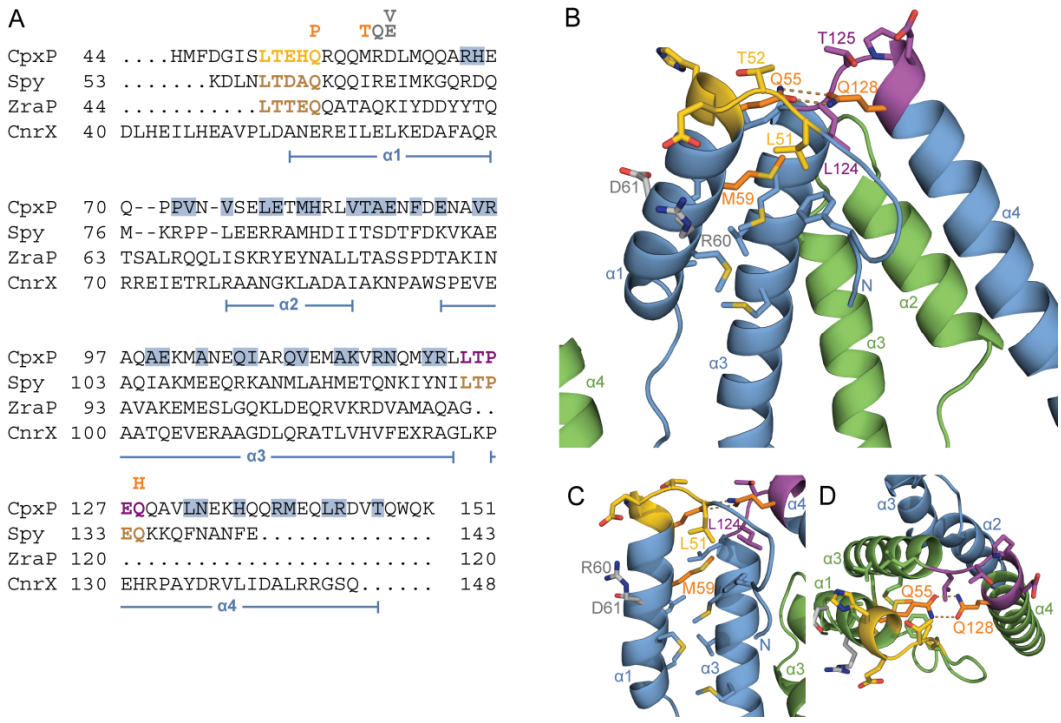


Figure 2-6. Loss-of-function mutations are located in and near the highly conserved LTXXQ motifs of CpxP. A) Structure-based sequence alignment of CpxP with Spy, ZraP, and CnrX, generated using Dali (19, 20) with numbering from the full protein sequences (36). The conserved N- and C-terminal LTXXQ motifs, located at residues 51-55 and 124-128 of CpxP, are shown in gold and purple, respectively. The LTXXQ motifs of Spy and ZraP are coloured in brown. Residues of chain A that contribute to the dimerization interface of CpxP are highlighted in blue, while the locations of helices α1-4 are illustrated below the alignment. Three of the six previously identified *cpxP* loss-of-function mutations (5), Q55P, M59T, and Q128H, which are predicted to impact the structure or stability of CpxP, are listed in orange above the CpxP sequence. The three remaining mutations, R60Q, D61E, and D61V, that are predicted to have functional implications for CpxP signalling or chaperone activity, are indicated above the CpxP sequence in grey. B-D) Cartoon representation of the CpxP structure illustrating the locations of the conserved LTXXQ motifs and residues conveying loss-of-function mutations. C) Hydrophobic region including L51 and M59. D) Hydrogen bonds between the conserved glutamine residues of the N- and C-terminal LTXXQ motifs (Q55 and Q128). Colour-coding is the same as for A) with chain A of CpxP in blue and chain B in green. Residues of the LTXXQ motifs, mutated residues, and residues of the hydrophobic region surrounding L51 and M59 are shown as sticks. Prominent residues, helices α1-4, and the N- and C-termini are labeled.

cluster to the N- and C-terminal LTXXQ motifs and a segment of α1 adjacent to the N-terminal LTXXQ motif (Figure 2-6A-D). Three of these mutations, M59T, Q55P, and Q128H, resulted in reduced levels of CpxP, which were returned to near normal levels in a *degP*-null strain (5). M59 resides in the α1 helix and forms part of a hydrophobic protein core beneath the N-terminal LTXXQ motif (Figure 2-6A-C). The introduction of the smaller, hydrophilic threonine side chain at this position in the M59T mutant would likely disrupt the hydrophobic packing of the α1 and α3 helices, thereby destabilizing the protein fold. Mutation of the conserved glutamine residue within each LTXXQ motif (Figure 2-6A, B and D) to Q55P or Q128H would eliminate the double hydrogen bond that links the two motifs and would therefore reduce the stability of the mutant protein. Additionally, the substitution to P55 may impose a conformational constraint on the protein backbone that might disrupt the α1 helix. Thus, it is likely that each of these three mutations lead to a loss of function through a reduction in protein stability.

Since levels of the CpxP(Q55P) and CpxP(Q128H) proteins were restored in the absence of DegP (5), we tested whether these mutations affected CpxP function in a *degP* mutant background. The ability of CpxP(Q55P) and CpxP(Q128H) to inhibit CpxA in both wild-type and *degP* mutant strains was examined. Plasmids overexpressing C-terminal β-lactamase fusions (previously shown not to affect CpxP function (5)) to either the wild-type, CpxP(Q55P), or CpxP(Q128H) proteins were transformed into wild-type and *degP*-null strains bearing a Cpx-regulated *lacZ* reporter gene. β-Galactosidase activity was measured as an indicator of Cpx transcriptional activation. As previously observed, overexpression of wild-type CpxP leads to a dramatic decrease in Cpx pathway activity in both strains (Figure 2-7, compare lanes 1 and 2 to 3 and 4). In contrast, the Q55P mutant protein was unable to inhibit Cpx pathway activity in either the wild-type strain background, where the Q55P mutant protein is degraded by DegP, or the *degP* mutant strain, where the mutant CpxP protein

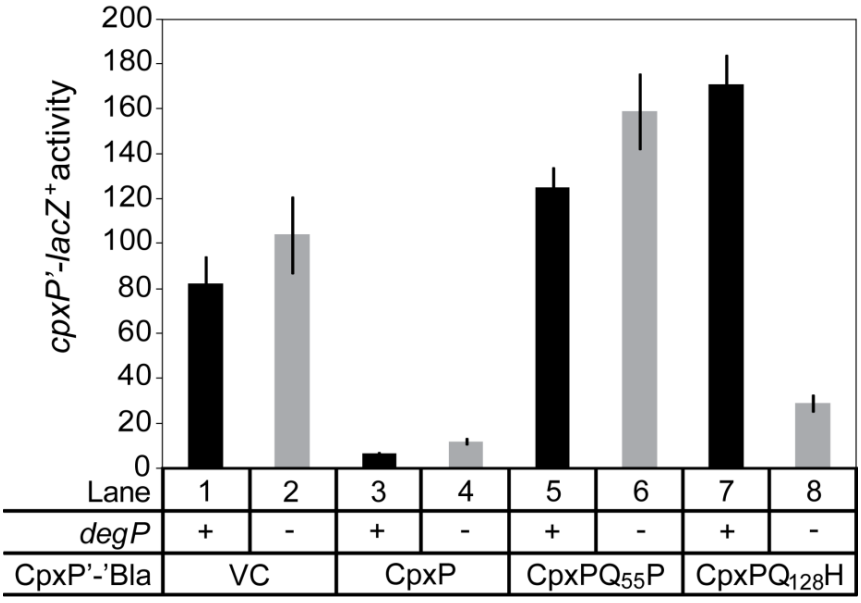


Figure 2-7. A CpxP(Q128H) mutant regains signalling activity in a *degP* mutant background. A chromosomal *cpxP'-lacZ*$^{+}$ fusion was used to determine the effect of overproduction of CpxP'-'Bla fusion proteins on the Cpx pathway in a *degP* null background. β-Galactosidase levels were measured from wild-type (odd lanes) and *degP* null strains (even lanes) carrying wild-type and mutant CpxP'-'Bla overexpression plasmids. Strains used in this experiment are as follows TR930 (lane 1), DB5 (lane 2), TR932 (lane 3), DB4 (lane 4), DB30 (lane 5), DB71 (lane 6), DB40 (lane 7), and DB75 (lane 8). VC refers to vector control.

level is higher (Figure 2-7, lanes 5 and 6) (5). This result suggests that, even in the absence of proteolysis, the structural perturbation predicted to be a consequence of the Q55P mutation disrupts protein function. The Q128H mutant was similarly unable to inhibit the Cpx response in a wild-type strain; however, in contrast to the Q55P mutant, it partially regained the ability to inhibit Cpx pathway activity in the *degP* mutant strain (Figure 2-7, lanes 7 and 8) (5). The Q128H mutation may allow for a weak interaction between the two LTXXQ motifs, *via* a single hydrogen bond between the substituted H128 and Q55 that could not be formed by the Q55P mutant, thus explaining the milder phenotype associated with the Q128H substitution. These data support the premise that the Q55P and Q128H mutations disrupt CpxP regulatory function by destabilizing critical structural elements of the conserved LTXXQ motifs.

Discussion

The molecular mechanisms by which the periplasmic protein CpxP acts to control signalling by the sensor kinase, CpxA, and facilitate DegP-mediated proteolysis of misfolded pilus proteins are not known. The crystal structure of truncated *E. coli* CpxP, determined to 2.85 Å resolution, demonstrates that CpxP forms an antiparallel dimer of interconnected α-helices, with each protomer assuming a bent, hooked hairpin fold. SAXS studies illustrated that the crystal structure is a good representation of the structure of CpxP in solution. Residues Q55, Q128, and M59 of CpxP are specifically required to maintain protein structural integrity, while residues R60 and D61 are solvent-exposed and could mediate interactions with other proteins.

CpxP has several potential protein binding partners that may interact with either of the discrete, charged surfaces. The highly basic concave cavity strongly suggests a role as a binding site for a negatively charged surface (Figure 2-4B). It is also possible that molecules could recognize the acidic and

hydrophobic patches on the convex surface of the CpxP structure. A large body of evidence suggests that CpxP interacts with the histidine kinase, CpxA (15, 47-49), and although direct binding has yet to be established, CpxA represents a likely candidate for interaction with CpxP. Previous studies have also proposed that CpxP could act as an adaptor or chaperone, associating with misfolded envelope proteins to escort them to DegP for degradation or refolding (22).

Three mutations in two neighbouring residues of CpxP, R60Q, D61E, and D61V, which are known to disrupt CpxP regulatory function but do not affect mutant protein levels (5), were found to be localized to a region of α1 close to the N-terminal LTXXQ motif (Figure 2-6A-C). These mutations are not predicted to alter CpxP structural stability; therefore residues R60 and D61 may be essential for interaction with the sensing domain of CpxA, or might play a role in recognition of other binding partners such as DegP or misfolded proteins. Given that the highly conservative substitution of aspartic acid for glutamic acid at residue 61 obliterates signalling activity, the role of this residue must be highly specific. Notably, the homologous protein, Spy, contains a glutamic acid residue at this position (Figure 2-6A), and does not function as an inhibitor of the Cpx response (47). Further investigation into the identity of any binding partners and the elucidation of their mechanism of recognition of CpxP will lead us closer to understanding the signalling, chaperone, and other possible activities of CpxP.

Searches of sequence and protein structure databases uncovered several proteins with a similar core fold to CpxP. Although the function of *E. coli* Spy, the closest homologue of CpxP with 26.5% sequence identity, remains unknown, its crystal structure reveals striking similarities. Spy (PDB ID 3OEO) forms a nearly identical dimer comprised of antiparallel α-helical hairpins, similar positioning of the LTXXQ motifs, and a basic concave face (RMSD of 1.4 Å over 91 aligned C_α atoms) (Figure 2-6A and Figure 2-8C and D) (28). A thorough examination of the differences in the structures of CpxP and Spy may reveal clues to their distinct

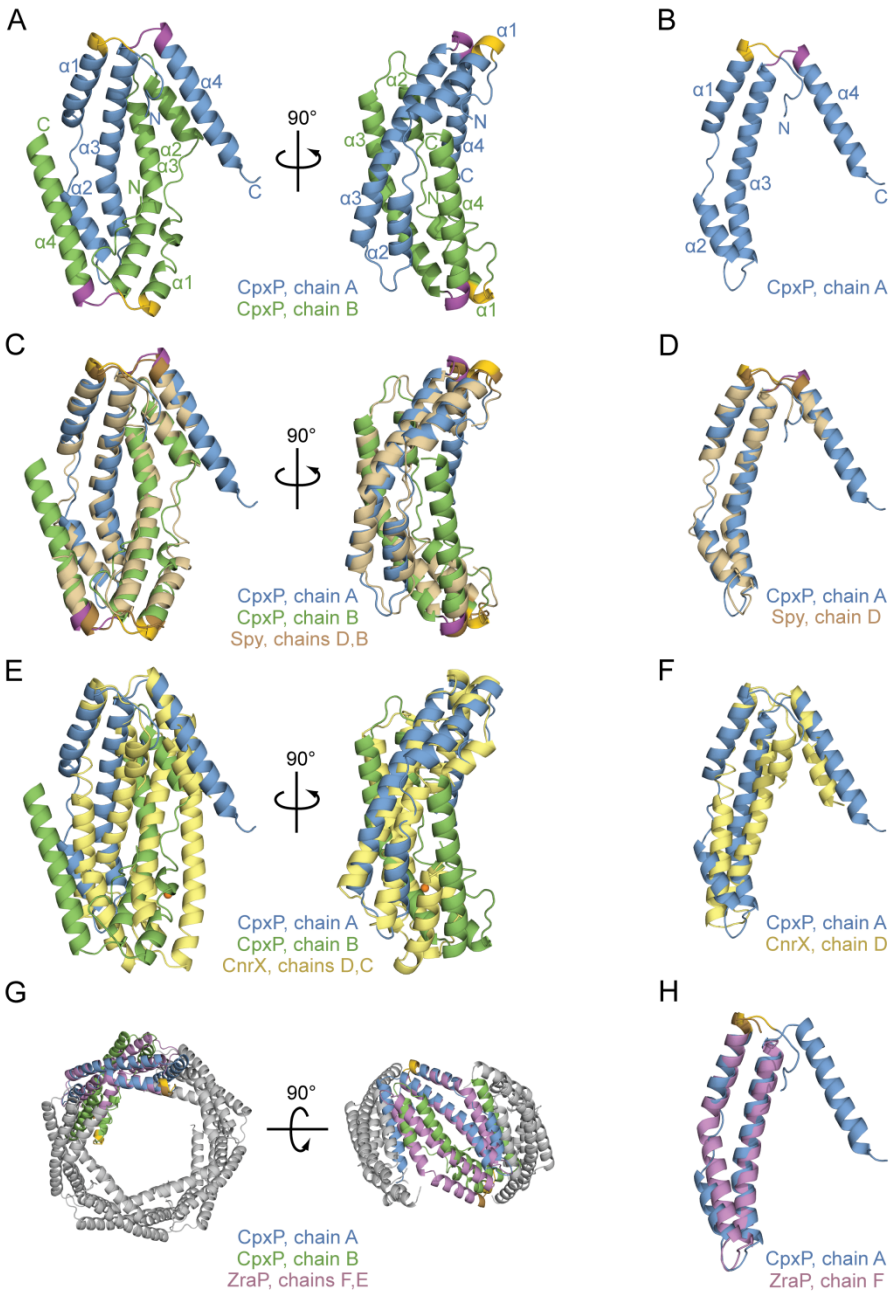


Figure 2-8. CpxP shares a similar core fold with Spy, CnrX, and ZraP. A) The crystal structure of the CpxP(40-151) dimer shown in cartoon representation in two orientations, rotated by 90°. B) Bent and hooked helical hairpin fold of the CpxP protomer, chain A. Chain A of CpxP is shown in blue, and chain B in green. The conserved N- and C-terminal LTXXQ motifs of CpxP are displayed in gold and purple, respectively. Helices α1-4 and the N- and C-termini are labeled. Structural alignments and 3-D superposition of the CpxP dimer (A) and protomer (B) with the biological units and individual protomers of Spy (PDB ID: 3OEO); light brown; shown in panels C-D), CnrX (3EPV); yellow; panels E-F), and ZraP or STM4172 (3LAY); light purple; G-H). The LTXXQ motifs of Spy and ZraP are coloured brown.

roles in bacterial stress response pathways.

Another dimeric periplasmic protein, CnrX (PDB ID 3EPV) (44), also bears strong structural resemblance to CpxP (RMSD of 3.0 Å over 101 aligned C_α atoms); however, they share only 8% sequence identity (Figure 2-6A and Figure 2-8E and F). Each CnrX monomer consists of a four-helix hooked-hairpin similar to that of CpxP and Spy, and was found to exist as a dimer in solution. However, CnrX forms a much straighter four-helix hooked-hairpin that does not incorporate a kink or bend, or the coil region between helices α1 and α2 that are present in CpxP and Spy. In addition, the N-terminus of CnrX contains the first α-helix of the structure whereas it forms a segment of random coil in CpxP and Spy. CnrX does not contain canonical LTXXQ motifs, but forms comparable turns to those seen in CpxP and Spy. Instead, CnrX retains only the conserved leucines, which also pack under the turn, suggesting a critical role for this residue in the LTXXQ-like diverging turn.

The structure of the zinc resistance-associated protein ZraP/STM4172 (PDB ID 3LAY) also adopts the bent hairpin fold common to CpxP and Spy (RMSD of 1.6 Å over 74 aligned C_α atoms with 18% sequence identity). This periplasmic protein assembles into a decameric α-helical barrel, with hairpins forming antiparallel arrangements analogous to the CpxP dimer (Figure 2-6A and Figure 2-8G and H). However, the kink or bend in ZraP is incorporated without the coil region between helices α1 and α2 of CpxP and Spy. Additionally, ZraP contains only one, N-terminal LTXXQ motif, and the structure does not reveal the presence of the hooks in the hooked hairpin fold found in CpxP, Spy or CnrX.

Analysis of the structures of CpxP, Spy, and ZraP reveals that the LTXXQ motif forms a structural element which incorporates a diverging turn in periplasmic stress response proteins with disparate functions, and it is likely that this role will be extended to the remaining members of this Pfam PF07813 family (http://pfam.sanger.ac.uk/family?entry=PF07813&type=Family). Furthermore,

the family may be expanded to include LTXXQ-like turns, such as those observed in CnrX.

Both CnrX and ZraP function in bacterial stress response pathways that are responsible for the efflux of toxic metals from the cell. ZraP is a metal-binding protein that is induced by a two-component system that responds to zinc or lead (32, 41), while CnrX is thought to act as a metal sensor that binds cobalt, nickel, and potentially copper in an extracytoplasmic function (ECF)-sigma factor signalling pathway (17, 18, 34, 38, 58). Interestingly, copper ions have been implicated as an inducer of the Cpx pathway (25, 63-65). In addition, the mRNA levels of *cpxP*, *zraP*, and *spy* were found to increase by 11-, 47-, and 21-fold, respectively, in response to zinc ions (31). It is unknown whether the upregulation of CpxP stems from a genuine function related to these metals, or if it is a consequence of general protein misfolding as a result of their presence. It is interesting, however, that CpxP crystallized in the presence of high zinc concentrations. Several zinc ions were found to be bound in the structure, although these ions mainly appear to mediate contacts between symmetry-related molecules in the crystal lattice and to stabilize amino acid side chains. While the biological significance of the candidate zinc sites in CpxP remains to be explored, the structural similarities between CpxP, CnrX, and ZraP may point to common periplasmic functions.

Two hypotheses regarding the fate of CpxP upon induction of the Cpx response are as follows: (i) the inducing cue might cause CpxP itself to become misfolded, leading to its proteolysis by DegP, (ii) alternatively, CpxP could interact with misfolded proteins to target them to DegP, where they would jointly be degraded (5, 22, 35). Size-exclusion chromatography, MALLS, SAXS, and CD experiments indicate that the shift to alkaline pH that triggers the Cpx response does not initiate a change in oligomeric state, or large conformational rearrangements in CpxP *in vitro*. These biophysical techniques demonstrate that

CpxP does not become a misfolded protein substrate suitable for DegP in the presence of natural Cpx pathway-inducing conditions *in vitro*. The absence of a structural change in CpxP is analogous to CnrX, for which no conformational changes were correlated with copper-binding, a pathway activating signal (38, 44). It is therefore likely that these antiparallel hooked-hairpin proteins sense or react to stress response cues through an alternative mechanism.

Here, we outline several sites, including the charged surfaces of CpxP and residues R60 and D61, which are likely involved in interactions with potential CpxP binding partners including CpxA, DegP, or misfolded proteins. We suggest that the conserved LTXXQ motifs in periplasmic stress response proteins have a structural role in the formation of diverging turns. And finally, we identify two additional periplasmic stress proteins that share common structural elements and similar core folds with CpxP and Spy, however, differences remain in the overall bent and hooked hairpin fold of CpxP and Spy that render them distinct.

References

1. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58: 1948-54

2. Ades SE, Grigorova IL, Gross CA. 2003. Regulation of the alternative sigma factor σ^E during initiation, adaptation and shutoff of the extracytoplasmic heat shock response in *Escherichia coli*. *J Bacteriol* 185: 2512-9

3. Andersen CA, Palmer AG, Brunak S, Rost B. 2002. Continuum secondary structure captures protein flexibility. *Structure* 10: 175-84

4. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl*

Acad Sci U S A 98: 10037-41

5. Buelow DR, Raivio TL. 2005. Cpx signal transduction is influenced by a conserved N-terminal domain in the novel inhibitor CpxP and the periplasmic protease DegP. *J Bacteriol* 187: 6622-30

6. Casadaban MJ. 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *Jounal of Molecular Biology* 104: 541-55

7. Collaborative Computational Project N. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50: 760-3

8. Danese PN, Silhavy TJ. 1997. The σ^E and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. *Genes and Development* 11: 1183-93

9. Danese PN, Silhavy TJ. 1998. CpxP, a stress-combative member of the Cpx regulon. *Journal of Bacteriology* 180: 831-9

10. Danese PN, Snyder WB, Cosma CL, Davis LJ, Silhavy TJ. 1995. The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. *Genes and Development* 9: 387-98

11. DiGiuseppe PA, Silhavy TJ. 2003. Signal detection and target gene induction by the CpxRA two-component system. *J Bacteriol* 185: 2432-40

12. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, et al. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 35: W522-5

13. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. 2004. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32: W665-7

14. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126-32

15. Fleischer R, Heermann R, Jung K, Hunke S. 2007. Purification, reconstitution, and characterization of the CpxRAP envelope stress system of *Escherichia coli*. *J Biol Chem* 282: 8583-93

16. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31: 3784-8

17. Grass G, Fricke B, Nies DH. 2005. Control of expression of a periplasmic nickel efflux pump by periplasmic nickel concentrations. *Biometals* 18: 437-48

18. Grass G, Grosse C, Nies DH. 2000. Regulation of the *cnr* cobalt and nickel resistance determinant from *Ralstonia* sp. strain CH34. *J Bacteriol* 182: 1390-8

19. Holm L, Rosenstrom P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38 Suppl: W545-9

20. Holm L, Sander C. 1996. Mapping the protein universe. *Science* 273: 595-603

21. Hunke S, Keller R, Muller VS. 2012. Signal integration by the Cpx-envelope stress system. *FEMS Microbiol Lett* 326: 12-22

22. Isaac DD, Pinkner JS, Hultgren SJ, Silhavy TJ. 2005. The extracytoplasmic adaptor protein CpxP is degraded with substrate by DegP. *Proc Natl Acad Sci U S A*

23. Jones CH, Danese PN, Pinkner JS, Silhavy TJ, Hultgren SJ. 1997. The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems. *EMBO Journal* 21: 6394-406

24. Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577-637

25. Kershaw CJ, Brown NL, Constantinidou C, Patel MD, Hobman JL. 2005. The expression profile of *Escherichia coli* K-12 in response to minimal, optimal and excess copper concentrations. *Microbiology* 151: 1187-98

26. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* 36: 1277-82

27. Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372: 774-97

28. Kwon E, Kim DY, Gross CA, Gross JD, Kim KK. 2010. The crystal structure *Escherichia coli* Spy. *Protein Sci* 19: 2252-9

29. Larsen RA, Thomas MG, Wood GE, Postle K. 1994. Partial suppression of an *Escherichia coli* TonB transmembrane domain mutation (ΔV17) by a missense mutation in ExbB. *Mol Microbiol* 13: 627-40

30. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26: 283-91

31. Lee LJ, Barrett JA, Poole RK. 2005. Genome-wide transcriptional response of chemostat-cultured *Escherichia coli* to zinc. *J Bacteriol* 187: 1124-34

32. Leonhartsberger S, Huber A, Lottspeich F, Bock A. 2001. The *hydH/G* genes from *Escherichia coli* code for a zinc and lead responsive two-component regulatory system. *J Mol Biol* 307: 93-105

33. Lerner MG, Carlson HA. 2006. APBS plugin for PyMOL. *University of Michigan, Ann Arbor*

34. Liesegang H, Lemke K, Siddiqui RA, Schlegel HG. 1993. Characterization of the inducible nickel and cobalt resistance determinant *cnr* from pMOL28 of *Alcaligenes eutrophus* CH34. *J Bacteriol* 175: 767-78

35. MacRitchie DM, Buelow, D.R., Price, N.L., Raivio, T.L. 2007. Two-component signaling and Gram negative envelope stress response systems. In *Bacterial signal transduction: network and drug targets*, ed. R Utsumi. Austin, TX: Landes Bioscience

36. Magrane M, Consortium U. 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011: bar009

37. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* 40: 658-74

38. Monchy S, Benotmane MA, Janssen P, Vallaeys T, Taghavi S, et al. 2007. Plasmids pMOL28 and pMOL30 of *Cupriavidus metallidurans* are specialized in the maximal viable response to heavy metals. *J Bacteriol* 189: 7417-25

39. Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240-55

40. Nevesinjac AZ, Raivio TL. 2005. The Cpx Envelope Stress Response Affects Expression of the Type IV Bundle-Forming Pili of Enteropathogenic *Escherichia coli*. *J Bacteriol* 187: 672-86

41. Noll M, Petrukhin K, Lutsenko S. 1998. Identification of a novel transcription regulator from *Proteus mirabilis*, PMTR, revealed a possible role of YJAI protein in balancing zinc in *Escherichia coli*. *J Biol Chem* 273: 21393-401

42. Otwinowski Z, Minor W. 1997. Processing of X-ray Diffraction Data Collected In Oscillation Mode. In *Methods in Enzymology*, ed. J Charles W. Carter, RM Sweet, pp. 307-26. New York: Academic Press

43. Pogliano JA, Lynch S, Belin D, Lin ECC, Beckwith J. 1997. Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. *Genes and Development* 11: 1169-82

44. Pompidor G, Maillard AP, Girard E, Gambarelli S, Kahn R, Coves J. 2008. X-ray structure of the metal-sensor CnrX in both the apo- and copper-bound forms. *FEBS Lett* 582: 3954-8

45. Potterton E, Briggs P, Turkenburg M, Dodson E. 2003. A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 59: 1131-7

46. Price NL, Raivio TL. 2009. Characterization of the Cpx regulon in *Escherichia coli* strain MC4100. *J Bacteriol* 191: 1798-815

47. Raivio TL, Laird MW, Joly JC, Silhavy TJ. 2000. Tethering of CpxP to the inner membrane prevents spheroplast induction of the Cpx envelope stress response. *Molecular Microbiology* 37: 1186-97

48. Raivio TL, Popkin DL, Silhavy TJ. 1999. The Cpx envelope stress response is controlled by amplification and feedback inhibition. *Journal of Bacteriology* 181: 5263-72

49. Raivio TL, Silhavy TJ. 1997. Transduction of envelope stress in *Escherichia coli* by the Cpx two-component system. *Journal of Bacteriology* 179: 7724-33

50. Silhavy TJ, Berman ML, Enquist LW. 1984. *Experiments with gene fusions*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press

51. Svergun DI. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* 25: 495-503

52. Svergun DI, Petoukhov MV, Koch MH. 2001. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80: 2946-53

53. Tamames B, Sousa SF, Tamames J, Fernandes PA, Ramos MJ. 2007. Analysis of zinc-ligand bond lengths in metalloproteins: trends and patterns. *Proteins* 69: 466-75

54. Terwilliger TC. 2000. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 56: 965-72

55. Terwilliger TC. 2003. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* 59: 38-44

56. Terwilliger TC, Berendzen J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55: 849-61

57. Theobald DL, Wuttke DS. 2006. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* 22: 2171-2

58. Tibazarwa C, Wuertz S, Mergeay M, Wyns L, van Der Lelie D. 2000. Regulation of the *cnr* cobalt and nickel resistance determinant of *Ralstonia eutropha* (*Alcaligenes eutrophus*) CH34. *J Bacteriol* 182: 1399-409

59. Vogt SL, Raivio TL. 2012. Just scratching the surface: an expanding view of the Cpx envelope stress response. *FEMS Microbiol Lett* 326: 2-11

60. Volkov VV, Svergun DI. 2003. Uniqueness of *ab initio* shape determination in small-angle scattering. *Journal of Applied Crystallography* 36: 860-4

61. Winn MD, Isupov MN, Murshudov GN. 2001. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* 57: 122-33

62. Wriggers W. 2010. Using Situs for the Integration of Multi-Resolution Structures. *Biophysical Reviews 2010* 2: 21-7

63. Yamamoto K, Ishihama A. 2005. Transcriptional response of *Escherichia coli* to external copper. *Mol Microbiol* 56: 215-27

64. Yamamoto K, Ishihama A. 2006. Characterization of copper-inducible promoters regulated by CpxA/CpxR in *Escherichia coli*. *Biosci Biotechnol Biochem* 70: 1688-95

65. Yamamoto K, Ogasawara H, Ishihama A. 2008. Involvement of multiple transcription factors for metal-induced *spy* gene expression in *Escherichia coli*. *J Biotechnol* 133: 196-200

CHAPTER 3

Characterization of the periplasmic sensing domain of the histidine kinase, CpxA

Introduction

The CpxAR two-component system (TCS) is one of several envelope stress responses that function in Gram-negative bacteria such as *Escherichia coli*. The inner membrane-spanning sensor histidine kinase (HK), CpxA, and cognate response regulator (RR), CpxR, mediate a broad response to envelope stress primarily centered on remediation of damage due to periplasmic protein misfolding (35).

The Cpx response is induced by a variety of stimuli including alkaline pH (8, 22), bacterial adhesion to abiotic surfaces (24), and the overexpression of pilus subunits such as BfpA (23) or PapE and PapG in the absence of their cognate chaperone (15). Additional activating signals include altered membrane composition resulting from a lack of phosphatidylethanolamine (21), and the overexpression of the outer membrane lipoprotein NlpE (32). A unifying feature of a number of these stresses is the resulting aggregation or accumulation of misfolded proteins at the periplasmic face of the inner membrane; however the way in which CpxA senses these signals is unknown.

The HK CpxA detects these specific inducing cues through a periplasmic sensor domain (29). The signal is transmitted through the membrane to a cytosolic histidine kinase core. CpxA then autophosphorylates at a conserved histidine residue and recruits CpxR, which accepts the phosphoryl group *via* a conserved aspartate residue. Activated CpxR functions as a transcription factor to control the expression of genes in the Cpx regulon. The expression of the periplasmic accessory protein, CpxP, is upregulated by the Cpx response (8). CpxP represses the pathway by inhibiting the autokinase activity of CpxA through a putative interaction with the sensing domain (9, 28). This negative regulation facilitates the return of the pathway to basal activity levels and further serves to maintain the TCS in an inactive state in the absence of stress signals (2, 28).

CpxR regulates the expression of numerous genes, several of which affect bacterial virulence (35). While envelope protein folding and degrading factors, cell wall remodelling enzymes, and CpxARP pathway constituents are up-regulated by the Cpx response, the expression of complexes such as pili and flagella are repressed (27, 37). In general, many of the positively regulated genes function in periplasmic protein maintenance; however it is evident that changes in the regulation of many genes are required in order to combat stress in the envelope.

The extracytoplasmic sensor domain of CpxA is, thus far, uncharacterized. Although HK periplasmic domains share low sequence identity, to date, only three types of structural folds have been identified. One structure reveals a periplasmic binding protein-like fold, several illustrate an all-α-helical arrangement, and the remainder display an α/β PAS-like PDC domain fold (5). Due to the low level of sequence conservation, primary sequence alignments do not provide much information regarding the organization of the CpxA sensor domain. Therefore, we sought to characterize the CpxA periplasmic domain using biochemical techniques and to obtain a high-resolution structure using X-ray crystallography in order to gain insight into the molecular mechanism for stimulus perception utilized by CpxA.

Size-exclusion chromatography and multi-angle laser light scattering experiments were used to observe the oligomeric state of the CpxA periplasmic domain. Additionally, circular dichroism spectra and melting curves were analyzed to determine the secondary structure composition and thermal stability of the domain. Lastly, crystallization conditions were screened to identify preliminary protein crystal hits. Our results reveal that the isolated CpxA periplasmic domain exists in an atypical tetramer-dimer equilibrium. Furthermore, we identify labile regions of this domain that might be involved in structural rearrangements. We suggest that the periplasmic sensor domain of

CpxA forms a PAS-like PDC fold, and have determined several excellent protein crystallization conditions that will be optimized to investigate this hypothesis.

Materials and Methods

Cloning of constructs based on the predicted periplasmic domain of CpxA

Four protein constructs were designed based on the predicted periplasmic domain of CpxA: CpxA(31-163), CpxA(31-154), CpxA(40-163), and CpxA(40-154) (Figure 3-1)[1]. The sequences corresponding to these four constructs were amplified from the *Escherichia coli* MC4100 genome using combinations of the PCR primers 5'-31 (5'-CGCGGATCCGATTCACGCCAGATGACCGA) or 5'-40 (5'-CGCGGATCCGATAGCGAACAGCGTCAGG), and 3'-154 (5'-GCGAATTCCTAGGATTGAGAACTGCTGGC) or 3'-163 (5'-GCGAATTCCTAGCGGTCAAACAGTAAGTT). The fragments were ligated into the BamHI and EcoRI restriction sites of the pGEX-6P-1 expression vector (GE Healthcare), which contains N-terminal glutathione S-transferase (GST) tag and PreScission™ Protease cleavage site sequences. The subclones were confirmed by DNA sequencing (Molecular Biology Services Unit, University of Alberta).

Selection, over-expression, and purification of a recombinant CpxA periplasmic domain construct

The four CpxA-containing plasmids were transformed into *E. coli* BL21(DE3) cells. Cells were cultured at 30°C in Luria-Bertani medium with 100 µg/mL ampicillin. Optimal CpxA over-expression was achieved by inducing cells

[1] The CpxA construct sequences were chosen and cloned by Roxana Malpica from the laboratory of Dr. Tracy Raivio, Department of Biological Sciences, University of Alberta.

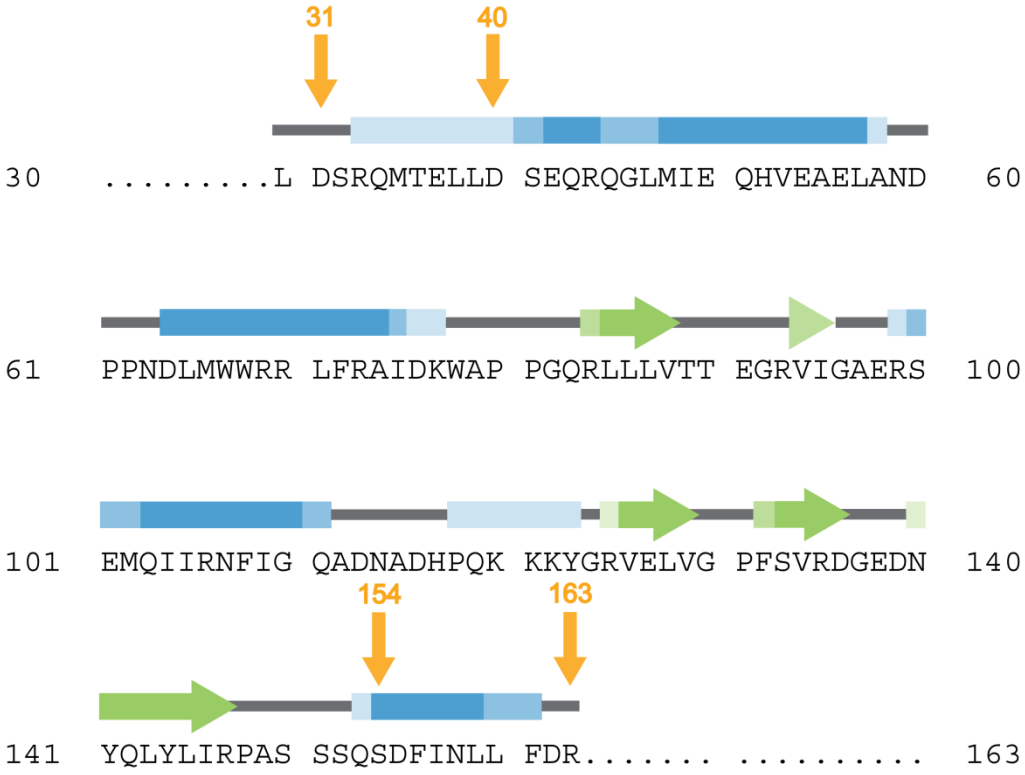


Figure 3-1. The predicted periplasmic sensor domain of CpxA. Sequence of the predicted periplasmic domain of CpxA, residues 30-163, as defined by the UniProtKB database (19). Secondary structure, predicted using the programs Porter (26), Jpred3 (7), and PsiPred v3.0 (16, 20), is shown above the sequence where blue boxes depict α-helices and green arrows represent β-strands. The three shades of colour, from light to dark, illustrate the number of programs from one to three, respectively, that predict an element of secondary structure for each residue. The orange arrows indicate the start and end points of the four CpxA constructs, CpxA(31-163), CpxA(31-154), CpxA(40-163), and CpxA(40-154), which were cloned, expressed, and partially purified.

that had reached an optical density at 600 nm of 0.8-0.9 with 0.05 mM isopropyl-β-D-thiogalactopyranoside (IPTG), followed by growth at a reduced temperature (22°C) for 19-21 h. The CpxA fusion constructs were purified at 4°C using a protocol modified from the GST Gene Fusion System Handbook (GE Healthcare).

Cells were harvested by centrifugation and the cell pellets were resuspended in buffer composed of 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and the protease inhibitors leupeptin, pepstatin, and phenylmethylsulfonyl fluoride. The cells were lysed by chicken egg white lysozyme and sonication, and the cell lysate was fractionated by centrifugation. The clarified lysate was mixed with Glutathione Sepharose™ 4B resin (GE Healthcare) pre-equilibrated in resuspension buffer, and the binding reaction was incubated for 1.5 h with nutating. The resin was washed with 10 volumes of resuspension buffer and the bound GST fusion protein was eluted using resuspension buffer containing 20 mM reduced glutathione at pH 7.5. The expression and solubility of the four constructs were assessed after each of the purification steps (Figure 3-2A). The largest construct, CpxA(31-163), was clearly best in both respects, and was therefore chosen for further studies.

For all CpxA preparations after this initial assessment, the resuspension buffer contained 500 mM NaCl, as a higher salt concentration was necessary for protein solubility after separation from the GST tag. The GST-CpxA(31-163) elutions from the GST affinity purification step were pooled, buffer exchanged into resuspension buffer, and concentrated to approximately 50 mL using a 10,000 MWCO centrifugal filter device (Millipore, Fisher Scientific). The GST tag was removed by PreScission™ Protease (GE Healthcare) over 12-16 h at 4°C, and the progress of the reaction was monitored by SDS-PAGE. The cleavage reaction was subsequently applied to a second and third GST affinity column to remove resilient proteolyzed GST and any remaining fusion protein. The column flow-through and washes containing CpxA(31-163) were pooled and concentrated

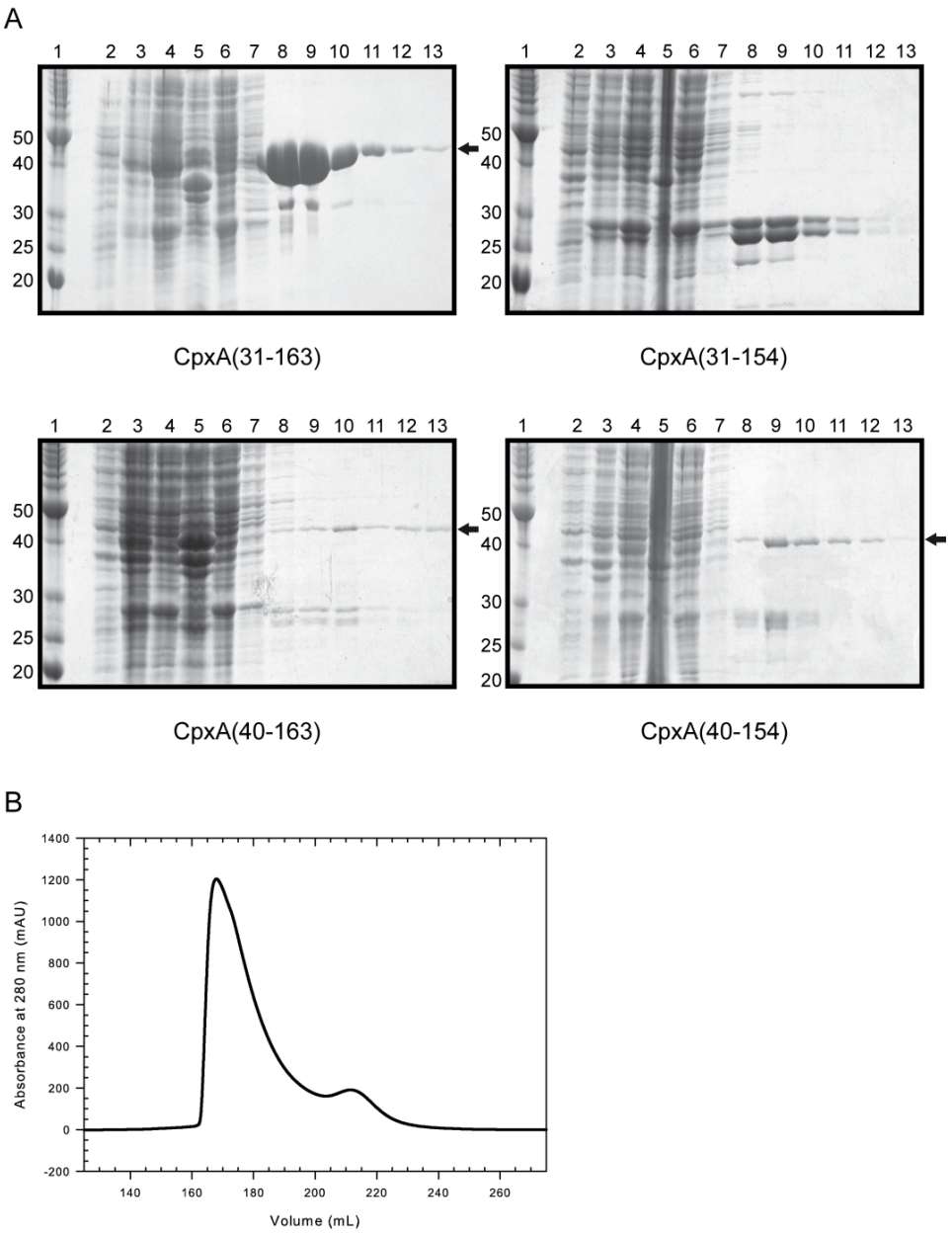


Figure 3-2. Purification of the CpxA periplasmic domain. A) Expression and initial purification steps for selection of a CpxA periplasmic domain construct. SDS-PAGE analysis of the four constructs, CpxA(31-163), CpxA(31-154), CpxA(40-163), and CpxA(40-154). Lanes: 1 – protein ladder with corresponding molecular weight designations on the left, 2 – pre-induction sample, 3 – post-induction sample, 4 – cleared lysate, 5 – insoluble fraction (pellet), 6 – GST affinity column unbound fraction, 7 – GST affinity column concentrated wash, 8-13 – GST affinity column eluted fractions. Black arrows point to the likely full-length CpxA fusion proteins. Some smaller degradation products are visible in the GST affinity column elutions for all of the constructs. B) Size-exclusion chromatography elution profile for CpxA(31-163). Both peaks, at approximately 167 mL and 212 mL, correspond to CpxA. The void volume, V_0, for the column is 100 ± 2 mL.

using a 3,000 MWCO spin concentrator (Millipore, Fisher Scientific). As a final purification step, CpxA(31-163) was passed over a HiLoad Superdex™ 75 26/60 size-exclusion column equilibrated in resuspension buffer on an ÄKTA™purifier (GE Healthcare). CpxA(31-163) eluted in two partially overlapping peaks, which were pooled separately and concentrated by ultrafiltration using a 3,000 MWCO spin concentrator (Figure 3-2B).

Size-exclusion chromatography with multi-angle laser light scattering (MALLS)

Concentrated samples from the two purified CpxA(31-163) gel filtration peaks (see above) were passed over a Superose™ 6 10/300 size-exclusion column (GE Healthcare) that was directly in-line with a DAWN EOS™ MALLS module (which itself was connected to a WyattQELS™ quasi-elastic light scattering unit) and an Optiplex rEX™ differential refractive index (dRI) detector (Wyatt Technology Corporation). The system was pre-equilibrated in buffer composed of 50 mM Tris-HCl pH 7.5 and 500 mM NaCl, and maintained at 25°C. Monomeric bovine serum albumin (BSA), purified from the monomer-dimer equilibrium using a Superdex™ 200 16/60 gel filtration column on an ÄKTA™purifier (GE Healthcare), was used as a standard to normalize the detectors and to determine the detector delay volume. Values for the molar mass and hydrodynamic radius of each protein species were determined from the light scattering data using the ASTRA software (Wyatt Technology Corporation).

Dynamic light scattering (DLS)

The physical state of CpxA(31-163) with respect to solubility and aggregation was qualitatively assessed in a range of different buffer conditions. Having observed earlier that the separate peaks isolated from the size-exclusion

column during purification re-equilibrated into a double peak shortly thereafter, it was determined that analysis of only one of the CpxA(31-163) peak samples was required. Ten salt concentrations were tested, covering a 10-fold range from 500 mM to 50 mM NaCl in 50 mM increments. For these experiments, 50 mM Tris-HCl pH 7.5 was used as the buffer as there is evidence from purification by size-exclusion chromatography and MALLS that CpxA(31-163) was soluble in this buffer. In addition, 12 pHs were tested, which covered a range from pH 5.8 to 8.0 in 0.2 pH unit increments. Phosphate buffers (100 mM) were used due to their wide effective range in addition to 500 mM NaCl. All samples were equilibrated for 10 minutes on ice, and then spun at maximum speed in a tabletop centrifuge for 10 minutes at 4°C to remove large aggregates that would otherwise overload the signal. Data were collected at 4°C using a Dynapro module equipped with a Temperature Controlled Microsampler (Protein Solutions/Wyatt Technology Corporation). A protein concentration was chosen to ensure a reliable reading of a minimum of 50 000 counts (light scattering intensity) for each measurement, and a minimum of 50 data points were collected for each sample. Results were evaluated using size distribution histograms generated by the Dynamics V5.25.44 software (Protein Solutions/Wyatt Technology Corporation). The relative peak size near a radius of hydration (R_h) of approximately 10 nm, the approximate size expected for proteins, was monitored in comparison to the signal at approximately 1000 nm, which corresponds to aggregate.

Circular dichroism (CD)

Purified CpxA(31-163) was dialyzed into phosphate buffer composed of 100 mM sodium phosphate pH 5.8 or 8.0, and 500 mM NaCl. Protein samples were used at a concentration of 0.2 mg/mL; this concentration was high enough to produce a good signal without reaching saturation of the detector (monitored

by observing high tension voltages). All CD data were recorded on a JASCO J-720 spectropolarimeter (JASCO, Inc.) using a QS 0.05 cm cylindrical quartz cuvette.

CD secondary structure spectra of CpxA(31-163) samples in phosphate buffer pH 5.8 or 8.0 were collected at 24°C in the far-UV range (250-190 nm) using a 0.1 nm step resolution. The "24°C" and "72°C" spectra were captured before and after, respectively, the melting curve analysis (below). The "guanidine hydrochloride" spectrum was collected on a CpxA sample that was subjected to 6 M guanidine hydrochloride (GndCl) and boiled for 30 minutes. Each spectrum is an average of 8 individual spectra.

CD melting curves were obtained from CpxA(31-163) samples in phosphate buffer pH 5.8. Data were collected at a wavelength of 222 nm to monitor the α-helix secondary structure signal as a function of temperature from 24 to 72°C. A decrease in the α-helix signal represents an increase in unfolding of α-helical secondary structure, and ideally a melting temperature (T_m) can be calculated as the temperature at which the signal, or mean residue ellipticity (MRE), at 222 nm is reduced to 50%. An RMS 6 water bath (LAUDA-Brinkmann) was used to control the temperature of the cell holder. The temperature of the protein sample inside the cuvette was also monitored using a thermocouple probe. After reaching each temperature, the cell was allowed to equilibrate for five minutes before measurements were made to ensure that the temperature of the protein sample reached the temperature of the cell. Three hundred readings were recorded for each measurement over 5 min, from which a buffer-corrected average MRE and standard deviation were calculated.

Data were processed using the program ACDP v2.8 (12). Buffer control data or spectra were subtracted from those of the sample to correct for possible contributions from the buffer and cuvette, and the spectropolarimeter baseline. The ellipticity (θ) data, in units of $1e^{-3}$° or mdeg, were then converted to units of mean residue ellipticity (MRE) (°cm^2 $dmol^{-1}$) using the molecular mass of

CpxA(31-163) including remnants of cloning and PreScission™ Protease cleavage sequences, which was calculated to be 15 867.8 g/mol using ProtParam (11). The secondary structure spectra were smoothed using an algorithm based on the local mean. Secondary structure deconvolution of the normalized CD spectra was conducted over a wavelength range from 200-240 nm using the Linear Combination (Fasman) method with five basis sets in ACDP (12), and the K2D3 program (18). An approximation of the T_m (for a portion of the CpxA construct) was determined from the point of inflection of the melting curve by taking the first and second derivatives of the plot data in Excel (Microsoft). Additional replicates and experiments performed after separation of the different oligomeric states would be required for more accurate values.

Preliminary crystallization

Purified CpxA(31-163) was dialyzed into crystallization buffer consisting of 50 mM HEPES pH 7.0 and 350 mM NaCl (a "safe" salt concentration taken from the DLS experiments) using a 3,000 MWCO spin concentrator (Millipore, Fisher Scientific). The concentration of the CpxA size-exclusion column peak samples was estimated to be 6.1 mg/mL for Peak 1, and 4.1 mg/mL for Peak 2 from the absorbance at 280 nm, using the theoretical extinction coefficient of 26 470 $M^{-1}cm^{-1}$ derived from the amino acid sequence of CpxA(31-163) plus extra residues using ProtParam (11). The two CpxA(31-163) samples were subjected to seven commercial sparse matrix crystallization screens using the vapour diffusion method: Emerald Wizard I and II (Emerald BioSystems), NeXtal Classics, Classics II, JCSG$^+$ ProComplex, and MBClass (QIAGEN), and PEG/Ion (Hampton Research). Sitting drop 96-well Intelli-plate® crystallization trays (Art Robbins Instruments) were set at room temperature (~21°C) with drops consisting of 0.2 µL protein solution at either 6.1 or 4.1 mg/mL and 0.2 µL well solution, with a 100 µL reservoir using a Gryphon robot equipped with a Phoenix nano dispenser

(Art Robbins Instruments). A number of crystal hits were identified, and crystals were harvested and flash frozen in liquid nitrogen without any additional cryoprotectant to screen for potential protein crystals. Diffraction data were collected at the home source rotating anode X-ray generator (Rigaku), and at the Canadian Light Source beamline CMCF 08ID-1 (Saskatoon, Canada). A full data set was collected from one crystal in condition JCSG$^+$ F3 (0.1 M Tris pH 8.5, 20% (v/v) MPD; final pH 8), which was processed with DENZO (25) but the data could not be scaled. The two crystals from condition JCSG$^+$ F3 that had been screened were melted and washed, then dissolved in 5 µL distilled water and vortexed before running on a 15% SDS-PAGE gel. The gel was stained with SYPRO® Orange protein gel stain (Invitrogen) and visualized by UV illumination.

All graphs were generated using SigmaPlot (Systat Software Inc.).

Results

The periplasmic domain construct, CpxA(31-163), forms higher order oligomers

Four constructs were generated from the predicted periplasmic sensor domain of CpxA (Figure 3-1). These fragments were expressed and partially purified in order to identify a construct that would be suitable for biochemical and structural studies (Figure 3-2A). The CpxA(31-163) construct yielded good over-expression levels, and the majority of the protein fragment was soluble. In contrast, no clear over-expression was observed for the CpxA(31-154) construct. CpxA(40-163) appeared to be moderately over-expressed, however it was nearly all found in the insoluble fraction. Over-expression of the CpxA(40-154) construct was difficult to discern, although there was some soluble protein present in the GST affinity column elutions. Of the four constructs tested, the largest one, CpxA(31-163), was clearly the best with respect to both over-expression levels and solubility, and was therefore chosen for all further studies.

In the final purification step, CpxA(31-163) was observed to elute from the size-exclusion chromatography column in two, slightly overlapping peaks (Figure 3-2B). The two peaks, which eluted at approximately 167 mL and 212 mL, were both confirmed to contain pure CpxA by SDS-PAGE analysis (not shown). Since full-length histidine kinases are known to function as dimers in the cell, the possibility that the periplasmic sensor domain of CpxA could dimerize on its own was explored. A comparison of the CpxA size-exclusion column traces to profiles of other proteins with known molecular masses proved inconclusive. Therefore, size-exclusion chromatography was used in combination with multi-angle laser light scattering (MALLS) in order to determine the oligomeric state(s) present in the CpxA(31-163) elution profile. Light scattering data were used to calculate the molar mass (M_z) over the elution peak of each sample. Bovine serum albumin (BSA) was used as a control for this experiment. The M_z for BSA was determined to be 66 910 ± 4010 g/mol, which is in agreement with the expected value of 66 463 g/mol (not shown). The two CpxA species, which had been pooled separately after the last step of purification, were observed to have re-equilibrated into a double peak shortly thereafter. Consequentially, the light scattering analysis was unavoidably performed on overlapping size-exclusion peaks and accurate results could not be determined. Approximate molar masses were, however, obtained for both CpxA(31-163) species. The first, larger peak was estimated to have a M_z of 58 770 ± 1760 g/mol, while the second, smaller peak had an approximate M_z of 34 130 ± 2050 g/mol (Figure 3-3). Given that the molar mass of the CpxA(31-163) construct with residual amino acid residues had a molar mass of 15 867.8 g/mol (ProtParam (11)), the two peaks corresponded closely, but not within error, to a dimer (Peak 2) and tetramer (Peak 1).

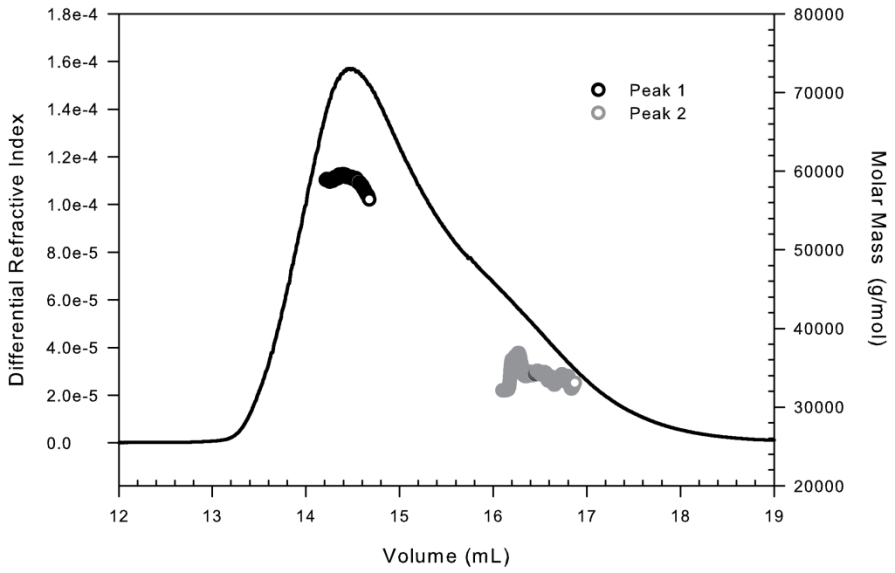


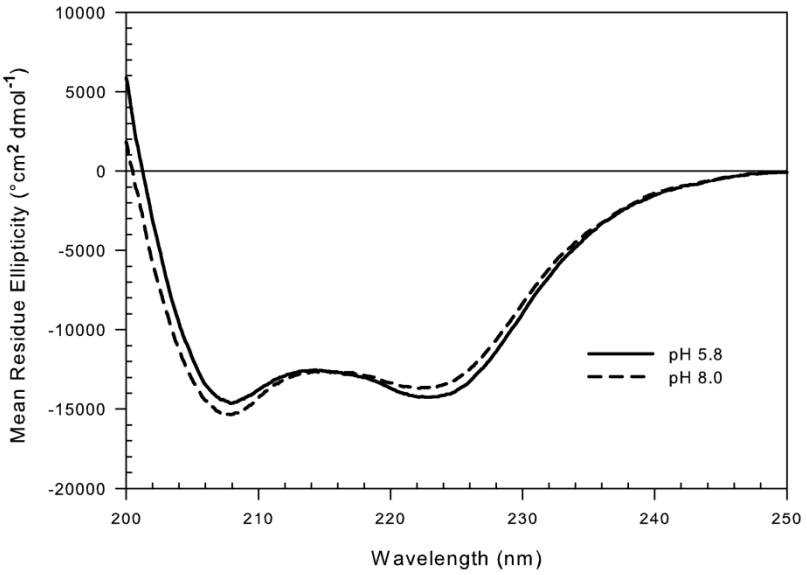
Figure 3-3. The CpxA sensor domain exists in two oligomeric states. Size-exclusion chromatography performed in combination with multi-angle laser light scattering for determination of the oligomeric state(s) of CpxA(31-163). Molar mass measurements are represented by black and grey circles superimposed over peaks 1 and 2, respectively, of the CpxA size-exclusion chromatography elution profile.

The CpxA periplasmic domain has a mixed α/β secondary structure composition that changes slightly with a shift to alkaline pH, an inducing cue for the Cpx pathway

Dynamic light scattering (DLS) experiments were carried out to determine whether the CpxA periplasmic domain construct was in a soluble, non-aggregated state over the entire range from pH 5.8, a condition where the Cpx pathway is inactive, to pH 8.0, where the stress response is active (8). In comparison to MALLS studies, DLS experiments do not generate an accurate value for molecular mass; however much less material is required, and this technique is useful for obtaining rapid, qualitative estimations of aggregation in a number of conditions. Size distribution histograms, representing the hydrodynamic radii (R_h) of the particles in each sample, were monitored for peaks at approximately 10 nm, which corresponds to the approximate size for soluble proteins, and approximately 1000 nm, which is consistent with aggregates of misfolded proteins. For all pHs tested between pH 5.8 and 8.0, there was no discernible aggregate peak near 1000 nm (not shown). These results indicated that CpxA(31-163) was stable in 100 mM phosphate buffers over the entire pH range from 5.8 to 8.0 in 500 mM NaCl, and these conditions could be used for future experiments.

To begin to characterize the structure of the predicted periplasmic domain of CpxA, circular dichroism (CD) secondary structure spectra were obtained in the far-UV range (250-200 nm) for CpxA(31-163) at pH 5.8 and 8.0 (Figure 3-4A). The curves revealed distinctive elements typical of folded proteins at both pHs. Although the curves were quite similar, small differences were observed in the vicinity of the characteristic α-helix signals at 208 and 222 nm. The spectrum for CpxA at pH 8.0 exhibited a weaker signal at 222 nm, which is indicative of lower α-helix secondary structure content. To further examine these spectra, two deconvolution methods were employed to calculate the

A



B

pH	Deconvolution method	α-helix (%)	β-strand (%)
5.8	Linear Combination	33	15
	K2D3	37	15
8.0	Linear Combination	30	18
	K2D3	33	13

C

Prediction program	α-helix (%)	β-strand (%)
Porter	42	17
Jpred3	37	18
PsiPred v3.0	38	16

Figure 3-4. The CpxA periplasmic sensor domain contains both α-helical and β-strand secondary structure elements. A) Far-UV circular dichroism spectra of CpxA(31-163) at pH 5.8 (solid line) and 8.0 (dashed line) collected at 24°C. The raw ellipticities were converted to normalized mean residue ellipticities and plotted as a function of wavelength from 250-200 nm. B) Summary of the percentages of α-helical and β-strand content calculated from the normalized CD spectra of CpxA(31-163) at pH 5.8 and 8.0 by the Linear Combination (Fasman) method in ACDP (12), and the program K2D3 (18). C) Summary of the percentages of α-helical and β-strand content predicted by the secondary structure prediction programs Porter (26), Jpred3 (7), and PsiPred v3.0 (16, 20) for the putative periplasmic sensor domain of CpxA, residues 31-163.

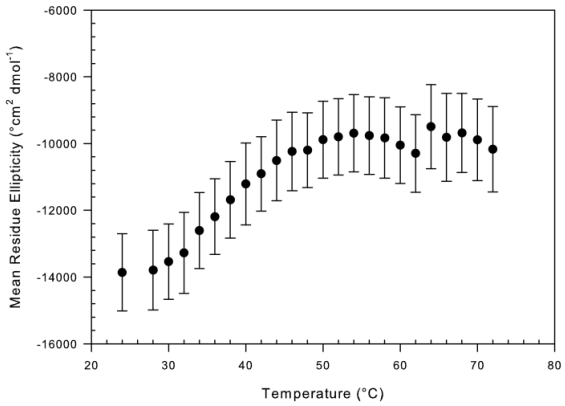
secondary structure composition of each curve (Figure 3-4B). Both algorithms verified that the α-helical content at pH 8.0 was slightly lower than that at pH 5.8. These small changes in secondary structure suggest that the CpxA sensor domain might undergo small structural rearrangements upon a shift to alkaline pH.

Additionally, it is of note that the proportions of secondary structure calculated by the CD spectrum deconvolution programs (Figure 3-4B) are in line with those predicted by the three secondary structure prediction programs used (Figures 3-1 and 3-4C). The Linear Combination (Fasman) method in ACDP (12), and the K2D3 program (18) computed between 30-37% α-helix content and 13-18% β-strand secondary structure content for the CpxA periplasmic domain from the CD data, while the secondary structure prediction programs Porter (26), Jpred3 (7), and PsiPred v3.0 (16, 20) predicted the sensor domain to contain between 37-42% α-helix, and 16-18% β-strand secondary structure content.

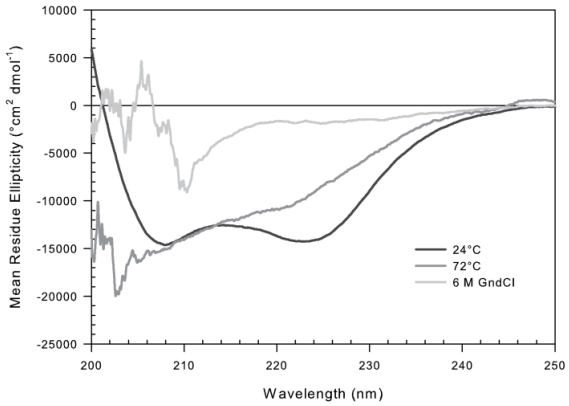
The β-strand secondary structure elements of the CpxA sensor domain are relatively thermostable

The CpxA(31-163) construct was tested further using melting curves to characterize the stability of the protein domain. The CD α-helix signal at 222 nm was monitored as a function of increasing temperature from 24 to 72°C over approximately 8 h (Figure 3-5A). The decrease in the signal measured at 222 nm corresponds to an increase in unfolding of α-helical secondary structure, and ideally a melting temperature (T_m), defined as the temperature at which the signal has decreased by half, could be calculated for the CpxA domain. Upon reaching 72°C, readings indicated that the protein had not undergone full denaturation. Due to a technical issue the temperature could not be increased further to complete the experiment. However, an approximation of the T_m (for a portion of the CpxA construct) was determined to be 34°C.

A



B



C

Condition	Deconvolution method	α-helix (%)	β-strand (%)
24°C	Linear Combination	33	15
	K2D3	37	15
72°C	Linear Combination	6	27
	K2D3	30	15
GndCl	Linear Combination	0	0
	K2D3	1	41

Figure 3-5. Analysis of the stability of CpxA(31-163). A) Melting curve of CpxA(31-163) at pH 5.8. An average mean residue ellipticity at 222 nm and a corresponding standard deviation were calculated from 300 readings recorded over 5 min at each temperature, which covers a range from 24-72°C. Only partial melting of the construct was achieved, however an estimate of the T_m for the unfolded portion of the protein was calculated to be 34°C. B) Far-UV CD spectra of CpxA(31-163) pH 5.8 collected at 24°C, 72°C, and in 6 M GndCl. Raw ellipticities were converted to normalized mean residue ellipticities and plotted as a function of wavelength from 250-200 nm. The "24°C" and "72°C" spectra were captured before and after, respectively, the melting curve analysis in Figure 5A). The "GndCl" spectrum was collected on a CpxA sample that was subjected to 6 M GndCl and boiled for 30 minutes, which illustrates a spectrum of fully unfolded protein. C) Summary of the percentages of α-helical and β-strand content calculated by the Linear Combination (Fasman) method in ACDP (12), and the program K2D3 (18) from the normalized CD spectra of CpxA(31-163) pH 5.8 at 24°C, 72°C, and in 6 M GndCl.

To confirm that the protein was not fully melted by 72°C, far-UV CD secondary structure spectra were recorded after the experiment on the sample that had incrementally increased in temperature from 24 to 72°C over approximately 8 h, on a fresh protein sample at 24°C, and a CpxA sample subjected to 6 M guanidine hydrochloride (GndCl) with boiling for 30 minutes (Figure 3-5B). The spectrum collected at 24°C exhibited the distinctive traits characteristic of a well folded protein, with minima at 208 and 222 nm that represent the major CD signals attributed to α-helical secondary structure. The spectrum recorded from the CpxA sample at 72°C revealed a loss of these minima, while the spectrum from CpxA in 6 M GndCl is typical of proteins composed of random coil, and corresponds to fully denatured protein. These results demonstrate that the CpxA periplasmic domain was partially unfolded at 72°C, having lost much of its α-helical character, although it still retained evidence of some secondary structure.

Secondary structure deconvolution was performed on the normalized CD spectra for CpxA(31-163) at 24°C, 72°C, and melted in 6 M GndCl using the Linear Combination (Fasman) (12), and the K2D3 (18) algorithms (Figure 3-5C). While the absolute values varied, both methods calculated that the CpxA sensor domain possessed a reduced α-helical content at 72°C compared to 24°C, while the β-strand elements were retained. Therefore, the CpxA(31-163) β-strand secondary structure was determined to be stable with increases in temperature to at least 72°C. Additionally, both the α-helix and β-strand elements were stable at 24°C, thus these studies demonstrated that the CpxA construct was robust enough for crystallization trials at room temperature.

Preliminary crystallization of CpxA(31-163)

DLS experiments were performed in order to qualitatively determine the minimal salt concentration to be used in the CpxA crystallization buffer. Peaks in

the size distribution histograms near an R_h of approximately 10 nm, corresponding to soluble, folded proteins, and 1000 nm, representing the range for aggregates, were analyzed (not shown). For samples with 500-300 mM NaCl, the peak at around 10 nm was consistently very large relative to a small peak at around 1000 nm. For samples with 250-50 mM NaCl the peak corresponding to aggregate was markedly larger, in comparison. Therefore, this study implies that CpxA(31-163) was in a soluble, non-aggregated state in salt concentrations ranging from 500 mM to 300 mM NaCl in Tris-HCl pH 7.5, and thus the CpxA periplasmic domain construct should be stable in crystallization buffers with salt concentrations as low as 300 mM NaCl.

In order to obtain high resolution structural information about the periplasmic sensor domain of CpxA, purified CpxA(31-163) was subjected to initial screening for protein crystals. In total, twelve hits were identified from seven commercial sparse matrix screens. The crystal morphology ranged from microcrystals to needles, plates and chunky crystals. Of the twelve, four of the hits could not be harvested for the next step of screening. These included tiny needles and microcrystals. The eight remaining crystal hits were screened by X-ray diffraction. Three of the harvested crystal hits had no detectable diffraction. Since these hits included needles and microcrystals, it is still possible that they were protein and the diffraction was not strong enough to detect. Of the five remaining hits, two were protein or very likely protein with weak diffraction, and three were protein with strong diffraction (Table 3-1, Figure 3-6). Although diffraction from the thin plate crystals in condition JCSG$^+$ F6 was very weak and the protein content was not conclusive, this condition was added to the list of protein hits due to its notable similarity to the condition with the best-diffracting crystals, JCSG$^+$ F3.

One full dataset was collected from a crystal in condition JCSG$^+$ F3 (Figure 3-7). The data were indexed but scaling was unsuccessful. The crystal was highly

Table 3-1. Initial CpxA(31-163) protein crystal hits

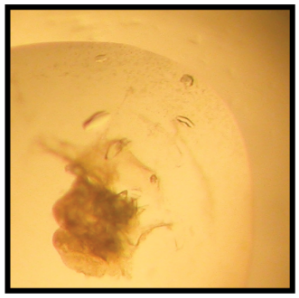
Screen Name and Code	Condition	Protein Concentration	Approximate Resolution	Preliminary Indexing	Additional Information
JCSG$^+$ F3	0.1 M Tris pH 8.5, 20% (v/v) MPD (final pH 8)	6.1 mg/mL	~3 Å (2 crystals)	P222	Dataset collected High mosaicity Preliminary unit cell: a = 37 Å, b = 43 Å, c = 189 Å, α=β=γ = 90°
Wizard II N26	30% (v/v) PEG-400; 0.1 M CHES pH 9.5	6.1 mg/mL	~3.3-3.5 Å (2 crystals)	P222	
Classics E3	35% (v/v) Dioxane	4.1 mg/mL chunky 6.1 mg/mL plates	~ 3-5 Å (3 crystals)		Different morphologies 4.1 mg/mL has better diffraction
JCSG$^+$ A12	0.2 M potassium nitrate, 20% (w/v) PEG 3350	6.1 mg/mL	~12 Å (1 crystal)		
JCSG$^+$ F6	0.1 M Bicine pH 8.5, 10% (v/v) MPD (final pH 9)	4.1 mg/mL	Very weak (1 crystal)		Inconclusive

JCSG+ F3
6.1 mg/mL
0.1 M Tris pH 8.5,
20% (v/v) MPD (final pH 8)

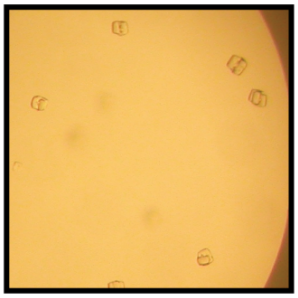
Wizard II N26
6.1 mg/mL
30% (v/v) PEG-400;
0.1 M CHES pH 9.5

Classics E3
4.1 mg/mL
35% (v/v) Dioxane

Classics E3
6.1 mg/mL
35% (v/v) Dioxane



JCSG+ A12
6.1 mg/mL
0.2 M potassium nitrate,
20% (w/v) PEG 3350



JCSG+ F6
4.1 mg/mL
0.1 M Bicine pH 8.5,
10% (v/v) MPD (final pH 9)

Figure 3-6. Identification of initial crystallization conditions for CpxA(31-163). The best six protein crystal hits in five different conditions are shown. JCSG$^+$ A12 and JCSG$^+$ F6 correspond to conditions that resulted in protein or very likely protein crystals with weak diffraction while conditions for JCSG$^+$ F3, Wizard II N26, and Classics E3 produced protein crystals with strong diffraction. The screen name and code, protein concentration used, and mother liquor are indicated below the images.

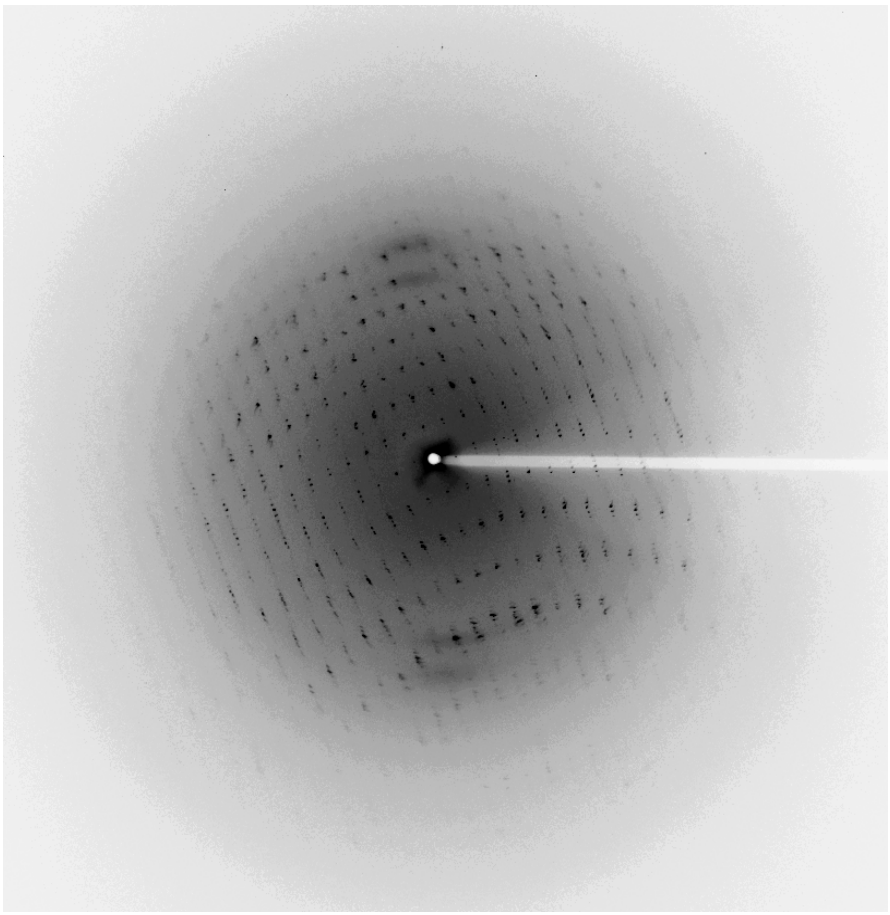


Figure 3-7. Preliminary X-ray diffraction from a crystal of CpxA(31-163). Representative image of initial diffraction obtained from a CpxA(31-163) crystal grown in condition JCSG$^+$ F3 (0.1 M Tris pH 8.5, 20% (v/v) MPD; final pH 8), with data extending out to approximately 3.2 Å resolution. No additional cryoprotectant was added to the crystallization solution. This image was collected at the Canadian Light Source beamline CMCF 08ID-1 (Saskatoon, Canada).

mosaic, which was to be expected for a crystal that was insufficiently cryoprotected. The crystal was assigned a preliminary lattice of primitive orthorhombic, and a preliminary space group of P222. A preliminary unit cell was determined to have the approximate dimensions a = 37 Å, b = 43 Å, c = 189 Å, α=β=γ = 90°. A brief search of structures in the RCSB PDB database yielded no structures with this approximate unit cell. Additionally, SDS-PAGE analysis of two melted screened crystals from condition JCSG$^+$ F3 revealed a band with a mobility that is consistent with CpxA(31-163) (not shown), therefore it is very likely that these highly-diffracting crystals contain the periplasmic domain of CpxA.

Discussion

The molecular foundation for signal sensing and transmission by the periplasmic domain of the histidine kinase, CpxA, remains unknown. In this study, a biochemical and structural investigation of the CpxA sensor domain was initiated. First, a periplasmic domain construct, CpxA(31-163), was selected, purified, and conditions in which the protein was soluble and thermally stable were established. This protein domain was observed to exist as two species, one twice the size of the other, which likely correspond to a tetramer-dimer equilibrium. The domain was folded at both pH 5.8 and 8.0, representative of the off- and on-states of the Cpx pathway, respectively (8). The sensor domain was found to have a mixed α/β secondary structure composition that decreased in α-helical content slightly upon a shift to alkaline pH. A portion of the domain, attributed to α-helical content, unfolded with an approximate T_m of 34°C, while the β-strand elements were thermostable. Strong protein crystal hits were obtained from preliminary crystallization trials, which will be subjected to further optimization.

Full-length HKs are known to function as homodimers. Self-association of the subunits is mediated by the dimerization domains, which comprise part of the cytoplasmic histidine kinase module in each molecule. Several examples in the literature, however, point to the weak association of isolated single PDC domains in solution. A size-exclusion chromatography study revealed that the periplasmic domain of CitA formed a monomer-dimer equilibrium in the presence and the absence of the ligand citrate, with the majority of the protein existing in the monomeric state in both cases (30). In addition, while equilibrium centrifugation experiments suggested that the sensor domain of PhoQ existed as a monomer in the presence and absence of divalent cations (36), a crystal structure of a functional PhoQ dimer was recently determined, providing evidence for an association (4). Similarly, results from gel filtration chromatography indicated that the periplasmic domain of DcuS was monomeric in the presence and absence of ligand, however analysis of sedimentation equilibrium analytical ultracentrifugation data demonstrated the existence of a monomer-dimer equilibrium (6). Together, these examples suggest that the intrinsic affinity for self-association of the isolated single PDC domains in solution is very low. In contrast, size-exclusion chromatography and multi-angle laser light scattering experiments reproducibly demonstrated that the periplasmic domain of CpxA existed in two oligomeric states, one of which was double the molar mass of the other (Figures 3-2B and 3-3). Although the inseparable and overlapping species introduced a source of error into the mass measurements, it is very likely that these oligomeric states correspond to a CpxA dimer-tetramer equilibrium. For a definitive answer, the species must be resolved completely using a high-resolution size-exclusion column and the results could be verified using protein crosslinking *in vivo* or native gel electrophoresis. Interestingly, there is some evidence for the tetramerization of other isolated PDC domains. In the assessment of the analytical ultracentrifugation data from DcuS, described above, the residuals from data analysis were improved slightly by including a

small contribution from a tetramer in the monomer-dimer equilibrium model (6). Regardless, the CpxA periplasmic domain presents an example of strong oligomerization, with the predominant state corresponding to the species of greater mass, that is unlike the examples of weak dimerization that have previously been described.

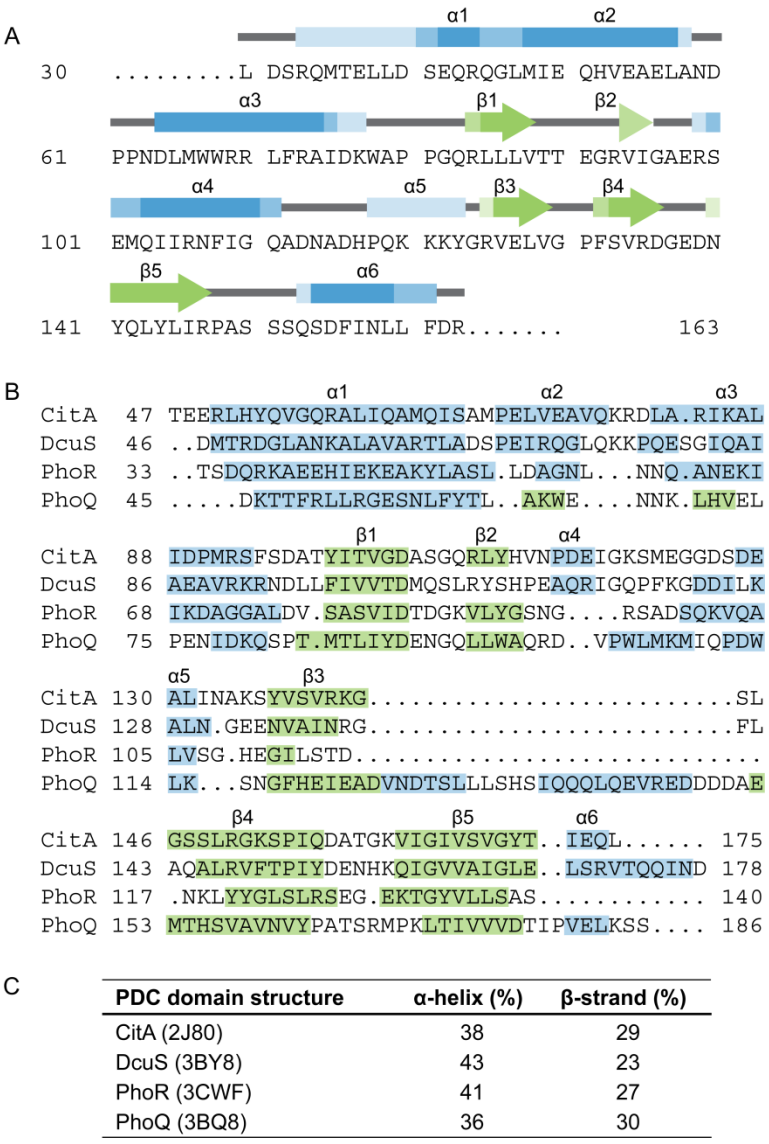
A conformational change in the extracytoplasmic PAS-like PDC domain has been proposed to be involved in the signalling mechanism for sensor HKs including CitA (31), DcuS (6), PhoQ (10, 34), and DctB (6, 38). Using far-UV CD, small differences were observed in the α-helical secondary structure composition of the CpxA sensor domain in response to alkaline pH, an activating signal for the Cpx stress pathway (8) (Figure 3-4A and 3-4B). Additionally, CD melting curves revealed that a portion of the CpxA periplasmic domain, attributed to α-helical content, had a very low T_m of 34°C (Figure 3-5). These results might reflect a pH- or temperature-dependent instability of particular α-helices in solution. If the α-helices affected in these experiments correspond to the N- and C-terminal secondary structure elements (Figure 3-1), then these changes may not occur in an intact molecule where they might be stabilized by direct connections with the transmembrane helices that flank the periplasmic domain. Alternatively, the α-helical content affected in these studies might be intrinsically labile and primed for structural rearrangements upon perception of signals such as alkaline pH. While previous studies have suggested that a conformational change occurs upon signal sensing in several extracytoplasmic HK domains, whether a structural change will be involved in CpxA signalling remains to be determined.

While the majority of HK periplasmic domains are predicted to adopt a PAS-like PDC fold (3, 33), several other sensor domain topologies exist. In order to gain insight into the sensory and signalling capabilities of the periplasmic domain of CpxA, primary, secondary, and tertiary structural analyses were

performed. Far-UV CD experiments confirmed that the CpxA sensor domain contains a mixed α/β composition, which is consistent with the secondary structure predictions generated by the programs Porter (26), Jpred3 (7), and PsiPred v3.0 (16, 20) (Figures 3-1 and 3-4). In addition, the proportions of α-helical and β-strand secondary structure elements that were calculated from deconvolutions of the CD spectra at pH 5.8 and 8.0 (Figure 3-4B) agree reasonably well not only with those from the secondary structure predictions made from the primary sequence (Figure 3-4C), but also with the percentages of α-helical and β-strand content from single PDC domains in the protein structure database, CitA, DcuS, PhoR, and PhoQ (Figure 3-8B and 3-8C). Furthermore, a comparison of the linear arrangement of secondary structure elements from the CpxA secondary structure predictions (Figure 3-8A) with that of the PDC domains (Figure 3-8B) reveals, generally, an analogous structural organization: αααββααββββα. The similarities observed in this evaluation imply that the CpxA sensor domain will likely form a PAS-like PDC fold.

CD melting curve data, paired with far-UV CD spectral analysis of the periplasmic domain of CpxA, also supports the formation of a PDC fold. The β-strand secondary structure elements were determined to be relatively thermostable (Figure 3-5). This is consistent with the putative protection attained through packing of the β-strands in the core of the PDC domain. Likewise, the α-helices surround the β-sheet in the PDC domain, which could result in an increased susceptibility of the helices to denaturation. The localization of the α-helices to the exterior of the domain would agree with the instability observed for some of the α-helical elements in the CpxA periplasmic segment (T_m ~34°C).

Given the evidence supporting the formation of a PDC fold, the CpxA sensor domain was subjected to crystallization screens with promising initial results. Strong diffraction was obtained from three protein hits, and another



PDC domain structure	α-helix (%)	β-strand (%)
CitA (2J80)	38	29
DcuS (3BY8)	43	23
PhoR (3CWF)	41	27
PhoQ (3BQ8)	36	30

Figure 3-8. The predicted periplasmic domain of CpxA shares a similar arrangement of secondary structure elements with other HK single PDC domains. A) Secondary structure summary for the putative periplasmic domain of CpxA. Predictions were made using the programs Porter (26), Jpred3 (7), and PsiPred v3.0 (16, 20). The three shades of colour, from light to dark, illustrate the number of programs from one to three, respectively, that predict an element of secondary structure for each residue. B) Structure-based sequence alignment of single PDC domain proteins CitA (PDB ID 2J80), DcuS (3BY8), PhoR (3CWF), and PhoQ (3BQ8). PhoQ contains a large insertion between the third and fourth beta strands of CitA. Residues comprising α-helical secondary structure elements are highlighted in blue, and β-strands in green, with elements labelled according to the structure of CitA (2J80). The structure-based alignment was generated, with minor manual modifications, using Dali (13, 14) with CitA (2J80) as the query molecule. Secondary structure was assigned by DSSP (1, 17), and numbering corresponds to the full protein sequence (19). C) Summary of the percentages of α-helical and β-strand content for the PDC domains in B).

two hits were protein or very likely protein and diffracted weakly (Table 3-1, Figure 3-6). These conditions will be optimized to improve the resolution and mosaicity of the crystals, and screens will be performed to select for adequate cryoprotectants.

Here, we suggest that the periplasmic sensor domain of CpxA forms a PAS-like PDC fold. We provide the first example of strong oligomerization of an isolated putative sensor HK PDC domain. We identify that the β-strand secondary structure elements are stable, while the α-helical component is much more labile and could possibly be involved in a structural rearrangement upon stress response activation by a signal such as alkaline pH. And finally, we establish crystallization conditions for the determination of a high resolution structure, which will ultimately provide insight into the fold and characteristics of the domain that are specific for its signalling function.

References

1. Andersen CA, Palmer AG, Brunak S, Rost B. 2002. Continuum secondary structure captures protein flexibility. *Structure* 10: 175-84

2. Buelow DR, Raivio TL. 2005. Cpx signal transduction is influenced by a conserved N-terminal domain in the novel inhibitor CpxP and the periplasmic protease DegP. *J Bacteriol* 187: 6622-30

3. Chang C, Tesar C, Gu M, Babnigg G, Joachimiak A, et al. 2010. Extracytoplasmic PAS-like domains are common in signal transduction proteins. *J Bacteriol* 192: 1156-9

4. Cheung J, Bingman CA, Reyngold M, Hendrickson WA, Waldburger CD. 2008. Crystal structure of a functional dimer of the PhoQ sensor domain. *J Biol Chem* 283: 13762-70

5. Cheung J, Hendrickson WA. 2010. Sensor domains of two-component regulatory systems. *Curr Opin Microbiol* 13: 116-23

6. Cheung J, Hendrickson WA. 2008. Crystal structures of C4-dicarboxylate ligand complexes with sensor domains of histidine kinases DcuS and DctB. *J Biol Chem* 283: 30256-65

7. Cole C, Barber JD, Barton GJ. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36: W197-201

8. Danese PN, Silhavy TJ. 1998. CpxP, a stress-combative member of the Cpx regulon. *Journal of Bacteriology* 180: 831-9

9. Fleischer R, Heermann R, Jung K, Hunke S. 2007. Purification, reconstitution, and characterization of the CpxRAP envelope stress system of *Escherichia coli*. *J Biol Chem* 282: 8583-93

10. Garcia Vescovi E, Soncini FC, Groisman EA. 1996. Mg^{2+} as an extracellular signal: environmental regulation of Salmonella virulence. *Cell* 84: 165-74

11. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31: 3784-8

12. Hofmann A. 2009. ACDP - a Java application for data processing and analysis of protein circular dichroism spectra. *Journal of Applied Crystallography* 42: 137-9

13. Holm L, Rosenstrom P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38 Suppl: W545-9

14. Holm L, Sander C. 1996. Mapping the protein universe. *Science* 273: 595-603

15. Jones CH, Danese PN, Pinkner JS, Silhavy TJ, Hultgren SJ. 1997. The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems. *EMBO Journal* 21: 6394-406

16. Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292: 195-202

17. Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577-637

18. Louis-Jeune C, Andrade-Navarro MA, Perez-Iratxeta C. 2011. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins* 80: 374-381

19. Magrane M, Consortium U. 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011: bar009

20. McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404-5

21. Mileykovskaya E, Dowhan W. 1997. The Cpx two-component signal transduction pathway is activated in *Escherichia coli* mutant strains lacking phosphatidylethanolamine. *Journal of Bacteriology* 179: 1029-34

22. Nakayama S-I, Watanabe H. 1995. Involvement of *cpxA*, a sensor of a two-component regulatory system, in the pH-dependent regulation of expression of *Shigella sonnei virF* gene. *Journal of Bacteriology* 177: 5062-9

23. Nevesinjac AZ, Raivio TL. 2005. The Cpx Envelope Stress Response Affects Expression of the Type IV Bundle-Forming Pili of Enteropathogenic *Escherichia coli*. *J Bacteriol* 187: 672-86

24. Otto K, Silhavy TJ. 2002. Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. *Proc Natl Acad Sci U S A* 99: 2287-92

25. Otwinowski Z, Minor W. 1997. Processing of X-ray Diffraction Data Collected In Oscillation Mode. In *Methods in Enzymology*, ed. J Charles W. Carter, RM Sweet, pp. 307-26. New York: Academic Press

26. Pollastri G, McLysaght A. 2005. Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 21: 1719-20

27. Price NL, Raivio TL. 2009. Characterization of the Cpx regulon in *Escherichia coli* strain MC4100. *J Bacteriol* 191: 1798-815

28. Raivio TL, Popkin DL, Silhavy TJ. 1999. The Cpx envelope stress response is controlled by amplification and feedback inhibition. *Journal of Bacteriology* 181: 5263-72

29. Raivio TL, Silhavy TJ. 1997. Transduction of envelope stress in *Escherichia coli* by the Cpx two-component system. *Journal of Bacteriology* 179: 7724-33

30. Reinelt S, Hofmann E, Gerharz T, Bott M, Madden DR. 2003. The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. *J Biol Chem* 278: 39189-96

31. Sevvana M, Vijayan V, Zweckstetter M, Reinelt S, Madden DR, et al. 2008. A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA. *J Mol Biol* 377: 512-23

32. Snyder WB, Davis LJB, Danese PN, Cosma CL, Silhavy TJ. 1995. Overproduction of NlpE, a new outer membrane lipoprotein, suppresses the toxicity of periplasmic LacZ by activation of the Cpx signal transduction pathway. *Journal of Bacteriology* 177: 4216-23

33. Szurmant H, White RA, Hoch JA. 2007. Sensor complexes regulating two-component signal transduction. *Curr Opin Struct Biol* 17: 706-15

34. Vescovi EG, Ayala YM, Di Cera E, Groisman EA. 1997. Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg^{2+} and Ca^{2+}. *J Biol Chem* 272: 1440-3

35. Vogt SL, Raivio TL. 2012. Just scratching the surface: an expanding view of the Cpx envelope stress response. *FEMS Microbiol Lett* 326: 2-11

36. Waldburger CD, Sauer RT. 1996. Signal detection by the PhoQ sensor-transmitter. Characterization of the sensor domain and a response-impaired mutant that identifies ligand-binding determinants. *J Biol Chem* 271: 26630-6

37. Weatherspoon-Griffin N, Zhao G, Kong W, Kong Y, Morigen, et al. 2011. The CpxR/CpxA two-component system up-regulates two Tat-dependent peptidoglycan amidases to confer bacterial resistance to antimicrobial peptide. *J Biol Chem* 286: 5529-39

38. Zhou YF, Nan B, Nan J, Ma Q, Panjikar S, et al. 2008. C4-dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. *J Mol Biol* 383: 49-61

CHAPTER 4

Discussion

In this chapter, recent work in the field will be highlighted, and possible mechanisms for signal detection by CpxA will be considered. The putative interactions between CpxA and CpxP, and CpxP and DegP will be discussed, and the possible exploitation of the Cpx stress response as a target for therapeutic applications will be presented.

Recent advances

Since this work was carried out, there have been remarkable developments in the elucidation of new structures and functions for CpxP and the structural homologues of CpxP identified in Chapter 2: Spy, ZraP, and CnrX. In the span of one month, a second structure of Spy and an additional structure of CpxP were published concurrently with our work, all of which reveal similar structural arrangements (12, 19, 25, 30). A great deal of progress has also been made with respect to the attribution of functions to these proteins.

Before last year, little was known about the function of the closest homologue of CpxP, the periplasmic spheroplast protein y (Spy). Studies had revealed that Spy expression was upregulated by both the Cpx and Bae envelope stress responses (20, 28, 29), which suggested a role in mediating a reaction to extracytoplasmic disturbances. Apart from experiments that established Spy was unable to inhibit the Cpx pathway in a manner similar to CpxP, limited functional information existed for Spy (21). In an innovative investigation examining the effects of introducing periplasmic protein folding mutants into *Escherichia coli*, the expression of Spy was observed to be greatly enhanced in the presence of the misfolded proteins (19). Further characterization revealed that Spy was a highly effective ATP-independent periplasmic chaperone that was capable of both reducing protein aggregation and facilitating periplasmic protein refolding. Based on these experiments, a model was proposed whereby Spy might bind to and shield unfolded regions of substrates, preventing their aggregation. The

apparent flexibility observed in the structure of Spy might assist in protein refolding and release. Additional findings suggested that Spy likely does not have a regulatory role comparable to CpxP; however CpxP exhibits weak chaperone activity *in vitro*.

In a separate study, CpxP was also found to demonstrate a mild chaperone activity that was deemed to be sufficient for titration of CpxP from CpxA by misfolded proteins upon Cpx pathway activation (30). In support, an elongated hydrophobic groove located between the symmetry-related protomers on the convex surface of CpxP was identified as the recognition site for misfolded substrates such as pilus proteins. Peptide array screens were also used to show that the purified CpxA periplasmic domain directly interacts with peptides encompassing helix α1 and the C-terminus of CpxP, and in reverse, purified CpxP interacts with the C-terminus of CpxA sensor domain peptides, and specifically with negative charges. From these results the authors suggest that CpxP inhibits the Cpx pathway, and protects CpxA from inducing signals, through an interaction between the α1 helix of the basic patch in CpxP and negative charges in the CpxA sensor domain.

In another report, the chaperone assay established for Spy, above, was used to compare the function of the newfound structural homologue, ZraP, to the activities of Spy and CpxP (1). ZraP was not only shown to possess ATP-independent, zinc-dependent periplasmic chaperone activity that reduced protein aggregation and assisted periplasmic protein refolding akin to Spy, but also demonstrated regulatory function as a repressor of the zinc responsive ZraSR TCS, similar to the role of CpxP.

Finally, through a comparison of structures obtained in the presence of the activating metals nickel and cobalt, and the non-activating metal zinc, the CnrX dimer was found to undergo small changes in molecular packing that suggested an allosteric mechanism for induction of the CnrYXH pathway in

Cupriavidus metallidurans (26). It would be interesting to explore whether the small changes observed in the structure of CpxP in response to pH (Chapter 2) are similar to those illustrated in CnrX.

Taken together, many of the examples of new developments outlined here were assisted by combining information from a new class of structurally similar proteins that could not be identified by sequence alone. Comparisons of these structural homologues have already provided insight into the role of the LTXXQ motif, and common periplasmic chaperone and signalling roles. Hopefully, research on this class of periplasmic proteins will expand to decipher the mechanisms that regulate their numerous functional roles.

Molecular nature of the signal sensed by CpxA

Although significant advances have been made in the number of sensing domain structures and in the identification of specific pathway stimuli, the molecular mechanisms through which the signals are recognized by many HKs, including CpxA, are still unknown. The wide variety of activating signals detected by CpxA could be sensed directly or indirectly through interactions with auxiliary signal transduction proteins.

Many of the HK PDC domains have been shown to bind small molecules or ions directly. These sensing domains share little primary sequence identity, and this sequence variability makes them versatile for signal perception. From comparisons of the CpxA sequence and secondary structure organization with alignments of several PDC domains, it is not immediately clear what part of CpxA might be involved in signal sensing (Figure 3-8). Data can be extracted from mutagenesis studies (10), and future structures of the CpxA periplasmic domain could be analyzed for variations to the core PDC fold that might indicate putative

ligand interaction sites and give us insight into the nature of the signal sensed by CpxA.

Additional support for the direct sensing of stimuli by CpxA comes from *in vitro* assays involving the misfolded maltose binding protein variant MalE219 (11). This mutant protein activated the Cpx response in a reconstituted system through an increased rate of phosphotransfer from CpxA to CpxR. Since no other factors were involved, the mutant misfolded protein was proposed to directly contact the CpxA periplasmic domain, inducing the Cpx pathway through an allosteric mechanism that is independent of CpxP and NlpE.

The outer membrane lipoprotein NlpE, which is required for activation of the Cpx response *via* adhesion to abiotic surfaces (18), has also been suggested to have the capacity to contact CpxA directly. With unfolding of a portion of NlpE, this lipoprotein was calculated to be long enough to reach and activate CpxA imbedded in the inner membrane (7).

Signals may also be detected by CpxA indirectly through a co-sensor such as CpxP. The PDC domain of the quorum sensor LuxQ senses stimuli through ligand binding to a constitutively associated protein, LuxP (16, 17). Therefore, it is possible that CpxP may also associate with CpxA for signal detection. Recently, chaperone activity has been demonstrated for CpxP, which implies that it can bind to misfolded proteins (19, 30). The convex surface of CpxP is comprised mainly of mixed hydrophobic and acidic patches (25, 30), and a particular hydrophobic region on this surface has been suggested to bind misfolded substrates (30). Thus, CpxP could possibly aid CpxA in the detection of misfolded periplasmic proteins.

To further complicate the elucidation of the signalling properties of CpxA, our studies indicate that the periplasmic domain of CpxA exists in both a dimeric and tetrameric state (Chapter 3). While this might represent an artifact due to

the isolated context of this domain, this unique and predominant tetrameric species should be explored in a signalling environment.

Many details pertaining to the mechanisms by which CpxA recognizes specific stimuli remain to be determined. Not only are the nature of the inducing cue and the oligomeric state of CpxA undefined, but whether CpxA senses signals directly, requires an alternative auxiliary factor for signal detection, or due to the multitude of stimuli that could activate this pathway, whether both are necessary to integrate Cpx pathway inducing signals is not yet clear.

The putative interaction between the periplasmic sensing domain of CpxA and CpxP

While numerous genetic and biochemical studies imply that CpxA and CpxP interact, as of yet, no direct association between the wild-type proteins has been established. The first suggestion that a periplasmic regulatory molecule might interact with CpxA came from an analysis of CpxA periplasmic gain-of-function mutations that resulted in the desensitization of CpxA to typical inducing signals. The phenotypes associated with these mutations alluded to the association, under non-inducing conditions, of CpxA with a periplasmic factor that would reduce the ratio of CpxA kinase to phosphatase activity (23). Additional studies demonstrated that inhibition of the Cpx pathway by CpxP was dependent on the CpxA periplasmic sensing domain (22). Further efforts revealed that Cpx signal transduction could not be completely activated when CpxP was fastened to the inner membrane *via* a maltose binding protein derivative, which suggests an inhibitory interaction between CpxA and CpxP (21). Accordingly, the autophosphorylation activity of CpxA was shown to be inhibited by approximately 50% upon incorporation of purified CpxP into reconstituted proteoliposomes containing purified CpxA (5). Recently, experiments using peptide libraries revealed that the purified CpxA periplasmic domain interacts

with peptides from the α1 helix and C-terminus of CpxP, while purified CpxP interacts with acidic residues in CpxA sensor domain peptides (30). The same group developed a new method called Membrane-SPINE, which integrated formaldehyde cross-linking and mass spectrometry analyses to identify protein interaction partners of membrane proteins *in vivo* (15). Under the Cpx-inducing conditions of elevated pH, the authors were able to identify CpxR and a few other proteins as interacting partners of membrane-bound CpxA from *Salmonella typhimurium*. However, there was no mention of an attempt to identify a CpxP-CpxA interaction under non-inducing conditions.

Our studies not only identified oppositely charged concave and convex surfaces of CpxP that might mediate interactions with other proteins or substrates, but also two solvent-exposed residues, R60 and D61, that convey *cpxP* loss-of-function phenotypes when mutated (inability to inhibit the Cpx response)(2) that could possibly be involved in an interaction with CpxA. These two residues reside in the α1 helix of CpxP, which is a region that was found to interact with the sensor domain of CpxA in peptide binding studies (30). The α1 helix of CpxP, therefore, presents a plausible site for intermolecular interactions with CpxA that should be the subject of further study. Although the results from many investigations have outlined a likely association between CpxP and CpxA, specific contact between the full-length proteins has yet to be demonstrated.

PDC fold domains appear to serve as versatile ligand-binding modules. However, there are no examples in the literature of associations that resemble a periplasmic protein-single PDC domain interaction. Taking into account the evidence that suggests that the CpxA sensing domain also adopts a PDC fold (Chapter 3), it is likely that CpxP utilizes a distinct mode of interaction from those seen previously. The molecular mechanisms by which the periplasmic protein CpxP acts to control signalling by the sensor kinase, CpxA, are not known. Further biochemical studies are necessary to provide confirmation of a direct

physical protein-protein interaction, however recent work provides some insight into the potentially novel mode of recognition of a periplasmic regulatory protein by a sensor HK.

The recognition of CpxP by DegP

DegP has long been linked with the Cpx pathway, with an established role in suppressing the toxicity resulting from misfolded proteins in the periplasm (3). DegP possesses both ATP-independent chaperone and serine protease activities. In the inactive state, DegP exists as a hexamer ($DegP_6$), but upon binding of unfolded protein substrates DegP assembles into large cages that can incorporate entire proteins into their hydrophobic core ($DegP_{12/24}$). Here, nearly folded proteins are stabilized by the DegP chaperone function, while unfolded proteins are degraded (24).

In the Cpx stress response pathway, CpxP is degraded by DegP under certain conditions such as an alkaline extracellular environment and the overexpression of pilus subunits (2, 8). Additionally, CpxP was demonstrated to be required for the degradation of misfolded pilus subunits by DegP (8). From these findings, CpxP has been proposed to have a second role as an adaptor that assists in the degradation of misfolded proteins by DegP during episodes of envelope stress (2, 8). It is unclear, however, as to whether CpxP is recognized by DegP in a specific or a non-specific manner.

It is possible that CpxP might become unfolded under the same conditions that induce other periplasmic protein misfolding, and as a consequence DegP might recognize CpxP as a denatured substrate (13). Alternatively, CpxP might bind misfolded proteins and these would be detected by DegP, while CpxP might be degraded as a consequence of its association with the misfolded proteins. Since several biophysical techniques demonstrated that

CpxP does not undergo a shift in oligomeric state or become unfolded in response to the inducing cue of alkaline pH *in vitro* (Chapter 2), the initial option is not likely. On the other hand, CpxP has been shown to possess moderate chaperone function (19, 30). It is plausible that CpxP can bind misfolded proteins through a hydrophobic patch on its convex surface and transport them to DegP for proteolysis (30).

Although the manner in which CpxP is recognized by DegP and targeted for proteolysis is still unknown, recent evidence supports a chaperone function rather than protein instability at alkaline pH. Whether CpxP is detected specifically by DegP through a target sequence or a particular charged surface of CpxP, or nonspecifically through recognition of the associated misfolded proteins remains to be explored. Further study may help to establish a role for CpxP not only as a negative regulator of the Cpx envelope stress response, but also possibly as an adaptor for DegP-mediated proteolysis of misfolded periplasmic proteins.

An updated overview of the Cpx stress response

While much of the Cpx envelope stress response pathway in *E. coli* remains the same as previously described (Figure 1-6) our work and recent studies have revealed new details pertaining to several aspects of the model (Figure 4-1).

First, the two crystal structures reveal that CpxP forms an antiparallel dimer of intertwined α-helices and these structures are representative of CpxP in solution (25, 30). The periplasmic domain of CpxA is predicted to adopt a PAS-like PDC fold similar to CitA, and exists as two species that likely correspond to a tetramer-dimer equilibrium (Chapter 3). It is not known whether the presence of these two species represents an artifact or physiologically relevant oligomeric

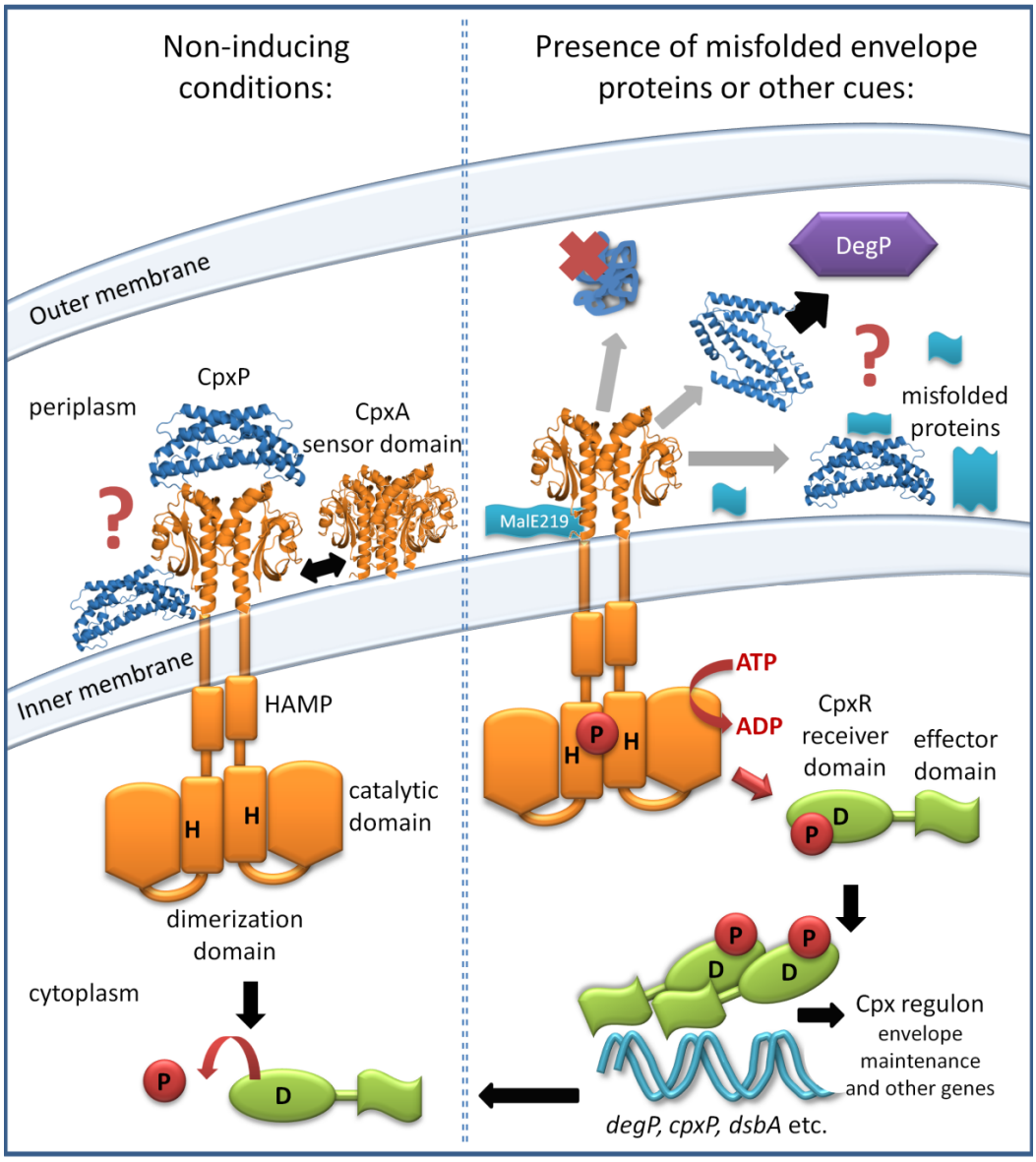


Figure 4-1. Revised model of the Cpx envelope stress response pathway. Model for activation and signal transduction through the Cpx two-component system. See text for details. "H" and "D" represent the conserved histidine and aspartate residues involved in phosphotransfer, respectively. "P" designates a phosphoryl group. The CpxA sensor domain is represented by the crystal structure of the CitA dimer (PDB ID 2J80).

states. It will be interesting to learn what oligomeric state(s) are captured in the various crystal forms observed in Chapter 3.

In non-activating conditions such as low pH (pH 5.8), the pathway is retained in an inactive state through a putative interaction of CpxP with the sensor domain of CpxA. Although the evidence supporting this interaction is steadily growing, a direct association between full-length proteins has not yet been demonstrated. However, studies reveal that acidic residues in CpxA may interact with at least the basic region of the α1 helix of CpxP (30).

The CpxA periplasmic domain senses a wide array of signals either directly or indirectly through an unknown mechanism. The misfolded protein MalE219 has been proposed to activate CpxA allosterically (11), while the auxiliary protein CpxP may act as a co-sensor to indirectly detect misfolded proteins.

Upon sensing an inducing cue, such as alkaline pH or the overexpression of pilus subunits, the pathway inhibition imposed by CpxP is removed. Three hypotheses have been proposed for the mechanism by which CpxP facilitates proteolysis of misfolded pilus proteins and is degraded by DegP upon activation of the Cpx stress response. Biophysical studies have indicated that CpxP maintains a dimeric state, and only very minor structural rearrangements are observed upon a shift to alkaline pH *in vitro* (Chapter 2). As CpxP is not unfolded by a mildly alkaline extracellular environment, it has been suggested that CpxP does not present to DegP as a severely misfolded protein substrate targeted for proteolysis under these conditions. It remains to be investigated whether the small pH-dependent structural adjustment in CpxP is important for its function. Recently, CpxP was discovered to exhibit mild chaperone activity, and an extended hydrophobic groove on its convex surface was identified as a site for binding of misfolded proteins (19, 30). It is plausible that CpxP could utilize its modest chaperone activity to bind misfolded proteins and protect them from

forming toxic aggregates in the periplasm, but does not carry out protein refolding. Either CpxP or the bound unfolded proteins might be recognized by DegP, which would assemble into a large cage to incorporate the CpxP-bound unfolded proteins into its proteolytic chamber for degradation.

Pathway activation may initiate through a conformational change in the CpxA sensing domain, comparable to the structural rearrangements proposed for other extracytoplasmic PDC domains. While the results need to be substantiated in the context of a full-length protein, α-helices in the isolated CpxA periplasmic domain appear to be unstable with respect to pH and temperature, and might reflect labile regions that could be involved in a structural change in this domain (Chapter 3).

Subsequently, the signal may be transduced through conformational changes in the transmembrane region and HAMP linker to induce the autokinase activity of CpxA. The phosphoryl group is transferred to CpxR, which regulates the expression of genes in the Cpx regulon. In the absence of stimuli, CpxA phosphatase activity and the inhibitory action of CpxP return the pathway to a basal level (as described in Chapter 1).

While significant progress has been made in a brief time, numerous questions remain unresolved. One in particular that immediately draws attention is the undefined role of the highly basic concave surface of CpxP. Another concerns the observation that CpxP is only able to account for ~50% reduction in CpxA autokinase activity in a reconstituted system with purified components (5). It is quite possible that another unidentified pathway constituent may play a part in the Cpx response. Additionally, since the basic surface of CpxP strongly insinuates a role as a binding site for a negatively charged particle, it is tempting to suggest that CpxP could accommodate negatively charged macromolecules such as phospholipids, or possibly contact with a membrane, through this highly basic cavity. It is possible that CpxP could

be recruited to and remain at the inner membrane, perhaps making contact with both the periplasmic domain of CpxA and the membrane, or even inducing a slight curvature in the membrane as a potential allosteric regulatory mechanism. Misfolded proteins, such as pilus subunits, could bind to hydrophobic patches on its convex surface and titrate it away from the membrane, which might activate CpxA allosterically through disruption of a direct interaction with CpxP or release of the slight membrane curvature. Altered membrane composition, including that of a reconstituted system, alkaline pH, or salt could affect the ability of CpxP to associate with the inner membrane, which would also induce the Cpx response. While this scenario does not explain the function of the similar basic surface of Spy (12), it may be interesting to consider. Further investigation will contribute to an understanding of the mechanisms of activation and regulation of the broad response to envelope stress mediated by the Cpx TCS in *E. coli*.

The Cpx response as a potential target of antibacterial drugs

Many TCSs have recognized roles in bacterial pathogenesis. Coupled with the evidence that no TCSs have been identified in higher eukaryotes, these pathways represent an ideal target for the development of novel antimicrobials. Although its direct role appears to be primarily related to envelope protein maintenance, the Cpx response also mediates bacterial virulence through its involvement in bacterial adhesion and pilus assembly (27). The effects of the Cpx pathway in pathogenesis have been demonstrated in a number of species, making this an attractive target for structure-based drug design in an age of bacterial multi-drug resistance. Efforts to develop inhibitors of signal transduction in TCSs are already underway. Initially, the kinase core of HKs was targeted due to its high conservation, however selectivity became an issue. The current focus is aimed at the RRs and the HK sensor domains, such as those involved in quorum sensing, for increased specificity (6). Our work could reveal

another opportunity for antimicrobial design directed towards the structurally conserved PDC fold of HK periplasmic domains. Although the Cpx pathway is not essential for bacterial survival (14), inhibitors are only required to control and reduce virulence rather than directly kill the bacterium.

Interestingly, eukaryotes have already exploited the Cpx system to mediate bacterial cell death. Mammalian peptidoglycan recognition proteins have been found to bind to the cell wall of Gram-positive bacteria and to the outer membrane of Gram-negative bacteria, inducing stress responses that respond to misfolded proteins: the CssR-CssS TCS in *Bacillus subtilis* and the CpxAR stress response in *E. coli*. The effects of hyper-activating these stress responses, including membrane depolarization and inhibition of protein, RNA, and DNA synthesis, kill the bacterial cell (4, 9). Thus, the molecular details outlined by our research could provide a platform for the development of novel drugs as antibacterial agents.

References

1. Appia-Ayme C, Hall A, Patrick E, Rajadurai S, Clarke TA, Rowley G. 2012. ZraP is a periplasmic molecular chaperone and a repressor of the zinc-responsive two-component regulator ZraSR. *Biochem J* 442: 85-93

2. Buelow DR, Raivio TL. 2005. Cpx signal transduction is influenced by a conserved N-terminal domain in the novel inhibitor CpxP and the periplasmic protease DegP. *J Bacteriol* 187: 6622-30

3. Danese PN, Snyder WB, Cosma CL, Davis LJ, Silhavy TJ. 1995. The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. *Genes and Development* 9: 387-98

4. Dziarski R, Kashyap DR, Gupta D. 2012. Mammalian peptidoglycan recognition proteins kill bacteria by activating two-component systems modulate microbiome and inflammation. *Microb Drug Resist* 18: 280-5

5. Fleischer R, Heermann R, Jung K, Hunke S. 2007. Purification, reconstitution, and characterization of the CpxRAP envelope stress system of *Escherichia coli*. *J Biol Chem* 282: 8583-93

6. Gotoh Y, Eguchi Y, Watanabe T, Okamoto S, Doi A, Utsumi R. 2010. Two-component signal transduction as potential drug targets in pathogenic bacteria. *Curr Opin Microbiol* 13: 232-9

7. Hirano Y, Hossain MM, Takeda K, Tokuda H, Miki K. 2007. Structural studies of the Cpx pathway activator NlpE on the outer membrane of *Escherichia coli*. *Structure* 15: 963-76

8. Isaac DD, Pinkner JS, Hultgren SJ, Silhavy TJ. 2005. The extracytoplasmic adaptor protein CpxP is degraded with substrate by DegP. *Proc Natl Acad Sci U S A*

9. Kashyap DR, Wang M, Liu LH, Boons GJ, Gupta D, Dziarski R. 2011. Peptidoglycan recognition proteins kill bacteria by activating protein-sensing two-component systems. *Nat Med* 17: 676-83

10. Keller R, Havemann J, Hunke S. 2011. Different regulatory regions are located on the sensor domain of CpxA to fine-tune signal transduction. *Res Microbiol* 162: 405-9

11. Keller RF, Hunke S. 2009. Misfolded maltose binding protein MalE219 induces the CpxRA envelope stress response by stimulating phosphoryl transfer from CpxA to CpxR. *Res Microbiol* 160: 396-400

12. Kwon E, Kim DY, Gross CA, Gross JD, Kim KK. 2010. The crystal structure *Escherichia coli* Spy. *Protein Sci* 19: 2252-9

13. MacRitchie DM, Buelow, D.R., Price, N.L., Raivio, T.L. 2007. Two-component signaling and Gram negative envelope stress response systems. In *Bacterial signal transduction: network and drug targets*, ed. R Utsumi. Austin, TX: Landes Bioscience

14. McEwen J, Silverman PM. 1980. Genetic analysis of *Escherichia coli* K-12 chromosomal mutants defective in expression of F-plasmid functions: identification of genes *cpxA* and *cpxB*. *Journal of Bacteriology* 144: 60-7

15. Muller VS, Jungblut PR, Meyer TF, Hunke S. 2011. Membrane-SPINE: an improved method to identify protein-protein interaction partners of membrane proteins in vivo. *Proteomics* 11: 2124-8

16. Neiditch MB, Federle MJ, Miller ST, Bassler BL, Hughson FM. 2005. Regulation of LuxPQ receptor activity by the quorum-sensing signal autoinducer-2. *Mol Cell* 18: 507-18

17. Neiditch MB, Federle MJ, Pompeani AJ, Kelly RC, Swem DL, et al. 2006. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell* 126: 1095-108

18. Otto K, Silhavy TJ. 2002. Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. *Proc Natl Acad Sci U S A* 99: 2287-92

19. Quan S, Koldewey P, Tapley T, Kirsch N, Ruane KM, et al. 2011. Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. *Nat Struct Mol Biol* 18: 262-9

20. Raffa RG, Raivio TL. 2002. A third envelope stress signal transduction pathway in *Escherichia coli*. *Mol Microbiol* 45: 1599-611

21. Raivio TL, Laird MW, Joly JC, Silhavy TJ. 2000. Tethering of CpxP to the

inner membrane prevents spheroplast induction of the Cpx envelope stress response. *Molecular Microbiology* 37: 1186-97

22. Raivio TL, Popkin DL, Silhavy TJ. 1999. The Cpx envelope stress response is controlled by amplification and feedback inhibition. *Journal of Bacteriology* 181: 5263-72

23. Raivio TL, Silhavy TJ. 1997. Transduction of envelope stress in *Escherichia coli* by the Cpx two-component system. *Journal of Bacteriology* 179: 7724-33

24. Sawa J, Heuck A, Ehrmann M, Clausen T. 2010. Molecular transformers in the cell: lessons learned from the DegP protease-chaperone. *Curr Opin Struct Biol* 20: 253-8

25. Thede GL, Arthur DC, Edwards RA, Buelow DR, et al. 2011. Structure of the periplasmic stress response protein CpxP. *J Bacteriol* 193: 2149-57

26. Trepreau J, Girard E, Maillard AP, de Rosny E, Petit-Haertlein I, et al. 2011. Structural basis for metal sensing by CnrX. *J Mol Biol* 408: 766-79

27. Vogt SL, Raivio TL. 2012. Just scratching the surface: an expanding view of the Cpx envelope stress response. *FEMS Microbiol Lett* 326: 2-11

28. Yamamoto K, Ishihama A. 2005. Transcriptional response of *Escherichia coli* to external copper. *Mol Microbiol* 56: 215-27

29. Yamamoto K, Ogasawara H, Ishihama A. 2008. Involvement of multiple transcription factors for metal-induced *spy* gene expression in *Escherichia coli*. *J Biotechnol* 133: 196-200

30. Zhou X, Keller R, Volkmer R, Krauss N, Scheerer P, Hunke S. 2011. Structural basis for two-component system inhibition and pilus sensing by the auxiliary CpxP protein. *J Biol Chem* 286: 9805-14

APPENDIX A

Structural investigation of the microRNA cluster, miR-17~92

A version of this chapter is published in:

Chaulk, S. G., G. L. Thede, O. A. Kent, Z. Xu, E. M. Gesner, R. A. Veldhoen, S. K. Khanna, I. S. Goping, A. M. MacMillan, J. T. Mendell, H. S. Young, R. P. Fahlman, and J. N. M. Glover. 2011. Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. RNA Biol **8**:1105-14.

Introduction

MicroRNAs (miRNAs) make up a large family of short (20-22 nucleotide), non-coding RNAs. They have been found to regulate the expression of a number of target genes that control various physiological and developmental processes in eukaryotic organisms; however the function of the majority of these miRNAs is unknown (1, 8).

In animals, the small miRNAs are formed from large transcripts that may be up to several kb long, called primary- (pri-)miRNAs. The pri-miRNAs are processed sequentially by two double strand-specific endonucleases, Drosha and Dicer (reviewed in (10)). In the nucleus, Drosha first crops miRNA-containing RNA hairpins approximately 70-100 nucleotides in length from the pri-miRNA. These hairpins are subsequently cleaved by Dicer in the cytoplasm and acted upon by a helicase to generate the mature miRNAs. The single-stranded miRNAs are incorporated into RNA-induced silencing complexes (RISCs) that bind to specific target messenger RNA (mRNA), forming a miRNA:mRNA duplex that regulates the mRNA gene expression.

miRNAs and their precursors have been found to be mis-regulated in many cancers. In particular, the miR-17~92 pri-miRNA, a polycistron containing six miRNAs, has been implicated in human B-cell lymphoma and lung cancers amongst other cancer types (6, 7, 15). As a human oncogene it has thus earned the alternative designation, oncomiR-1 (12, 14).

As of yet, there is no tertiary structural information to describe this large class of functionally unique RNAs. To date, the role of RNA structure in miRNA biogenesis has only been considered with regard to the secondary structural features that are required for processing of transcripts by Drosha and Dicer (17, 18). The potential significance of expressing a number of miRNAs together as one entity has not been explored. Here, we have initiated an investigation of the

tertiary structure of miR-17~92 by employing biochemical and structural methods.

Using size-exclusion chromatography and native gel electrophoretic mobility shift assays, the miR-17~92 miRNA cluster was shown to adopt a more compact fold in the presence compared to the absence of physiologically relevant levels of Mg^{2+}, often an essential factor for RNA folding. This compact pri-miRNA was resistant to degradation by ribonucleases, supporting a model for formation of a compact Mg^{2+}-dependent tertiary fold.

Next, ribonuclease footprinting experiments revealed that several of the miRNA-containing hairpins at the 5' end of the pri-miRNA transcript, corresponding to miR-17, 19a and 20, were solvent exposed and accessible to ribonucleases for cleavage. In contrast, the miR-18a, 19b and 92-containing hairpins were not cleaved. Upon removal of the hairpins in the 5' region of the cluster, the previously protected miR-19b and miR-92-containing hairpins were also subject to ribonuclease digestion. In parallel, the miR-19b and miR-92 hairpins were found to be protected from hydroxy radical cleavage in the full-length cluster. Together, these experiments suggest that in the compact, folded pri-miRNA, the 5' region shields the 3' miR-19b and miR-92 hairpins from enzymatic digestion through the formation of tertiary structure.

In an *in vitro* Drosha processing assay, the miR-17, miR-19a and miR-20-containing hairpins were found to be cropped from the miR-17~92 cluster more efficiently than the miR-18a, miR-19b and miR-92-contiaining stem-loops. In addition, processing of the miR-92 hairpin was demonstrated to be less efficient in the wild-type cluster than in a mutant where the miR-92 stem-loop had previously been determined to be solvent accessible. These findings are consistent with results from miR-17~92 cluster processing in human HEK293T cells, where the largest increase in miRNA expression between the wild-type cluster and the mutant cluster just mentioned was observed for miR-92. These

processing assays suggest that tertiary structure within the intact cluster limits the accessibility of Drosha and impedes expression of the internalized miRNAs relative to the more surface-exposed miRNA-containing hairpins.

miR-92 has previously been shown to target and down-regulate integrin α5 (ITGA5) mRNA (3). Therefore, an experiment was performed to monitor the ITGA5 mRNA levels in HEK293T cells expressing the intact miR-17~92 cluster or the mutant cluster in which the miR-92 stem-loop is solvent accessible. Expression of the mutant cluster resulted in a lower integrin mRNA level than that observed for the intact pri-miRNA. This suggests that perturbations to the cluster structure expose the miR-92 miRNA hairpin for processing, which results in increased repression of the target mRNA.

The nuclear exosome is proposed to degrade pri-miRNA transcripts once the first miRNA-containing hairpin is removed by Drosha (2, 13). The expression levels of miR-17~92 cluster miRNAs were ascertained from cells with and without reduced nuclear exosome levels (achieved through RNA interference). As expected, increased miR-92 expression was detected in cells with lower nuclear exosome levels. Thus, the nuclear exosome may play a role in miRNA biogenesis for the miR-17~92 cluster. Taken together, these results point to the existence of a buried 3' core in the miR-17~92 cluster that contains the miR-19b and miR-92 RNA hairpins and is protected from Drosha by the 5' region.

In order to directly visualize the folded structure of miR-17~92, we imaged the polycistron using single particle electron microscopy (EM) of negatively-stained samples. The EM reconstruction of miR-17~92 revealed a compact globular particle with a central core of density. The 3D structure of a 3' truncation mutant lacking a portion of the predicted core was more unfolded, confirming that the deleted core region functions as a structural scaffold for packing of the remainder of the RNA.

Materials and Methods

Electron microscopy and image processing

Annealed RNAs (0.2 mg mL^{-1}) were applied to carbon-coated and glow-discharged 400 mesh Cu/Rh Maxtaform grids (Ted Pella, Inc.). The grids were subsequently washed with 16% ammonium molybdate and 2% uranyl acetate. Grid preparation using only ammonium molybdate or uranyl acetate revealed no significant differences in the appearance of the RNA particles. The grids were imaged under low electron dose conditions on a Tecnai F20 electron microscope set to 200 keV (Philips/FEI, situated in the Microscopy and Imaging Facility, University of Calgary, Alberta, Canada). The images were recorded at 50,000x magnification on Kodak SO-163 film over a defocus range from –1 µm to –2.5 µm. The micrographs were digitized (Nikon Super Coolscan 9000) and pixel averaging was applied to reach a resolution of 5.08 Å/pixel.

Image processing and single particle reconstruction made use of both EMAN (11) and SPIDER (5) software packages for parallel processing of the data. miR-17~92 particles (17,120 in total), and miR-17~19b particles (6,574 in total) were selected from the micrographs using semi-automatic boxing methods (boxer, EMAN). The particles were centred (cenalignint, EMAN), and masked (proc2d, EMAN; MA, SPIDER). The images were then aligned and subjected to reference-free classification and averaging (startnrclasses, EMAN; AP CA, RT SQ, AS R, SPIDER). For miR-17~92, 175 class averages were generated using each of the programs, EMAN and SPIDER (350 independent class averages). The two sets of class averages were combined and reduced into a common set of 45 class averages (startnrclasses, EMAN), generated from a subset of the miR-17~92 particles (12,880 particles selected based on self-clustering during the classification process). Euler angles were assigned to the class averages in EMAN and SPIDER and initial 3D models were constructed using common lines in Fourier space (startAny, EMAN; OP, BP RP, SPIDER). The initial models were

refined through 5 iterations (refine, EMAN), and both reconstructions were then low pass filtered to 20 Å resolution, aligned, and averaged together (proc3d, align3d, avg3d, EMAN). Projections of this averaged model were made (makeboxref.py, EMAN; PJ 3Q, SPIDER), to initiate a round of reference-based classification, alignment, initial reconstruction, refinement, and averaging (additional programs used include classesbymra, classalignall, EMAN; AP SH, VO MQ, SPIDER). This last step was repeated to generate the final model. An identical procedure was used for reconstruction of the truncation mutant, miR-17~19b. The distribution of class averages that contributed to the final structures was found to be uniform around the Euler sphere. The robustness of this 3D reconstruction procedure was tested by using incorrect starting models for reference-based reconstruction of both miR-17~92 and miR-17~19b (Figure A-3). The handedness of the 3D reconstructions was arbitrarily assigned and does not affect the conclusions drawn from the structures. The resolution of the final reconstructions were determined to be 21 Å for miR-17~92 and 27 Å for miR-17~19b, by calculating the Fourier shell correlation (FSC) between two independent half data sets using the 0.5 FSC criterion (eotest, EMAN) (4) (Figure A-2A). Particle sizes were determined by iterative centering and aligning the entire particle sets (EMAN, cenalignint). Finally, the reconstructions were contoured to their correct molecular volumes for viewing in Chimera (16).

Results and Discussion

miR-17~92 adopts a folded, globular structure

Grids of freshly annealed *in vitro* transcribed miR-17~92 clusters were negatively stained and imaged by EM, revealing homogeneous fields of spherical particles in the electron micrographs (Figure A-1A). Approximately 17,000 particles were selected and classified into 45 groups for single particle

reconstruction. Representative class averages are shown in Figure A-1B. These class averages were subsequently processed to generate a 3D structure. The final 3D reconstruction of miR-17~92, to 21 Å resolution, revealed a globular asymmetric particle approximately 108 Å in diameter with surface ridges and depressions (Figures A-1C, A-2 and A-3). The size of the particle is consistent with the predicted volume for a 260 kDa RNA with a compact globular fold (9), and is in agreement with results from size exclusion chromatography (not shown). The surface features surround a central density, which is more easily visible when the structure is contoured at a volume that corresponds to 70% of its molecular mass (approximately 3σ density cutoff) (Figures A-1C and A-2B). The EM reconstruction, while of insufficient resolution to delineate the exact position of RNA secondary structure elements, clearly demonstrates that the miR-17~92 cluster adopts a highly folded, globular structure. The highly folded structure of the cluster, taken together with the ribonuclease protection data, implies that the core region identified at the 3' end of the pri-miRNA sequence is contained within the central density of the EM reconstruction, while the nuclease-accessible miRNA hairpins are likely exposed on the surface.

We further examined the importance of the 3' core region to the overall organization of the cluster by analyzing the structure of a 3' truncation mutant by EM. This mutant, miR-17~19b, is identical to the full-length cluster but lacks the 3' 67 nucleotides including the miR-92 hairpin. Electron micrographs of negatively stained miR~17-19b revealed that while the mutant particles are globular, they are larger than the full-length cluster particles with a ~137 Å diameter, which is indicative of a less compact structure. A 3D reconstruction of miR-17~19b, to 27 Å-resolution, was generated from approximately 6600 particles (Figures A-1C, A-2 and A-3). The 3D reconstruction illustrated that the mutant does not contain the central density seen in the wild-type RNA. The missing core density and an overall opening of the structure are readily apparent when comparing equivalent slices through the reconstructions of the full-length

cluster and the truncation mutant (Figure A-2B). It is evident from the larger diameter and the more open structure of the mutant that deletion of this 3' region disrupts hierarchical folding and scaffolding effects in the cluster. Furthermore, the opening of the mutant structure correlates with the increases in Drosha processing observed for the miRNA-containing hairpins that would normally be internalized within the wild-type structure.

This work demonstrates that the miR-17~92 cluster adopts a discrete, specifically packed globular fold, and establishes a role for tertiary structure interactions in this structured miRNA precursor. Disruption of these interactions impacts miRNA biogenesis, and ultimately the regulation of mRNA gene expression.

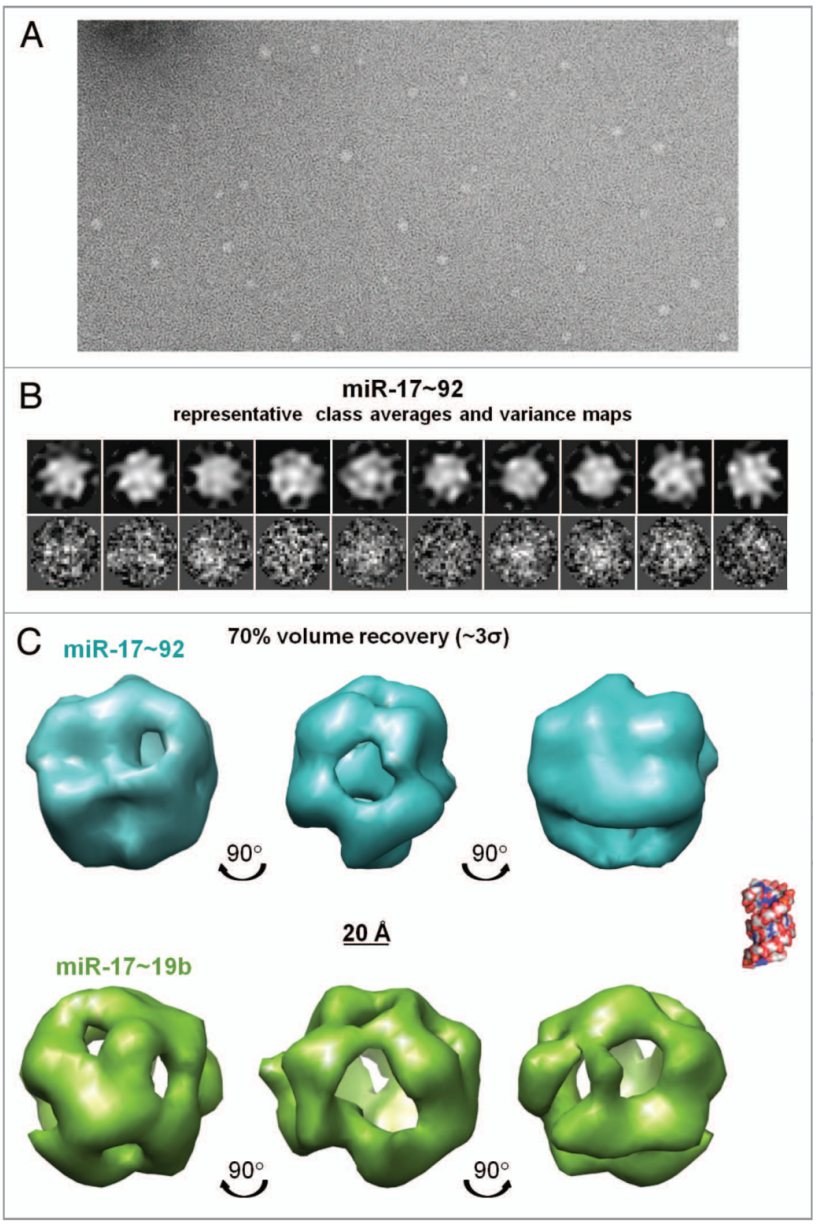


Figure A-1. Electron microscopy and single particle reconstruction of miR-17~92 and miR17~19b. A) A field of negatively stained miR-17~92 particles from a typical electron micrograph. B) Representative class averages and variance maps for the 3D reconstruction of the full-length miR-17~92 cluster. C) Comparison of the 3D reconstruction of miR-17~92 (blue) with the 3' truncation mutant miR-17~19b (green). Both reconstructions are contoured to density thresholds that represent 70% volume recovery for their respective molecular masses (~3σ). This density threshold was chosen to highlight the central density in the structure. Images represent different views of the 3D reconstructions rotated by 90° around the vertical axis as indicated. A 20 Å scale bar and a space filling model of a fourteen base pair dsRNA (PDB ID 1RNA) are included for reference.

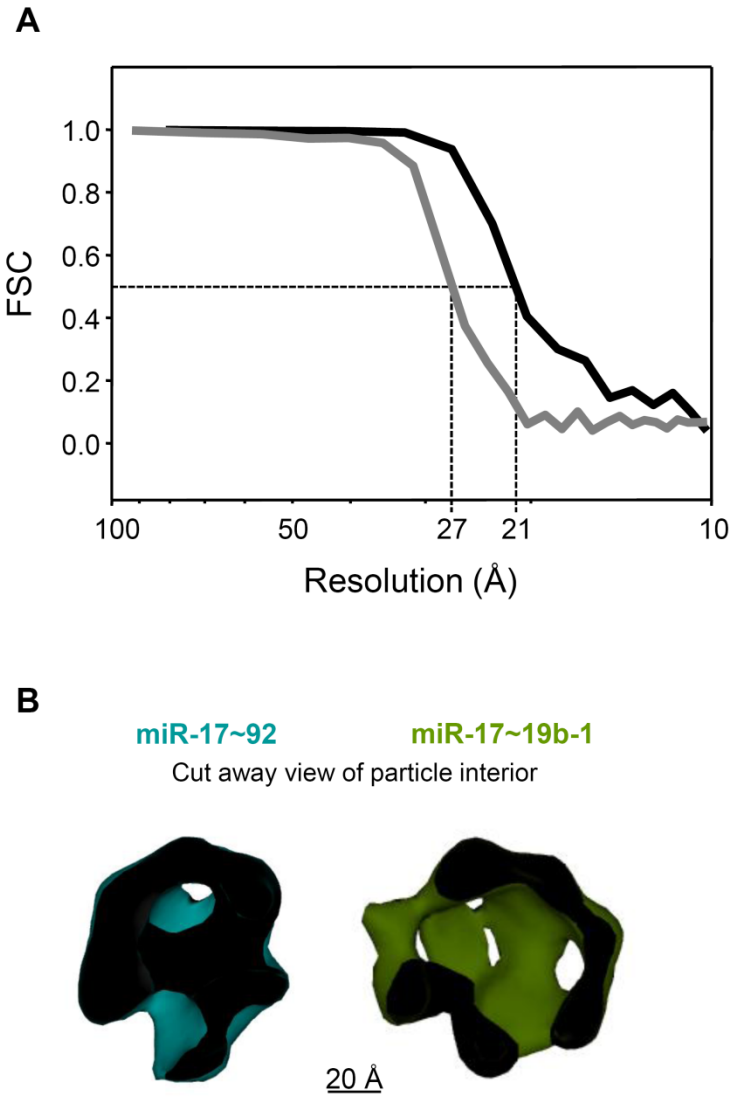


Figure A-2. Accompanying information for single particle reconstructions of miR-17~92 and miR-17~19b. A) FSC curves for miR-17~92 (black curve) and the 3' truncation mutant, miR-17~19b (grey curve). Using the conservative criterion of a 0.5 threshold, the cut-off values are 21 Å (miR-17~92) and 27 Å (miR-17~19b). The final reconstructions were low pass filtered at these resolution limits. B) Sections through the centre of the reconstructions of the miR-17~92 (blue) and the 3' truncation miR-17~19b (green) clusters. The central density is present in the wild-type and absent in the truncated cluster. The reconstruction of the 3' truncation miR-17~19b cluster is larger and more open than the wild-type cluster, suggesting significant perturbation of the RNA tertiary structure.

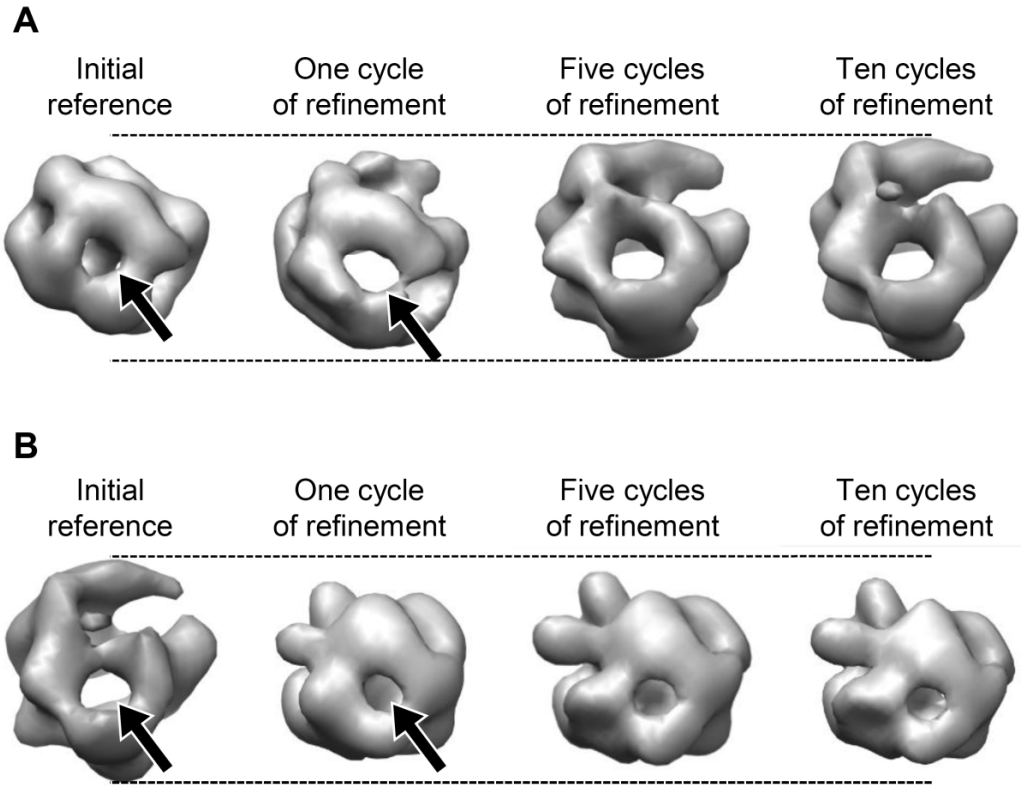


Figure A-3. Reference-based single particle reconstructions of the full-length miR-17~92 cluster and the 3' truncation mutant, miR-17~19b. The mutant is identical to the full-length cluster but lacks the 3' 67 nucleotides including the miR92a hairpin. In order to test the alignment, classification, and 3D reconstruction by angular reconstitution, we performed independent reconstructions using incorrect initial models (reference-based reconstruction). A) The 3D structure of the full-length cluster (miR-17~92) was used as an initial reference for 3D reconstruction of the 3' truncation mutant (miR-17~19b). The methods employed were similar to those previously described, with the exception that the first round of classification was based on projection matching with the starting reference model. After one round of refinement the central density seen in the initial model was absent from the reconstruction of the 3' truncation mutant (*arrows*), and by five cycles of refinement the reconstruction of the 3' truncation mutant was larger and more open than the initial model. B) In the second test, the 3D structure of the 3' truncation mutant (miR-17~19b) was used as initial reference for 3D reconstruction of the full-length cluster (miR-17~92). After one round of refinement the central density absent from the initial model was present in the reconstruction of the full-length cluster (*arrows*) and the reconstruction of the full-length cluster was smaller and more compact than the initial model. Note that the final reconstructions in A) and B) strongly resemble those presented in the manuscript, yet they are not identical. This indicates that an incorrect initial model does not necessarily bias the final 3D reconstruction; however, it does ultimately impact the quality of the reconstruction.

References

1. Abbott AL. 2011. Uncovering new functions for microRNAs in *Caenorhabditis elegans*. *Curr Biol* 21: R668-71

2. Ballarino M, Pagano F, Girardi E, Morlando M, Cacchiarelli D, et al. 2009. Coupled RNA processing and transcription of intergenic primary microRNAs. *Mol Cell Biol* 29: 5632-8

3. Bonauer A, Carmona G, Iwasaki M, Mione M, Koyanagi M, et al. 2009. MicroRNA-92a controls angiogenesis and functional recovery of ischemic tissues in mice. *Science* 324: 1710-3

4. Bottcher B, Wynne SA, Crowther RA. 1997. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386: 88-91

5. Frank J, Radermacher M, Penczek P, Zhu J, Li Y, et al. 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116: 190-9

6. Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, et al. 2005. A polycistronic microRNA cluster, *miR-17-92*, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res* 65: 9628-32

7. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, et al. 2005. A microRNA polycistron as a potential human oncogene. *Nature* 435: 828-33

8. Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM, Zhang GZ. 2011. Biological functions of microRNAs: a review. *J Physiol Biochem* 67: 129-39

9. Hyeon C, Dima RI, Thirumalai D. 2006. Size, shape, and flexibility of RNA structures. *J Chem Phys* 125: 194905

10. Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* 10: 126-39

11. Ludtke SJ, Baldwin PR, Chiu W. 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128: 82-97

12. Mendell JT. 2008. miRiad roles for the miR-17-92 cluster in development and disease. *Cell* 133: 217-22

13. Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I, Proudfoot NJ. 2008. Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* 15: 902-9

14. Olive V, Jiang I, He L. 2010. *mir-17-92*, a cluster of miRNAs in the midst of the cancer network. *Int J Biochem Cell Biol* 42: 1348-54

15. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, et al. 2004. Identification and characterization of a novel gene, *C13orf25*, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Res* 64: 3087-95

16. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-12

17. Ritchie W, Legendre M, Gautheret D. 2007. RNA stem-loops: to be or not to be cleaved by RNAse III. *RNA* 13: 457-62

18. Starega-Roslan J, Koscianska E, Kozlowski P, Krzyzosiak WJ. 2011. The role of the precursor structure in the biogenesis of microRNA. *Cell Mol Life Sci* 68: 2859-71