Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jeffrey W. Mercante Date

Graduate Division of Biological and Biomedical Sciences

Program in Microbiology and Molecular Genetics

Structure and Function of the *Escherichia coli* RNA-Binding Global Regulatory Protein CsrA

By

Jeffrey W. Mercante

Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences

Program in Microbiology and Molecular Genetics

_____
Tony Romeo
Advisor

_____
Charles P. Moran
Committee Member

_____
William Shafer
Committee Member

_____
Daniel Kalman
Committee Member

_____
Philip N. Rather
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

_____
Date

Structure and Function of the *Escherichia coli* RNA-Binding Global Regulatory Protein CsrA

By

Jeffrey W. Mercante
B.A., Southern Methodist University, 1998
M.S., Louisiana State University, 2002

Advisor: Tony Romeo, Ph.D.

An Abstract of
A dissertation submitted to the Faculty of the Graduate
School of Emory University in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Division of Biological and Biomedical Sciences
Program in Microbiology and Molecular Genetics

2009

ABSTRACT

Structure and Function of the *Escherichia coli* RNA-Binding Global Regulatory Protein CsrA

By

Jeffrey W. Mercante

In *Escherichia coli* K12, the small RNA-binding protein, CsrA, is the central component of a pathway that has pleiotropic effects on cellular processes such as carbon metabolism, motility and biofilm formation. CsrA binds to the 5'-untranslated leader of mRNA transcripts and, in repressed messages, inhibits ribosome access and thus translation. At the outset of this project, little was known about the structure of CsrA except that it was a homodimer. Therefore, a comprehensive alanine-scanning mutagenesis of the protein was undertaken to determine which amino acids were important for RNA-binding and regulation of gene expression. Two regions of the protein were identified where mutations caused a dramatic affect on CsrA activity. Region 1 was found at the N-terminus (β_1, residues 2-7) of the protein while region 2 was situated close to the C-terminus (β_5, residues 40-47). When these regions were mapped to a recently determined 3D CsrA structure two RNA-binding subdomains were defined on opposite sides of the bilaterally symmetrical protein; each subdomain consisted of β_1 from one subunit positioned adjacent and parallel to β_5 from the other subunit in three-dimensional space. Critical amino acids that make up the binding surfaces, including the most important residue, R44, were found to be highly conserved in all species examined. Construction of a heterodimer CsrA protein (HD-CsrA) that contains only a single normally functioning RNA-binding surface and EMSA studies revealed that CsrA has the capacity to bind one or two independent RNA molecules. Furthermore, CsrA can interact with two RNA target sites within the same RNA oligonucleotide simultaneously (dual binding). A distance of 18 nucleotides was established as optimal spacing between targets to allow for dual binding. Finally, a native *E. coli* CsrA target that contains multiple binding sites, *glgCAP*, was repressed ~14-fold more efficiently *in vitro* from the wild-type CsrA dimer than by HD-CsrA. When one CsrA target site was deleted in the *glgCAP* leader, repression by wild-type CsrA decreased while HD-CsrA regulation was unchanged, thus establishing that dual binding has biological relevance.

Structure and Function of the *Escherichia coli* RNA-Binding Global Regulatory Protein CsrA

By

Jeffrey W. Mercante
B.A., Southern Methodist University, 1998
M.S., Louisiana State University, 2002

Advisor: Tony Romeo, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Division of Biological and Biomedical Sciences
Program in Microbiology and Molecular Genetics

2009

ACKNOWLEDGEMENTS

I am grateful for the support and encouragement of many people without whom this journey would not have been possible. First I want to thank my mentor, Tony Romeo, who has been a steady voice of reason, wisdom and encouragement through these years. His analytical mind and unsinkable optimism guided me through many less-than-successful experiments and he taught me to wade through and distill the most important aspects of my research into a coherent, original body of work. I may never learn to write as succinctly as him, but at least I know now to what I should aspire.

I want to thank my dissertation committee – Drs. Charles Moran, William Shafer, Daniel Kalman and Philip Rather - for the time and effort they have put forth and the commitment they have made to my graduate training. Thank you all for setting high standards for me and my work.

I am grateful to the members of the Romeo Lab, both past and present, including Kazushi “Hands of Gold” Suzuki, Yoshi “Where’s the Beer” Itoh, Amy “Tennis shoe allergy” Kirby and Adrianne “I Brake for Pandas” Edwards. I especially want to thank Carlos “Bling-Bling ‘Los’” Goller who spent long hours in lab with me trying to figure out the true meaning of good science. Your company and council are sorely missed.

I want to also thank Keith Wilkinson and Monica Taylor for the opportunity to serve on the DSAC committee representing the Microbiology and Molecular Genetics Program. I truly enjoyed dabbling in the politics and boardroom discussions that make the GDBBS what it is.

I am grateful for my loving, loyal family, including my mother Josie, and my father Francis who always ask about my studies and who sit uncomplainingly as I explain my ongoing projects in painful detail. I truly appreciate their undying support and patience.

And last, but certainly not least, I thank my wife, Alex, who has sustained me through my studies and late nights even as she has worked to succeed in her own projects. I could not have done this without you. You are my counselor, confidant, and best friend. Thank you for standing with me as I trudged through the hard times. I wish you luck in your degree and I want you to know that I will always be there to support you in your endeavors, I love you with all my heart.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

Chapter 1. Introduction

Rationale and specific aims

RNA-binding proteins (RBPs) are key mediators of gene expression in both prokaryotes and eukaryotes. Their importance is underscored by their essentiality (Sayed, Matsuyama *et al.*, 1999; Grall, Livny *et al.*, 2009) or associated mutant phenotypes (Romeo, Gong *et al.*, 1993; Pannekoek, Huis in 't Veld *et al.*, 2009) and by the number of diseases linked with their dysfunction (Lukong, Chang *et al.*, 2008; Christiansen, Kolte *et al.*, 2009) or absence in humans (Katsanou, Milatos *et al.*, 2009). RBPs recognize and bind RNA through a variety of protein modules or motifs, some of which are evolutionarily conserved while others are specific to particular families or domains. To understand the mechanism by which RBPs bind their targets and effect gene expression, we must identify the molecular interactions required for protein-RNA association. It is additionally imperative that we try to grasp the macromolecular context or RNA structural requirements that allow binding to occur, and how these rules are applied in biological systems.

In this series of experiments, we sought to define the protein and RNA structural requirements that allow the *Escherichia coli* K12 global regulator, CsrA, to specifically bind RNA target sites and regulate gene expression. The specific aims of this dissertation were:

1. To elucidate at the amino acid level the structural and functional properties of the *E. coli* RNA-binding protein CsrA that allow protein-RNA interactions.

2. To clarify the exact role of RNA binding in CsrA-mediated translational regulation.
3. To characterize the RNA target site spacing necessary for proper CsrA dual site binding.
4. To define the purpose of the CsrA dual RNA-binding surface arrangement and whether it has biological significance.

Background and significance

Central to the survival and success of all organisms is the ability to react and adapt to changing environmental conditions. At the core of this response is the capacity to control gene expression. Both eukaryotic and prokaryotic gene expression entails a hierarchy that may include transcriptional and/or post-transcriptional regulatory mechanisms (Snyder and Champness, 2007). Among the various regulatory tools available to a cell for this purpose are extensive numbers of RNA-binding proteins (RBPs), which, when bound to RNA are referred to as ribonucleoprotein (RNP) complexes or particles (Lunde, Moore *et al.*, 2007; Glisovic, Bachorik *et al.*, 2008). RBPs are involved in most aspects of RNA metabolism from simple bacterial systems, where transcription and translation are coupled, to complex eukaryotic pathways that require highly orchestrated RNA splicing events and nuclear export prior to translation (Abaza and Gebauer, 2008; Alberts, 2008). The importance of RBPs is highlighted by their ubiquity; they are found not only in archaea, bacteria and eukaryota, but also in phage or viruses specific for the later two domains (Rana and Jeang, 1999; Lim, Downey

et al., 2001; Dreyfuss, Kim *et al.*, 2002; Guo, Xue *et al.*, 2003; Keiler, 2008; Scofield and Lynch, 2008).

While many RBPs have been extensively studied in viruses and prokaryotes, the majority of known RBPs have been characterized in eukaryotes (Finn, Tate *et al.*, 2008). Among other factors, it is eukaryotic cellular compartmentalization separating transcription and translation that allows for additional levels of regulatory control. And the differences between prokaryotic and eukaryotic translational regulation extend well beyond compartmentalization; for example, eubacterial translation relies on 16S ribosomal recognition of the Shine-Dalgarno (SD) sequence that is properly situated upstream of the start codon while eukaryotic mRNA must assemble a 5'-cap structure and 3'-poly(A) tail which then recruits the ribosomal pre-initiation complex (PIC) to scan for the correct triplet (Ganoza, Kiel *et al.*, 2002; Sonenberg and Hinnebusch, 2009). Unlike the bacterial system where regulation centers around access to the SD, any factor that inhibits one of the requisite eukaryotic steps can inhibit translation initiation (Kozak, 2005). In comparison to lower organisms, this increase in complexity leads to increased numbers of dedicated RBPs to carry out additional specialized functions (Ganoza, Kiel *et al.*, 2002). And while numerous differences in the translational pathway exist between phylogenetic domains, many key steps are performed by similar or highly conserved proteins. For example, the prokaryotic Initiation Factor 1 (IF1) has significant similarity to its eukaryotic counterpart, eIF1A (Ganoza, Kiel *et al.*, 2002). Thus a more complete understanding of post-transcriptional regulation requires familiarity with elements of both prokaryotic and eukaryotic systems.

Post-transcriptional processes modulated by RNA-binding proteins

Eukaryotic RBPs important for post-transcriptional regulation

RNA-binding proteins are involved in every step of the post-transcriptional lives of messenger RNA (mRNA), including splicing, trafficking, translation initiation, and termination and decay (Kim and Dreyfuss, 2001; Abaza and Gebauer, 2008; Bolognani and Perrone-Bizzozero, 2008). The following paragraphs outline and describe RBPs essential for these cellular translationally related processes, focusing initially on eukaryote-specific non-ribosomal components.

Splicing

In eukaryotes, the spliceosome is a dynamic ribonucleoprotein complex of at least 5 small nuclear RNAs (snRNAs) and numerous protein factors responsible for the post-transcriptional processing of pre-mRNA. Through the removal of introns and concomitant assemblage of exons, a near-translationally ready mRNA is produced for export from the nucleus (Matlin, Clark *et al.*, 2005; Wang and Burge, 2008). Several RBPs are found to contact mRNA within the spliceosomal complex including SR (serine/arginine-rich) and SR-related protein families that are critical for splice site selection where they recognize enhancer elements (ESE) on the pre-mRNA (Blencowe, Bowman *et al.*, 1999). The large family of heteronuclear ribonucleoproteins (hnRNPs) also bind mRNA at cis-acting sequences called silencer elements (ESS) where they antagonize SR protein activity (Matlin, Clark *et al.*, 2005; Wang and Burge, 2008; Long and Caceres, 2009). The core of the spliceosome is composed of 5 small nuclear

ribonucleo-proteins (snRNPs) which directly bind pre-mRNA and perform the central function of splicing (Alberts, 2008).

Nuclear export

Several different pathways exist for the trafficking or export of RNA from the eukaryotic nucleus, but they all require RBPs to target the processed message to the nuclear pore complex (NPC) which occurs in concert with transcription and splicing (Dreyfuss, Kim *et al.*, 2002). In addition to their requirement for pre-mRNA splicing, several hnRNP (heterogeneous nuclear ribonucleoprotein) and SR family proteins also function as nuclear shuttling proteins, and other RBPs are recruited to the processed transcript creating a messenger ribonuceleo-protein (mRNP). Among these other RBPs are splicing-dependent/sequence-independent proteins such as Y14, specific export factors like Aly/REF, components of the exon-exon junction complex (EJC) such as RNPS1 and the poly(A) binding protein (PABP) that interacts with the message 3'-poly(A) tail (Kim and Dreyfuss, 2001; Reed and Magni, 2001; Dreyfuss, Kim *et al.*, 2002; Keiler, 2008). PABPs actually serve several functions in eukaryotic mRNA metabolism; they are important for translation initiation, termination and decay in addition to nuclear export (Mangus, Evans *et al.*, 2003).

Translation initiation

Central to most eukaryotic translation initiation is the formation of a cap-binding protein complex (CBC) at the extreme 5'-end of the transcript. After nuclear export, the

CBC is replaced by initiation factor eIF4E leading to assembly of the eIF4F complex and subsequent ribosomal recruitment and translation (Pestova, Kolupaeva *et al.*, 2001; Reed and Magni, 2001; Wells, 2006). An alternative to cap-mediated translation is initiation at an RNA internal ribosome entry site (IRES). In this mechanism, the eIF4G protein is proteolytically cleaved, binds the IRES and recruits the necessary translational machinery (Lopez de Quinto and Martinez-Salas, 2000; Pestova, Kolupaeva *et al.*, 2001). The ribosome's capacity for translation is also directly tied to its RNA-binding protein components, which make up approximately one third of its mass (Draper, 1995; Liljas and Garber, 1995). A review of translational regulation by ribosomal proteins is reserved for a subsequent section in this chapter focused on prokaryotic translation. mRNAs themselves can be targeted for localized expression in eukaryotes through RBP interaction. For instance, the translational repressor protein ZBP1 appears to target mRNA to the cell periphery through its affinity for a 54 nucleotide (nt) cis-acting "zipcode" sequence (Rodriguez, Czaplinski *et al.*, 2008). Additional control can be exerted through accessory factors acting at the 5' or 3'-untranslated region (UTR). One well-characterized example is the binding of iron responsive proteins (IRP) to iron responsive elements (IRE) found in mRNA stem-loop structures. IREs are typically located near the 5'-cap where the presence of low intracellular iron causes IRP-IRE complex formation and inhibits the binding of the 40S ribosomal subunit (Thomson, Rogers *et al.*, 1999; Wilkie, Dickson *et al.*, 2003; Alberts, 2008). IRPs bound distal to the site of ribosome loading, but not in the ORF, also negatively affect expression by stalling the process of ribosome scanning, and in non-mammalian eukaryotic cells, causing premature termination (Rouault and Klausner, 1996; Paraskeva, Gray *et al.*,

1999). Studies have also revealed a cis-acting 3'-sequence essential for proper translational control during embryogenesis known as the cytoplasmic polyadenylation element (CPE) (Mendez and Richter, 2001). The CPE is a U-rich mRNA sequence recognized by the CPEB protein which can either repress translation in the unphosphorylated state, or facilitate poly(A) tail elongation and subsequent translation initiation when phosphorylated (Mendez and Richter, 2001; Mazumder, Seshadri *et al.*, 2003).

Translation termination and mRNA decay

Translation termination and mRNA decay are also highly orchestrated events in eukaryotes, and most essential RBPs needed for normal termination interact directly or indirectly with ribosomal RNA, such as release factors eRF1 and eRF3 (Frolova, Merkulova *et al.*, 2000). Several proteins that influence translation termination and/or decay are known to bind directly to the transcript 3'-UTR. Among these factors are SBP2, which recognizes 3'-UTR selenocysteine insertion signal (SECIS) stem loops and recruits additional factors that reprogram the translating ribosome to incorporate selenocysteine at UGA stop codons instead of terminating translation (Hoffmann and Berry, 2005). The human double-stranded RNA-binding Staufen1 (Stau1) protein targets Stau1-binding sequences (SBS) in the 3'-UTR where it facilitates increased decay of normally terminated transcripts, termed Stau1-mediated decay (SMD) (Ramos, Grunert *et al.*, 2000; Kim, Furic *et al.*, 2005; Kim, Furic *et al.*, 2007; Gong, Kim *et al.*, 2009). Similar to SMD is nonsense-mediated decay (NMD), a quality control pathway for the

targeted exclusion of mRNAs containing premature stop codons, which is also controlled by RBPs. Several NMD protein factors directly interact with mutant mRNA, including UPF2, MLN51, EIF4AIII and RNPS1, allowing the assembly of a core NMD complex (Chang, Imam *et al.*, 2007).

AU-rich elements (AREs) are widely distributed sequences typically found in the transcript 3'-UTR that influence mRNA turnover when bound by ARE-binding proteins (Wilusz and Wilusz, 2004). Upon binding to their ligands, ARE-binding proteins can target the mRNA for degradation by the exosome, or they can facilitate deadenylation or decapping (Wilusz and Wilusz, 2004). Control of mRNA stability through the ARE is differential; mRNA decay is increased by binding of some proteins, such as KSRP, BRF, TTP and AUF1, while other factors stabilize the transcript, like NF90 and Hu proteins (Wilusz, Wormington *et al.*, 2001; Wilusz and Wilusz, 2004; Bolognani and Perrone-Bizzozero, 2008). Similar to AREs, GU-rich elements (GREs) have been identified as 3'-UTR cis-acting regulatory sequences capable of affecting mRNA stability of short lived T cell transcripts due to binding by RBPs (Vlasova and Bohjanen, 2008). The highly conserved CELF (CUGBP and embryonically lethal abnormal vision-type RNA binding protein 3-like factors) family of Sm-like Lsm proteins, including CUGBP1 and EDEN-BP, target GREs in a wide variety of species, facilitating rapid mRNA decay (Vlasova and Bohjanen, 2008). This later hetero-heptameric RBP associates with single nucleotide-rich 3'-regions of nuclear poly(A) RNA and cytoplasmic deadenylated mRNAs; the Lsm1-7 configuration appears to promote mRNA decay while the Lsm2-8 arrangement may lead to transcript stabilization (Wilusz, Wormington *et al.*, 2001; Wilusz and Wilusz, 2005).

Prokaryotic RBPs important for post-transcriptional regulation

Ribosomal and ribosome-associated RBPs

Eubacteria actively manage cellular rRNA:ribosomal protein (r-protein) stoichiometry (and therefore ribosomal activity and translation) in response to nutritional and energetic cues. The P1 promoter of the *E. coli* K12 *rrn* operons is subject to both stringent regulation (via ppGpp) and bacterial growth rate (Keener and Nomura, 1996). rRNA and r-proteins are assembled in roughly stoichiometric proportions, thus r-protein synthesis is controlled indirectly in response to rRNA abundance. This is accomplished through a negative feedback loop where excess r-proteins inhibit the expression of their own messages by binding a conserved sequence in the 5'-UTR (Keener and Nomura, 1996; Kaberdin and Blasi, 2006; Babitzke, Baker *et al.*, 2009). Interestingly, while r-proteins associate intimately with rRNA they do not interact directly with the translating mRNA when they are part of the bacterial translational complex (Laursen, Sorensen *et al.*, 2005). The exception to this rule is protein S1 whose primary function within the ribosome is dependent on its mRNA-binding capacity (Sengupta, Agrawal *et al.*, 2001). S1 contacts an 11 nt AU-rich region upstream from the SD and appears to be important for targeting the 30S subunit to the proper translational start position (Boni, Isaeva *et al.*, 1991; Laursen, Sorensen *et al.*, 2005). Fifty-two r-protein genes are organized into 19 operons including several polycistrons which allows a handful of r-proteins to control the expression of many related factors (Keener and Nomura, 1996). This is typically accomplished in operons through translational coupling where downstream gene products bind and repress translation at the first gene in the operon which leads to decreased

expression of downstream genes (Keener and Nomura, 1996; Kozak, 2005; Kaberdin and Blasi, 2006).

Two mechanisms have been described for r-protein-mediated translational repression, including competition with the ribosome for access to the SD sequence, and entrapment of the 30S subunit in an inactive complex near the SD (Keener and Nomura, 1996; Schlax, Xavier *et al.*, 2001; Babitzke, Baker *et al.*, 2009). The former occlusion mechanism is used to regulate expression of the *E. coli* L35 (*infC-rpmI-rplT),* spc *(rplNXE-rpsNH-rplFR-rpsE-rpmD-rplO-secY-rpmJ),* and possibly the S10 (*rpsJ-rplC-rplD-rplW-rplB-rpsS-rplV-rpsC-rplP-rpmC-rpsQ-rplN*) and L10 (*rplA-rplJ)* operons while the later entrapment mechanism is utilized by the α (*rpsMKD-rpoD-rplQ*) and S15 (*rpsO-pnp)* operons (Chiaruttini, Milet *et al.*, 1996; Keener and Nomura, 1996; Laursen, Sorensen *et al.*, 2005; Kaberdin and Blasi, 2006; Iben and Draper, 2008; Babitzke, Baker *et al.*, 2009). Interestingly, the str (*rpsL-rpsG-fusA-tufA*) operon is probably retroregulated wherein binding of the primary repressor protein, S7, occurs downstream of the first gene in the operon, S12, which probably effects expression by decreasing stability of the entire transcript (Keener and Nomura, 1996). Regulation of the L11 (*rplK-rplA-rplJ-rplL-rpoB-rpoC*) and S1 (*rpsA*) operons are also repressed by their gene products, however the mechanism is not understood (Keener and Nomura, 1996). Ribosome-associated factors also exhibit autoregulation of their gene products. Initiation factor 2 (IF2), a central protein in the formation of the 70S ribosome, appears to tightly bind its own mRNA downstream from the start of translation, possibly forming a complex that interferes with the actively translating ribosome, however, this mechanism has not been fully characterized (Laursen, de *et al.*, 2002). Autoregulation of *thrS* expression through

binding of its gene product, threonyl-tRNA synthetase, has been well defined. The threonyl-tRNA synthetase protein binds a structured operator sequence in its own leader that mimics the anticodon loop of $tRNA^{Thr}$, thereby blocking 30S ribosome loading and translation initiation (Springer, Plumbridge *et al.*, 1985; Patte, 1996; Romby and Springer, 2003; Babitzke, Baker *et al.*, 2009).

Non-ribosomal RBPs

Several eubacterial non-ribosomal RBPs have been well characterized at the regulatory and/or structural levels; they include factors for transcription termination/antitermination, such as LicT and BglG, as well as those for post-transcriptional control. However, this section will concentrate on translational factors that influence ribosome loading or translation initiation and message stability.

Cold shock proteins

The mechanism underlying the cold shock response in *E. coli* and *Bacillus subtilis* has been a topic of heated discussion and disagreement over the past 20 years (Gualerzi, Giuliodori *et al.*, 2003). A sudden decrease in temperature from 37^{o} to below 20^{o} C leads to a rapid acclimatization phase where most normal protein expression is repressed while a small subset of genes, the cold shock proteins (CSP), are upregulated (Nogueira and Springer, 2000). The most abundant CSP is CspA, whose mRNA is stabilized up to 150-fold upon temperature shift (Fang, Jiang *et al.*, 1997; Gualerzi, Giuliodori *et al.*, 2003). Similar to CspB, C and D of *B. subtilis*, the *E. coli* CspA is an RBP now thought to

interact, albeit nonspecifically, with the 5'-UTR of its own RNA at low temperatures where it destabilizes RNA secondary structures that are believed to inhibit ribosome access and translation (Graumann, Wendrich *et al.*, 1997; Jiang, Hou *et al.*, 1997; Graumann and Marahiel, 1998). Most CSPs appear to have a putative RNA-binding cold shock domain (CSD) similar to the eukaryotic Y-box proteins. While a putative conserved cold shock "downstream box" (DB) was identified in the *cspA* coding region, it is still not a widely accepted RNA-binding target responsible for interaction with CSPs (Jiang, Hou *et al.*, 1997; Graumann and Marahiel, 1998).

AcnA and AcnB aconitases

The *E. coli* AcnA and AcnB proteins are bifunctional aconitase/RBPs similar in function to the eukaryotic iron responsive protein, IRP1 (Prodromou, Artymiuk *et al.*, 1992; Tang and Guest, 1999). While both enzymes have catalytic activity, AcnA is the major stationary phase isoenzyme that responds to iron and oxidative stress and AcnB is expressed during exponential growth and is the major enzyme in the citric acid cycle that catalyzes the isomerization of isocitrate and citrate (Cunningham, Gruer *et al.*, 1997; Tang and Guest, 1999; Varghese, Tang *et al.*, 2003). Post-transcriptional regulation occurs under low iron conditions when both apo-enzymes bind to paired stem loop motifs in the 3'-UTR of their respective mRNAs, stabilizing the transcripts and increasing protein synthesis (Tang and Guest, 1999). AcnA and AcnB also differentially regulate expression of superoxide dismutase (SOD), presumably by binding and destabilizing the *sod* transcript (Tang, Quail *et al.*, 2002).

In *Salmonella enterica,* apo-AcnB indirectly activates expression of FliC, the major flagellum subunit, by binding and apparently destabilizing the message for the FtsH protease. Decreased FtsH levels lead to a cascade involving σ^{32} and DnaK, ultimately reducing flagellum biosynthesis (Tang, Guest *et al.*, 2004). The single *B. subtilis* aconitase, CitB, is homologous to the AcnA family of cytoplasmic bacterial aconitases. *In vitro*, CitB binds to IRE-like sequences in RNAs that code for proteins in iron metabolism (*feuA*/*B*) and electron transport (*qoxD*) (Alen and Sonenshein, 1999). CitB protein was also found to specifically bind the 3'-UTR of mRNA for the sporulation specific transcriptional regulator GerE, which is required for late stage sporulation events (Serio, Pechter *et al.*, 2006). A mutagenized CitB protein with decreased RNA-binding capacity caused defects in sporulation.

Pseudomonas Crc

In most cases, genetic regulation by carbon catabolite repression occurs at the transcriptional level (Gorke and Stulke, 2008). However, the *Pseudomonas putida* Crc (catabolite repression control) protein is a post-transcriptional global regulator. While it is not known to which metabolic signal(s) it responds, Crc directly or indirectly controls pathways for carbon metabolism, the synthesis of nitrogenated compounds, benzoate degradation, hydrocarbon assimilation and biofilm formation (Collier, Hager *et al.*, 1996; Hester, Lehman *et al.*, 2000; Hester, Madhusudhan *et al.*, 2000; Moreno, Ruiz-Manzano *et al.*, 2007; Moreno and Rojo, 2008; O'Toole, Gibbs *et al*., 2000). Recently, a possible general mechanism by which Crc regulates gene expression was uncovered; Crc binds to the 5'-UTR of the transcript for AlkS, a key regulator of the alkane degradation pathway,

and BenR, a transcription factor in the benzoate degradation pathway, thereby decreasing expression (Moreno, Ruiz-Manzano *et al.*, 2007; Moreno and Rojo, 2008). Surprisingly, Crc binding appeared to inhibit translation of *alkS* without affecting the transcript half-life while BenR mRNA stability was presumably decreased in parallel to its expression; the exact mechanism of this regulation, however, is not known (Yuste and Rojo, 2001).

Lactococcus group II Intron maturase, LtrA

The *Lactococcus lactis* L1.LtrB group II intron encodes a maturase protein, LtrA, which specifically and with high affinity binds to a highly structured sequence in its own 5'-UTR where it helps fold the RNA into a catalytically active structure (Matsuura, Noah *et al.*, 2001; Singh, Saldanha *et al.*, 2002). The LtrA RNA-binding site also overlaps the SD allowing LtrA to compete with the ribosome for access to the RBS and start codon, leading to translational down-regulation (Singh, Saldanha *et al.*, 2002). Thus, LtrA pulls "double-duty" where it both autoregulates its own expression and assists in the formation of an enzymatically active intron.

Bacteriophage coat proteins

Qβ, MS2 and PP7 are a well-characterized family of single-stranded RNA coliphages that use a conserved post-transcriptionally mechanism to regulate gene expression (Witherell and Uhlenbeck, 1989; Fouts, True *et al.*, 1997; Lim, Downey *et al.*, 2001; Babitzke, Baker *et al.*, 2009). The coat protein of these phages serves a dual purpose both as the structural component of the viral capsid and as an RBP that represses expression of the phage replicase gene (Chao, Patskovsky *et al.*, 2008). Repression

occurs when the coat protein binds to a hairpin loop at the translational initiation region preventing ribosome access to the SD sequence (Bernardi and Spahr, 1972; Witherell and Uhlenbeck, 1989; Lim and Peabody, 2002). Surprisingly, the sequence of the single stranded loop was found not to be critical for protein-RNA association in PP7 or M2, however, the overall hairpin structure, and in particular a bulge in the double-stranded stem, was found to be essential for tight binding (Romaniuk, Lowary *et al.*, 1987; Lim and Peabody, 2002). In a similar way, phage Qβ coat protein mainly recognized the structure of the hairpin and not its absolute sequence, although without the requisite stem bulge (Witherell and Uhlenbeck, 1989).

Phage T4 autoregulatory proteins

Bacteriophage T4 encodes three proteins capable of translationally repressing their own synthesis: protein 32 (gp32), protein 43 (gp43) and RegA (Babitzke, Baker *et al.*, 2009). gp32 and gp43 are primarily DNA binding proteins; gp32 is a single-stranded DNA-binding protein required for DNA synthesis, recombination, repair and possibly transcription and viral packaging while gp43 is a DNA-polymerase (Nossal, 1992; Shamoo, Tam *et al.*, 1993; Mosig, 1998; Pavlov and Karam, 2000). Not surprisingly, gp32 and gp43 interact with each other during DNA synthesis (Mosig, 1998). Both proteins also inhibit ribosome loading through binding of RNA secondary structural elements upstream from the RBS that do not overlap the SD (Shamoo, Tam *et al.*, 1993; Pavlov and Karam, 2000). gp32 interacts with an RNA pseudoknot found between 40 and 70 nt upstream of the start codon; nucleation at that site leads to cooperative binding of multiple gp32 proteins through a downstream unstructured region, eventually

overlapping the initiation codon (McPheeters, Stormo *et al.*, 1988; Shamoo, Tam *et al.*, 1993). gp43 appears to occlude ribosome binding by contacting an operator sequence just upstream of the RBS consisting of a simple stem-loop (Pavlov and Karam, 2000; Petrov and Karam, 2002). gp43 RNA-binding was found to be sensitive to the RNA stem composition (presumably as it relates to structure), the presence of a single stranded nucleotide pair, and the stem 3'-tail length (Pavlov and Karam, 2000). The T4 RegA protein is known to regulate at least nine T4 genes, in addition to itself, during the phage life cycle (Brown, Brown *et al.*, 1997). However, a RegA consensus binding target is not clear from studies of regulated transcripts; while RegA does bind the region surrounding the SD sequence, no apparent secondary structure is required (Unnithan, Green *et al.*, 1990; Brown, Brown *et al.*, 1997). Studies suggest the protein is a dimer in solution but binds as a monomer; the AUG initiation codon is essential for protein interaction while a UU pair and poly(A) tract are highly conserved in most targets (Unnithan, Green *et al.*, 1990; Phillips, Gordon *et al.*, 1996; Brown, Brown *et al.*, 1997).

B. subtilis Tryptophan regulation and TRAP

The *trp* RNA-binding attenuation protein (TRAP) is a key regulator of tryptophan metabolism and transport in *B. subtilis*. TRAP is an intracellular sensor that responds to tryptophan concentrations through transcriptional and post-transcriptional regulation of the appropriate *trp* operon genes (Gollnick, Babitzke *et al.*, 2005). TRAP is a ring shaped homo-undecameric protein that specifically recognizes single-stranded RNA triplet repeat sequences separated by a non-conserved 2-3 nt spacer (Hopcroft, Manfredo *et al.*, 2004; Shevtsov, Chen *et al.*, 2005). Transcriptional control is applied when a

sufficient cellular concentration of tryptophan accumulates to activate TRAP, which then binds to the nascent *trpEDCFBA* transcript causing the formation of an intrinsic transcriptional terminator. The release of TRAP from the RNA results in an antiterminator structure that allows polymerase read-through (Du and Babitzke, 1998).

TRAP also post-transcriptionally regulates expression of various tryptophan biosynthetic genes; at least two different mechanisms are employed which rely on TRAP-RNA binding (Gollnick, Babitzke *et al.*, 2005). *trpE* transcripts not prematurely terminated by the attenuation method are subjected to translational control whereby TRAP binding at the same leader target sequence causes the formation of a downstream hairpin that sequesters the SD and blocks ribosome loading and translation (Merino, Babitzke *et al.*, 1995; Du and Babitzke, 1998). The second mechanism involves direct binding of activated TRAP to sequences in and around the translational initiation region leading to direct competition with the ribosome for access to the SD. Genes from three separate operons employ this method of regulation, they include *trpG* (*pabA*), involved in folic acid biosynthesis, *trpP* (*yhaG*), a putative tryptophan transporter, and *ycbK*, a gene with homology to known efflux pumps (Du, Tarpey *et al.*, 1997; Sarsero, Merino *et al.*, 2000; Sarsero, Merino *et al.*, 2000; Yakhnin, Zhang *et al.*, 2004; Yakhnin, Yakhnin *et al.*, 2006).

CsrA and the E. coli K12 Csr system-Overview

The carbon storage regulatory system (Csr) in *E. coli* K12 is a global gene control pathway with pleiotropic effects on cell physiology and metabolism. The Csr circuit is critical for regulating cellular activity; the key mediator of gene expression, CsrA,

represses stationary phase pathways for glycogen synthesis and catabolism as well as biofilm formation while activating exponential phase processes such as glycolysis, acetate metabolism, motility and biofilm dispersal (Sabnis, Yang *et al.*, 1995; Wei, Brun-Zinkernagel *et al.*, 2001; Wang, Dubey *et al.*, 2005). The central component of the system, CsrA, is a homodimeric RNA-binding protein, which was discovered by T. Romeo while searching for factors that control the expression of the glycogen biosynthesis genes *glgCAP* and *glgBX* (Romeo, Gong *et al.*, 1993). Other elements of the system include CsrB and CsrC, which are small untranslated RNAs (sRNA) that possess multiple CsrA-binding GGA repeat elements and act as a sink for CsrA by sequestering the protein; BarA/UvrY is a two-component signal transduction system that directs CsrB and CsrC transcription; and CsrD is a protein factor that targets CsrB and CsrC for degradation by RNaseE (Liu, Gui *et al.*, 1997; Liu and Romeo, 1997; Suzuki, Wang *et al.*, 2002; Weilbacher, Suzuki *et al.*, 2003; Suzuki, Babitzke *et al.*, 2006). A summary of the *E. coli* K12 Csr pathway based on a previously described model is outlined in Fig. 1-1.

Noncoding sRNAs CsrB and CsrC

The noncoding sRNA CsrB was discovered through pull-down studies because it tightly associated with a His-tagged CsrA protein (Liu, Gui *et al.*, 1997). The functionally similar CsrC sRNA was identified in a genetic screen for factors that affect glycogen biosynthesis (Weilbacher, Suzuki *et al.*, 2003). CsrB and CsrC antagonize CsrA activity by binding multiple protein dimers, thereby removing active CsrA from the free cellular pool. CsrB (369 nt) is the major sRNA, it contains 18-22 putative CsrA

target sites and is capable of binding ~18 CsrA subunits (Liu, Gui *et al.*, 1997; Babitzke, Baker *et al.*, 2009). Curiously, negatively stained CsrA-CsrB complexes appear globular by transmission electron microscopy as opposed to an extended "beads-on-a-string" configuration (Liu, Gui *et al.*, 1997). CsrC (245 nt) is two-thirds the size of CsrB and contains half the number of putative CsrA target sites (Weilbacher, Suzuki *et al.*, 2003; Babitzke, Baker *et al.*, 2009). CsrA exhibits cooperative binding and high avidity for both CsrB (~1 nM) and CsrC (~9 nM), yet the sRNAs display no structural similarity besides their Rho-independent terminators and repeated CsrA target sites which are typically found in single-stranded regions of predicted hairpin loops or bulges (Liu, Gui *et al.*, 1997; Weilbacher, Suzuki *et al.*, 2003; Babitzke and Romeo, 2007). CsrA binding sites are concentrated at the 5'-end of the CsrC RNA in comparison to CsrB, however the significance of this is unclear (Weilbacher, Suzuki *et al.*, 2003). The function of CsrC and CsrB appear to be redundant, as their individual and combined effects on glycogen accumulation, biofilm formation and motility are all CsrA-dependent (Weilbacher, Suzuki *et al.*, 2003). Both CsrB and CsrC expression are activated indirectly by CsrA via BarA/UvrY, however, CsrC transcription is less dependent on UvrY compared with CsrB (Gudapaty, Suzuki *et al.*, 2001; Suzuki, Wang *et al.*, 2002; Weilbacher, Suzuki *et al.*, 2003). Sharing a common regulatory element (UvrY) in combination with the Csr circuit feedback loop (Fig. 1-1) also means that deletion of one sRNA results in a compensatory increase in expression of the other (Weilbacher, Suzuki *et al.*, 2003).

BarA/UvrY two component system

BarA/UvrY is a pH-dependent (Mondragon, Franco *et al.*, 2006) two-component signal transduction system (TCS) that activates expression of CsrB and CsrC (Suzuki, Wang et al., 2002; Weilbacher, Suzuki *et al.*, 2003). The homologous *Pseudomonas flourescens* pair, GacS/GacA, were originally found to operate upstream in the same pathway as their CsrA homologue RsmA (Blumer, Heeb *et al.*, 1999); BarA/UvrY were subsequently characterized in *E. coli* K12 to directly activate transcription of CsrB and CsrC (Suzuki, Wang *et al.*, 2002). BarA is in a unique subclass of tripartite sensor kinases that phosphorylate their cognate DNA-binding response regulators (in this case, UvrY) by an ATP-His-Asp-His phosphorelay (Nagasawa, Tokishita *et al.*, 1992; Pernestig, Melefors *et al.*, 2001). Both CsrA and UvrY increase BarA transcription, however, neither BarA nor UvrY affect expression of CsrA (Suzuki, Wang *et al.*, 2002). As mentioned previously, transcription of CsrB and, to a lesser extent CsrC, are dependent upon CsrA and UvrY; this opens the possibility for an alternative mechanism of CsrC regulation (Suzuki, Wang *et al.*, 2002; Weilbacher, Suzuki *et al.*, 2003; Jonas, Tomenius *et al.*, 2006). In addition, while CsrB expression requires CsrA, approximately 30% of CsrB/C transcription is BarA-independent; the components of this alternate UvrY-activating pathway are currently unknown (Suzuki, Wang *et al.*, 2002). Additionally, the intra or extracellular signal sensed by BarA is also unclear.

One accessory input to the BarA/UvrY TCS comes from *sdiA*, which ecodes a LuxR family DNA-binding transcriptional regulator (Suzuki, Wang *et al.*, 2002). The regulator likely participates in interspecies communication (quorum sensing) possibly through the binding of acylhomoserine lactones (AHL) and indole (Ahmer, 2004; Lee,

Jayaraman *et al.*, 2007), however, further studies are needed to confirm this interaction. Initial DNA microarray and RT-PCR studies described UvrY as a possible direct target for SdiA (Wei, Lee *et al.*, 2001). Subsequent promoter trap experiments confirmed this and further analyses revealed that SdiA positively regulates CsrB transcription and biofilm formation in a UvrY-dependent manner (Suzuki, Wang *et al.*, 2002; Van Houdt, Aertsen *et al.*, 2006).

The RNA-binding protein CsrA

The Csr system regulates gene expression post-transcriptionally through the action of the 61 amino acid CsrA protein, which binds specifically to conserved sequences found predominately in the 5'-untranslated leader of regulated transcripts (Liu, Yang *et al.*, 1995; Wei, Brun-Zinkernagel *et al.*, 2001; Baker, Morozov *et al.*, 2002). In most cases, CsrA binds multiple target sites in the RNA, at least one of which overlaps the SD, thereby preventing ribosome loading and translation initiation (Baker, Morozov *et al.*, 2002; Dubey, Baker *et al.*, 2003; Wang, Dubey *et al.*, 2005; Baker, Eory *et al.*, 2007). Studies have shown in all but one instance (*hfq*), CsrA-dependent gene repression results in part from decreased mRNA stability (Liu, Yang *et al.*, 1995; Baker, Eory *et al.*, 2007). However, CsrA binding does not greatly affect the stability of either CsrB or CsrC RNA (Gudapaty, Suzuki *et al.*, 2001; Weilbacher, Suzuki *et al.*, 2003). The exact mechanism of CsrA-mediated mRNA decay has not been demonstrated, but may be due to a passive decay process where CsrA prevents ribosome access along with subsequent mRNA protection that is a function of the actively translating complex. Thus, increased opportunity for RNaseE cleavage at internal sites in the mRNA leads to rapid decay

(Yarchuk, Jacques *et al.*, 1992; Vytvytska, Moll *et al.*, 2000; Babitzke and Romeo, 2007). Little is known about the mechanism of CsrA-dependent gene activation except that messages appear to be stabilized by protein binding (Sabnis, Yang *et al.*, 1995; Wei, Shin *et al.*, 2000; Wei, Brun-Zinkernagel *et al.*, 2001). A more thorough discussion of CsrA-regulated genes as well as an analysis of CsrA structure and function is provided below.

CsrD protein and CsrB/C decay

CsrD is a protein factor of the Csr system that specifically targets CsrB and CsrC sRNA for decay by RNaseE. K. Suzuki and K. Jonas discovered CsrD independently during random transposon mutagenesis screens for regulators of CsrB expression (Jonas, Tomenius *et al.*, 2006; Suzuki, Babitzke *et al.*, 2006). CsrD is not a nuclease, however a mutant is characterized by a dramatic stabilization of both sRNAs, increasing the half life of CsrB and CsrC from 1.4 min and 2.2 min, respectively, to >30 min. (Suzuki, Babitzke *et al.*, 2006). Consistant with what is known of the Csr regulatory circuit, the same mutation also causes a decrease in CsrB and CsrC transcription which results from their increased stability; in effect, the overexpression of both sRNAs titrates active CsrA from the cellular pool for a longer period of time, thus decreasing their own CsrA-dependent transcription. CsrD contains two transmembrane domains that likely anchor the protein at the cell membrane, a HAMP (histidine kinases, adenylyl cyclases, methyl binding proteins and phosphatases)-like domain, and degenerate GGDEF and EAL domains. Recently, GGDEF and EAL containing proteins have been characterized to synthesize and hydrolyze the bacterial second messenger cyclic-di-GMP, respectively (see Romling and Amikam, 2006 for review). However, CsrD was not found to synthesize or

hydrolyze cyclic-di-GMP and site-directed mutagenesis of both putative active sites did not affect protein activity; thus, CsrD represents the first known example of a GGDEF/EAL protein with a cyclic-di-GMP-independent function (Suzuki, Babitzke *et al.*, 2006).

CsrA-regulated genes in *E. coli* K12

Glycogen metabolism in *E. coli* is highly regulated through both allosteric and genetic means (Preiss and Romeo, 1989; Romeo and Preiss, 1989). Three enzymes are essential for glycogen biosynthesis, ADP-glucose pyrophosphorylase, glycogen synthase, and glycogen branching enzyme, which are encoded by *glgC, glgA and glgB*, respectively. CsrA is known to negatively regulate expression of these genes in addition to *pckA (*phosphoenolpyruvate carboxykinase) (Romeo, Gong *et al*., 1993), as well as those for glycogen breakdown which include *glgX*, the glycogen debranching enzyme, *glgP*, glycogen phosphorylase, and *glgS*, whose gene product stimulates glycogen synthesis (Liu, Yang *et al.*, 1995; Yang, Liu *et al.*, 1996). CsrA also regulates expression of enzymes involved in central carbon metabolism, specifically the Embden-Meyerhoff pathway, such as phosphoglucomutase, fructose-1, 6,-bisphosphate, phosphoenolpyruvate synthetase, glucose-6-phosphate isomerase, triose-phosphate isomerase, enolase, pyruvate kinase A and F, phosphofructokinase I and II, and those for acetate metabolism, acetyl-coenzyme A synthetase, and the glyoxylate shunt, isocitrate lyase (Sabnis, Yang et al., 1995; Yang, Liu et al., 1996; Wei, Shin et al., 2000). By far the best characterized CsrA-regulated gene in this metabalon is *glgC* whose 5'-UTR contains 4 CsrA binding sites (Romeo, Gong *et al.*, 1993; Liu and Romeo, 1997; Baker, Morozov *et al.*, 2002;

Mercante, Edwards *et al*., submitted). Two of the four target sites have been well defined; one site overlaps the SD while the other is located 28 nt upstream in an apparent hairpin loop (Baker, Morozov *et al.*, 2002). Through toeprint analysis CsrA was shown to physically compete with the ribosome for access to the transcript and S30-driven *in vitro* transcription-translation studies proved that this interaction inhibited *glgC'-'lacZ* translation (Liu and Romeo, 1997; Baker, Morozov *et al.*, 2002). In addition, the absence of translation destabilizes the message by presumably allowing RNase access to the untranslated and therefore unprotected transcript (Yarchuk, Jacques *et al.*, 1992; Liu, Yang *et al.*, 1995).

FlhDC is a hetero-oligomeric zinc-associated protein complex that is the master transcriptional activator of flagellum biosynthesis and a regulator of anaerobic respiration and other factors (Pruss, Campbell *et al.*, 2003; Wang, Fleming *et al.*, 2006). CsrA has been experimentally shown to bind *flhDC* and positively regulate expression (Wei, Brun-Zinkernagel *et al.*, 2001); however, while the *flhDC* 5'-UTR is a direct CsrA target, identification of exact binding sites has proven elusive due to extensive RNA secondary structure. In a manner opposite from regulation of *glgCAP* expression, CsrA-binding stabilizes the *flhDC* transcript; the steady-state level of *flhDC* mRNA is ~ 3-fold higher in a *csrA* wild type strain compared to its isogenic mutant (Wei, Brun-Zinkernagel *et al.*, 2001). Phenotypically, a *csrA* transposon mutant produces no visible flagella and is completely non-motile in tryptone medium containing a low percentage of agarose, although growth is unaffected (Wei, Brun-Zinkernagel *et al.*, 2001).

The proteins encoded by the *pgaABCD* operon are responsible for the synthesis of a cell-bound extracellular β-1,6-*N*-acetyl-D-glucosamine polysaccharide adhesin

(PGA) that is essential for *E. coli* biofilm formation (Wang, Preston *et al.*, 2004). Expression of *pgaABCD* (Wang, Dubey *et al.*, 2005) as well as a divergently transcribed GGDEF encoding gene, *ycdT*, is post-transcriptionally regulated by CsrA (Jonas, Edwards *et al.*, 2008). CsrA exhibited a specific, high-affinity interaction with the *pga* 5'-untranslated leader which was revealed to contain 6 CsrA target sequences by footprint, toeprint and boundary analyses. One target site overlaps the SD sequence while another is found at the initiation codon; this is the greatest number of binding sites yet discovered in a CsrA-regulated mRNA (Wang, Dubey *et al.*, 2005). Similar to *glgC*, CsrA competed with the ribosome for binding to *pgaA* mRNA which was subsequently stabilized in a CsrA mutant strain (Wang, Dubey *et al.*, 2005). Biochemically, a CsrA mutation causes an ~9 fold increase in *pgaA'-'lacZ* expression, a ~3-fold increase in PGA production and a subsequent increase in biofilm formation at 26^{o} C (Wang, Dubey *et al.*, 2005).

Under conditions of carbon starvation or upon entering stationary phase *E. coli* induces expression of a gene for the uptake and utilization of extracellular peptides, *cstA* (Schultz and Matin, 1991). *cstA* was originally identified as a CsrA target by *in silico* analysis for probable binding sites (Dubey, Baker *et al.*, 2003). *cstA* expression was derepressed in a *csrA* mutant ~2-fold; derepression increased to 5- to 10-fold upon addition of glucose. CsrA bound specifically to *cstA* mRNA and studies employing nucleolytic probing, footprinting and boundary analysis revealed up to 4 locations, 2 in the 5'-UTR and 2 in the ORF, that were bound by CsrA; importantly, site II overlapped the *cstA* SD sequence (Dubey, Baker *et al.*, 2003). Toeprinting and S30-coupled transcription-translation studies showed that CsrA repressed *cstA* translation in the same

fashion as *glgC*, by blocking ribosome access to the RBS, thus facilitating mRNA decay. The study by Dubey *et. al*., (2003) on *cstA* was important because it experimentally verified earlier claims of the multimeric composition of CsrA (Baker, Morozov *et al.*, 2002) proving by chemical crosslinking, SDS-PAGE and MALDI-TOF mass spectrometry that CsrA was a homodimer (Dubey, Baker *et al.*, 2003).

The Hfq protein is an RNA-binding global regulatory protein found in a variety of bacteria, including *E. coli*, *Salmonella sp.* (Vogel, 2009), *Vibrio paraheamolyticus* (Nakano, Takahashi *et al.*, 2008), *Francisella tularensis* (Meibom, Forslund *et al.*, 2009), *Synechocystis* sp. (Dienst, Duhring *et al.*, 2008) *Neisseria* sp. (Fantappie, Metruccio *et al.*, 2009) and many other genera (Brennan and Link, 2007). Hfq is a homohexameric Lsm-type RNA chaperone that facilitates sRNA-mRNA base-pairing and subsequent post-transcriptional regulation via two RNA-binding surfaces on opposite faces of its ring-like structure (Nogueira and Springer, 2000; Schumacher, Pearson *et al.*, 2002; Sauter, Basquin *et al.*, 2003; Thore, Mayer *et al.*, 2003). Hfq transcriptional expression is controlled by normal housekeeping promoters (σ^{70}) as well as by those for heat-shock induction (σ^{32}) (Valentin-Hansen, Eriksen *et al.*, 2004) and Hfq autoregulates its own expression post-transcriptionally (Vecerek, Moll *et al.*, 2005). Bioinformatics initially identified a putative CsrA target site overlapping the *hfq* SD sequence that was confirmed by gel shift analysis and ribonuclease probing (Baker, Eory *et al.*, 2007). RNA toeprinting revealed that CsrA competitively blocked ribosome access to the RBS leading to translational repression which had a ~2-fold effect on *hfq'-'lacZ* expression (Baker, Eory *et al.*, 2007). Curiously, one binding site for Hfq itself overlaps that of the sole CsrA target on the *hfq* message, and addition of both proteins to *in vitro* transcription

reactions using mRNA from an *hfq'-'GFP* fusion resulted in additive translational repression. Significantly, unlike every other CsrA-repressed message, binding by CsrA did not destabilize the *hfq* mRNA. Thus, *hfq* represents a new model by which CsrA represses gene expression without affecting mRNA stability.

As discussed above, the GGDEF and EAL protein domains have within the last 10 years been recognized for their ability to synthesize and hydrolyze cyclic di-GMP (c-di-GMP), respectively (Romling and Amikam, 2006). Cyclic di-GMP is a bacterial second messenger that regulates a growing number of cellular processes, including cellulose and fimbrae production, motility, biofilm formation, and various virulence associated traits (see Romling and Amikam, 2006 for a review of c-di-GMP signaling). Recently, mRNAs of two genes of unknown function containing GGDEF domains, *ydeH* and *ycdT*, were found to be down regulated after an *in vivo* "pulse" of CsrA from an arabinose-inducible multicopy vector (Jonas, Edwards *et al.*, 2008). Both proteins appeared to actively synthesize c-di-GMP *in vivo*, negatively regulating *E. coli* and *S. enterica* flagella-based swimming motility. However, only overexpression of YdeH led to a decrease in *E. coli* flagella production while deletion of either gene resulted in a slight increase in motility. The proteins were confirmed to be important for signaling as mutations in the conserved GGDEF region abolished their effect on motility. Both genes contain dual CsrA binding sites with one target overlapping the SD sequence. CsrA was demonstrated to bind each transcript with high affinity and, while mRNA half-lives were not calculated, real-time RT-PCR and transcriptional fusions showed that both messages were destabilized without effects on transcription, as opposed to CsrA regulation of *hfq* which has a mild, indirect negative effect on transcription (Jonas, Edwards *et al.*, 2008).

YdeH also has an apparent effect on biofilm formation mediated through the *pgaABCD* operon, although experiments are ongoing (Jonas, Edwards *et al.*, 2008; Panuri, Goller *et al*., in progress). Additionally, several other GGDEF and EAL-encoding genes were suggested as part of the Csr regulon, including *yddV*, *dos*, *yliF*, *yhjK*, *yliE*, *yjcC* and *csrD* (Jonas, Edwards *et al.*, 2008). These data are consistent with the Csr system as a repressor of sessility and an activator of motility.

Csr/Rsm Circuitry in other bacteria

The Csr circuit is most clearly understood in the non-pathogenic model bacterium *E. coli* K12. However, orthologous Csr circuits and components have been identified in a large number of eubacteria. Studies of the Csr system in most other species have centered on its role in virulence and/or pathogenesis. Though BarA/UvrY homologs are common, research has not always identified CsrA and this TCS within the same species. Therefore this section will focus on those systems that minimally contain the central component of the network, CsrA, sometimes referred to as RsmA/E for regulator of secondary metabolites.

***Salmonella enterica* serovar Typhimurium**

The *Salmonella* Csr circuitry is closely related to that of *E. coli*. *S. enterica* (serovar Typhimurim) contains a CsrA homolog identical to its *E. coli* cousin, as well as orthologs of BarA (Altier, Suyemoto *et al.*, 2000), UvrY (SirA) (Johnston, Pegues *et al.*, 1996), CsrB (Altier, Suyemoto *et al.*, 2000) and CsrC (Fortune, Suyemoto *et al.*, 2006). Unlike in *E. coli*, CsrA does not play a prominent role in *Salmonella* carbon metabolism

(Lawhon, Frye *et al.*, 2003), however, it has been shown that glycogen accumulation is correlated with biofilm formation in *S. enteritidis* and CsrA represses biofilm formation in the close relative *S. enterica* (Bonafonte, Solano *et al.*, 2000). Thus, additional studies are required to clarify the role of CsrA in *Salmonella* carbon metabolism. Both *csrA* knockout and overexpression appear to negatively influence *Salmonella* invasion and pathogenecity through indirect regulation of the *Salmonella* pathogenecity island I (SPI-1)-encoded transcriptional regulator HilA which then causes a cascade of virulence gene expression, activating *invF*, *orgA*, *prgH* and *etc*; *invF* in turn activates expression of *etc* and the secreted factors *sopB* and *sspC/sipC* (Bajaj, Lucas *et al.*, 1996; Altier, Suyemoto *et al.*, 2000) (see Altier, 2005; Jones, 2005 for reviews). *Salmonella* intestinal cell invasion is dependent upon its type III secretion apparatus and various effector proteins encoded on SPI-1 and SPI-2. CsrA was shown to also influence the expression of the invasion genes *invF* and *sspC* independent of *hilA*, possibly by binding and stabilizing their respective messages, however, neither *in vitro* regulation nor direct interaction has ever been shown (Altier, Suyemoto *et al.*, 2000). *hilA* is found within SPI-1 and is transcriptionally activated by *hilD, hilC* and phosphorylated SirA, which directly binds the *hilA* transcript (Bajaj, Hwang *et al.*, 1995; Bajaj, Lucas *et al.*, 1996; Schechter, Damrauer *et al.*, 1999; Lucas and Lee, 2001; Teplitski, Goodier *et al.*, 2003). In addition, both SirA and CsrA were found to activate expression of *hilC* and *hilD* (Altier, Suyemoto *et al.*, 2000; Teplitski, Goodier *et al.*, 2003; Ellermeier, Ellermeier *et al.*, 2005), however only SirA has ever been shown to directly bind *hilC* promoter DNA (Teplitski, Goodier *et al.*, 2003). Consistent with the established *E. coli* model, SirA activates expression of *csrB* and *csrC*, and although SirA-*csrB* interaction was experimentally verified, binding

to *csrC* could not be shown (Altier, Suyemoto *et al.*, 2000; Altier, Suyemoto *et al.*, 2000; Fortune, Suyemoto *et al.*, 2006; Teplitski, Goodier *et al.*, 2006). Unlike the *E. coli* circuit, CsrA does not regulate *hilA* expression through the activity of the BarA/SirA TCS (Altier, Suyemoto *et al.*, 2000; Teplitski, Goodier *et al.*, 2003; Fortune, Suyemoto *et al.*, 2006); indeed, one study suggests CsrA does not regulate expression of the sensor kinase BarA in *Salmonella* spp. (Altier, Suyemoto *et al.*, 2000). In addition, *csrA* mutation was proposed to decrease the stability of CsrC as revealed by quantitative PCR, however a mechanism for this affect was not investigated (Fortune, Suyemoto *et al.*, 2006).

A minor subject of interest and discussion has been the relationship between the Csr system and *Salmonella* motility. SirA was found to negatively regulate expression of the *flhDC* master operon (Teplitski, Goodier *et al.*, 2003) and multiple flagellar genes including *fliA*, *motB*, *flgA* and *fliC* (Goodier and Ahmer, 2001), however the authors speculate most SirA regulation of *flhDC* could be indirect (Goodier and Ahmer, 2001). This indirect mechanism may involve SirA activation of *csrB* expression which titrates active CsrA leading to *flhDC* deactivation (Teplitski, Goodier *et al.*, 2003). Recent microarray analysis revealed that CsrA positively affects the expression of up to 16 flagellar biosynthesis genes in addition to 7 for chemotaxis; CsrA is also required for the production of flagella and a motile phenotype (Lawhon, Frye *et al.*, 2003). However, experiments further describing the direct interaction of CsrA with flagellar transcripts, as was established in *E. coli*, have not been conducted in *Salmonella* spp. In addition, SirA has been shown to affect *Salmonella* biofilm formation by directly binding and regulating expression of the *fim* operon for type I fimbrae (Teplitski, Al-Agely *et al.*, 2006). Finally, the expression of the secreted effector protein AvrA, which interferes with the

eukaryotic NFκB pathway, is thought to be regulated post-transcriptionally by CsrA, but more work will be required to confirm this observation (Kerrinnes, Zelas *et al.*, 2008). It should be noted that while *Salmonella* spp. do encode orthologous components of the *E. coli* Csr system, no study to date has shown direct binding of CsrA to any *Salmonella* mRNA or biochemical proof of post-transcriptional control by this system.

Legionella pneumophila

Legionella pneumophila are water-borne γ-proteobacteria that inhabit single-cell protozoa such as amoeba and, when aerosolized, can infect human alveolar macrophages leading to the development of Legionnaire's disease. *L. pneumophila* contains most known components of the Csr system including LetS (BarA), LetA (UvrY) (Hammer, Tateda *et al.*, 2002), two small non-coding RNAs, RsmY and RsmZ (Kulkarni, Cui *et al.*, 2006; Hovel-Miner, Pampou *et al.*, 2009; Rasis and Segal, 2009; Sahr, Bruggemann *et al.*, 2009) that function in a similar manner as *E. coli* CsrB and CsrC, and a CsrA ortholog that can functionally replace the *E. coli* version (Fettes, Forsbach-Birk *et al.*, 2001; Molofsky and Swanson, 2003). In addition, a regulator of *csrA* expression, PmrA, has also been described that may bind to a conserved sequence in the *csrA* 5' DNA leader and activate transcription (Rasis and Segal, 2009). However, physical PmrA-*csrA* association has not yet been demonstrated. Interestingly, all *L. pneumophila* strains characterized to date contain a minimum of 3 (strain Lens) and a maximum of 5 (strains Paris and Corby) non-identical *csrA* genes (Finn, Tate *et al.*, 2008); the purpose for this seeming redundancy is not known but hints at the critical role for this protein in the *Legionella* life cycle. The transmissive and replicative phenotypes of *L. pneumophila* are

tightly regulated by the Csr system (Fettes, Forsbach-Birk *et al.*, 2001; Molofsky and Swanson, 2003; Rasis and Segal, 2009; Sahr, Bruggemann *et al.*, 2009) as well as by σ^s (Bachman and Swanson, 2004; Hovel-Miner, Pampou *et al.*, 2009) and a recently described enhancer protein, LetE (Hammer, Tateda *et al.*, 2002; Bachman and Swanson, 2004) (see Molofsky and Swanson, 2004 for a review). It was discovered early on that *L. pneumophila* CsrA is required for intracellular growth but not for the initial infection process; CsrA is only expressed during the replicative and not the transmissive phase (Molofsky and Swanson, 2003). A *csrA* mutant is characterized by premature expression of transmissive/stationary phase traits such as motility and cytotoxicity (Molofsky and Swanson, 2003; McNealy, Forsbach-Birk *et al.*, 2005) while constituitive or overexpression suppresses virulence-associated traits such as resistance to heat and osmotic stress (Molofsky and Swanson, 2003; Forsbach-Birk, McNealy *et al.*, 2004). CsrA appears to mediate *L. pneumophila* phenotype switching partly through its effects on expression of the flagellar structural gene *flaA* (Fettes, Forsbach-Birk *et al.*, 2001; McNealy, Forsbach-Birk *et al.*, 2005) and the flagellar sigma factor FliA (Fettes, Forsbach-Birk *et al.*, 2001; Molofsky and Swanson, 2003; McNealy, Forsbach-Birk *et al.*, 2005), which is required for several of the aforementioned traits in addition to evasion from phagosome-lysosome fusion (Molofsky and Swanson, 2003).

The LetA/S TCS was found to relieve CsrA-mediated repression allowing for expression of the transmissive phenotype (Hammer, Tateda *et al.*, 2002; Molofsky and Swanson, 2003). Similar to *E. coli* BarA/UvrY, LetA/S functions primarily through activation of *rsmY* and *rsmZ* which were experimentally confirmed by three different groups simultaneously (Hovel-Miner, Pampou *et al.*, 2009; Rasis and Segal, 2009; Sahr,

Bruggemann *et al.*, 2009) after they were predicted by *in silico* analysis (Kulkarni, Cui *et al.*, 2006). Like CsrB and CsrC, RsmY and RsmZ, contain conserved GGA target sequences displayed in the single-stranded region of hairpin loops (Sahr, Bruggemann *et al.*, 2009); the *L. pneumophila* sRNAs are also functionally similar to their *E. coli* counterparts because overexpression in *E. coli* leads to glycogen accumulation (Sahr, Bruggemann *et al.*, 2009). Consistent with the *E. coli* model, LetA has been shown to directly bind RsmY and RsmZ DNA (Rasis and Segal, 2009; Sahr, Bruggemann *et al.*, 2009). Additionally, their overexpression leads to downregulation of replicative genes, and upregulation of transmissive genes (Sahr, Bruggemann *et al.*, 2009). CsrA directly binds both sRNAs (Sahr, Bruggemann *et al.*, 2009) and it has been presumed that CsrA also binds to the RNAs for at least 3 secreted effector proteins, YlfA, YlfB and VipA (Rasis and Segal, 2009), and regulates their expression, however, further studies are sorely needed to confirm these findings. While very little is known about LetE, it appears to enhance FlaA production by an unknown mechanism, possibly by increasing *flaA* mRNA stability (Bachman and Swanson, 2004). The stationary phase sigma factor, σ^S, has also been implicated in *L. pneumophila* Csr regulation; RpoS was shown to affect RsmY and RsmZ transcription (Rasis and Segal, 2009) while CsrA may repress *rpoS* expression (Forsbach-Birk, McNealy *et al.*, 2004). One additional element that may contribute to Csr regulation of replicative vs. transmissive traits is the amino acid starvation sensing protein, RelA. Preliminary experiments suggest that after active replication in the host cell, *L. pneumophila* RelA synthesizes ppGpp that activates LetA/S, thus relieving CsrA repression and allowing the bacterium to enter the

transmissive phase (Hammer, Tateda *et al.*, 2002; Molofsky and Swanson, 2003). However, further studies are required to confirm this effect.

Vibrio spp.

The *Vibrio* spp. Csr system has not been extensively studied; however, a connection to quorum sensing is documented in at least one study (Lenz, Miller *et al.*, 2005). The *V. cholerae* Csr circuit is composed of a CsrA homolog, the two component system VarS/VarA (BarA/UvrY) and 3 small, noncoding RNAs, CsrB, CsrC and CsrD (Lenz, Miller *et al.*, 2005). *V. cholerae* CsrA and sRNAs have been shown to functionally replace the *E. coli* versions and regulate glycogen accumulation. VarS/VarA and CsrA were first identified in a screen for regulators of *V. cholerae* quorum sensing that influenced light production from a *V. harveyi luxCDABE* reporter. Subsequent *in silico* analysis led to the discovery of all three sRNAs. The *V. cholerae* VarS/VarA-CsrA/B,C,D circuit functions in a manner similar to the *E. coli* Csr system; additionally, the pathway runs parallel to the two characterized quorum sensing relays based on CqsS and LuxP/Q. An unknown growth-dependent signal induces a regulatory cascade whereby VarS phosphorylates VarA which then activates expression of CsrB,C and D. These sRNAs apparently bind CsrA and modulate its activity; CsrA indirectly activates LuxO that ultimately regulates the expression of HapR responsible for protease production and virulence gene expression (Lenz, Miller *et al.*, 2005). Through this LuxO-dependent circuit, the *V. cholerae* Csr system was also shown to regulate expression of *vpsT* and *vpsL* which encode a regulator and a structural gene for exopolysaccharide production, respectively. A similar mechanism is assumed to exist

that regulates *V. vulnificus* biofilm as well. Interestingly, in contrast to *V. vulnificus* isolates from water, blood or wounds, those from oysters were found to be *csrA*- and therefore proficient for biofilm formation (Jones, Warner *et al.*, 2008).

***Erwinia* spp.**

Erwinia carotovora subspecies *carotovora* is a plant pathogen that causes soft-rotting disease through secretion of various extracellular enzymes, including pectate lysase (Pel), polygalacteronase (Peh), cellulose (Cel) and protease (Prt) (Chatterjee, Cui *et al.*, 1995). The *Erwinia* CsrA homolog, RsmA, represses these virulence factors and overall pathogenecity (Chatterjee, Cui *et al.*, 1995; Cui, Chatterjee *et al.*, 1995). In addition to RsmA, *Erwinia* spp. encode a BarA/UvrY (ExpS/ExpA) (Cui, Chatterjee *et al.*, 2001) TCS and a single CsrB-like sRNA, RsmB (Liu, Cui *et al.*, 1998). These three common Csr components can all functionally replace their *E. coli* counterparts; RsmA and RsmB have opposing affects on *Erwinia* sp. glycogen accumulation (Cui, Chatterjee *et al.*, 1995; Liu, Cui *et al.*, 1998) while ExpA stimulated CsrB expression. *Erwinia* sp. also contain a relatively uncharacterized protein factor, RsmC, (Cui, Mukherjee *et al.*, 1999) not found in *E. coli* that activates RsmA production but represses RsmB transcription. The PmrA-PmrB two component system which controls virulence and cell surface properties was also characterized in *Erwinia* spp., however, unlike in *L. pneumophila*, PmrA does not appear to affect expression of any component of the Csr system (Hyytiainen, Sjoblom *et al.*, 2003). Through similarity to *E. coli* CsrA, RsmA is believed to post-transciptionally regulate expression of the aforementioned extracellular enzymes as well as HrpL (Chatterjee, Cui *et al.*, 2002), an alternative sigma factor

required for synthesis of the type III secretion apparatus and the hypersensitive response in plants. However, unlike in *E. coli*, RsmA stabilizes the RsmB transcript (Chatterjee, Cui *et al.*, 2002). While RsmA undoubtedly has a large effect on gene regulation in *Erwinia* spp., experiments to confirm post-transcriptional processes have shown little convincing evidence and no study has yet demonstrated direct binding of RsmA to any m RNA transcript.

rsmA expression itself may be mediated by RpoS (Mukherjee, Cui *et al.*, 1998) and is sensitive to 2 classes of AHL quorum sensing signaling molecules, both produced by *Erwinia* spp. (Koiv and Mae, 2001; Chatterjee, Cui *et al.*, 2005; Cui, Chatterjee *et al.*, 2005). ExpR1 and ExpR2 are putative AHL acceptors of the LuxR family found in most *Erwinia* strains. These regulators were recently described to bind non-overlapping operator sites on the *rsmA* DNA leader in the absence of their respective AHL and activate transcription (Chatterjee, Cui *et al.*, 2005; Cui, Chatterjee *et al.*, 2005; Cui, Chatterjee *et al.*, 2006; Sjoblom, Brader *et al.*, 2006). The presence of AHL neutralizes this effect, thus decreasing *rsmA* expression and upregulating virulence factor production.

***Pseudomonas* spp.**

The *Pseudomonads* are a diverse family within the γ-proteobacteria composed of animal and plant-colonizing members. The Csr system has been best characterized in two *Pseudomonas* species: *P. fluorescens*, a plant beneficial root-colonizing biocontrol strain that suppresses diseases caused by soil-borne fungi (Blumer, Heeb *et al.*, 1999), and *P. aeurginosa*, an opportunistic human pathogen that is commonly found in the lungs of patients with cystic fibrosis and other immune compromising diseases (Pessi, Williams

et al., 2001). Two CsrA paralogs, RsmA (Blumer, Heeb *et al.*, 1999) and RsmE (Haas, Keel *et al.*, 2002; Reimmann, Valverde *et al.*, 2005) have been described in *P. fluorescens* along with the GacS/GacA (BarA/UvrY) two component system and 3 sRNAs, RsmX (Kay, Dubuis *et al.*, 2005), RsmY (Valverde, Heeb *et al.*, 2003) and RsmZ (Heeb, Blumer *et al.*, 2002), that directly interact with both RsmA (Valverde, Heeb *et al.*, 2003) and RsmE (Kay, Dubuis *et al.*, 2005). Similar to the *E. coli* Csr circuit, transcription of *rsmX*, *rsmY* and *rsmZ* appear to be GacS/GacA-dependent (Heeb, Blumer *et al.*, 2002; Kay, Dubuis *et al.*, 2005), and RsmY expression (and probably that of the other sRNAs) is influenced indirectly by RsmA (Valverde, Heeb *et al.*, 2003) in a GacA-dependent manner. However, this final assertion has been discounted by at least one study that erroneously choose to examine a GacA'-'LacZ translational fusion for regulation by RsmA (Valverde, Heeb *et al.*, 2003); recall in *E. coli* that CsrA indirectly influences UvrY activity independent of its expression (Suzuki, Wang *et al.*, 2002). The stability of RsmY (and presumably RsmX and RsmZ) was also found to be dependent on its repeated CsrA target motifs; deletion of several GGA trinucleotides led to a decrease in sRNA half-life (Valverde, Lindell *et al.*, 2004). Both RsmA (Blumer, Heeb *et al.*, 1999) and RsmE (Haas, Keel *et al.*, 2002) bind target sequences within mRNAs for anti-fungal enzymes such as hydrogen cyanide (HCN) and an extracellular protease (AprA) at the 5'-untranslated leader (Blumer, Heeb *et al.*, 1999; Blumer and Haas, 2000) thereby repressing their expression.

The *P. aeuruginosa* Csr circuit consists of a single CsrA ortholog (RsmA) (Pessi, Williams *et al.*, 2001), the two component system GacS/GacA, two untranslated RNAs RsmY (Kay, Humair *et al.*, 2006) and RsmZ (Heurlier, Williams *et al.*, 2004; Burrowes,

Abbas *et al.*, 2005), and 2 additional factors, LadS (Ventre, Goodman *et al.*, 2006) and RetS (Goodman, Kulasekara *et al.*, 2004; Ventre, Goodman *et al.*, 2006), that appear to have opposing affects on the expression of RsmZ. In addition, Hfq has been shown to interact directly with the single-stranded region of RsmY between CsrA-bound hairpin loops, thus stabilizing the sRNA (Sorger-Domenigg, Sonnleitner *et al.*, 2007). The role of RsmA in *P. aeurginosa* pathogenecity has been studied in some detail; RsmA appears to be important for initial invasion and colonization of airway epithelium (Mulcahy, O'Callaghan *et al.*, 2006; Mulcahy, O'Callaghan *et al.*, 2008), although *in vitro* and *in vivo* studies conflict, and an *rsmA* mutant has decreased fitness and mortality but displays increased pulmonary inflammation and chronic persistence compared to the wild-type PAO1in a mouse model of infection (Mulcahy, O'Callaghan *et al.*, 2008). Specifically, RsmA is expressed in stationary phase and at high cell densities (Pessi, Williams *et al.*, 2001) where it represses expression of staphylolytic activity (LasA protease), HCN, pyocyanine pigment (Pessi, Williams *et al.*, 2001), pyoluteorin (Plt) antibiotic production (Huang, Zhang *et al.*, 2008), genes for iron utilization (Burrowes, Baysse *et al.*, 2006), and a multi-drug efflux pump (MexAB-OprM) (Burrowes, Baysse *et al.*, 2006). In contrast, RsmA upregulates genes for motility (Burrowes, Baysse *et al.*, 2006) and the type III secretion system toxic effector proteins ExoS, ExoT, PopB, PopD and PcrV (Mulcahy, O'Callaghan *et al.*, 2006).

CsrA in various other bacteria

By simple sequence comparison a large number of different eubacteria (currently ~300) are known to contain CsrA-like protein coding genes (Finn, Tate *et al.*, 2008),

however, only a handful of individual species have been studied in depth. In addition to the previously described examples, brief descriptions of Csr circuitry have been made in at least 12 other genera: The CsrA protein along with 2 noncoding CsrB and CsrC-like RNAs have been characterized in the the fecal-oral human pathogen *Yersinia pseudotuberculosis* to positively regulate motility through binding of *flhDC* mRNA and indirectly control factors necessary for host cell invasion (Heroven, Bohme *et al.*, 2008). In the human gastric pathogen, *Helicobacter pylori*, CsrA is required for adaptation to environmental conditions. Loss of *csrA* decreases bacterial resistance to oxidative stress, heat shock and increases sensitivity to acidic surroundings; not surprisingly, a *csrA* mutant is attenuated for virulence in a mouse model of infection (Barnard, Loughlin *et al.*, 2004; Wang, Alamuri *et al.*, 2006). *csrA* strains are also less motile compared to wild type and expression of key flagellar genes, *flaA* and *flaB,* is decreased, although their mRNA stability is unaffected (Barnard, Loughlin *et al.*, 2004). Interestingly, ectopically expressed *H. pylori* CsrA could not complement an *E. coli* mutant, however, plasmid-based protein expression was not verified.

Serratia marscescens, a nosocomial infection-associated human pathogen, contains the CsrA, CsrB and CsrD orthologs, RsmA, RsmB and PigX, respectively, which regulate swarming motility, AHL and surfactant production (Ang, Horng *et al.*, 2001; Williamson, Fineran *et al.*, 2008). While an *rsmA* knockout in *Proteus mirabilis* was not possible, overexpression inhibited swarming motility as well as the virulence factors haemolysin, protease and urease (Liaw, Lai *et al.*, 2003). The respiratory pathogen *Haemophilus influenza* synthesizes a lipooligosaccharide (LOS) phosphorylcholin (PC) moiety that is important for colonization and persistence. A CsrA

ortholog in this species may regulate carbohydrate flux similar to *E. coli*, and thus control the precursors for this LOS modification (Wong and Akerley, 2005). *Xanthomonas campestris* is a phytopathogen that employs RsmA to regulate expression of virulence factors essential for pathogenesis and disease. These determinants include extracellular amylase, endoglucanase, polysaccharides, type III secretion system components and effectors. An *rsmA* mutant of *X. campestris* is therefore attenuated and unable to replicate inside host cells (Chao, Wei *et al.*, 2008).

The model Gram-positive spore-forming bacterium *Bacillus subtilis* is the first Firmicute in which *csrA* has been characterized. *csrA* is the second gene in a two gene σ^A –dependent operon that is found at the end of a larger operon containing genes for flagellum biosynthesis. Both *csrA* overexpression and a knockout exhibited reduced *B. subtilis* motility on agar plates (Yakhnin, Pandit *et al.*, 2007). The gene for the major flagellin protein, *hag*, is located directly downstream from *csrA*. CsrA represses Hag expression by binding to conserved GGA target sequences in the *hag* 5'-untranslated leader, including one that overlaps the SD sequence, thus blocking ribosome loading (Yakhnin, Pandit *et al.*, 2007). CsrA and CsrB from the insect pathogen *Photorhabdus luminescens* were recently found to complement their respective *E. coli* mutants for regulation of glycogen synthesis (Krin, Derzelle *et al.*, 2008). CsrA in this bacterium appears to regulate virulence factors including genes for AI-2 synthesis and transport, proteolytic degradation and antibiotic synthesis enzymes. CsrA of *Campylobacter jejuni*, a leading cause of human gastroenteritis, also regulates pathogenecity. A *csrA* mutant strain is defective for swarming, sensitive to oxidative stress and does not adhere well to

intestinal epithelium; however, similar to *L. pneumophila*, a *C. jejuni csrA* mutant displays a large increase in cellular invasion (Fields and Thompson, 2008).

Protein RNA-binding families

As demonstrated above, a large number of RNA-binding proteins exist in both eukaryotes and prokaryotes that are important for post-translational processes; however, the critical structural motifs which confer RNA-binding specificity upon these proteins are relatively few by comparison (Lunde, Moore *et al.*, 2007; Glisovic, Bachorik *et al.*, 2008). The position, orientation and/or combination of such motifs in a single protein imparts flexibility to RNA-binding protein construction (Lunde, Moore *et al.*, 2007; Glisovic, Bachorik *et al.*, 2008) in much the same way a seemingly disparate line of automobiles (composed of a minivan, a truck and a sedan) may use the same engine, frame or chassis components to assemble vehicles with completely different purposes. Following a modified categorization used by Chen and Varani (2005), the following section outlines those RNA-binding motif families which have been studied at the molecular level listing examples and describing their interaction with RNA with an emphasis on proteins involved in post-transcriptional regulation where possible.

Small basic arginine-rich motifs (ARM)

Although not truly a definable motif, many bacterial and viral proteins contain stretches of arginine and/or lysine residues which mediate interaction with RNA (Smith, Calabro *et al.*, 2000; Chen and Varani, 2005). These domains adopt vastly different conformations dependent on their particular RNA ligand; in some cases a single domain

can recognize multiple structurally unique target scaffolds via different protein configurations (Burd and Dreyfuss, 1994; Smith, Calabro *et al.*, 2000). For instance, the HIV transcriptional regulator, Tat, contains a stretch of basic amino acids, among which a single arginine is critical for RNA-binding (Calnan, Tidor *et al.*, 1991; Puglisi, Tan *et al.*, 1992). The interaction of Tat protein with the TAR (*trans*-activation response region) RNA hairpin is mediated by three nucleotides that stabilize arginine and allow it to hydrogen bond with backbone phosphates and a single guanine (G26) nucleotide (Puglisi, Tan *et al.*, 1992). In fact, the Tat basic region only becomes structured upon binding TAR where it also induces a conformational change in the RNA (Calnan, Biancalana *et al.*, 1991; Puglisi, Tan *et al.*, 1992). Solution NMR of the Bovine immunodeficiency virus (BIV) version of Tat-TAR prove that the unstructured peptide takes on a β-hairpin conformation (Calnan, Tidor *et al.*, 1991) when bound at the major groove of TAR RNA (Puglisi, Chen *et al.*, 1995); and studies show the HIV version adopts similar extended conformations when bound to TAR regardless of the peptide's original CD-spectra (Tan and Frankel, 1995). The HIV Rev peptide, however, which also binds at the major groove of a stem-loop structure, takes an all alpha helical conformation when bound to its cognate Rev response element (RRE) (Battiste, Mao *et al.*, 1996).

An interesting case of TAR induced conformational change is Jembrana disease virus (JDV) Tat protein, which is able to bind TAR sites from both BIV and HIV (Smith, Calabro *et al.*, 2000). Mutagenesis and 3D-NMR studies suggest that JDV Tat utilizes different arginine residues to bind BIV and HIV TAR sequences. And consistent with the idea of flexibility of small basic arginine-rich domains, the peptide adopts an extended structure when bound to HIV TAR but takes a β-hairpin conformation when

interacting with BIV TAR (Smith, Calabro *et al.*, 2000). This principal was used to examine ARM-synthetic RNA interactions *in vitro*; the study demonstrated that RNA ligands and ARMs can cross-recognize subsets of each other, and thus there are multiple mechanisms by which small arginine-rich domains and their cognate RNAs interact (Bayer, Booth *et al.*, 2005).

All helical proteins

The small RNA-binding Rop (repressor of primer) protein, also known as Rom, is encoded by the classic *E. coli* plasmid ColE1 as a regulator of plasmid copy number (Cesareni, Helmer-Citterich *et al.*, 1991). Rop stabilizes the pairing of a small 108 nt transcript, RNA I, with the 5'-end of the primer precursor, RNA II, thereby affecting the folding pathway of RNA II. The alternate folding pathway of RNA II prevents subsequent RNA-DNA hybrid formation ($\beta\alpha$ pairing is favored instead of $\alpha\beta$) with the plasmid origin and inhibits leading strand synthesis and DNA I polymerase-dependent replication (Cesareni, Helmer-Citterich *et al.*, 1991). The Rop protein itself is a 4 helix bundle homodimer composed of 2 helix-turn-helix monomers that form a coiled-coil structure wrapped around a hydrophobic core with 2-fold symmetry (Predki, Nayak *et al.*, 1995). The residues important for RNA-binding are aligned down one face of the dimer and form a symmetrical binding site encompassing a highly electropositive surface composed of helix 1 and 1'. Alanine substitution studies defined the Phe-14/Phe-14' pair as essential for interaction with RNA hairpin loops and binding is believed to occur by base stacking (Struble, Ladner *et al.*, 2008) or contact with the backbone. Four (total of 8 from both helices) hydrophilic amino acids (Lys-3, Asn-10, Gln-18 and Lys-25) that

surround the central Phe pair are also critical for RNA-binding; they likely interact with the ribose-phosphate backbone (Predki, Nayak *et al.*, 1995). Surprisingly, Rop-RNA binding is structure and not sequence-dependent, as mutations in the RNA which change the primary sequence but conserve the size of the single-stranded stem loop at 6 or 8, but not 7 nt, still allow protein binding.

The abundant eukaryotic protein domain SAM (sterile alpha motif) was originally characterized as a protein-protein interaction motif, but recent studies have identified a variant with the capacity to bind stem-loop RNA structures (Oberstrass, Lee *et al.*, 2006). The X-ray and NMR solution structures of the *Saccharomyces cerevisiae* post-transcriptional regulator Vts1 SAM domain were solved revealing a rigid, 5-helix bundle that has both sequence and structure-specific RNA-binding requirements (Aviv, Amborski *et al.*, 2006; Oberstrass, Lee *et al.*, 2006). Vts1 specifically binds and caps the CNGGN-pentaloop presented in an RNA hairpin as well as binding in the RNA helix major groove (Aviv, Amborski *et al.*, 2006; Oberstrass, Lee *et al.*, 2006); the C-terminal end of helix α_1, the N-terminal portion of helix α_5 and the loop connecting α_1 and α_2 directly bind RNA (Aviv, Amborski *et al.*, 2006). The majority of contacts are made between nucleic acid phosphate groups and residue side chains of Lys, His, Arg and to backbone amides of Leu and Gly (Aviv, Amborski *et al.*, 2006). The G3 base of the pentaloop sits in a shallow pocket lined by hydrophobic residues and also makes base-specific contacts with an Arg backbone carbonyl and a Tyr side chain (Aviv, Amborski *et al.*, 2006; Oberstrass, Lee *et al.*, 2006). Unlike ARM motifs, this all-helical domain is structurally rigid, displaying the same topology in both the bound and unbound states when it recognizes both sequence and structure of the RNA target.

α/β protein domains

Single-stranded RNA-binding motifs containing both α-helical and β-strand components make up the majority of RNA-binding domains. A short list of α/β protein motifs includes the RRM (RNA-recognition motif), KH domain type I and II (K-homology), dsRBD (double-stranded RNA-binding domain), the Piwi/Argonaute/Zwille domain (PAZ), RGG box (Arg-Gly-Gly), OB-fold (oligonucleotide/oligosaccharide binding) and DEAD-box (Burd and Dreyfuss, 1994; Draper, 1995; Theobald, Mitton-Fry *et al.*, 2003; Chen and Varani, 2005; Lunde, Moore *et al.*, 2007; Glisovic, Bachorik *et al.*, 2008). By far the most abundant α/β motif is the RRM (Burd and Dreyfuss, 1994; Chen and Varani, 2005), also known as the RBD or RNP motif, which is composed of two short conserved 6 and 8 amino acid stretches (RNP consensus sequences) named RNP1 and RNP2 separated by 30 residues in a 80-100 amino acid domain (Draper, 1995). Over 10,000 RRMs have been identified which participate in virtually all eukaryotic post-transcriptional processes (Lunde, Moore *et al.*, 2007). The RRM follows a βαββαβ topology where 2 α-helices pack perpendicular to and against a 4-stranded anti-parallel β-sheet (Burd and Dreyfuss, 1994; Lunde, Moore *et al.*, 2007). Residues from RNP1 and RNP2, which are found paired in the two central β-strands, β_1 and β_3 (Burd and Dreyfuss, 1994), are typically in direct contact with RNA along the face of the β-sheet. For example, 2 of the 3 RRM domains from HuD protein crystallized with an 11 nt AU-rich element (ARE) revealed that solvent exposed Phe and Tyr side chains found in RNP1 and RNP2, respectively, make stacking interactions with specific nucleotide bases while Lys (or Arg in related proteins) interacts with the phosphate backbone (Wang and Tanaka

Hall, 2001). RRM-containing proteins confer nucleotide specificity through spacing of their conserved motifs via single-stranded flexible linkers, such is the case with both HuD and the U1A spliceosomal protein (Nagai, 1996). Other, linker-dependent interactions can also be made with RNA, however, these typically serve to stabilize the binding of both RRMs.

The KH (heterogeneous nuclear ribonucleoprotein K) domain is a ~70 amino acid ssRNA/DNA-binding motif composed of 3 α-helices packed against a 3-stranded β-sheet (Valverde, Edwards *et al.*, 2008). Two KH variations exist, type I and type II, which are distinguished by their topology; the type I fold is normally found in eukaryotes and follows the βααββββα structure while type II folds are eubacterial-specific with a αββααβ arrangement (Valverde, Edwards *et al.*, 2008). These folds differ in overall structure especially with regards to sequence and size of variable loops between secondary structures. Common features of both folds include a cleft formed at the Gly-X-X-Gly motif found between α_1, α_2, and β_2 and the variable loop (Lunde, Moore *et al.*, 2007; Valverde, Edwards *et al.*, 2008) that binds single-stranded RNA or DNA in an extended conformation across its length. The structure of two type II KH folds from the prokaryotic transcriptional anti-terminator NusA bound to a short BoxC hairpin RNA demonstrates that protein binding disrupts the stem-loop, enabling RNA contact across both KH domains of the protein (Beuth, Pennell *et al.*, 2005). However, the structure and orientation of both KH domains is quite stable as protein conformation is unchanged by RNA interaction (Valverde, Edwards *et al.*, 2008). The nucleic acid is positioned on an α/β platform in a cleft formed by the Gly-X-X-Gly and the β_2 strand (Lunde, Moore *et al.*, 2007; Valverde, Edwards *et al.*, 2008); multiple types of interactions stabilize the

NusA-RNA complex, including stacking interactions of hydrophobic residues with adenine bases, structure specific binding of charged residues to the phosphate backbone and numerous hydrogen bonds to RNA bases (Beuth, Pennell *et al.*, 2005). RNA specificity and recognition are due mainly to stacking of Iso and Leu with two of the adenines and to the interaction of an unspecified protein component with the 2'-hydroxyl of an adenine, which would be lost in the presence of DNA.

Zinc finger domain

Zinc finger domains (ZnF) are highly abundant DNA-binding motifs in eukaryotes that have recently been shown to interact tightly and specifically with RNA (Chen and Varani, 2005). These domains are typically ~30 amino acids in length and composed of 2 short β-strands and a single α-helix with positionally conserved Cys and His residues that coordinate a single zinc ion to stabilize the protein fold (Lu, Searles *et al.*, 2003). ZnFs are classified into three types based on the number and arrangement of residues used to coordinate zinc: Cys-Cys-His-His (CCHH), Cys-Cys-Cys-His (CCCH), and Cys-Cys-His-Cys (CCHC). Amino acid side chains of the CCHH domain appear to specifically interact with nucleotide bases while the CCCH and CCHC domains recognize RNA structure through an induced fit mechanism that relies on hydrogen bonding of base Watson-Crick edges to the protein backbone (Hall, 2005; Lunde, Moore *et al.*, 2007). The ZnF motif was first recognized in the eukaryotic transcription factor TFIIIA, and studies within the past 6 years have revealed that while the modes of RNA and DNA binding are different, the protein shares at least 1 of 9 zinc fingers for binding both DNA and RNA (Clemens, Wolf *et al.*, 1993; Searles, Lu *et al.*, 2000; Lunde, Moore

et al., 2007). dTFIIIA contains 3 ZnFs of the CCHH-type which bind 5S rRNA by two distinct mechanisms; the first employed by ZnF 5 involves protein side chains from α-helix-based residues that interact with the backbone phosphates in the major groove of an RNA stem (Lu, Searles *et al.*, 2003; Hall, 2005). The second mechanism, used by ZnF 4 and 6, requires binding of main chain and side chain atoms found in residues of the N-terminal portion of the α-helix to the base and backbone atoms of an unpaired guanine (in ZnF 4) or to unpaired cytosine and adenine (in ZnF 6), both found in the stem of hairpin loops (Lu, Searles *et al.*, 2003; Hall, 2005).

Multimeric RNA-binding motifs

The members of the multimeric RNA-binding motif family encode a repeated structural domain involved in RNA binding. While a single domain in these proteins is able to bind RNA (usually at low affinity), the increased number and proper arrangement of motifs typically endows high affinity and increased specificity to the protein (Chen and Varani, 2005). Protein multimerization can be accomplished in two ways: repeated subunits can be translated in one continuous peptide which then folds as separate subunits to associate as a multimeric protein, as in the case of human Pumilio (Gupta, Nair *et al.*, 2008), or a single motif subunit can be individually expressed that assembles into a homo or heteromeric structure, represented by the Sm-like bacterial Hfq (Schumacher, Pearson *et al.*, 2002) and eukaryotic Lsm proteins (Achsel, Brahms *et al.*, 1999), respectively (Wilusz and Wilusz, 2005). A well characterized example of the latter is the toroid-shaped *B. subtilis* homo-undecameric TRAP protein, discussed above, that binds to repeated (G/U)AG triplets separated by 2 nt spacers found in the 5'-untranslated leaders

of regulated messages, such as *trpEDCFBA*. Each TRAP subunit is composed of a 7 stranded antiparallel β-sheet made up of β-strands from two adjacent polypeptides (Antson, Otridge *et al.*, 1995). TRAP is activated by binding L-tryptophan in a deep cleft between adjacent subunits (Antson, Otridge *et al.*, 1995) which likely causes a structural change to allow for RNA interaction; however, this mechanism is not fully understood. Additionally, L-tryptophan has been shown to only activate RNA-binding in the subunits to which it is bound (Li and Gollnick, 2002). The cognate RNA binds and conforms to the outside of the TRAP protein ring where the RNA triplet repeats come in contact with edges of the β-sheet instead of at their surface (Antson, Dodson *et al.*, 1999). Crystal structures of TRAP-RNA complexes reveal that the RNA-binding surface is formed by residues from two adjacent subunits; the second and third triplet nucleotides appear to be more important for protein interaction than the first G or U nucleotide. Conserved Lys, Arg and Glu make direct or water-mediated hyrdrogen bonds to the bases of both the second and third triplet positions while a single Phe stacks with the third G nucleotide. TRAP specificity for RNA is at least partially explained by a direct hydrogen bond between Phe and a ribose 2'-hydroxyl (Antson, Dodson *et al.*, 1999).

The CsrA protein motif

The RNA-binding capacity of the CsrA protein was originally described in *E. coli* K12 where it bound and regulated expression from the *glgCAP* 5'-untranslated leader (Liu and Romeo, 1997). Since then, CsrA has been shown to regulate a variety of different genes and pathways in other eubacterial families including the Firmicutes (Yakhnin, Pandit *et al.*, 2007). As described above, CsrA directly binds to conserved

target sequences found in regulated messages and in 2 noncoding small RNAs as determined by the electrophoretic mobility shift assay (EMSA) (Liu, Gui *et al.*, 1997; Liu and Romeo, 1997; Baker, Morozov *et al.*, 2002; Dubey, Baker *et al.*, 2003; Weilbacher, Suzuki *et al.*, 2003; Wang, Dubey *et al.*, 2005; Jonas, Tomenius *et al.*, 2006; Baker, Eory *et al.*, 2007; Yakhnin, Pandit *et al.*, 2007). A SELEX search for RNA targets revealed that CsrA bound with highest affinity to the conserved sequence RUACARGGAUGU (Dubey, Baker *et al.*, 2005). The primary sequence of CsrA, however, shows little similarity to most known RNA-binding motifs or domains and thus gives no indication as to the region(s) required for protein-RNA interaction. A breakthrough occurred in 2005 when two separate structural studies were published, one in *E. coli* (Gutierrez, Li *et al.*, 2005) and the other in *Pseudomonas putida* (Rife, Schwarzenbacher *et al.*, 2005), showing that CsrA was a completely novel class of protein fold. A single bilaterally symmetrical CsrA protein was composed of 2 opposing antiparallel β-sheets formed by the interdigitation of 2 separate polypeptide strands, producing a barrel-like structure. Each protein monomer has a β_1-β_2-β_3-β_4-β_5-α topology and the N-terminus of the protein, which is highly conserved in all species (Mercante, Suzuki *et al.*, 2006), is buried in a shallow hydrophobic pocket close to the lone, solvent exposed C-terminal α-helix. Hydrogen bond interactions hold the 5-stranded β-sheets together which are composed of the central 3 β-strands of one monomer (β_2-β_3-β_4) flanked by the terminal β-strands of the other monomer (β_1'-β_5') giving a stable hybrid configuration (β_1'-β_2-β_3-β_4-β_5'). These hybrid β-sheets enclose a hydrophobic core lined by nonpolar residue side chains from both monomers. Subsequent structure-function studies (Heeb, Blumer *et al.*, 2002) in *P. aeurginosa* and *P. fluorescens* with EMSAs suggested that a conserved Arginine

residue at codon 44 was essential for RNA-binding; however an RNA-binding surface has not been described.

REFERENCES

Abaza, I. and F. Gebauer (2008). "Trading translation with RNA-binding proteins." RNA **14**(3): 404-9.f

Achsel, T., H. Brahms, B. Kastner, A. Bachi, M. Wilm and R. Luhrmann (1999). "A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro." EMBO J **18**(20): 5789-802.

Ahmer, B. M. (2004). "Cell-to-cell signalling in *Escherichia coli* and *Salmonella enterica*." Mol Microbiol **52**(4): 933-45.

Alberts, B. (2008). "Molecular biology of the cell." 5th. from http://www.loc.gov/catdir/toc/ecip0710/2007005475.html.

Alen, C. and A. L. Sonenshein (1999). "*Bacillus subtilis* aconitase is an RNA-binding protein." Proc Natl Acad Sci U S A **96**(18): 10412-7.

Altier, C. (2005). "Genetic and environmental control of *Salmonella* invasion." J Microbiol **43 Spec No**: 85-92.

Altier, C., M. Suyemoto and S. D. Lawhon (2000). "Regulation of *Salmonella enterica* serovar Typhimurium invasion genes by *csrA*." Infect Immun **68**(12): 6790-7.

Altier, C., M. Suyemoto, A. I. Ruiz, K. D. Burnham and R. Maurer (2000). "Characterization of two novel regulatory genes affecting *Salmonella* invasion gene expression." Mol Microbiol **35**(3): 635-46.

Ang, S., Y. T. Horng, J. C. Shu, P. C. Soo, J. H. Liu, W. C. Yi, H. C. Lai, K. T. Luh, S. W. Ho and S. Swift (2001). "The role of RsmA in the regulation of swarming motility in *Serratia marcescens*." J Biomed Sci **8**(2): 160-9.

Antson, A. A., E. J. Dodson, G. Dodson, R. B. Greaves, X. Chen and P. Gollnick (1999). "Structure of the *trp* RNA-binding attenuation protein, TRAP, bound to RNA." Nature **401**(6750): 235-42.

Antson, A. A., J. Otridge, A. M. Brzozowski, E. J. Dodson, G. G. Dodson, K. S. Wilson, T. M. Smith, M. Yang, T. Kurecki and P. Gollnick (1995). "The structure of *trp* RNA-binding attenuation protein." Nature **374**(6524): 693-700.

Aviv, T., A. N. Amborski, X. S. Zhao, J. J. Kwan, P. E. Johnson, F. Sicheri and L. W. Donaldson (2006). "The NMR and X-ray structures of the *Saccharomyces cerevisiae* Vts1 SAM domain define a surface for the recognition of RNA hairpins." J Mol Biol **356**(2): 274-9.

Babitzke, P., C. S. Baker and T. Romeo (2009). "Regulation of Translation Initiation by RNA Binding Proteins." Annu Rev Microbiol.

Babitzke, P. and T. Romeo (2007). "CsrB sRNA family: sequestration of RNA-binding regulatory proteins." Curr Opin Microbiol **10**(2): 156-63.

Bachman, M. A. and M. S. Swanson (2004). "Genetic evidence that *Legionella pneumophila* RpoS modulates expression of the transmission phenotype in both the exponential phase and the stationary phase." Infect Immun **72**(5): 2468-76.

Bachman, M. A. and M. S. Swanson (2004). "The LetE protein enhances expression of multiple LetA/LetS-dependent transmission traits by *Legionella pneumophila*." Infect Immun **72**(6): 3284-93.

Bajaj, V., C. Hwang and C. A. Lee (1995). "*hilA* is a novel *ompR/toxR* family member that activates the expression of *Salmonella typhimurium* invasion genes." Mol Microbiol **18**(4): 715-27.

Bajaj, V., R. L. Lucas, C. Hwang and C. A. Lee (1996). "Co-ordinate regulation of *Salmonella typhimurium* invasion genes by environmental and regulatory factors is mediated by control of *hilA* expression." Mol Microbiol **22**(4): 703-14.

Baker, C. S., L. A. Eory, H. Yakhnin, J. Mercante, T. Romeo and P. Babitzke (2007). "CsrA inhibits translation initiation of *Escherichia coli hfq* by binding to a single site overlapping the Shine-Dalgarno sequence." J Bacteriol **189**(15): 5472-81.

Baker, C. S., I. Morozov, K. Suzuki, T. Romeo and P. Babitzke (2002). "CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*." Mol Microbiol **44**(6): 1599-610.

Barnard, F. M., M. F. Loughlin, H. P. Fainberg, M. P. Messenger, D. W. Ussery, P. Williams and P. J. Jenks (2004). "Global regulation of virulence and the stress response by CsrA in the highly adapted human gastric pathogen *Helicobacter pylori*." Mol Microbiol **51**(1): 15-32.

Battiste, J. L., H. Mao, N. S. Rao, R. Tan, D. R. Muhandiram, L. E. Kay, A. D. Frankel and J. R. Williamson (1996). "Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex." Science **273**(5281): 1547-51.

Bayer, T. S., L. N. Booth, S. M. Knudsen and A. D. Ellington (2005). "Arginine-rich motifs present multiple interfaces for specific binding by RNA." RNA **11**(12): 1848-57.

Bernardi, A. and P. F. Spahr (1972). "Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17." Proc Natl Acad Sci U S A **69**(10): 3033-7.

Beuth, B., S. Pennell, K. B. Arnvig, S. R. Martin and I. A. Taylor (2005). "Structure of a *Mycobacterium tuberculosis* NusA-RNA complex." EMBO J **24**(20): 3576-87.

Blencowe, B. J., J. A. Bowman, S. McCracken and E. Rosonina (1999). "SR-related proteins and the processing of messenger RNA precursors." Biochem Cell Biol **77**(4): 277-91.

Blumer, C. and D. Haas (2000). "Mechanism, regulation, and ecological role of bacterial cyanide biosynthesis." Arch Microbiol **173**(3): 170-7.

Blumer, C., S. Heeb, G. Pessi and D. Haas (1999). "Global GacA-steered control of cyanide and exoprotease production in *Pseudomonas fluorescens* involves specific ribosome binding sites." Proc Natl Acad Sci U S A **96**(24): 14073-8.

Bolognani, F. and N. I. Perrone-Bizzozero (2008). "RNA-protein interactions and control of mRNA stability in neurons." J Neurosci Res **86**(3): 481-9.

Bonafonte, M. A., C. Solano, B. Sesma, M. Alvarez, L. Montuenga, D. Garcia-Ros and C. Gamazo (2000). "The relationship between glycogen synthesis, biofilm formation and virulence in *Salmonella enteritidis*." FEMS Microbiol Lett **191**(1): 31-6.

Boni, I. V., D. M. Isaeva, M. L. Musychenko and N. V. Tzareva (1991). "Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1." Nucleic Acids Res **19**(1): 155-62.

Brennan, R. G. and T. M. Link (2007). "Hfq structure, function and ligand binding." Curr Opin Microbiol **10**(2): 125-33.

Brown, D., J. Brown, C. Kang, L. Gold and P. Allen (1997). "Single-stranded RNA recognition by the bacteriophage T4 translational repressor, RegA." J Biol Chem **272**(23): 14969-74.

Burd, C. G. and G. Dreyfuss (1994). "Conserved structures and diversity of functions of RNA-binding proteins." Science **265**(5172): 615-21.

Burrowes, E., A. Abbas, A. O'Neill, C. Adams and F. O'Gara (2005). "Characterisation of the regulatory RNA RsmB from *Pseudomonas aeruginosa* PAO1." Res Microbiol **156**(1): 7-16.

Burrowes, E., C. Baysse, C. Adams and F. O'Gara (2006). "Influence of the regulatory protein RsmA on cellular functions in *Pseudomonas aeruginosa* PAO1, as revealed by transcriptome analysis." Microbiology **152**(Pt 2): 405-18.

Calnan, B. J., S. Biancalana, D. Hudson and A. D. Frankel (1991). "Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition." Genes Dev **5**(2): 201-10.

Calnan, B. J., B. Tidor, S. Biancalana, D. Hudson and A. D. Frankel (1991). "Arginine-mediated RNA recognition: the arginine fork." Science **252**(5010): 1167-71.

Cesareni, G., M. Helmer-Citterich and L. Castagnoli (1991). "Control of ColE1 plasmid replication by antisense RNA." Trends Genet **7**(7): 230-5.

Chang, Y. F., J. S. Imam and M. F. Wilkinson (2007). "The nonsense-mediated decay RNA surveillance pathway." Annu Rev Biochem **76**: 51-74.

Chao, J. A., Y. Patskovsky, S. C. Almo and R. H. Singer (2008). "Structural basis for the coevolution of a viral RNA-protein complex." Nat Struct Mol Biol **15**(1): 103-5.

Chao, N. X., K. Wei, Q. Chen, Q. L. Meng, D. J. Tang, Y. Q. He, G. T. Lu, B. L. Jiang, X. X. Liang, J. X. Feng, B. Chen and J. L. Tang (2008). "The *rsmA*-like gene *rsmA*(Xcc) of *Xanthomonas campestris* pv. *campestris* is involved in the control of various cellular processes, including pathogenesis." Mol Plant Microbe Interact **21**(4): 411-23.

Chatterjee, A., Y. Cui and A. K. Chatterjee (2002). "Regulation of *Erwinia carotovora hrpL*(Ecc) (sigma-L(Ecc)), which encodes an extracytoplasmic function subfamily of sigma factor required for expression of the HRP regulon." Mol Plant Microbe Interact **15**(9): 971-80.

Chatterjee, A., Y. Cui and A. K. Chatterjee (2002). "RsmA and the quorum-sensing signal, N-[3-oxohexanoyl]-L-homoserine lactone, control the levels of *rsmB* RNA in *Erwinia carotovora* subsp. *carotovora* by affecting its stability." J Bacteriol **184**(15): 4089-95.

Chatterjee, A., Y. Cui, H. Hasegawa, N. Leigh, V. Dixit and A. K. Chatterjee (2005). "Comparative analysis of two classes of quorum-sensing signaling systems that control production of extracellular proteins and secondary metabolites in *Erwinia carotovora* subspecies." J Bacteriol **187**(23): 8026-38.

Chatterjee, A., Y. Cui, Y. Liu, C. K. Dumenyo and A. K. Chatterjee (1995). "Inactivation of *rsmA* leads to overproduction of extracellular pectinases, cellulases, and proteases in *Erwinia carotovora* subsp. *carotovora* in the absence of the

starvation/cell density-sensing signal, N-(3-oxohexanoyl)-L-homoserine lactone." Appl Environ Microbiol **61**(5): 1959-67.

Chen, Y. and G. Varani (2005). "Protein families and RNA recognition." FEBS J **272**(9): 2088-97.

Chiaruttini, C., M. Milet and M. Springer (1996). "A long-range RNA-RNA interaction forms a pseudoknot required for translational control of the IF3-L35-L20 ribosomal protein operon in *Escherichia coli*." EMBO J **15**(16): 4402-13.

Christiansen, J., A. Kolte, T. Hansen and F. Nielsen (2009). "IGF2 mRNA-binding protein 2 - biological function and putative role in type 2 diabetes." J Mol Endocrinol.

Clemens, K. R., V. Wolf, S. J. McBryant, P. Zhang, X. Liao, P. E. Wright and J. M. Gottesfeld (1993). "Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein." Science **260**(5107): 530-3.

Collier, D. N., P. W. Hager and P. V. Phibbs, Jr. (1996). "Catabolite repression control in the Pseudomonads." Res Microbiol **147**(6-7): 551-61.

Cui, Y., A. Chatterjee and A. K. Chatterjee (2001). "Effects of the two-component system comprising GacA and GacS of *Erwinia carotovora* subsp. *carotovora* on the production of global regulatory *rsmB* RNA, extracellular enzymes, and harpinEcc." Mol Plant Microbe Interact **14**(4): 516-26.

Cui, Y., A. Chatterjee, H. Hasegawa and A. K. Chatterjee (2006). "*Erwinia carotovora* subspecies produce duplicate variants of ExpR, LuxR homologs that activate *rsmA* transcription but differ in their interactions with N-acylhomoserine lactone signals." J Bacteriol **188**(13): 4715-26.

Cui, Y., A. Chatterjee, H. Hasegawa, V. Dixit, N. Leigh and A. K. Chatterjee (2005). "ExpR, a LuxR homolog of *Erwinia carotovora* subsp. *carotovora*, activates transcription of *rsmA*, which specifies a global regulatory RNA-binding protein." J Bacteriol **187**(14): 4792-803.

Cui, Y., A. Chatterjee, Y. Liu, C. K. Dumenyo and A. K. Chatterjee (1995). "Identification of a global repressor gene, *rsmA*, of *Erwinia carotovora* subsp. *carotovora* that controls extracellular enzymes, N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting *Erwinia* spp." J Bacteriol **177**(17): 5108-15.

Cui, Y., A. Mukherjee, C. K. Dumenyo, Y. Liu and A. K. Chatterjee (1999). "*rsmC* of the soft-rotting bacterium *Erwinia carotovora* subsp. *carotovora* negatively controls extracellular enzyme and harpin(Ecc) production and virulence by modulating levels of regulatory RNA (*rsmB*) and RNA-binding protein (RsmA)." J Bacteriol **181**(19): 6042-52.

Cunningham, L., M. J. Gruer and J. R. Guest (1997). "Transcriptional regulation of the aconitase genes (*acnA* and *acnB*) of *Escherichia coli*." Microbiology **143 (Pt 12)**: 3795-805.

Dienst, D., U. Duhring, H. J. Mollenkopf, J. Vogel, J. Golecki, W. R. Hess and A. Wilde (2008). "The cyanobacterial homologue of the RNA chaperone Hfq is essential for motility of *Synechocystis* sp. PCC 6803." Microbiology **154**(Pt 10): 3134-43.

Draper, D. E. (1995). "Protein-RNA recognition." Annu Rev Biochem **64**: 593-620.

Dreyfuss, G., V. N. Kim and N. Kataoka (2002). "Messenger-RNA-binding proteins and the messages they carry." Nat Rev Mol Cell Biol **3**(3): 195-205.

Du, H. and P. Babitzke (1998). "*trp* RNA-binding attenuation protein-mediated long distance RNA refolding regulates translation of *trpE* in *Bacillus subtilis*." J Biol Chem **273**(32): 20494-503.

Du, H., R. Tarpey and P. Babitzke (1997). "The *trp* RNA-binding attenuation protein regulates TrpG synthesis by binding to the *trpG* ribosome binding site of *Bacillus subtilis*." J Bacteriol **179**(8): 2582-6.

Dubey, A. K., C. S. Baker, T. Romeo and P. Babitzke (2005). "RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction." RNA **11**(10): 1579-87.

Dubey, A. K., C. S. Baker, K. Suzuki, A. D. Jones, P. Pandit, T. Romeo and P. Babitzke (2003). "CsrA regulates translation of the *Escherichia coli* carbon starvation gene, *cstA*, by blocking ribosome access to the *cstA* transcript." J Bacteriol **185**(15): 4450-60.

Ellermeier, C. D., J. R. Ellermeier and J. M. Slauch (2005). "HilD, HilC and RtsA constitute a feed forward loop that controls expression of the SPI1 type three secretion system regulator *hilA* in *Salmonella enterica* serovar Typhimurium." Mol Microbiol **57**(3): 691-705.

Fang, L., W. Jiang, W. Bae and M. Inouye (1997). "Promoter-independent cold-shock induction of *cspA* and its derepression at 37 degrees C by mRNA stabilization." Mol Microbiol **23**(2): 355-64.

Fantappie, L., M. M. Metruccio, K. L. Seib, F. Oriente, E. Cartocci, F. Ferlicca, M. M. Giuliani, V. Scarlato and I. Delany (2009). "The RNA chaperone Hfq is involved

in stress response and virulence in *Neisseria meningitidis* and is a pleiotropic regulator of protein expression." Infect Immun **77**(5): 1842-53.

Fettes, P. S., V. Forsbach-Birk, D. Lynch and R. Marre (2001). "Overexpresssion of a *Legionella pneumophila* homologue of the *E. coli* regulator *csrA* affects cell size, flagellation, and pigmentation." Int J Med Microbiol **291**(5): 353-60.

Fields, J. A. and S. A. Thompson (2008). "*Campylobacter jejuni* CsrA mediates oxidative stress responses, biofilm formation, and host cell invasion." J Bacteriol **190**(9): 3411-6.

Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer and A. Bateman (2008). "The Pfam protein families database." Nucleic Acids Res **36**(Database issue): D281-8.

Forsbach-Birk, V., T. McNealy, C. Shi, D. Lynch and R. Marre (2004). "Reduced expression of the global regulator protein CsrA in *Legionella pneumophila* affects virulence-associated regulators and growth in *Acanthamoeba castellanii*." Int J Med Microbiol **294**(1): 15-25.

Fortune, D. R., M. Suyemoto and C. Altier (2006). "Identification of CsrC and characterization of its role in epithelial cell invasion in *Salmonella enterica* serovar Typhimurium." Infect Immun **74**(1): 331-9.

Fouts, D. E., H. L. True and D. W. Celander (1997). "Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor." Nucleic Acids Res **25**(22): 4464-73.

Frolova, L. Y., T. I. Merkulova and L. L. Kisselev (2000). "Translation termination in eukaryotes: polypeptide release factor eRF1 is composed of functionally and structurally distinct domains." RNA **6**(3): 381-90.

Ganoza, M. C., M. C. Kiel and H. Aoki (2002). "Evolutionary conservation of reactions in translation." Microbiol Mol Biol Rev **66**(3): 460-85, table of contents.

Glisovic, T., J. L. Bachorik, J. Yong and G. Dreyfuss (2008). "RNA-binding proteins and post-transcriptional gene regulation." FEBS Lett **582**(14): 1977-86.

Gollnick, P., P. Babitzke, A. Antson and C. Yanofsky (2005). "Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis*." Annu Rev Genet **39**: 47-68.

Gong, C., Y. K. Kim, C. F. Woeller, Y. Tang and L. E. Maquat (2009). "SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs." Genes Dev **23**(1): 54-66.

Goodier, R. I. and B. M. Ahmer (2001). "SirA orthologs affect both motility and virulence." J Bacteriol **183**(7): 2249-58.

Goodman, A. L., B. Kulasekara, A. Rietsch, D. Boyd, R. S. Smith and S. Lory (2004). "A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in *Pseudomonas aeruginosa*." Dev Cell **7**(5): 745-54.

Gorke, B. and J. Stulke (2008). "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients." Nat Rev Microbiol **6**(8): 613-24.

Grall, N., J. Livny, M. Waldor, M. Barel, A. Charbit and K. L. Meibom (2009). "Pivotal role of the *Francisella tularensis* heat shock sigma factor RpoH." Microbiology.

Graumann, P., T. M. Wendrich, M. H. Weber, K. Schroder and M. A. Marahiel (1997). "A family of cold shock proteins in *Bacillus subtilis* is essential for cellular

growth and for efficient protein synthesis at optimal and low temperatures." Mol Microbiol **25**(4): 741-56.

Graumann, P. L. and M. A. Marahiel (1998). "A superfamily of proteins that contain the cold-shock domain." Trends Biochem Sci **23**(8): 286-90.

Gualerzi, C. O., A. M. Giuliodori and C. L. Pon (2003). "Transcriptional and post-transcriptional control of cold-shock genes." J Mol Biol **331**(3): 527-39.

Gudapaty, S., K. Suzuki, X. Wang, P. Babitzke and T. Romeo (2001). "Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*." J Bacteriol **183**(20): 6017-27.

Guo, R., H. Xue and L. Huang (2003). "Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo." Mol Microbiol **50**(5): 1605-15.

Gupta, Y. K., D. T. Nair, R. P. Wharton and A. K. Aggarwal (2008). "Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity." Structure **16**(4): 549-57.

Gutierrez, P., Y. Li, M. J. Osborne, E. Pomerantseva, Q. Liu and K. Gehring (2005). "Solution structure of the carbon storage regulator protein CsrA from *Escherichia coli*." J Bacteriol **187**(10): 3496-501.

Haas, D., C. Keel and C. Reimmann (2002). "Signal transduction in plant-beneficial rhizobacteria with biocontrol properties." Antonie Van Leeuwenhoek **81**(1-4): 385-95.

Hall, T. M. (2005). "Multiple modes of RNA recognition by zinc finger proteins." Curr Opin Struct Biol **15**(3): 367-73.

Hammer, B. K., E. S. Tateda and M. S. Swanson (2002). "A two-component regulator induces the transmission phenotype of stationary-phase *Legionella pneumophila*." Mol Microbiol **44**(1): 107-18.

Heeb, S., C. Blumer and D. Haas (2002). "Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0." J Bacteriol **184**(4): 1046-56.

Heroven, A. K., K. Bohme, M. Rohde and P. Dersch (2008). "A Csr-type regulatory system, including small non-coding RNAs, regulates the global virulence regulator RovA of *Yersinia pseudotuberculosis* through RovM." Mol Microbiol **68**(5): 1179-95.

Hester, K. L., J. Lehman, F. Najar, L. Song, B. A. Roe, C. H. MacGregor, P. W. Hager, P. V. Phibbs, Jr. and J. R. Sokatch (2000). "Crc is involved in catabolite repression control of the *bkd* operons of *Pseudomonas putida* and *Pseudomonas aeruginosa*." J Bacteriol **182**(4): 1144-9.

Hester, K. L., K. T. Madhusudhan and J. R. Sokatch (2000). "Catabolite repression control by *crc* in 2xYT medium is mediated by posttranscriptional regulation of *bkdR* expression in *Pseudomonas putida*." J Bacteriol **182**(4): 1150-3.

Heurlier, K., F. Williams, S. Heeb, C. Dormond, G. Pessi, D. Singer, M. Camara, P. Williams and D. Haas (2004). "Positive control of swarming, rhamnolipid synthesis, and lipase production by the posttranscriptional RsmA/RsmZ system in *Pseudomonas aeruginosa* PAO1." J Bacteriol **186**(10): 2936-45.

Hoffmann, P. R. and M. J. Berry (2005). "Selenoprotein synthesis: a unique translational mechanism used by a diverse family of proteins." Thyroid **15**(8): 769-75.

Hopcroft, N. H., A. Manfredo, A. L. Wendt, A. M. Brzozowski, P. Gollnick and A. A. Antson (2004). "The interaction of RNA with TRAP: the role of triplet repeats and separating spacer nucleotides." J Mol Biol **338**(1): 43-53.

Hovel-Miner, G., S. Pampou, S. P. Faucher, M. Clarke, I. Morozova, P. Morozov, J. J. Russo, H. A. Shuman and S. Kalachikov (2009). "SigmaS controls multiple pathways associated with intracellular multiplication of *Legionella pneumophila*." J Bacteriol **191**(8): 2461-73.

Huang, X., X. Zhang and Y. Xu (2008). "PltR expression modulated by the global regulators GacA, RsmA, LasI and RhlI in *Pseudomonas* sp. M18." Res Microbiol **159**(2): 128-36.

Hyytiainen, H., S. Sjoblom, T. Palomaki, A. Tuikkala and E. Tapio Palva (2003). "The PmrA-PmrB two-component system responding to acidic pH and iron controls virulence in the plant pathogen *Erwinia carotovora* ssp. *carotovora*." Mol Microbiol **50**(3): 795-807.

Iben, J. R. and D. E. Draper (2008). "Specific interactions of the L10(L12)4 ribosomal protein complex with mRNA, rRNA, and L11." Biochemistry **47**(9): 2721-31.

Jiang, W., Y. Hou and M. Inouye (1997). "CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone." J Biol Chem **272**(1): 196-202.

Johnston, C., D. A. Pegues, C. J. Hueck, A. Lee and S. I. Miller (1996). "Transcriptional activation of *Salmonella typhimurium* invasion genes by a member of the phosphorylated response-regulator superfamily." Mol Microbiol **22**(4): 715-27.

Jonas, K., A. N. Edwards, R. Simm, T. Romeo, U. Romling and O. Melefors (2008). "The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins." Mol Microbiol **70**(1): 236-57.

Jonas, K., H. Tomenius, U. Romling, D. Georgellis and O. Melefors (2006). "Identification of YhdA as a regulator of the *Escherichia coli* carbon storage regulation system." FEMS Microbiol Lett **264**(2): 232-7.

Jones, B. D. (2005). "*Salmonella* invasion gene regulation: a story of environmental awareness." J Microbiol **43 Spec No**: 110-7.

Jones, M. K., E. B. Warner and J. D. Oliver (2008). "*csrA* inhibits the formation of biofilms by *Vibrio vulnificus*." Appl Environ Microbiol **74**(22): 7064-6.

Kaberdin, V. R. and U. Blasi (2006). "Translation initiation and the fate of bacterial mRNAs." FEMS Microbiol Rev **30**(6): 967-79.

Katsanou, V., S. Milatos, A. Yiakouvaki, N. Sgantzis, A. Kotsoni, M. Alexiou, V. Harokopos, V. Aidinis, M. Hemberger and D. L. Kontoyiannis (2009). "The RNA-binding protein Elavl1/HuR is essential for placental branching morphogenesis and embryonic development." Mol Cell Biol **29**(10): 2762-76.

Kay, E., C. Dubuis and D. Haas (2005). "Three small RNAs jointly ensure secondary metabolism and biocontrol in *Pseudomonas fluorescens* CHA0." Proc Natl Acad Sci U S A **102**(47): 17136-41.

Kay, E., B. Humair, V. Denervaud, K. Riedel, S. Spahr, L. Eberl, C. Valverde and D. Haas (2006). "Two GacA-dependent small RNAs modulate the quorum-sensing response in *Pseudomonas aeruginosa*." J Bacteriol **188**(16): 6026-33.

Keener, J. and M. Nomura (1996). Regulation of ribosome synthesis. Escherichia coli and Salmonella: Cellular and Molecular Biology. F. C. Neidhardt, R. I. Curtis, J. L. Ingrahamet al. Washington, D.C., American Society for Microbiology. **2:** 1417-1431.

Keiler, K. C. (2008). "Biology of trans-translation." Annu Rev Microbiol **62**: 133-51.

Kerrinnes, T., Z. B. Zelas, W. Streckel, F. Faber, E. Tietze, H. Tschape and S. Yaron (2008). "CsrA and CsrB are required for the post-transcriptional control of the virulence-associated effector protein AvrA of *Salmonella enterica*." Int J Med Microbiol.

Kim, V. N. and G. Dreyfuss (2001). "Nuclear mRNA binding proteins couple pre-mRNA splicing and post-splicing events." Mol Cells **12**(1): 1-10.

Kim, Y. K., L. Furic, L. Desgroseillers and L. E. Maquat (2005). "Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay." Cell **120**(2): 195-208.

Kim, Y. K., L. Furic, M. Parisien, F. Major, L. DesGroseillers and L. E. Maquat (2007). "Staufen1 regulates diverse classes of mammalian transcripts." EMBO J **26**(11): 2670-81.

Koiv, V. and A. Mae (2001). "Quorum sensing controls the synthesis of virulence factors by modulating *rsmA* gene expression in *Erwinia carotovora* subsp. *carotovora*." Mol Genet Genomics **265**(2): 287-92.

Kozak, M. (2005). "Regulation of translation via mRNA structure in prokaryotes and eukaryotes." Gene **361**: 13-37.

Krin, E., S. Derzelle, K. Bedard, M. Adib-Conquy, E. Turlin, P. Lenormand, M. F. Hullo, I. Bonne, N. Chakroun, C. Lacroix and A. Danchin (2008). "Regulatory role of UvrY in adaptation of Photorhabdus luminescens growth inside the insect." Environ Microbiol **10**(5): 1118-34.

Kulkarni, P. R., X. Cui, J. W. Williams, A. M. Stevens and R. V. Kulkarni (2006). "Prediction of CsrA-regulating small RNAs in bacteria and their experimental verification in *Vibrio fischeri*." Nucleic Acids Res **34**(11): 3361-9.

Laursen, B. S., A. S. S. A. de, J. Hedegaard, J. M. Moreno, K. K. Mortensen and H. U. Sperling-Petersen (2002). "Structural requirements of the mRNA for intracistronic translation initiation of the enterobacterial *infB* gene." Genes Cells **7**(9): 901-10.

Laursen, B. S., H. P. Sorensen, K. K. Mortensen and H. U. Sperling-Petersen (2005). "Initiation of protein synthesis in bacteria." Microbiol Mol Biol Rev **69**(1): 101-23.

Lawhon, S. D., J. G. Frye, M. Suyemoto, S. Porwollik, M. McClelland and C. Altier (2003). "Global regulation by CsrA in *Salmonella typhimurium*." Mol Microbiol **48**(6): 1633-45.

Lee, J., A. Jayaraman and T. K. Wood (2007). "Indole is an inter-species biofilm signal mediated by SdiA." BMC Microbiol **7**: 42.

Lenz, D. H., M. B. Miller, J. Zhu, R. V. Kulkarni and B. L. Bassler (2005). "CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*." Mol Microbiol **58**(4): 1186-202.

Li, P. T. and P. Gollnick (2002). "Using hetero-11-mers composed of wild type and mutant subunits to study tryptophan binding to TRAP and its role in activating RNA binding." J Biol Chem **277**(38): 35567-73.

Liaw, S. J., H. C. Lai, S. W. Ho, K. T. Luh and W. B. Wang (2003). "Role of RsmA in the regulation of swarming motility and virulence factor expression in *Proteus mirabilis*." J Med Microbiol **52**(Pt 1): 19-28.

Liljas, A. and M. Garber (1995). "Ribosomal proteins and elongation factors." Curr Opin Struct Biol **5**(6): 721-7.

Lim, F., T. P. Downey and D. S. Peabody (2001). "Translational repression and specific RNA binding by the coat protein of the *Pseudomonas* phage PP7." J Biol Chem **276**(25): 22507-13.

Lim, F. and D. S. Peabody (2002). "RNA recognition site of PP7 coat protein." Nucleic Acids Res **30**(19): 4138-44.

Liu, M. Y., G. Gui, B. Wei, J. F. Preston, 3rd, L. Oakford, U. Yuksel, D. P. Giedroc and T. Romeo (1997). "The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*." J Biol Chem **272**(28): 17502-10.

Liu, M. Y. and T. Romeo (1997). "The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein." J Bacteriol **179**(14): 4639-42.

Liu, M. Y., H. Yang and T. Romeo (1995). "The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability." J Bacteriol **177**(10): 2663-72.

Liu, Y., Y. Cui, A. Mukherjee and A. K. Chatterjee (1998). "Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. *carotovora* that controls production of extracellular enzymes and secondary metabolites." Mol Microbiol **29**(1): 219-34.

Long, J. C. and J. F. Caceres (2009). "The SR protein family of splicing factors: master regulators of gene expression." Biochem J **417**(1): 15-27.

Lopez de Quinto, S. and E. Martinez-Salas (2000). "Interaction of the eIF4G initiation factor with the aphthovirus IRES is essential for internal translation initiation in vivo." RNA **6**(10): 1380-92.

Lu, D., M. A. Searles and A. Klug (2003). "Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition." Nature **426**(6962): 96-100.

Lucas, R. L. and C. A. Lee (2001). "Roles of *hilC* and *hilD* in regulation of *hilA* expression in *Salmonella enterica* serovar Typhimurium." J Bacteriol **183**(9): 2733-45.

Lukong, K. E., K. W. Chang, E. W. Khandjian and S. Richard (2008). "RNA-binding proteins in human genetic disease." Trends Genet **24**(8): 416-25.

Lunde, B. M., C. Moore and G. Varani (2007). "RNA-binding proteins: modular design for efficient function." Nat Rev Mol Cell Biol **8**(6): 479-90.

Mangus, D. A., M. C. Evans and A. Jacobson (2003). "Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression." Genome Biol **4**(7): 223.

Matlin, A. J., F. Clark and C. W. Smith (2005). "Understanding alternative splicing: towards a cellular code." Nat Rev Mol Cell Biol **6**(5): 386-98.

Matsuura, M., J. W. Noah and A. M. Lambowitz (2001). "Mechanism of maturase-promoted group II intron splicing." EMBO J **20**(24): 7259-70.

Mazumder, B., V. Seshadri and P. L. Fox (2003). "Translational control by the 3'-UTR: the ends specify the means." Trends Biochem Sci **28**(2): 91-8.

McNealy, T. L., V. Forsbach-Birk, C. Shi and R. Marre (2005). "The Hfq homolog in *Legionella pneumophila* demonstrates regulation by LetA and RpoS and interacts with the global regulator CsrA." J Bacteriol **187**(4): 1527-32.

McPheeters, D. S., G. D. Stormo and L. Gold (1988). "Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA." J Mol Biol **201**(3): 517-35.

Meibom, K. L., A. L. Forslund, K. Kuoppa, K. Alkhuder, I. Dubail, M. Dupuis, A. Forsberg and A. Charbit (2009). "Hfq, a novel pleiotropic regulator of virulence-associated genes in *Francisella tularensis*." Infect Immun **77**(5): 1866-80.

Mendez, R. and J. D. Richter (2001). "Translational control by CPEB: a means to the end." Nat Rev Mol Cell Biol **2**(7): 521-9.

Mercante, J., K. Suzuki, X. Cheng, P. Babitzke and T. Romeo (2006). "Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance." J Biol Chem **281**(42): 31832-42.

Merino, E., P. Babitzke and C. Yanofsky (1995). "*trp* RNA-binding attenuation protein (TRAP)-*trp* leader RNA interactions mediate translational as well as transcriptional regulation of the *Bacillus subtilis trp* operon." J Bacteriol **177**(22): 6362-70.

Molofsky, A. B. and M. S. Swanson (2003). "*Legionella pneumophila* CsrA is a pivotal repressor of transmission traits and activator of replication." Mol Microbiol **50**(2): 445-61.

Molofsky, A. B. and M. S. Swanson (2004). "Differentiate to thrive: lessons from the *Legionella* pneumophila life cycle." Mol Microbiol **53**(1): 29-40.

Mondragon, V., B. Franco, K. Jonas, K. Suzuki, T. Romeo, O. Melefors and D. Georgellis (2006). "pH-dependent activation of the BarA-UvrY two-component system in *Escherichia coli*." J Bacteriol **188**(23): 8303-6.

Moreno, R. and F. Rojo (2008). "The target for the *Pseudomonas putida* Crc global regulator in the benzoate degradation pathway is the BenR transcriptional regulator." J Bacteriol **190**(5): 1539-45.

Moreno, R., A. Ruiz-Manzano, L. Yuste and F. Rojo (2007). "The *Pseudomonas putida* Crc global regulator is an RNA binding protein that inhibits translation of the AlkS transcriptional regulator." Mol Microbiol **64**(3): 665-75.

Mosig, G. (1998). "Recombination and recombination-dependent DNA replication in bacteriophage T4." Annu Rev Genet **32**: 379-413.

Mukherjee, A., Y. Cui, W. Ma, Y. Liu, A. Ishihama, A. Eisenstark and A. K. Chatterjee (1998). "RpoS (sigma-S) controls expression of *rsmA*, a global regulator of secondary metabolites, harpin, and extracellular proteins in *Erwinia carotovora*." J Bacteriol **180**(14): 3629-34.

Mulcahy, H., J. O'Callaghan, E. P. O'Grady, C. Adams and F. O'Gara (2006). "The posttranscriptional regulator RsmA plays a role in the interaction between

Pseudomonas aeruginosa and human airway epithelial cells by positively regulating the type III secretion system." Infect Immun **74**(5): 3012-5.

Mulcahy, H., J. O'Callaghan, E. P. O'Grady, M. D. Macia, N. Borrell, C. Gomez, P. G. Casey, C. Hill, C. Adams, C. G. Gahan, A. Oliver and F. O'Gara (2008). "*Pseudomonas aeruginosa* RsmA plays an important role during murine infection by influencing colonization, virulence, persistence, and pulmonary inflammation." Infect Immun **76**(2): 632-8.

Nagai, K. (1996). "RNA-protein complexes." Curr Opin Struct Biol **6**(1): 53-61.

Nagasawa, S., S. Tokishita, H. Aiba and T. Mizuno (1992). "A novel sensor-regulator protein that belongs to the homologous family of signal-transduction proteins involved in adaptive responses in *Escherichia coli*." Mol Microbiol **6**(6): 799-807.

Nakano, M., A. Takahashi, Z. Su, N. Harada, K. Mawatari and Y. Nakaya (2008). "Hfq regulates the expression of the thermostable direct hemolysin gene in *Vibrio parahaemolyticus*." BMC Microbiol **8**: 155.

Nogueira, T. and M. Springer (2000). "Post-transcriptional control by global regulators of gene expression in bacteria." Curr Opin Microbiol **3**(2): 154-8.

Nossal, N. G. (1992). "Protein-protein interactions at a DNA replication fork: bacteriophage T4 as a model." FASEB J **6**(3): 871-8.

Oberstrass, F. C., A. Lee, R. Stefl, M. Janis, G. Chanfreau and F. H. Allain (2006). "Shape-specific recognition in the structure of the Vts1p SAM domain with RNA." Nat Struct Mol Biol **13**(2): 160-7.

Pannekoek, Y., R. Huis in 't Veld, C. T. Hopman, A. A. Langerak, D. Speijer and A. van der Ende (2009). "Molecular characterization and identification of proteins

regulated by Hfq in *Neisseria meningitidis*." FEMS Microbiol Lett **294**(2): 216-24.

Paraskeva, E., N. K. Gray, B. Schlager, K. Wehr and M. W. Hentze (1999). "Ribosomal pausing and scanning arrest as mechanisms of translational regulation from cap-distal iron-responsive elements." Mol Cell Biol **19**(1): 807-16.

Patte, J. (1996). Biosynthesis of Threonine and Lysine. Escherichia coli and Salmonella: Cellular and Molecular Biology. F. C. Neidhardt, R. I. Curtis, J. L. Ingrahamet al. Washington, D.C., American Society for Microbiology. **2:** 528-541.

Pavlov, A. R. and J. D. Karam (2000). "Nucleotide-sequence-specific and non-specific interactions of T4 DNA polymerase with its own mRNA." Nucleic Acids Res **28**(23): 4657-64.

Pernestig, A. K., O. Melefors and D. Georgellis (2001). "Identification of UvrY as the cognate response regulator for the BarA sensor kinase in *Escherichia coli*." J Biol Chem **276**(1): 225-31.

Pessi, G., F. Williams, Z. Hindle, K. Heurlier, M. T. Holden, M. Camara, D. Haas and P. Williams (2001). "The global posttranscriptional regulator RsmA modulates production of virulence determinants and N-acylhomoserine lactones in *Pseudomonas aeruginosa*." J Bacteriol **183**(22): 6676-83.

Pestova, T. V., V. G. Kolupaeva, I. B. Lomakin, E. V. Pilipenko, I. N. Shatsky, V. I. Agol and C. U. Hellen (2001). "Molecular mechanisms of translation initiation in eukaryotes." Proc Natl Acad Sci U S A **98**(13): 7029-36.

Petrov, V. M. and J. D. Karam (2002). "RNA determinants of translational operator recognition by the DNA polymerases of bacteriophages T4 and RB69." Nucleic Acids Res **30**(15): 3341-8.

Phillips, C. A., J. Gordon and E. K. Spicer (1996). "Bacteriophage T4 regA protein binds RNA as a monomer, overcoming dimer interactions." Nucleic Acids Res **24**(21): 4319-26.

Predki, P. F., L. M. Nayak, M. B. Gottlieb and L. Regan (1995). "Dissecting RNA-protein interactions: RNA-RNA recognition by Rop." Cell **80**(1): 41-50.

Preiss, J. and T. Romeo (1989). "Physiology, biochemistry and genetics of bacterial glycogen synthesis." Adv Microb Physiol **30**: 183-238.

Prodromou, C., P. J. Artymiuk and J. R. Guest (1992). "The aconitase of *Escherichia coli*. Nucleotide sequence of the aconitase gene and amino acid sequence similarity with mitochondrial aconitases, the iron-responsive-element-binding protein and isopropylmalate isomerases." Eur J Biochem **204**(2): 599-609.

Pruss, B. M., J. W. Campbell, T. K. Van Dyk, C. Zhu, Y. Kogan and P. Matsumura (2003). "FlhD/FlhC is a regulator of anaerobic respiration and the Entner-Doudoroff pathway through induction of the methyl-accepting chemotaxis protein Aer." J Bacteriol **185**(2): 534-43.

Puglisi, J. D., L. Chen, S. Blanchard and A. D. Frankel (1995). "Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex." Science **270**(5239): 1200-3.

Puglisi, J. D., R. Tan, B. J. Calnan, A. D. Frankel and J. R. Williamson (1992). "Conformation of the TAR RNA-arginine complex by NMR spectroscopy." Science **257**(5066): 76-80.

Ramos, A., S. Grunert, J. Adams, D. R. Micklem, M. R. Proctor, S. Freund, M. Bycroft, D. St Johnston and G. Varani (2000). "RNA recognition by a Staufen double-stranded RNA-binding domain." EMBO J **19**(5): 997-1009.

Rana, T. M. and K. T. Jeang (1999). "Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA." Arch Biochem Biophys **365**(2): 175-85.

Rasis, M. and G. Segal (2009). "The LetA-RsmYZ-CsrA regulatory cascade, together with RpoS and PmrA, post-transcriptionally regulates stationary phase activation of *Legionella pneumophila* Icm/Dot effectors." Mol Microbiol.

Reed, R. and K. Magni (2001). "A new view of mRNA export: separating the wheat from the chaff." Nat Cell Biol **3**(9): E201-4.

Reimmann, C., C. Valverde, E. Kay and D. Haas (2005). "Posttranscriptional repression of GacS/GacA-controlled genes by the RNA-binding protein RsmE acting together with RsmA in the biocontrol strain *Pseudomonas fluorescens* CHA0." J Bacteriol **187**(1): 276-85.

Rife, C., R. Schwarzenbacher, D. McMullan, P. Abdubek, E. Ambing, H. Axelrod, T. Biorac, J. M. Canaves, H. J. Chiu, A. M. Deacon, M. DiDonato, M. A. Elsliger, A. Godzik, C. Grittini, S. K. Grzechnik, J. Hale, E. Hampton, G. W. Han, J. Haugen, M. Hornsby, L. Jaroszewski, H. E. Klock, E. Koesema, A. Kreusch, P. Kuhn, S. A. Lesley, M. D. Miller, K. Moy, E. Nigoghossian, J. Paulsen, K. Quijano, R. Reyes, E. Sims, G. Spraggon, R. C. Stevens, H. van den Bedem, J.

Velasquez, J. Vincent, A. White, G. Wolf, Q. Xu, K. O. Hodgson, J. Wooley and I. A. Wilson (2005). "Crystal structure of the global regulatory protein CsrA from *Pseudomonas putida* at 2.05 A resolution reveals a new fold." Proteins **61**(2): 449-53.

Rodriguez, A. J., K. Czaplinski, J. S. Condeelis and R. H. Singer (2008). "Mechanisms and cellular roles of local protein synthesis in mammalian cells." Curr Opin Cell Biol **20**(2): 144-9.

Romaniuk, P. J., P. Lowary, H. N. Wu, G. Stormo and O. C. Uhlenbeck (1987). "RNA binding site of R17 coat protein." Biochemistry **26**(6): 1563-8.

Romby, P. and M. Springer (2003). "Bacterial translational control at atomic resolution." Trends Genet **19**(3): 155-61.

Romeo, T., M. Gong, M. Y. Liu and A. M. Brun-Zinkernagel (1993). "Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties." J Bacteriol **175**(15): 4744-55.

Romeo, T. and J. Preiss (1989). "Genetic regulation of glycogen biosynthesis in *Escherichia coli*: in vitro effects of cyclic AMP and guanosine 5'-diphosphate 3'-diphosphate and analysis of in vivo transcripts." J Bacteriol **171**(5): 2773-82.

Romling, U. and D. Amikam (2006). "Cyclic di-GMP as a second messenger." Curr Opin Microbiol **9**(2): 218-28.

Rouault, T. A. and R. D. Klausner (1996). "Iron-sulfur clusters as biosensors of oxidants and iron." Trends Biochem Sci **21**(5): 174-7.

Sabnis, N. A., H. Yang and T. Romeo (1995). "Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene csrA." J Biol Chem **270**(49): 29096-104.

Sahr, T., H. Bruggemann, M. Jules, M. Lomma, C. Albert-Weissenberger, C. Cazalet and C. Buchrieser (2009). "Two small ncRNAs jointly govern virulence and transmission in *Legionella pneumophila*." Mol Microbiol.

Sarsero, J. P., E. Merino and C. Yanofsky (2000). "A *Bacillus subtilis* gene of previously unknown function, *yhaG*, is translationally regulated by tryptophan-activated TRAP and appears to be involved in tryptophan transport." J Bacteriol **182**(8): 2329-31.

Sarsero, J. P., E. Merino and C. Yanofsky (2000). "A *Bacillus subtilis* operon containing genes of unknown function senses tRNATrp charging and regulates expression of the genes of tryptophan biosynthesis." Proc Natl Acad Sci U S A **97**(6): 2656-61.

Sauter, C., J. Basquin and D. Suck (2003). "Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*." Nucleic Acids Res **31**(14): 4091-8.

Sayed, A., S. Matsuyama and M. Inouye (1999). "Era, an essential *Escherichia coli* small G-protein, binds to the 30S ribosomal subunit." Biochem Biophys Res Commun **264**(1): 51-4.

Schechter, L. M., S. M. Damrauer and C. A. Lee (1999). "Two AraC/XylS family members can independently counteract the effect of repressing sequences upstream of the *hilA* promoter." Mol Microbiol **32**(3): 629-42.

Schlax, P. J., K. A. Xavier, T. C. Gluick and D. E. Draper (2001). "Translational repression of the *Escherichia coli* alpha operon mRNA: importance of an mRNA conformational switch and a ternary entrapment complex." J Biol Chem **276**(42): 38494-501.

Schultz, J. E. and A. Matin (1991). "Molecular and functional characterization of a carbon starvation gene of *Escherichia coli*." J Mol Biol **218**(1): 129-40.

Schumacher, M. A., R. F. Pearson, T. Moller, P. Valentin-Hansen and R. G. Brennan (2002). "Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein." EMBO J **21**(13): 3546-56.

Scofield, D. G. and M. Lynch (2008). "Evolutionary diversification of the Sm family of RNA-associated proteins." Mol Biol Evol **25**(11): 2255-67.

Searles, M. A., D. Lu and A. Klug (2000). "The role of the central zinc fingers of transcription factor IIIA in binding to 5 S RNA." J Mol Biol **301**(1): 47-60.

Sengupta, J., R. K. Agrawal and J. Frank (2001). "Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA." Proc Natl Acad Sci U S A **98**(21): 11991-6.

Serio, A. W., K. B. Pechter and A. L. Sonenshein (2006). "*Bacillus subtilis* aconitase is required for efficient late-sporulation gene expression." J Bacteriol **188**(17): 6396-405.

Shamoo, Y., A. Tam, W. H. Konigsberg and K. R. Williams (1993). "Translational repression by the bacteriophage T4 gene 32 protein involves specific recognition of an RNA pseudoknot structure." J Mol Biol **232**(1): 89-104.

Shevtsov, M. B., Y. Chen, P. Gollnick and A. A. Antson (2005). "Crystal structure of *Bacillus subtilis* anti-TRAP protein, an antagonist of TRAP/RNA interaction." Proc Natl Acad Sci U S A **102**(49): 17600-5.

Singh, R. N., R. J. Saldanha, L. M. D'Souza and A. M. Lambowitz (2002). "Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation." J Mol Biol **318**(2): 287-303.

Sjoblom, S., G. Brader, G. Koch and E. T. Palva (2006). "Cooperation of two distinct ExpR regulators controls quorum sensing specificity and virulence in the plant pathogen *Erwinia carotovora*." Mol Microbiol **60**(6): 1474-89.

Smith, C. A., V. Calabro and A. D. Frankel (2000). "An RNA-binding chameleon." Mol Cell **6**(5): 1067-76.

Snyder, L. and W. Champness (2007). Molecular genetics of bacteria. Washington, D.C., ASM Press.

Sonenberg, N. and A. G. Hinnebusch (2009). "Regulation of translation initiation in eukaryotes: mechanisms and biological targets." Cell **136**(4): 731-45.

Sorger-Domenigg, T., E. Sonnleitner, V. R. Kaberdin and U. Blasi (2007). "Distinct and overlapping binding sites of *Pseudomonas aeruginosa* Hfq and RsmA proteins on the non-coding RNA RsmY." Biochem Biophys Res Commun **352**(3): 769-73.

Springer, M., J. A. Plumbridge, J. S. Butler, M. Graffe, J. Dondon, J. F. Mayaux, G. Fayat, P. Lestienne, S. Blanquet and M. Grunberg-Manago (1985). "Autogenous control of *Escherichia coli* threonyl-tRNA synthetase expression in vivo." J Mol Biol **185**(1): 93-104.

Struble, E. B., J. E. Ladner, D. M. Brabazon and J. P. Marino (2008). "New crystal structures of ColE1 Rom and variants resulting from mutation of a surface exposed residue: Implications for RNA-recognition." Proteins **72**(2): 761-8.

Suzuki, K., P. Babitzke, S. R. Kushner and T. Romeo (2006). "Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E." Genes Dev **20**(18): 2605-17.

Suzuki, K., X. Wang, T. Weilbacher, A. K. Pernestig, O. Melefors, D. Georgellis, P. Babitzke and T. Romeo (2002). "Regulatory circuitry of the CsrA/CsrB and BarA/UvrY systems of *Escherichia coli*." J Bacteriol **184**(18): 5130-40.

Tan, R. and A. D. Frankel (1995). "Structural variety of arginine-rich RNA-binding peptides." Proc Natl Acad Sci U S A **92**(12): 5282-6.

Tang, Y. and J. R. Guest (1999). "Direct evidence for mRNA binding and post-transcriptional regulation by *Escherichia coli* aconitases." Microbiology **145 (Pt 11)**: 3069-79.

Tang, Y., J. R. Guest, P. J. Artymiuk, R. C. Read and J. Green (2004). "Post-transcriptional regulation of bacterial motility by aconitase proteins." Mol Microbiol **51**(6): 1817-26.

Tang, Y., M. A. Quail, P. J. Artymiuk, J. R. Guest and J. Green (2002). "*Escherichia coli* aconitases and oxidative stress: post-transcriptional regulation of *sodA* expression." Microbiology **148**(Pt 4): 1027-37.

Teplitski, M., A. Al-Agely and B. M. Ahmer (2006). "Contribution of the SirA regulon to biofilm formation in *Salmonella enterica* serovar Typhimurium." Microbiology **152**(Pt 11): 3411-24.

Teplitski, M., R. I. Goodier and B. M. Ahmer (2003). "Pathways leading from BarA/SirA to motility and virulence gene expression in *Salmonella*." J Bacteriol **185**(24): 7257-65.

Teplitski, M., R. I. Goodier and B. M. Ahmer (2006). "Catabolite repression of the SirA regulatory cascade in *Salmonella enterica*." Int J Med Microbiol **296**(7): 449-66.

Theobald, D. L., R. M. Mitton-Fry and D. S. Wuttke (2003). "Nucleic acid recognition by OB-fold proteins." Annu Rev Biophys Biomol Struct **32**: 115-33.

Thomson, A. M., J. T. Rogers and P. J. Leedman (1999). "Iron-regulatory proteins, iron-responsive elements and ferritin mRNA translation." Int J Biochem Cell Biol **31**(10): 1139-52.

Thore, S., C. Mayer, C. Sauter, S. Weeks and D. Suck (2003). "Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya." J Biol Chem **278**(2): 1239-47.

Unnithan, S., L. Green, L. Morrissey, J. Binkley, B. Singer, J. Karam and L. Gold (1990). "Binding of the bacteriophage T4 RegA protein to mRNA targets: an initiator AUG is required." Nucleic Acids Res **18**(23): 7083-92.

Valentin-Hansen, P., M. Eriksen and C. Udesen (2004). "The bacterial Sm-like protein Hfq: a key player in RNA transactions." Mol Microbiol **51**(6): 1525-33.

Valverde, C., S. Heeb, C. Keel and D. Haas (2003). "RsmY, a small regulatory RNA, is required in concert with RsmZ for GacA-dependent expression of biocontrol traits in *Pseudomonas fluorescens* CHA0." Mol Microbiol **50**(4): 1361-79.

Valverde, C., M. Lindell, E. G. Wagner and D. Haas (2004). "A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of *Pseudomonas fluorescens*." J Biol Chem **279**(24): 25066-74.

Valverde, R., L. Edwards and L. Regan (2008). "Structure and function of KH domains." FEBS J **275**(11): 2712-26.

Van Houdt, R., A. Aertsen, P. Moons, K. Vanoirbeek and C. W. Michiels (2006). "N-acyl-L-homoserine lactone signal interception by *Escherichia coli*." FEMS Microbiol Lett **256**(1): 83-9.

Varghese, S., Y. Tang and J. A. Imlay (2003). "Contrasting sensitivities of *Escherichia coli* aconitases A and B to oxidation and iron depletion." J Bacteriol **185**(1): 221-30.

Vecerek, B., I. Moll and U. Blasi (2005). "Translational autocontrol of the *Escherichia coli hfq* RNA chaperone gene." RNA **11**(6): 976-84.

Ventre, I., A. L. Goodman, I. Vallet-Gely, P. Vasseur, C. Soscia, S. Molin, S. Bleves, A. Lazdunski, S. Lory and A. Filloux (2006). "Multiple sensors control reciprocal expression of *Pseudomonas aeruginosa* regulatory RNA and virulence genes." Proc Natl Acad Sci U S A **103**(1): 171-6.

Vlasova, I. A. and P. R. Bohjanen (2008). "Posttranscriptional regulation of gene networks by GU-rich elements and CELF proteins." RNA Biol **5**(4): 201-7.

Vogel, J. (2009). "A rough guide to the non-coding RNA world of *Salmonella*." Mol Microbiol **71**(1): 1-11.

Vytvytska, O., I. Moll, V. R. Kaberdin, A. von Gabain and U. Blasi (2000). "Hfq (HF1) stimulates *ompA* mRNA decay by interfering with ribosome binding." Genes Dev **14**(9): 1109-18.

Wang, G., P. Alamuri and R. J. Maier (2006). "The diverse antioxidant systems of *Helicobacter pylori*." Mol Microbiol **61**(4): 847-60.

Wang, S., R. T. Fleming, E. M. Westbrook, P. Matsumura and D. B. McKay (2006). "Structure of the *Escherichia coli* FlhDC complex, a prokaryotic heteromeric regulator of transcription." J Mol Biol **355**(4): 798-808.

Wang, X., A. K. Dubey, K. Suzuki, C. S. Baker, P. Babitzke and T. Romeo (2005). "CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*." Mol Microbiol **56**(6): 1648-63.

Wang, X., J. F. Preston, 3rd and T. Romeo (2004). "The *pgaABCD* locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation." J Bacteriol **186**(9): 2724-34.

Wang, X. and T. M. Tanaka Hall (2001). "Structural basis for recognition of AU-rich element RNA by the HuD protein." Nat Struct Biol **8**(2): 141-5.

Wang, Z. and C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." RNA **14**(5): 802-13.

Wei, B., S. Shin, D. LaPorte, A. J. Wolfe and T. Romeo (2000). "Global regulatory mutations in *csrA* and *rpoS* cause severe central carbon stress in *Escherichia coli* in the presence of acetate." J Bacteriol **182**(6): 1632-40.

Wei, B. L., A. M. Brun-Zinkernagel, J. W. Simecka, B. M. Pruss, P. Babitzke and T. Romeo (2001). "Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*." Mol Microbiol **40**(1): 245-56.

Wei, Y., J. M. Lee, D. R. Smulski and R. A. LaRossa (2001). "Global impact of *sdiA* amplification revealed by comprehensive gene expression profiling of *Escherichia coli*." J Bacteriol **183**(7): 2265-72.

Weilbacher, T., K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke and T. Romeo (2003). "A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*." Mol Microbiol **48**(3): 657-70.

Wells, D. G. (2006). "RNA-binding proteins: a lesson in repression." J Neurosci **26**(27): 7135-8.

Wilkie, G. S., K. S. Dickson and N. K. Gray (2003). "Regulation of mRNA translation by 5'- and 3'-UTR-binding factors." Trends Biochem Sci **28**(4): 182-8.

Williamson, N. R., P. C. Fineran, W. Ogawa, L. R. Woodley and G. P. Salmond (2008). "Integrated regulation involving quorum sensing, a two-component system, a GGDEF/EAL domain protein and a post-transcriptional regulator controls swarming and RhlA-dependent surfactant biosynthesis in *Serratia*." Environ Microbiol **10**(5): 1202-17.

Wilusz, C. J. and J. Wilusz (2004). "Bringing the role of mRNA decay in the control of gene expression into focus." Trends Genet **20**(10): 491-7.

Wilusz, C. J. and J. Wilusz (2005). "Eukaryotic Lsm proteins: lessons from bacteria." Nat Struct Mol Biol **12**(12): 1031-6.

Wilusz, C. J., M. Wormington and S. W. Peltz (2001). "The cap-to-tail guide to mRNA turnover." Nat Rev Mol Cell Biol **2**(4): 237-46.

Witherell, G. W. and O. C. Uhlenbeck (1989). "Specific RNA binding by Q beta coat protein." Biochemistry **28**(1): 71-6.

Wong, S. M. and B. J. Akerley (2005). "Environmental and genetic regulation of the phosphorylcholine epitope of *Haemophilus influenzae* lipooligosaccharide." Mol Microbiol **55**(3): 724-38.

Yakhnin, H., P. Pandit, T. J. Petty, C. S. Baker, T. Romeo and P. Babitzke (2007). "CsrA of *Bacillus subtilis* regulates translation initiation of the gene encoding the flagellin protein (*hag*) by blocking ribosome binding." Mol Microbiol **64**(6): 1605-20.

Yakhnin, H., A. V. Yakhnin and P. Babitzke (2006). "The trp RNA-binding attenuation protein (TRAP) of *Bacillus subtilis* regulates translation initiation of *ycbK*, a gene encoding a putative efflux protein, by blocking ribosome binding." Mol Microbiol **61**(5): 1252-66.

Yakhnin, H., H. Zhang, A. V. Yakhnin and P. Babitzke (2004). "The *trp* RNA-binding attenuation protein of *Bacillus subtilis* regulates translation of the tryptophan transport gene *trpP* (*yhaG*) by blocking ribosome binding." J Bacteriol **186**(2): 278-86.

Yang, H., M. Y. Liu and T. Romeo (1996). "Coordinate genetic regulation of glycogen catabolism and biosynthesis in *Escherichia coli* via the CsrA gene product." J Bacteriol **178**(4): 1012-7.

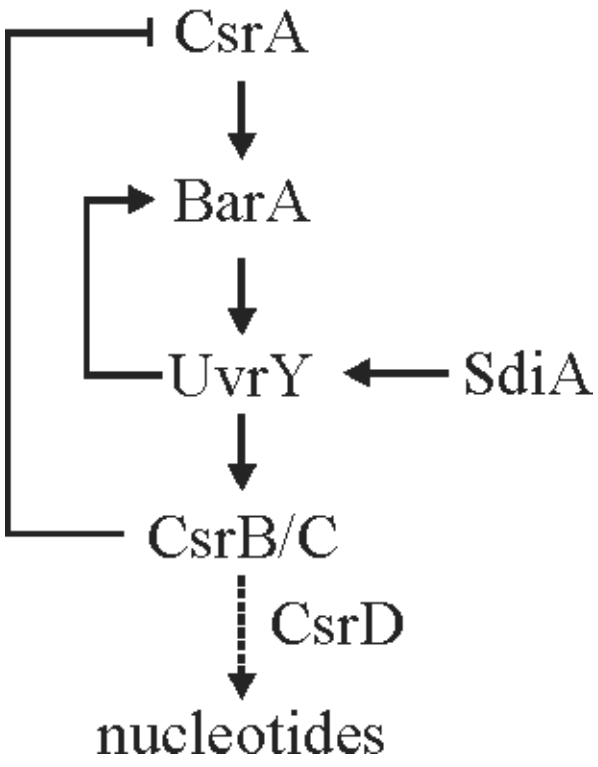
Yarchuk, O., N. Jacques, J. Guillerez and M. Dreyfus (1992). "Interdependence of translation, transcription and mRNA degradation in the *lacZ* gene." J Mol Biol **226**(3): 581-96.

Yuste, L. and F. Rojo (2001). "Role of the *crc* gene in catabolic repression of the *Pseudomonas putida* GPo1 alkane degradation pathway." J Bacteriol **183**(21): 6197-206.

FIGURE LEGEND

Fig. 1-1. An outline of the *E. coli* K12 Csr circuitry. CsrA indirectly activates expression of the BarA sensor kinase that phosphorylates the UvrY response regulator, which then directly activates transcription of the CsrB and CsrC sRNAs. CsrB and CsrC bind to multiple CsrA dimers in a globular complex that titrates free CsrA from the cellular pool, thus repressing CsrA activity. UvrY directly activates BarA transcription, and SdiA positively affects CsrB and CsrC transcription in a UvrY-dependent manner.

Fig. 1-1.



Chapter 2. Comprehensive Alanine-scanning Mutagenesis of *Escherichia coli* CsrA Defines Two Subdomains of Critical Functional Importance

Jeffrey Mercante[1], Kazushi Suzuki[1], Xiaodong Cheng[2], Paul Babitzke[3], and Tony Romeo[1,4]

[1]Microbiology and Immunology and [2]Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30322.

[3]Department of Microbiology and Cell Science. P.O. Box 110700. University of Florida. Gainesville, FL 32611-0700.

[4]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802

This chapter consists of a manuscript that was published in *J Biol Chem*. 2006 Oct 20; 281(42):31832-42. The manuscript was written by Jeffrey Mercante, Paul Babitzke and Tony Romeo. All experiments were performed by Jeffrey Mercante.

SUMMARY

CsrA is a small, 61 amino acid RNA-binding protein of *E. coli* K12 and a wide array of related eubacteria. CsrA post-transcriptionally regulates gene expression by binding to the 5'-untranslated leaders of target transcripts and inhibiting ribosome loading. A comprehensive alanine-scanning mutagenesis of CsrA revealed amino acids that are important for regulating the expression of *glgC'*, *pgaA'* and *flhDC'-'lacZ* fusions and for binding to a high affinity SELEX-derived RNA target. The ability of all CsrA site-directed mutants to regulate gene expression was positively correlated with their RNA-binding capacity. Important amino acids were distributed in 2 regions found at opposite ends of the protein primary sequence; one region is located at the CsrA extreme N-terminus (region 1, residues 2-7) and the other is closer to the C-terminus (region 2, residues 40-47). Recent structural studies revealed that CsrA is a bisymmetrical homodimer and an entirely new class of protein fold made up of 2 interdigitated peptide monomers. These structures enabled *in silico* mapping of both critical regions onto a three-dimensional model of CsrA, which showed that region 1 (located in β_1) from one subunit is found adjacent and parallel to region 2 (located in β_5') from the other subunit, and vice versa (β_1' is found adjacent and parallel to β_5). Therefore, it was deduced that CsrA contains 2 identical RNA-binding subdomains located on opposite sides of the dimer. And each functional, RNA-binding subdomain is composed of residues from both monomers, the most important being R44. The presence of 2 binding surfaces per dimer helps explain previous data showing that the globular structure of a CsrA-CsrB complex contains ~20 CsrA subunits associated with ~10 RNA target sites.

ABSTRACT

The RNA-binding protein CsrA (carbon storage regulator) of *Escherichia coli* is a global regulator of gene expression, and is representative of the CsrA/RsmA family of bacterial proteins. These proteins act by regulating mRNA translation and stability, and are antagonized by binding to small noncoding RNAs. While the RNA target sequence and structure for CsrA binding have been well defined, little information exists concerning the protein requirements for RNA recognition. The three-dimensional structures of three CsrA/RsmA proteins were recently solved, revealing a novel protein fold consisting of two interdigitated monomers. Here, we performed comprehensive alanine-scanning mutagenesis on *csrA* of *E. coli* and tested the 58 resulting mutants for regulation of glycogen accumulation, motility, and biofilm formation. Quantitative effects of these mutations on expression of *glgCA'-'lacZ, flhDC'-'lacZ,* and *pgaA'-'lacZ* translational fusions were also examined, and eight of the mutant proteins were purified and tested for RNA binding. These studies identified two regions of amino acid sequence that were critical for regulation and RNA binding, located within the first (β_1, residues 2-7) and containing the last (β_5, residues 40-47) β-strands of CsrA. The β_1 and β_5 strands of opposite monomers lie adjacent and parallel to each other in the three-dimensional structure of this protein. Given the symmetry of the CsrA dimer, these findings imply that two distinct RNA binding surfaces or functional subdomains lie on opposite sides of the protein.

INTRODUCTION

CsrA is a 61 amino acid RNA-binding protein originally identified in *Escherichia coli* K-12 (Romeo, Gong *et al.*, 1993). This protein is an important regulator of gene expression for the species in which it has been studied. At least 159 putative CsrA homologues are distributed among 130 different eubacterial species (Altschul, Gish *et al.*, 1990; Bateman, Coin *et al.*, 2004) with several species encoding more than one homologue (e.g. *Coxiella burnetii*, *Legionella pneumophila*, *Pseudomonas fluorescens,* and *Pirellula* sp.). In *E. coli*, CsrA represses glycogen biosynthesis, gluconeogenesis, peptide transport and biofilm formation while activating glycolysis, acetate metabolism and motility (Sabnis, Yang *et al.*, 1995; Wei, Brun-Zinkernagel *et al.*, 2001; Dubey, Baker *et al.*, 2003). CsrA is also known to be critical for the regulated expression of virulence factors in pathogens of both animals and plants. For example: *Legionella pneumophila* CsrA is essential for the repression of transmission-phase genes during infection of amoebae and alveolar macrophages (Molofsky and Swanson, 2003); *Salmonella enterica* CsrA regulates pathogenecity island (SPI1) genes required for intestinal epithelial cell invasion (Altier, Suyemoto *et al.*, 2000); *Erwinia carotovora* RsmA represses expression of extracellular enzymes needed for plant pathogenecity and development of soft-rot disease as well as production of homoserine-lactone quorum sensing signals (Cui, Chatterjee *et al.*, 1995).

In *E. coli*, CsrA controls gene expression post-transcriptionally, binding specifically and with high affinity to the 5'-untranslated leader of several mRNAs, including *cstA, glgCAP*, *flhDC* and *pgaABCD*, and affecting translation and RNA

stability (Liu and Romeo, 1997; Wei, Brun-Zinkernagel *et al.*, 2001; Baker, Morozov *et al.*, 2002; Dubey, Baker *et al.*, 2003; Wang, Dubey *et al.*, 2005). The activity of CsrA is antagonized by the untranslated RNAs CsrB and CsrC, which contain multiple binding sites (~18 and ~9, respectively) that permit sequestration of CsrA (Liu, Gui *et al.*, 1997; Weilbacher, Suzuki *et al.*, 2003). Regulation of gene expression by CsrA-regulated in other species (i.e. *Pseudomonas*, *Erwinia*, *Salmonella* and *Vibrio*) is believed to occur in a similar manner as in *E. coli* (Liu, Cui *et al.*, 1998; Altier, Suyemoto *et al.*, 2000; Heurlier, Williams *et al.*, 2004; Lenz, Miller *et al.*, 2005; Fortune, Suyemoto *et al.*, 2006). The affinity of CsrA for a particular RNA target depends on both the RNA sequence and secondary structure (Baker, Morozov *et al.*, 2002; Valverde, Lindell *et al.*, 2004; Wang, Dubey *et al.*, 2005). Until recently, no protein structural or functional information existed to describe how CsrA might recognize a particular RNA target with high affinity.

A major step in understanding CsrA structure was taken when three CsrA (RsmA) proteins were independently solved, all of which exhibited the same overall topology (Gutierrez, Li *et al.*, 2005; Rife, Schwarzenbacher *et al.*, 2005; Heeb, Kuehne *et al.*, 2006). These studies proved CsrA to be a novel class of RNA-binding protein and confirmed previous work that suggested CsrA forms a homodimer (Dubey, Baker *et al.*, 2003). They also highlighted the unusual way in which the dimer is formed: two interlocking CsrA monomers produce a hydrophobic core composed of 10 β-strands and 2 wing-like α-helices. As described previously (Heeb, Kuehne *et al.*, 2006), a CsrA dimer is a barrel-like structure stabilized non-covalently by an extensive network of hydrogen bonds from backbone amino and carboxyl groups. With the benefit of a CsrA

three-dimensional (3-D) structure onto which any particular amino acid change can be modeled, it is now possible to map protein regions essential for RNA binding and/or regulation of gene expression.

The goal of the present study was to examine the amino acid requirements for CsrA-RNA interaction and gene regulation by the systematic mutation of every non-alanine codon within the *E. coli* gene. Alanine replacement has an extensive history in structure-function studies (Baase, Eriksson *et al.*, 1992; Wells, Cunningham *et al.*, 1993; Kelly, Goujon *et al.*, 1994; MacLennan, Rice *et al.*, 1998; Sasaki and Sutoh, 1998; Wiesmann and de Vos, 2001; Norton, Pennington *et al.*, 2004) and is well suited for use in site-directed mutagenesis because it is abundant in many types of secondary structures and displays an uncharged and non-intrusive side chain that typically is compatible with native protein folding (Cunningham and Wells, 1989). *In vivo* examination of CsrA-repressed (*glgCAP, pgaABCD*) and activated (*flhDC*) operons in the presence of this series of mutants uncovered two regions within the primary structure of CsrA that are essential for its roles in repression and activation of gene expression. Region 1 was located at the extreme N-terminus (residues 2-7); region 2 was closer to the C-terminus (residues 40-47). Four CsrA mutant proteins from each of these regions (L2A, L4A, R6A, R7A, and V40A, V42A, R44A, I47A) were found to be defective in binding to a 16 nucleotide RNA probe containing a single high affinity CsrA target site. Interestingly, critical regions 1 and 2 were aligned in parallel and adjacent to each other within the three-dimensional space of the protein. In contrast to certain other multimeric RNA-binding proteins, such as the mammalian U1A and Nova (Jovine, Oubridge *et al.*, 1996; Symmons, Jones *et al.*, 2000), bacteriophage coat protein MS2 (van den Worm,

Stonehouse *et al.*, 1998) and *E. coli* AspRS (De Guzman, Turner *et al.*, 1998; Eiler, Dock-Bregeon *et al.*, 1999) wherein individual monomers comprise a complete functional domain, CsrA requires that each monomer contribute N- and C-terminal portions to the creation of a functional region. This further implies that a symmetrical CsrA homodimer contains two critical surfaces or subdomains located on opposite sides of the molecule. A model for CsrA structure that integrates these data and provides a basis for understanding how CsrA interacts with its RNA targets is presented and discussed.

EXPERIMENTAL PROCEDURES

Strains, Plasmids, Bacteriophage and Growth Conditions—All strains, plasmids and bacteriophage used in this study are listed in Table 1s in the supplemental data (SD). For routine culture of bacteria, strains were grown in Luria-Bertani (LB) medium (Miller, 1972) at 37 ^{o}C with shaking. Cultures for biofilm assays were grown in LB medium at 26 ^{o}C for 24 hours without shaking. Motility assays were carried out as previously described in semisolid tryptone medium (1% tryptone, 0.5% NaCl, pH 7.4, 0.35% agar) at 30 ^{o}C (Wei, Brun-Zinkernagel *et al.*, 2001). Glycogen accumulation was assessed on Kornberg medium (1.1% K_2HPO_4, 0.85% KH_2PO_4, 0.6% yeast extract, 1% glucose pH 6.8, 1.5% agar). Antibiotics were added to media as needed, at the following concentrations: ampicillin, 100 μg ml^{-1}; kanamycin, 100 μg ml^{-1}; chloramphenicol, 25 μg ml^{-1}; tetracycline, 10 μg ml^{-1}.

Glycogen Accumulation, Biofilm and Motility Assays—Endogenous glycogen accumulation was examined by iodine-vapor staining of colonies or patches grown

overnight at 37 ^{o}C on Kornberg medium (Romeo, Gong *et al.*, 1993). Quantitative biofilm formation assays were performed by staining adherent cells with crystal violet (Jackson, Suzuki *et al.*, 2002). These assays were performed at least twice, with four replicas per experiment, and the resulting averages and standard error values were determined. Motility was assessed as described previously (Wei, Brun-Zinkernagel *et al.*, 2001).

β-galactosidase Activity and Total Protein Assays—β-galactosidase activity was determined using the method described previously (Griffith and Wolf, 2002), with minor modifications. Briefly, cultures were diluted 1:500 in LB medium and grown with shaking at 37 ^{o}C in 2 ml polypropylene Deepwell™ 96-well plates (Nunc, Rochester, NY). At 24 hr the plate culture was placed on ice and chloramphenicol or tetracycline was added to a final concentration of 100 μg ml^{-1}, or 10 μg ml^{-1} respectively. Aliquots of each culture were removed to measure cell density (A_{600nm}) and for total protein determination. Into each well of a second deepwell plate was added 1 ml Z buffer (40 mM $NaH_2PO_4 \cdot H_2O$, 60 mM $Na_2HPO_4 \cdot 7H_2O$, 40 mM $MgSO_4 \cdot 7H_2O$, 50 mM β-mercaptoethanol), 20 μl 0.1% SDS, 40 μl CH_3Cl, and 50 μl of the original culture. Components were mixed by pipeting, CH_3Cl was allowed to settle to the bottom of the well and 100 μl of permeabilized cells was transferred to a 96-well microtiter plate. Either o-nitrophenyl-β-D-galactopyranoside (ONPG) or 4-methylumbelliferyl β-D-galactopyranoside (MUG) were added to the permeabilized cells at 666 μM or 500 μM, respectively. The ONPG reaction was stopped after 15 min by the addition of 50 μl of 1 M Na_2CO_3 and the sample plate was read on a Biotech Synergy HT microplate reader (Winooski, VT) at A_{420nm} and A_{550nm}. Samples with MUG were monitored for 1-5 hr in a

microplate reader, periodically reading fluorescence using an excitation wavelength of 360 nm ± 40 nm and an emission wavelength of 460 nm ± 40 nm. *In vivo* expression was normalized as a percent of empty vector control levels (labeled 100%) for the *glgCA'-'lacZ* and *pgaA'-'lacZ* fusions and as a percent of wild-type CsrA levels (labeled 100%) for the *flhDC'-'lacZ* fusion.

Total cellular protein was measured using the bicinchroninic acid (BCA) assay (Smith, Krohn *et al.*, 1985) as developed by Pierce Biotechnology (Rockford, IL). The current study employed a 96-well microplate reader to monitor A_{562nm} and bovine serum albumin as a protein standard (Pierce Biotechnology).

Construction of Chromosomal pgaA'-'lacZ and flhD'-'lacZ Translational Fusions—LacZ fusion plasmids pFDCZ6 and pPGAZ4 were converted from ampicillin resistant (Ap^R) to chloramphenicol resistant (Cm^R) by subcloning the *cat* gene into a deletion created in the coding region of *bla*. Primers for PCR amplification of the *cat* gene from pKD3 are listed in Table 2s in the SD; they include sites for directional cloning using *Pst*I and *Sca*I restriction enzymes. pFDCZ6 and pPGAZ4 were cut with *Eco*RI, made blunt ended using T4 DNA polymerase (NEB, Ipswich, MA) and subsequently cut with *Pst*I, which allowed ligation to the amplified *cat* gene. The resulting gene fusions in plasmids pFDCZ6CAT312 and pPGAZ4CAT2321, which were Ap^S and Cm^R, were integrated into the *E. coli* CF7789 chromosome using the λInCh1 system, as described previously (Boyd, Weiss *et al.*, 2000). Chromosomal fusions were confirmed by colony PCR according to the λInCh1 protocol.

Construction of CsrA-His$_6$ Site-Directed Mutants—Each non-alanine amino acid residue of *E. coli* CsrA was substituted with alanine, providing a library of plasmids expressing 58 site-directed mutants of CsrA-His$_6$ (Table 1s in the SD). Mutations were constructed using the Quickchange II Site-Directed Mutagenesis® system (Stratagene, La Jolla, CA) employing the template plasmid pCSRH6-19 and the primers listed in Table 2s in the SD. All mutations were confirmed by DNA sequence analysis performed at SeqWright DNA Technology Service (Houston, TX).

CsrA-His$_6$ Protein Purification—Native CsrA-His$_6$ and eight mutant proteins were purified from *E. coli* srain CF7789, using nickel affinity chromatography according to a previously described method (Qiaexpressionist, Qiagen, Valencia, CA). Briefly, 1 L of culture was grown overnight in LB + 1% glucose. The cells were concentrated by centrifugation, resuspended (1 g per ml) in wash buffer (10 mM imidazole and 1 mg ml^{-1} lysozyme) and lysed by sonication. The resulting lysate was cleared by centrifugation and combined with 1 ml of Ni-NTA for 1 hr to allow binding of CsrA-His$_6$ to the nickel-NTA beads. The lysate-Ni-NTA mix was applied to a gravity flow column and the beads were washed with 10 ml of wash buffer containing 50 mM imidazole. CsrA-His$_6$ was eluted with 5 ml wash buffer containing 250 mM imidazole. Protein purity was assessed by SDS-PAGE (15%) and Coomassie blue staining, followed by densitometry using the Biorad Universal Hood and QuantityOne software package (Hercules, CA). Purity of the CsrA wild-type and alanine mutant proteins was as follows: wild-type ≥ 98%; L2A ≥ 97%; L4A ≥ 97%; R6A ≥ 98%; R7A ≥ 98%; V40A ≥ 90%; V42A ≥ 90%; R44A ≥ 95%; I47A ≥ 98%.

Western Analyses of CsrA-His$_6$—Cultures for western blot analyses were grown at 37 ^{o}C for 24 hr with shaking. Cells from 1 ml of culture were concentrated and resuspended in lysis buffer (90 mM Tris-HCl, 2% SDS, pH 6.8). Samples were boiled for 3 min, cell debris was removed by centrifugation, and the supernatant solution saved and assayed for total protein. Five μl SDS-PAGE loading buffer (135 mM Tris-HCl, 3% SDS, 0.03% bromophenol blue, 30% glycerol, pH 6.8) was added to 100 μg of total protein (~10 μl) and proteins were separated by SDS-PAGE (15%). Gels were equilibrated in Towbin buffer + 20% methanol (Towbin, Staehelin *et al.*, 1979) and electroblotted overnight to a nitrocellulose membrane (0.2 μm). Blots were washed and probed with His-probe™ HRP (Pierce Biotechnology, Rockford, IL) at 1:2000 in tris-buffered saline with Tween20 and developed with SuperSignal® West Pico chemiluminescent substrate (Pierce Biotechnology) according to the manufacturer's recommendations.

Three-Dimensional Modeling—The coordinates for the 3-D structure of CsrA from *Yersinia enterocolitica* were obtained from the Protein Data Bank (http://www.pdb.org, accession code 2BTI). Images were rendered using the open source modeling software PyMol (Delano, 2002).

RNA Gel Mobility Shift Assay—Gel mobility shift assays were carried out according to previously published procedures (Wang, Dubey *et al.*, 2005). All calculations were based on the mass of CsrA-His$_6$ as a homodimer (15,357 Da). The RNA probe was designed based on a high-affinity consensus CsrA binding target as determined by SELEX (systematic evolution of ligands by exponential enrichment) analysis (Wang, Dubey *et al.*, 2005). The minimal CsrA target sequence, 5'-

GGCACAAGGAUGUGCC-3', was synthesized by Integrated DNA Technologies (Coralville, IA) and 5'-end-labelled using T4 polynucleotide kinase and [γ-^{32}P]-ATP. Radiolabeled RNA probe was gel purified, suspended in TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA), heated to 85 ^{o}C and slowly cooled to room temperature. Increasing concentrations of CsrA-His_6 wild type or mutant protein were combined with 30 pM labeled RNA probe in 10 μl binding reactions (10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 100 mM KCl, 3.25 ng total yeast RNA, 10 mM DTT, 10% glycerol, 4 U RNase Inhibitor – Ambion, Austin, TX) for 30 min at 37 ^{o}C to allow CsrA-RNA complex formation. Binding reactions were fractionated using native PAGE (12%) and radioactive bands were visualized with a Molecular Dynamics phosphoimager (Uppsala, Sweden). Bound and unbound RNA species were quantified with ImageQuant software (Molecular Dynamics) and an apparent equilibrium binding constant (K_d) was calculated for CsrA-RNA complex formation according to a previously described cooperative binding equation (Yakhnin, Trimble *et al.*, 2000), as adapted here:

$$RNA_b = (Y_{max} \times ((CsrA_f / K_d)^n)) / (1 + ((CsrA_f / K_d)^n))$$

where Y_{max} is the maximum possible bound fraction (100%) of RNA (RNA_b); K_d is the concentration of free protein ($CsrA_f$) at which RNA_b reaches 50% bound. Isolated CsrA proteins were assumed to be 100% active for these calculations. The cooperativity of binding is described by the Hill coefficient, n.

RESULTS

Phenotypic effects of CsrA site-directed mutations: glycogen accumulation, motility and biofilm formation─CsrA post-transcriptionally regulates several cellular

processes in *E. coli* K-12 (Wei, Brun-Zinkernagel *et al.*, 2001; Baker, Morozov *et al.*, 2002; Jackson, Suzuki *et al.*, 2002; Dubey, Baker *et al.*, 2003; Wang, Dubey *et al.*, 2005). To determine which CsrA residues were important for functionality *in vivo*, every non-alanine codon of a His-tagged *csrA* gene (Liu, Gui *et al.*, 1997) was changed to alanine. The resulting library of 58 mutant *csrA*-expressing plasmids was introduced into the *csrA*-deficient strain *E. coli* K-12 TR1-5 MG1655 and tested for effects on the repression of glycogen accumulation and biofilm formation, and activation of motility relative to the wild-type His-tagged CsrA protein and an empty vector control. Regulation of these three processes was highly dependent on the position of the mutation (Fig. 2-1, *B*, *D* and *F*). Substitutions that caused the greatest effects on glycogen accumulation and motility were mainly clustered in two regions: the N-terminus (residues 2-7) and close to the C-terminus (40-47). Additionally, mutation of certain amino acids between these two regions (11-14 and 22) affected these phenotypes. Biofilm formation was also elevated (regulation was defective) in strains containing *csrA* mutations in these two main regions. However, this phenotype appeared to be more sensitive to changes throughout the CsrA sequence, and mutations in two other regions (residues 18-23 and 31-38) also increased biofilm formation.

Effects of CsrA alanine substitutions on expression of chromosomally-encoded glgCA'-, flhDC'-, and pgaA'-lacZ translational fusions—We next determined the effects of CsrA alanine substitutions on the regulation of the *glgCAP*, *flhDC*, and *pgaABCD* operons using *glgCA'-'lacZ* , *flhDC'-'lacZ* and *pgaA'-'lacZ* translational fusions (Fig. 2-1, *A*, *C*, and *E*). The gene expression effects of alanine substitutions were well correlated with the phenotypic effects on glycogen accumulation, motility and biofilm formation.

Once again, the extreme N-terminus (residues 2-7), and a region closer to the C-terminus (residues 40-47) were important. Alanine substitution of T11, L12, I14 and V22 also led to defects in regulation. Based on a composite graph developed from these findings (Fig. 2-6, "*In vivo* regulatory defect"), mutations in two regions affected expression of all three fusions: region 1 (L2, I3, L4, R6, R7) and region 2 (V40, V42, R44, E46 and I47). Only modest defects were exhibited by mutation of T5, H43 or E45 in these regions. In addition, while *pgaA'-'lacZ* expression exhibited greater effects of most of the alanine mutations when compared to *glgCA'-'lacZ* or *flhDC'-'lacZ*, regulation of all fusions followed the same trends (Fig. 2-1). The *pgaABCD* leader contains six confirmed CsrA binding sites, more than any other known CsrA-regulated operon (Wang, Dubey *et al.*, 2005). Perhaps efficient binding at all six sites is needed for full repression, and therefore a small reduction in CsrA binding affinity may have greater potential to derepress *pgaABCD* expression.

Surprisingly, at least one alanine substitution, T19A, caused CsrA to become a stronger regulator of gene expression from all operons that were tested (Fig. 2-1). N35A also appeared to make CsrA a slightly better regulator of gene expression from *glgCA'-* and *flhDC'-'lacZ*.

Steady-state protein levels of CsrA site-directed mutant proteins—Altered CsrA function *in vivo* might reflect a change in either the intrinsic activity or intracellular concentration of a mutant protein. Therefore, the steady-state levels of mutant proteins were determined by western blot analysis (Fig. 2-2). This analysis revealed that the defective proteins I3A, T11A, L12A, I14A, V22A and E46A failed to accumulate properly. Levels of most of the mutant proteins (59%) were within two-fold of the wild

type protein. A substantial percentage of mutants exhibited either very low (≤ 25%) or very high (≥ 175%) levels. Interestingly, of the 3 mutants that accumulated to ≥ 175% of wild type protein levels, two were functionally impaired (R6A, R44A). Conversely, among the 9 proteins that exhibited ≤ 25% of wild type protein levels, 5 displayed ≥ 50% of wild type activity (Q52A, V18A, K55A, V20A and V34A).

It is noteworthy that while total protein (SDS-extractable) for R44A was high, the quantity of soluble protein isolated by nickel-affinity chromatography was similar to that of the wild-type (data not shown). This suggests that R44A may have formed aggregates that were relatively resistant to proteolysis in the cell, which were removed by the centrifugation during purification. Such behavior was not reported for the *Pseudomonas* R44A mutant protein, but may have been missed because the yield of this protein was determined only after purification under native conditions (Heeb, Kuehne *et al.*, 2006).

RNA-binding studies of eight CsrA mutant proteins—To determine whether the *in vivo* regulatory defects of selected mutant proteins were due to defective RNA binding, eight of the mutant proteins were purified and examined by gel mobility shift assays (Fig. 2-3). Proteins I3A, T11A, L12A, I14A, V22A and E46A, whose defects were likely to have been caused by altered accumulation (Fig. 2-2), were avoided. Mutant proteins L2A, L4A, R6A, R7A, V40A, V42A, R44A and I47A were purified by nickel-affinity chromatography and tested for binding to a 5' end-labeled 16 nucleotide RNA probe, which contained a high-affinity binding site (Wang, Dubey *et al.*, 2005). We chose this minimal RNA sequence for the binding studies, rather than using a native transcript containing multiple CsrA target sites, in order to simplify the analyses and focus on direct protein-RNA binding, while attempting to avoid cooperativity effects. Binding of wild

type CsrA and the 8 mutant proteins to the small RNA oligomer led to the formation of shifted complexes (Fig. 2-3*A-I*, *left panels*). Binding curves, Hill coefficients (n) and apparent equilibrium binding constants (K_d) for the wild type and mutant proteins (Fig. 2-3*A-I*, *right panels*) were calculated as described in the EXPERIMENTAL PROCEDURES. The affinities of these mutant proteins for RNA were substantially decreased. Their apparent K_d values were increased from 10- to 150-fold with respect to the wild type CsrA protein and their affinities for RNA were, in the order from highest to lowest: wild-type > R7A > R6A > L4A > V40A > I47A > L2A > V42A > R44A. In addition, the observed binding affinities of these proteins were roughly correlated with their abilities to regulate gene expression *in vivo* (Fig. 2-4). The cooperativity of binding (n) was also increased for all mutant proteins tested as compared with wild type. Currently, we cannot account for this difference in cooperativity.

Mapping critical amino acid residues of CsrA onto a three-dimensional model─Critical regions 1 and 2 were mapped to the recently published, high resolution 3-D structure of CsrA (RsmA) from *Yersinia enterocolitica* (Fig. 2-5)(Heeb, Kuehne *et al.*, 2006). CsrA from *Escherichia* and *Yersinia* are 95% identical, differing only at amino acids 58-60, which have not been structurally defined in any model (Gutierrez, Li *et al.*, 2005; Rife, Schwarzenbacher *et al.*, 2005; Heeb, Kuehne *et al.*, 2006) and which did not affect regulation (Fig. 2-1). The important residues of region 1 were all located within the first β-strand (β_1) of the protein. Region 2 residues were distributed from T_3 through β_5 to the beginning of the α helix (Fig. 2-5, *A* and *C* and Fig. 2-6, "Secondary Structure"). As revealed by the crystal structure, strand β_1 of one CsrA monomer is located adjacent and parallel to strand β_5 from the other monomer (Fig. 2-5, *A* and *C*). This suggests that

region 1 of one monomer and region 2 of the other monomer together define a functional subdomain and alanine substitutions created in either strand alter the same critical area of the dimer. The side chains from L2, L4, V40 and V42 all appear in hydrophobic surroundings (Fig. 2-5, *F* and *G*). Mutation of these residues likely compromise the core by creating space that results in improper folding or affects solvent access. Specifically, the side chain of V40 is directed into the hydrophobic core (Fig. 2-5*G*) while L2, L4 and V42 form a separate but adjacent hydrophobic pocket (Fig. 2-5*F*) that may stabilize the interactions between β_1', β_2' and β_5 (and β_1, β_2 and β_5'). The solvent-exposed side chains of R7, I47 and R44 are all oriented in the same direction on each side of the dimer and do not appear to be involved in intra or intermolecular contacts (Fig. 2-5, *D* and *E*). R6 and E46 are also exposed to solvent and appear to be connected via a salt bridge (data not shown), as suggested previously (Heeb, Kuehne *et al.*, 2006).

Other alanine substitutions outside of region 1 and 2, which compromised (T11A, L12A, I14A and V22A) or improved (T19A and N35A) regulation were also mapped onto the structural model (Fig. 2-5, *D* and *G*). Residues I14 and V22 clearly point into and stabilize the hydrophobic core Fig. 2-5*G*). Leucine-12 (Fig. 2-5*G*) is positioned at the boundary between the core proper and the extended hydrophobic pocket created by L2, L4 and V40 (Fig. 2-5*F*). According to the 3-D model, T11 is solvent exposed and potentially makes a polar contact with T21 (Fig. 2-5*D*). Additionally, the polar side chain of T19 is solvent exposed and is located directly adjacent to N35 (Fig. 2-5*D*). These two residues appear to interact with each other both directly and through a water molecule (data not shown). At present, it is difficult to explain why alanine substitutions of the latter residues (T19A and N35A) should cause a gain of function.

Sequence alignment of critical residues across thirty representative species─Orthologs of CsrA are well represented in the Proteobacteria, but are also known from the Actinobacteria, Thermotogae, Planctomycetes, Spirochaetes and Firmicutes (Bateman, Coin *et al.*, 2004). A sequence alignment of 30 CsrA representatives, including at least 1 member from each of the 6 Phyla listed above, is shown in Fig. 2-6. Among the representative sequences listed here and in the collection of 126 CsrA orthologs at the Sanger Pfam alignment database (http://pfam.wustl.edu/cgi-bin/getdesc?name=CsrA), 14 CsrA amino acid residues (23%) are ≥ 80% identical across all species. Conservation analysis was performed on the CsrA alignment using the algorithm AMAS (Analysis of Multiply Aligned Sequences) (Livingstone and Barton, 1993), as it is implemented in the JalView Sequence alignment viewer (Clamp, Cuff *et al.*, 2004) (Fig. 2-6, "Conservation"). AMAS grades positional physio-chemical amino acid conservation on a scale from 0-10 where 10 represents complete conservation. Based on this analysis, 21 (34%) of the CsrA residues aligned here and 23 (38%) residues in the Pfam CsrA alignment database are positionally identical (scored 10). Highly conserved residues (scoring 7-9) were distributed throughout the CsrA primary structure, but were especially concentrated at the extreme N-terminus, residues 2-7. A majority of the identical or highly conserved residues found between regions 1 and 2 (i.e. β_2-β_4) are located within constrained environments where they may contribute to the stability of the 3-D structure. These positions show a large number of neutral substitutions (i.e., L→I or V→L) (Kimura, 1979). These residues include L12, I14, V18, V20, V22, V25, V30, I32, I34, and A36, which point into the hydrophobic core, and G15 and P37 which are found within loops 2 and 4, respectively.

The two regions that were critical for regulation and RNA binding are highlighted in Fig. 2-6 (black boxes) and a composite graph ("*In vivo* regulatory defect") allows a direct comparison of sequence conservation and these data. There was a strong correlation between amino acids that were identical or highly conserved in regions 1 and 2 and those that were important for CsrA function. The residues labeled region 1, found entirely within the first β-strand, were given an AMAS score of 7 or above. This value increased to 9+ when comparing the entire Pfam group of 126 (data not shown). Specifically, positions 3 and 4 contained neutral substitutions to aliphatic amino acids (I, V or L), while positions 5 and 7 encoded polar (S or T) and basic residues (K or R), respectively. Both L2 and R6 were 100% identical and L4 was ≥ 96% identical across all species in the database. Unlike region 1, every residue within region 2 was not highly conserved; however, amino acids that proved to be critical for CsrA function (V40, V42, R44, E46 and I47) scored 8 or above by AMAS. Both R44 and E46 were ≥ 98% identical across all species in the Pfam database (data not shown). Positions V40, V42 and I47 were between 63-73% identical and contained mostly neutral substitutions to aliphatic amino acids.

DISCUSSION

The present study has taken advantage of recent advances in our understanding of CsrA structure in order to define its structure-function relationship. CsrA represents a novel class of RNA-binding protein, and our experiments define a new type of subdomain for RNA interaction and regulation of gene expression. The CsrA homodimer contains two of these functional subdomains, located on opposite sides of the protein. Each subdomain is composed of primary sequences (regions 1 and 2) contributed by each

polypeptide chain of the dimer (Fig. 2-5, *A* and *C*). The β-strands of these two regions are positioned adjacent and parallel to each other within the dimer. As discussed below, this structure appears to account for the stoichiometry of CsrA within the CsrA-CsrB ribonucleoprotein complex, which heretofore could not be explained.

A careful examination of the 3-D structure of CsrA permitted predictions of the manner in which specific alanine substitutions affected protein function. The residues most likely to interact directly with RNA are R7, R44 and I47. The side chains of these critical amino acids are solvent exposed and are not associated with other residues in the CsrA dimer. They are clustered on each side of the protein within areas of positive surface potential (Fig. 2-5*B*) forming two possible RNA binding surfaces (Fig. 2-5, *D* and *E*). R44A substitution caused the greatest defect in RNA-binding and eliminated regulation by CsrA. These results were consistent with a previous study (Heeb, Kuehne *et al.*, 2006) demonstrating that this amino acid was important for RsmA mediated regulation of several *Pseudomonas aeruginosa* phenotypes as well as RNA binding. I47A also exhibited low affinity for RNA and retained only ~5% of its ability to regulate gene expression. Of these proteins, R7A showed the mildest RNA binding defect and regulated gene expression at ~30% of wild-type levels. Originally, the side chains of T5 (Heeb, Kuehne *et al.*, 2006), K26, R31, N28 (Gutierrez, Li *et al.*, 2005; Heeb, Kuehne *et al.*, 2006), V30, E10 or Q29 (Gutierrez, Li *et al.*, 2005) were proposed to be components of an RNA binding surface based on their conservation, solvent accessibility, charge or proximity to R44. However, we found no evidence for critical involvement of these residues in regulation of gene expression. These cases highlight the pitfalls of using structural data alone to deduce the important features of a novel protein fold.

The basic residue R6 appears to span the gap between β_1 and β_5' to make an important structural contact with the side chain of E46. This bond probably stabilizes the relative positions of β_1 and β_5', in turn correctly orienting R7, R44 and I47. We cannot eliminate the possibility that R6 might also contact RNA directly. The importance of the R6-E46 interaction is also supported by the finding that these amino acids are ≥ 98% identical across the 126 species listed in Pfam. Mutation of R6 to alanine probably abolishes this critical inter-strand connection, perhaps increasing the flexibility of β_5' and β_1 and altering the hydrophobic pocket located around L2, L4 and V42. It is interesting to note that R6 is the main outlier when comparing RNA binding affinity with *in vivo* regulation (Fig. 2-4). Thus, relative to other substitutions, R6A was able to bind RNA with a greater affinity than its *in vivo* function would predict. This protein also accumulated ~2-fold greater than wild type (Fig. 2-2). A possible explanation for both observations is that this protein might form protease-resistant aggregates in the cell, which accumulate but do not participate in regulation *in vivo*. The tendency of CsrA to aggregate has been documented since the protein was first purified (Liu, Gui *et al.*, 1997; Gutierrez, Li *et al.*, 2005).

The extended hydrophobic pocket formed by L2, L4 and V42 that was disrupted by alanine substitution may serve as an RNA binding surface. There is precedence for this type of mechanism; human splicing factor protein U1A binds nucleic acids via variable loops, but also creates an enlarged hydrophobic core when it packs RNA bases against non-polar residues found on a conserved β-sheet (Allain, Howe *et al.*, 1997). It is not uncommon for non-polar amino acids to participate in nucleic acid binding, and occasionally hydrophobic interactions are the central mechanism of association. For

instance, Nova-2 KH3 binds RNA via an extended α/β platform that is composed entirely of hydrophobic amino acids. Interestingly, both human U1A, which recognizes RNA structure and charge distribution, and Nova-2 KH3 preferentially bind RNA stem-loops, similar to CsrA (Allain, Howe *et al.*, 1997; Lewis, Musunuru *et al.*, 2000).

Not every amino acid residue within the two critical regions of CsrA was vital for regulation of gene expression (e.g. T5 and S41 of regions 1 and 2, respectively), and some residues were more important than others (R7A exhibited the mildest regulatory defect, R44A the most severe). Interestingly, the majority of the important residues among the region 2 mutations exhibited an alternating pattern, consistent with their location within a β-strand. Residues directed to one side of the strand were critical (V40, V42, R44 and E46) while others (E39, S41, H43 and E45) were not.

While the previous X-ray and NMR analyses suggested similar structures for the three CsrA orthologs, all of these proteins contained amino-terminal modifications and none was documented to be biologically active (Gutierrez, Li *et al.*, 2005; Rife, Schwarzenbacher *et al.*, 2005; Heeb, Kuehne *et al.*, 2006). Studies with N-terminally-tagged CsrA derivatives (T. Romeo, data not shown) and results from the current study demonstrate that the N-terminus of CsrA is critical for its function. The N-terminal 4 amino acids in almost every species are hydrophobic (Fig. 2-6) and the N-terminal fusions, which introduced charged or polar residues almost certainly affect function (Cunningham and Wells, 1989). In fact, modeling of the *Yersinia enterocolitica* and *Pseudomonas aeruginosa* CsrA (PDB accession code 1VPZ) structures revealed 2 and 3 possible non-native hydrogen bonds, respectively, between the backbones of β_1 and β_4' in proteins containing these tags (data not shown). Thus, the structure of the native protein

around the important β_1, β_4'-T-β_5' region may be slightly different than the existing models.

The finding that a symmetric CsrA dimer contains two functional subdomains on opposite sides of the protein may explain the longstanding observation that CsrB RNA, which contains ~18 CsrA target sequences, binds to ~18 CsrA subunits (9 functional dimers) to form a globular complex (Liu, Gui *et al.*, 1997). Due to the flexibility of RNA, CsrA could theoretically bridge two target sites, e.g. within two hairpin loops. Additional studies will be required to clearly define the precise stoichiometry of this and other CsrA-RNA complexes. Furthermore, CsrA itself may react to RNA in an adaptable, plastic manner, as suggested by solution NMR studies, which found that almost every amide signal of CsrA exhibits a shift upon addition of RNA (Gutierrez, Li *et al.*, 2005). The dimeric structure of CsrA also dictates that for each mutation that was constructed, two alanine substitutions resulted in the homodimer. While pairs of substitutions may serve to further destabilize the core when they involve an interior hydrophobic residue, those that affect CsrA-RNA contact may actually have removed some ambiguity from the results by decreasing RNA binding affinity simultaneously at both subdomains.

RNA binding proteins utilize several distinct secondary structures to construct a wide array of tertiary folds that mediate protein-RNA contact (Siomi and Dreyfuss, 1997; Varani and Nagai, 1998; Theobald, Mitton-Fry *et al.*, 2003; Stefl, Skrisovska *et al.*, 2005). Detailed structural information exists for protein-RNA complexes formed by several motif classes including the RRM (RNA recognition motif) (Maris, Dominguez *et al.*, 2005; Stefl, Skrisovska *et al.*, 2005), KH (Messias and Sattler, 2004), Sm-like

(Wilusz and Wilusz, 2005), and OB fold (Theobald, Mitton-Fry *et al.*, 2003). The 5β-1α configuration of CsrA, however, does not fit within any of these protein families. The RRM is a 4β-2α structure where the central two strands in a 4-stranded anti-parallel sheet are involved in RNA binding (Maris, Dominguez *et al.*, 2005); the OB fold is characterized by a flattened β-barrel composed of 2 β-sheets where β-strands 2 and 3 interact with nucleic acid (Theobald, Mitton-Fry *et al.*, 2003). In both the RRM and OB a β-sheet comprises the primary RNA binding surface displaying aromatic side chains involved in intermolecular stacking. However, the OB-fold also stabilizes the RNA-protein complex via packing interactions with nonpolar residues and the aliphatic regions of polar side chains (Oubridge, Ito *et al.*, 1994; Theobald, Mitton-Fry *et al.*, 2003; Messias and Sattler, 2004). The CsrA functional region defined in this study contains hydrophobic, acidic and basic, but not aromatic residues, which appear to be critical for binding and biological function.

The number and arrangement of RNA recognition motifs in a protein is critical for proper function. Such motifs can function individually (e.g., human HNRNPC and U1-70K) or in multiples (e.g., *Drosophila* ELAV and human U2B") (Adam, Nakagawa *et al.*, 1986; Varani and Nagai, 1998). Tandem motifs can bind single stranded RNA across the face of both motifs, as exemplified by poly(A)-binding protein (Deo, Bonanno *et al.*, 1999), or within clefts formed between the motifs, as in *Drosophila* Sxl (Handa, Nureki *et al.*, 1999). The former type of protein-RNA relationship is also seen for the *T. thermophilus* SerRS dimer, where a single tRNA binds across the surface of both protein subunits (Biou, Yaremchuk *et al.*, 1994), and for the *Bacillus subtilis* TRAP undecamer, a toroid-shaped protein that contacts each of a series of RNA triplet repeats in the *trp*

leader transcript using sequential subunit interfaces around its circumference (Antson, Dodson *et al.*, 1999; Gollnick, Babitzke *et al.*, 2005). CsrA may employ its two distinct subdomains simultaneously in binding to RNA. Ultimately, a CsrA-RNA co-crystal or solution NMR structure will be necessary to define the critical contacts between CsrA and its RNA substrates and would provide new insights into protein-RNA molecular interactions.

ACKNOWLEDGEMENTS

We thank Alex Yakhnin for technical assistance and help with binding equations.

REFERENCES

Adam, S. A., T. Nakagawa, M. S. Swanson, T. K. Woodruff and G. Dreyfuss (1986). "mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence." Mol Cell Biol **6**(8): 2932-43.

Allain, F. H., P. W. Howe, D. Neuhaus and G. Varani (1997). "Structural basis of the RNA-binding specificity of human U1A protein." Embo J **16**(18): 5764-72.

Altier, C., M. Suyemoto and S. D. Lawhon (2000). "Regulation of *Salmonella enterica* serovar Typhimurium invasion genes by *csrA*." Infect Immun **68**(12): 6790-7.

Altier, C., M. Suyemoto, A. I. Ruiz, K. D. Burnham and R. Maurer (2000). "Characterization of two novel regulatory genes affecting *Salmonella* invasion gene expression." Mol Microbiol **35**(3): 635-46.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.

Antson, A. A., E. J. Dodson, G. Dodson, R. B. Greaves, X. Chen and P. Gollnick (1999). "Structure of the *trp* RNA-binding attenuation protein, TRAP, bound to RNA." Nature **401**(6750): 235-42.

Baase, W. A., A. E. Eriksson, X. J. Zhang, D. W. Heinz, U. Sauer, M. Blaber, E. P. Baldwin, J. A. Wozniak and B. W. Matthews (1992). "Dissection of protein structure and folding by directed mutagenesis." Faraday Discuss(93): 173-81.

Baker, C. S., I. Morozov, K. Suzuki, T. Romeo and P. Babitzke (2002). "CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*." Mol Microbiol **44**(6): 1599-610.

Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats and S. R. Eddy (2004). "The Pfam protein families database." Nucleic Acids Res **32**(Database issue): D138-41.

Biou, V., A. Yaremchuk, M. Tukalo and S. Cusack (1994). "The 2.9 A crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA(Ser)." Science **263**(5152): 1404-10.

Boyd, D., D. S. Weiss, J. C. Chen and J. Beckwith (2000). "Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system." J Bacteriol **182**(3): 842-7.

Clamp, M., J. Cuff, S. M. Searle and G. J. Barton (2004). "The Jalview Java alignment editor." Bioinformatics **20**(3): 426-7.

Cui, Y., A. Chatterjee, Y. Liu, C. K. Dumenyo and A. K. Chatterjee (1995). "Identification of a global repressor gene, *rsmA*, of *Erwinia carotovora* subsp. *carotovora* that controls extracellular enzymes, N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting *Erwinia* spp." J Bacteriol **177**(17): 5108-15.

Cunningham, B. C. and J. A. Wells (1989). "High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis." Science **244**(4908): 1081-5.

Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." Proc Natl Acad Sci U S A **97**(12): 6640-5.

De Guzman, R. N., R. B. Turner and M. F. Summers (1998). "Protein-RNA recognition." Biopolymers **48**(2-3): 181-95.

Delano, W. L. (2002). The PyMol Molecular Graphics System. San Carlos, CA, USA, DeLano Scientific.

Deo, R. C., J. B. Bonanno, N. Sonenberg and S. K. Burley (1999). "Recognition of polyadenylate RNA by the poly(A)-binding protein." Cell **98**(6): 835-45.

Dubey, A. K., C. S. Baker, K. Suzuki, A. D. Jones, P. Pandit, T. Romeo and P. Babitzke (2003). "CsrA regulates translation of the *Escherichia coli* carbon starvation gene, *cstA*, by blocking ribosome access to the *cstA* transcript." J Bacteriol **185**(15): 4450-60.

Eiler, S., A. Dock-Bregeon, L. Moulinier, J. C. Thierry and D. Moras (1999). "Synthesis of aspartyl-tRNA(Asp) in *Escherichia coli*--a snapshot of the second step." Embo J **18**(22): 6532-41.

Fortune, D. R., M. Suyemoto and C. Altier (2006). "Identification of CsrC and characterization of its role in epithelial cell invasion in *Salmonella enterica* serovar Typhimurium." Infect Immun **74**(1): 331-9.

Gollnick, P., P. Babitzke, A. Antson and C. Yanofsky (2005). "Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis*." Annu Rev Genet **39**: 47-68.

Griffith, K. L. and R. E. Wolf, Jr. (2002). "Measuring beta-galactosidase activity in bacteria: cell growth, permeabilization, and enzyme assays in 96-well arrays." Biochem Biophys Res Commun **290**(1): 397-402.

Gudapaty, S., K. Suzuki, X. Wang, P. Babitzke and T. Romeo (2001). "Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*." J Bacteriol **183**(20): 6017-27.

Gutierrez, P., Y. Li, M. J. Osborne, E. Pomerantseva, Q. Liu and K. Gehring (2005). "Solution structure of the carbon storage regulator protein CsrA from *Escherichia coli*." J Bacteriol **187**(10): 3496-501.

Handa, N., O. Nureki, K. Kurimoto, I. Kim, H. Sakamoto, Y. Shimura, Y. Muto and S. Yokoyama (1999). "Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein." Nature **398**(6728): 579-85.

Heeb, S., S. A. Kuehne, M. Bycroft, S. Crivii, M. D. Allen, D. Haas, M. Camara and P. Williams (2006). "Functional analysis of the post-transcriptional regulator RsmA reveals a novel RNA-binding site." J Mol Biol **355**(5): 1026-36.

Heurlier, K., F. Williams, S. Heeb, C. Dormond, G. Pessi, D. Singer, M. Camara, P. Williams and D. Haas (2004). "Positive control of swarming, rhamnolipid synthesis, and lipase production by the posttranscriptional RsmA/RsmZ system in *Pseudomonas aeruginosa* PAO1." J Bacteriol **186**(10): 2936-45.

Jackson, D. W., K. Suzuki, L. Oakford, J. W. Simecka, M. E. Hart and T. Romeo (2002). "Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*." J Bacteriol **184**(1): 290-301.

Jovine, L., C. Oubridge, J. M. Avis and K. Nagai (1996). "Two structurally different RNA molecules are bound by the spliceosomal protein U1A using the same recognition strategy." Structure **4**(5): 621-31.

Kelly, P. A., L. Goujon, A. Sotiropoulos, H. Dinerstein, N. Esposito, M. Edery, J. Finidori and M. C. Postel-Vinay (1994). "The GH receptor and signal transduction." Horm Res **42**(4-5): 133-9.

Kimura, M. (1979). "The neutral theory of molecular evolution." Sci Am **241**(5): 98-100, 102, 108 passim.

Lenz, D. H., M. B. Miller, J. Zhu, R. V. Kulkarni and B. L. Bassler (2005). "CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*." Mol Microbiol **58**(4): 1186-202.

Lewis, H. A., K. Musunuru, K. B. Jensen, C. Edo, H. Chen, R. B. Darnell and S. K. Burley (2000). "Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome." Cell **100**(3): 323-32.

Liu, M. Y., G. Gui, B. Wei, J. F. Preston, 3rd, L. Oakford, U. Yuksel, D. P. Giedroc and T. Romeo (1997). "The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*." J Biol Chem **272**(28): 17502-10.

Liu, M. Y. and T. Romeo (1997). "The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein." J Bacteriol **179**(14): 4639-42.

Liu, Y., Y. Cui, A. Mukherjee and A. K. Chatterjee (1998). "Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. *carotovora* that controls production of extracellular enzymes and secondary metabolites." Mol Microbiol **29**(1): 219-34.

Livingstone, C. D. and G. J. Barton (1993). "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation." Comput Appl Biosci **9**(6): 745-56.

MacLennan, D. H., W. J. Rice, A. Odermatt and N. M. Green (1998). "Structure-function relationships in the Ca(2+)-binding and translocation domain of SERCA1: physiological correlates in Brody disease." Acta Physiol Scand Suppl **643**: 55-67.

Maris, C., C. Dominguez and F. H. Allain (2005). "The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression." Febs J **272**(9): 2118-31.

Messias, A. C. and M. Sattler (2004). "Structural basis of single-stranded RNA recognition." Acc Chem Res **37**(5): 279-87.

Miller, J. H. (1972). Experiments in molecular genetics. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Molofsky, A. B. and M. S. Swanson (2003). "*Legionella pneumophila* CsrA is a pivotal repressor of transmission traits and activator of replication." Mol Microbiol **50**(2): 445-61.

Norton, R. S., M. W. Pennington and H. Wulff (2004). "Potassium channel blockade by the sea anemone toxin ShK for the treatment of multiple sclerosis and other autoimmune diseases." Curr Med Chem **11**(23): 3041-52.

Oubridge, C., N. Ito, P. R. Evans, C. H. Teo and K. Nagai (1994). "Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin." Nature **372**(6505): 432-8.

Rife, C., R. Schwarzenbacher, D. McMullan, P. Abdubek, E. Ambing, H. Axelrod, T. Biorac, J. M. Canaves, H. J. Chiu, A. M. Deacon, M. DiDonato, M. A. Elsliger, A. Godzik, C. Grittini, S. K. Grzechnik, J. Hale, E. Hampton, G. W. Han, J. Haugen, M. Hornsby, L. Jaroszewski, H. E. Klock, E. Koesema, A. Kreusch, P. Kuhn, S. A. Lesley, M. D. Miller, K. Moy, E. Nigoghossian, J. Paulsen, K. Quijano, R. Reyes, E. Sims, G. Spraggon, R. C. Stevens, H. van den Bedem, J. Velasquez, J. Vincent, A. White, G. Wolf, Q. Xu, K. O. Hodgson, J. Wooley and I. A. Wilson (2005). "Crystal structure of the global regulatory protein CsrA from *Pseudomonas putida* at 2.05 A resolution reveals a new fold." Proteins **61**(2): 449-53.

Romeo, T., M. Gong, M. Y. Liu and A. M. Brun-Zinkernagel (1993). "Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties." J Bacteriol **175**(15): 4744-55.

Sabnis, N. A., H. Yang and T. Romeo (1995). "Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene *csrA*." J Biol Chem **270**(49): 29096-104.

Sasaki, N. and K. Sutoh (1998). "Structure-mutation analysis of the ATPase site of *Dictyostelium discoideum* myosin II." Adv Biophys **35**: 1-24.

Silhavy, T. J., M. L. Berman, L. W. Enquist and Cold Spring Harbor Laboratory. (1984). Experiments with gene fusions. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Siomi, H. and G. Dreyfuss (1997). "RNA-binding proteins as regulators of gene expression." Curr Opin Genet Dev **7**(3): 345-53.

Smith, P. K., R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson and D. C. Klenk (1985). "Measurement of protein using bicinchoninic acid." Anal Biochem **150**(1): 76-85.

Stefl, R., L. Skrisovska and F. H. Allain (2005). "RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle." EMBO Rep **6**(1): 33-8.

Symmons, M. F., G. H. Jones and B. F. Luisi (2000). "A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation." Structure **8**(11): 1215-26.

Theobald, D. L., R. M. Mitton-Fry and D. S. Wuttke (2003). "Nucleic acid recognition by OB-fold proteins." Annu Rev Biophys Biomol Struct **32**: 115-33.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence

weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.

Towbin, H., T. Staehelin and J. Gordon (1979). "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications." Proc Natl Acad Sci U S A **76**(9): 4350-4.

Valverde, C., M. Lindell, E. G. Wagner and D. Haas (2004). "A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of *Pseudomonas fluorescens*." J Biol Chem **279**(24): 25066-74.

van den Worm, S. H., N. J. Stonehouse, K. Valegard, J. B. Murray, C. Walton, K. Fridborg, P. G. Stockley and L. Liljas (1998). "Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments." Nucleic Acids Res **26**(5): 1345-51.

Varani, G. and K. Nagai (1998). "RNA recognition by RNP proteins during RNA processing." Annu Rev Biophys Biomol Struct **27**: 407-45.

Wang, X., A. K. Dubey, K. Suzuki, C. S. Baker, P. Babitzke and T. Romeo (2005). "CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*." Mol Microbiol **56**(6): 1648-63.

Wei, B. L., A. M. Brun-Zinkernagel, J. W. Simecka, B. M. Pruss, P. Babitzke and T. Romeo (2001). "Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*." Mol Microbiol **40**(1): 245-56.

Weilbacher, T., K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke and T. Romeo (2003). "A novel sRNA component of

the carbon storage regulatory system of *Escherichia coli*." Mol Microbiol **48**(3): 657-70.

Wells, J. A., B. C. Cunningham, G. Fuh, H. B. Lowman, S. H. Bass, M. G. Mulkerrin, M. Ultsch and A. M. deVos (1993). "The molecular basis for growth hormone-receptor interactions." Recent Prog Horm Res **48**: 253-75.

Wiesmann, C. and A. M. de Vos (2001). "Nerve growth factor: structure and function." Cell Mol Life Sci **58**(5-6): 748-59.

Wilusz, C. J. and J. Wilusz (2005). "Eukaryotic Lsm proteins: lessons from bacteria." Nat Struct Mol Biol **12**(12): 1031-6.

Yakhnin, A. V., J. J. Trimble, C. R. Chiaro and P. Babitzke (2000). "Effects of mutations in the L-tryptophan binding pocket of the Trp RNA-binding attenuation protein of *Bacillus subtilis*." J Biol Chem **275**(6): 4519-24.

FIGURE LEGEND

Figure 2-1. Phenotypic and regulatory effects of CsrA alanine replacement mutations. Glycogen accumulation, biofilm formation and motility phenotypes, and expression of chromosomally-encoded *glgCA*'-'*lacZ, pgaA*'-'*lacZ* and *flhDC*'-'*lacZ* translational fusions, respectively, were examined in the presence of each alanine-scanning *csrA* mutant. The amino acid residue that was changed to alanine is indicated in each panel. The phenotypic effects were tested in strain TRMG1655 (*csrA::kan*). *A*, *C*, *E*, Expression from a *glgCA*'-'*lacZ* fusion in strain TRKSGA18 (*A*), *pgaA*'-'*lacZ* fusion in strain TRCFPGACAT2321 (*C*), and *flhDC*'-'*lacZ* fusion in strain TRCFFDCAT312 (*E*) are shown. All values are shown relative to the wild type (WT) and vector controls, located to the right. *B*, Intracellular glycogen levels were assessed by staining with iodine vapors after growth on Kornberg media with glucose. *D*, Biofilm formation

was assessed by staining adherent cells with crystal violet, and motility was determined in semi-solid tryptone media (see EXPERIMENTAL PROCEDURES). Values for expression analyses are reported as average ± standard error. All experiments were performed at least twice and each time in triplicate or quadruplicate.

Figure 2-2. Quantitative immunoblotting of TRMG1655 (*csrA::kan*) expressing wild type and 58 alanine-scanning mutant CsrA proteins. *E. coli* whole cell lysates were examined by SDS-PAGE and western analysis, using a nickel conjugated HRP to detect CsrA-His_6 (see EXPERIMENTAL PROCEDURES). CsrA mutant protein levels were measured by densitometry and are reported as a percent of wild type (WT). Experiments were repeated at least twice and values are reported as average ± standard error.

Figure 2-3. Gel mobility shift assays for the binding of wild type and mutant CsrA proteins to a 16 nucleotide high-affinity RNA target. 5'end-labeled RNA (30 pM) was incubated with increasing concentrations of CsrA, shown below each lane. The positions of bound (B) and free (F) RNA are indicated. *A-I*, Graphs to the right of each gel shift display the best-fit curve, apparent equilibrium binding constant (K_d), and Hill coefficient (n) for each protein (see EXPERIMENTAL PROCEDURES). *J*, Graphical comparison of CsrA mutant best fit binding curves. Graphpad Prism version 4 for Mac (San Diego, CA) software was used for calculations and graphical displays.

Figure 2-4. Correlation between the "*in vivo* regulatory defect" (Fig. 2-6) and the equilibrium binding constant (K_d) (Fig. 2-3) of 8 CsrA mutant proteins. The correlation coefficient (R^2) of a linear best fit comparison of these values was 0.63.

Figure 2-5. Critical residues of CsrA mapped onto the 3-D model of *Yersinia enterocolitica* CsrA. *A*, Ribbon diagram of CsrA highlighting functional regions 1 and 2 (red) identified in the present study. The individual CsrA monomers are colored white and green. The second functional region found on the opposite side of the homodimer is not highlighted. *B*, Identical view angle of CsrA, as shown in panel *A*, displaying electrostatic surface potential. Electropositive areas are represented in blue while electronegative is in red. *C*, CsrA displaying critical residues found in regions 1(L2, L4, R6, R7) and 2 (V40, V42, R44 and I47). The corresponding critical residues found on the opposite side of the protein are not illustrated. *D*, Front view of a semi-transparent CsrA displaying residues that are likely involved directly in RNA-binding (R44, I47 and R7), and several residues found outside region 1 and 2 (T11, T19, T21 and N35) that affect regulation. *E*, Bottom up view of the likely RNA binding residues listed in panel *D*. *F*, *G*, Top down view illustrating amino acids that compose a separate but adjacent hydrophobic pocket (L2, L4, and V42) (*F*) and those that point inward and stabilize the CsrA hydrophobic core (L12, I14, V22, V40) (*G*).

Figure 2-6. Comparison of CsrA orthologs from 6 Phyla, highlighting regions that were crucial for RNA binding and *in vivo* regulation with predicted amino acid functions. Secondary structure and residue conservation are displayed above and below the aligned sequences, respectively. The "*In vivo* regulatory defect" for each mutant was calculated from an equal weighted mean of the gene expression data from the three reporter fusions (Fig. 2-1).

Mutant CsrA proteins that were experimentally tested for RNA binding, CsrA primary structure that was important for function, and functional predictions for individual amino acids (P) are as shown in the legend at the bottom of this figure. Conservation analysis was performed as described in the "RESULTS" using the algorithm AMAS (a score of 10 indicates complete conservation) and displayed using JalView (Clamp, Cuff *et al.*, 2004). CsrA protein sequences were collected from the Sanger Pfam database (Bateman, Coin *et al.*, 2004) and aligned with ClustalW (Thompson, Higgins *et al.*, 1994). (*) represents positions that show 100% identity, and (+) signifies an AMAS score of 10.

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

Strain, plasmid or bacteriophage	Description	Source or reference
E. coli K12 Strains		
MG1655	Prototrophic	Michael Cashel
TR1-5 MG1655[a]	MG1655 *csrA*::*kan*	(Romeo, Gong *et al.*, 1993)
CF7789	MG1655 Δ*lacI-Z* (*Mlu*I)	Michael Cashel
CAGKSGA18	CF7789 *zfh*-3131::Tn*10* Φ(*glgA*::*lacZ*) (λ*plac*Mu15)	(Gudapaty, Suzuki *et al.*, 2001)
CFFDCAT312	CF7789 Δ(*att-lom*)::*bla*::*cam* Φ(*flhDC'-'lacZ*)6(Hyb) Amp^S Kan^S Cam^R	This study
CFPGACAT2321	CF7789 Δ(*att-lom*)::*bla*::*cam* Φ(*pgaA'-'lacZ*)4(Hyb) Amp^S Kan^S Cam^R	This study
RGTWMG1655	MG1655 Δ*csrB::cam* Δ*csrC::tet*	(Weilbacher, Suzuki *et al.*, 2003)
$TRCFB^-C^-$	CF7789 *csrA*::*kan* Δ*csrB::cam* Δ*csrC::tet*	This study
Plasmids		
pKK223-3	Expression vector under IPTG control; Amp^R	Pharmacia Corp.
pMLB1034	*'laZ* translational fusion vector; Amp^R	(Silhavy, Berman *et al.*, 1984)

pCSRH6-19	pKK223-3 expressing C-terminal His_6-tag CsrA; Amp^R	(Liu, Gui *et al.*, 1997)
pFDCZ6	pMLB1034 Φ(*flhDC'-'lacZ*); Amp^R	(Wei, Brun-Zinkernagel *et al.*, 2001)
pPGAZ4	pMLB1034 Φ(*pgaA'-'lacZ*); Amp^R	(Wang, Dubey *et al.*, 2005)
pFDCZ6CAT	pFDCZ6 Δ*bla*::*cat*	This study
pPGAZ4CAT	pPGAZ4 Δ*bla*::*cat*	This study
pKD3	Source of *cat* gene	(Datsenko and Wanner, 2000)
58 alanine mutant CsrA-His_6 expressing plasmids (pCSRL2A through pCSRY61A) were constructed, all named according to the following nomenclature:		
pCSRX#A	*csrA* X#A C-terminal His_6 tag, derived from pCSRH6-19	This study
Where X is the original CsrA amino acid and # is the residue position. The first 2 mutants are listed below as examples.		
pCSRL2A	*csrA* L2A C-terminal His_6 tag, derived from pCSRH6-19	This study
pCSRI3A	*csrA* I3A C-terminal His_6 tag, derived from pCSRH6-19	This study
Bacteriophage		
P1*vir*	Strictly lytic P1	Carol Gross
λInCh1	For genomic insertions; Kan^R	(Boyd, Weiss *et al.*, 2000)

[a]Strain designations containing the prefix TR indicate that the mutant allele *csrA*::*kan* was introduced by P1*vir* transduction.

Table 2-2. Oligonucleotide primers used in this study

Primer name	Primer sequence, 5' to 3'
Conversion of pFDCZ6 and pPGAZ4 from Amp^R to Cam^R	
pKD3 PstIL	CTTCTGCAGGTGTAGGCTGGAGCTGCTTC

pKD3 ScaIR	GATGCAAGTACTCATATGAATATCCTCCTTAGTT

Sequence analysis of CsrA-His_6 site-directed mutants

csrAseq-2L	GAGCGGATAACAATTTCACACA
csrAseq-2R	TTCTCTCATCCGCCAAAACA

Site-directed mutagenesis of CsrA-His_6

csrA-L2A-L	GGAAACAGAATTCCCATGGCGATTCTGACTCGTCGAGTTGG
csrA-L2A-R	CCAACTCGACGAGTCAGAATCGCCATGGGAATTCTGTTTCC
csrA-I3A-L	GGAAACAGAATTCCCATGCTGGCGCTGACTCGTCGAGTTGG
csrA-I3A-R	CCAACTCGACGAGTCAGCGCCAGCATGGGAATTCTGTTTCC
csrA-L4A-L	CAGAATTCCCATGCTGATTGCGACTCGTCGAGTTGGTG
csrA-L4A-R	CACCAACTCGACGAGTCGCAATCAGCATGGGAATTCTG
csrA-T5A-L	CCATGCTGATTCTGGCGCGTCGAGTTGGTGAGACC
csrA-T5A-R	GGTCTCACCAACTCGACGCGCCAGAATCAGCATGG
csrA-R6A-L	CCATGCTGATTCTGACTGCGCGAGTTGGTGAGACCCTCATG
csrA-R6A-R	CATGAGGGTCTCACCAACTCGCGCAGTCAGAATCAGCATGG
csrA-R7A-L	GCTGATTCTGACTCGTGCGGTTGGTGAGACCCTCATGATTGG
csrA-R7A-R	CCAATCATGAGGGTCTCACCAACCGCACGAGTCAGAATCAGC
csrA-V8A-L	CTGATTCTGACTCGTCGAGCGGGTGAGACCCTC
csrA-V8A-R	GAGGGTCTCACCCGCTCGACGAGTCAGAATCAG
csrA-G9A-L	CTGACTCGTCGAGTTGCGGAGACCCTCATGATTGG
csrA-G9A-R	CCAATCATGAGGGTCTCCGCAACTCGACGAGTCAG
csrA-E10A-L	GCTGATTCTGACTCGTCGAGTTGGTGCGACCCTCATGATTGG
csrA-E10A-R	CCAATCATGAGGGTCGCACCAACTCGACGAGTCAGAATCAGC

csrA-T11A-L	CGTCGAGTTGGTGAGGCGCTCATGATTGGGGATGAGG
csrA-T11A-R	CCTCATCCCCAATCATGAGCGCCTCACCAACTCGACG
csrA-L12A-L	CGTCGAGTTGGTGAGACCGCGATGATTGGGGATGAGGTC
csrA-L12A-R	GACCTCATCCCCAATCATCGCGGTCTCACCAACTCGACG
csrA-M13A-L	GAGTTGGTGAGACCCTCGCGATTGGGGATGAGGTC
csrA-M13A-R	GACCTCATCCCCAATCGCGAGGGTCTCACCAACTC
csrA-I14A-L	CGAGTTGGTGAGACCCTCATGGCGGGGGATGAGGTCACC
csrA-I14A-R	GGTGACCTCATCCCCCGCCATGAGGGTCTCACCAACTCG
csrA-G15A-L	GTGAGACCCTCATGATTGCGGATGAGGTCACCGTG
csrA-G15A-R	CACGGTGACCTCATCCGCAATCATGAGGGTCTCAC
csrA-D16A-L	GACCCTCATGATTGGGGCGGAGGTCACCGTGACAG
csrA-D16A-R	CTGTCACGGTGACCTCCGCCCCAATCATGAGGGTC
csrA-E17A-L	CCTCATGATTGGGGATGCGGTCACCGTGACAGTTTTAGG
csrA-E17A-R	CCTAAAACTGTCACGGTGACCGCATCCCCAATCATGAGG
csrA-V18A-L	CTCATGATTGGGGATGAGGCGACCGTGACAGTTTTAGG
csrA-V18A-R	CCTAAAACTGTCACGGTCGCCTCATCCCCAATCATGAG
csrA-T19A-L	GGGATGAGGTCGCGGTGACAGTTTTAGG
csrA-T19A-R	CCTAAAACTGTCACCGCGACCTCATCCC
csrA-V20A-L	GGATGAGGTCACCGCGACAGTTTTAGGG
csrA-V20A-R	CCCTAAAACTGTCGCGGTGACCTCATCC
csrA-T21A-L	GGATGAGGTCACCGTGGCGGTTTTAGGGGTAAAGG
csrA-T21A-R	CCTTTACCCCTAAAACCGCCACGGTGACCTCATCC
csrA-V22A-L	GGATGAGGTCACCGTGACAGCGTTAGGGGTAAAGGGCAACC

csrA-V22A-R	GGTTGCCCTTTACCCCTAACGCTGTCACGGTGACCTCATCC
csrA-L23A-L	GGTCACCGTGACAGTTGCGGGGGTAAAGGGCAACC
csrA-L23A-R	GGTTGCCCTTTACCCCCGCAACTGTCACGGTGACC
csrA-G24A-L	GGTCACCGTGACAGTTTTAGCGGTAAAGGGCAACCAGGTACG
csrA-G24A-R	CGTACCTGGTTGCCCTTTACCGCTAAAACTGTCACGGTGACC
csrA-V25A-L	GGTCACCGTGACAGTTTTAGGGGCGAAGGGCAACCAGGTACG
csrA-V25A-R	CGTACCTGGTTGCCCTTCGCCCCTAAAACTGTCACGGTGACC
csrA-K26A-L	CGTGACAGTTTTAGGGGTAGCGGGCAACCAGGTACGTATTGG
csrA-K26A-R	CCAATACGTACCTGGTTGCCCGCTACCCCTAAAACTGTCACG
csrA-G27A-L	CGTGACAGTTTTAGGGGTAAAGGCGAACCAGGTACGTATTGG
csrA-G27A-R	CCAATACGTACCTGGTTCGCCTTTACCCCTAAAACTGTCACG
csrA-N28A-L	GTTTTAGGGGTAAAGGGCGCGCAGGTACGTATTGGCGTAAATGC
csrA-N28A-R	GCATTTACGCCAATACGTACCTGCGCGCCCTTTACCCCTAAAAC
csrA-Q29A-L	GTTTTAGGGGTAAAGGGCAACGCGGTACGTATTGGCGTAAATGC
csrA-Q29A-R	GCATTTACGCCAATACGTACCGCGTTGCCCTTTACCCCTAAAAC
csrA-V30A-L	GGTAAAGGGCAACCAGGCGCGTATTGGCGTAAATGCC
csrA-V30A-R	GGCATTTACGCCAATACGCGCCTGGTTGCCCTTTACC
csrA-R31A-L	GGCAACCAGGTAGCGATTGGCGTAAATGCCCCGAAGG
csrA-R31A-R	CCTTCGGGGCATTTACGCCAATCGCTACCTGGTTGCC
csrA-I32A-L	GGCAACCAGGTACGTGCGGGCGTAAATGCCCCG
csrA-I32A-R	CGGGGCATTTACGCCCGCACGTACCTGGTTGCC
csrA-G33A-L	GCAACCAGGTACGTATTGCGGTAAATGCCCCGAAGG
csrA-G33A-R	CCTTCGGGGCATTTACCGCAATACGTACCTGGTTGC

csrA-V34A-L	CCAGGTACGTATTGGCGCGAATGCCCCGAAGG
csrA-V34A-R	CCTTCGGGGCATTCGCGCCAATACGTACCTGG
csrA-N35A-L	CAGGTACGTATTGGCGTAGCGGCCCCGAAGGAAGTTTC
csrA-N35A-R	GAAACTTCCTTCGGGGCCGCTACGCCAATACGTACCTG
csrA-P37A-L	CGTATTGGCGTAAATGCCGCCAAGGAAGTTTCTGTTCACC
csrA-P37A-R	GGTGAACAGAAACTTCCTTGGCGGCATTTACGCCAATACG
csrA-K38A-L	GCGTAAATGCCCCGGCGGAAGTTTCTGTTCACCGTG
csrA-K38A-R	CACGGTGAACAGAAACTTCCGCCGGGGCATTTACGC
csrA-E39A-L	CGTAAATGCCCCGAAGGCGGTTTCTGTTCACCG
csrA-E39A-R	CGGTGAACAGAAACCGCCTTCGGGGCATTTACG
csrA-V40A-L	CGTAAATGCCCCGAAGGAAGCGTCTGTTCACCGTGAAGAG
csrA-V40A-R	CTCTTCACGGTGAACAGACGCTTCCTTCGGGGCATTTACG
csrA-S41A-L	GCCCCGAAGGAAGTTGCGGTTCACCGTGAAGAG
csrA-S41A-R	CTCTTCACGGTGAACCGCAACTTCCTTCGGGGC
csrA-V42A-L	CCCGAAGGAAGTTTCTGCGCACCGTGAAGAGATCTACC
csrA-V42A-R	GGTAGATCTCTTCACGGTGCGCAGAAACTTCCTTCGGG
csrA-H43A-L	CGAAGGAAGTTTCTGTTGCGCGTGAAGAGATCTACCAGC
csrA-H43A-R	GCTGGTAGATCTCTTCACGCGCAACAGAAACTTCCTTCG
csrA-R44A-L	CGAAGGAAGTTTCTGTTCACGCGGAAGAGATCTACCAGCGTATC
csrA-R44A-R	GATACGCTGGTAGATCTCTTCCGCGTGAACAGAAACTTCCTTCG
csrA-E45A-L	GGAAGTTTCTGTTCACCGTGCGGAGATCTACCAGCGTATCC
csrA-E45A-R	GGATACGCTGGTAGATCTCCGCACGGTGAACAGAAACTTCC
csrA-E46A-L	GGAAGTTTCTGTTCACCGTGAAGCGATCTACCAGCGTATCCAGG

csrA-E46A-R	CCTGGATACGCTGGTAGATCGCTTCACGGTGAACAGAAACTTCC
csrA-I47A-L	CTGTTCACCGTGAAGAGGCGTACCAGCGTATCCAGGC
csrA-I47A-R	GCCTGGATACGCTGGTACGCCTCTTCACGGTGAACAG
csrA-Y48A-L	CTGTTCACCGTGAAGAGATCGCGCAGCGTATCCAGGCTG
csrA-Y48A-R	CAGCCTGGATACGCTGCGCGATCTCTTCACGGTGAACAG
csrA-Q49A-L	CCGTGAAGAGATCTACGCGCGTATCCAGGCTGAAAAATCC
csrA-Q49A-R	GGATTTTTCAGCCTGGATACGCGCGTAGATCTCTTCACGG
csrA-R50A-L	CCGTGAAGAGATCTACCAGGCGATCCAGGCTGAAAAATCCC
csrA-R50A-R	GGGATTTTTCAGCCTGGATCGCCTGGTAGATCTCTTCACGG
csrA-I51A-L	CGTGAAGAGATCTACCAGCGTGCGCAGGCTGAAAAATCCCAGC
csrA-I51A-R	GCTGGGATTTTTCAGCCTGCGCACGCTGGTAGATCTCTTCACG
csrA-Q52A-L	GAGATCTACCAGCGTATCGCGGCTGAAAAATCCCAGC
csrA-Q52A-R	GCTGGGATTTTTCAGCCGCGATACGCTGGTAGATCTC
csrA-E54A-L	GCGTATCCAGGCTGCGAAATCCCAGCAGTCC
csrA-E54A-R	GGACTGCTGGGATTTCGCAGCCTGGATACGC
csrA-K55A-L	CCAGCGTATCCAGGCTGAAGCGTCCCAGCAGTCCAGTTACC
csrA-K55A-R	GGTAACTGGACTGCTGGGACGCTTCAGCCTGGATACGCTGG
csrA-S56A-L	CGTATCCAGGCTGAAAAAGCGCAGCAGTCCAGTTACC
csrA-S56A-R	GGTAACTGGACTGCTGCGCTTTTTCAGCCTGGATACG
csrA-Q57A-L	CCAGGCTGAAAAATCCGCGCAGTCCAGTTACCATCATC
csrA-Q57A-R	GATGATGGTAACTGGACTGCGCGGATTTTTCAGCCTGG
csrA-Q58A-L	CAGGCTGAAAAATCCCAGGCGTCCAGTTACCATCATCATC
csrA-Q58A-R	GATGATGATGGTAACTGGACGCCTGGGATTTTTCAGCCTG

csrA-S59A-L	GCTGAAAAATCCCAGCAGGCGAGTTACCATCATCATC
csrA-S59A-R	GATGATGATGGTAACTCGCCTGCTGGGATTTTTCAGC
csrA-S60A-L	GCTGAAAAATCCCAGCAGTCCGCGTACCATCATCATCATCATC
csrA-S60A-R	GATGATGATGATGATGGTACGCGGACTGCTGGGATTTTTCAGC
csrA-Y61A-L	CCCAGCAGTCCAGTGCGCATCATCATCATCATCATTAAGG
csrA-Y61A-R	CCTTAATGATGATGATGATGATGCGCACTGGACTGCTGG

Figure 2-1

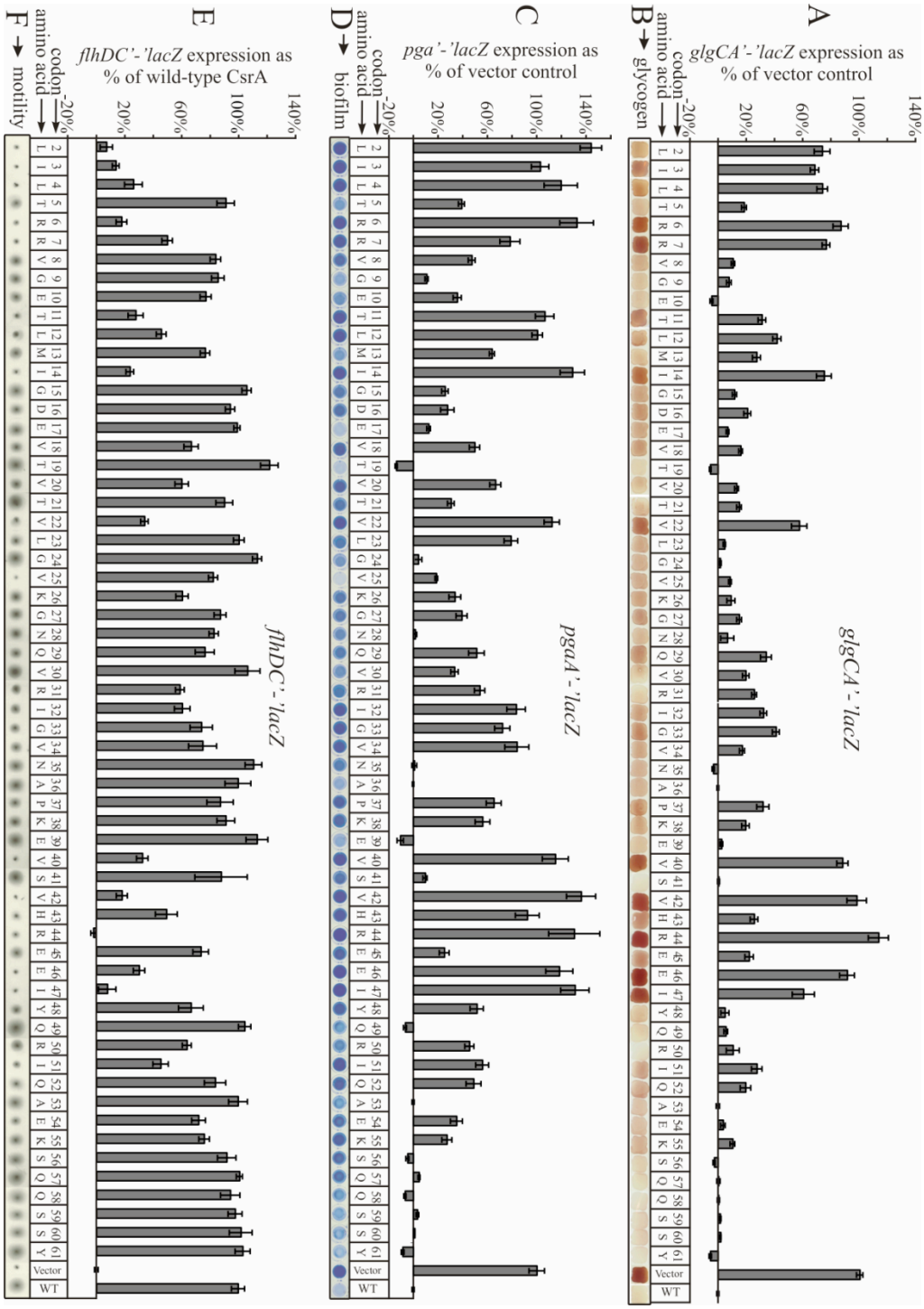

A
glgCA'-'lacZ expression as % of vector control
glgCA' - 'lacZ
B
glycogen
C
pga'-'lacZ expression as % of vector control
pgaA' - 'lacZ
D
biofilm
E
flhDC'-'lacZ expression as % of wild-type CsrA
flhDC' - 'lacZ
F
motility
codon
amino acid
20%
60%
100%
140%
WT
Vector

Figure 2-2

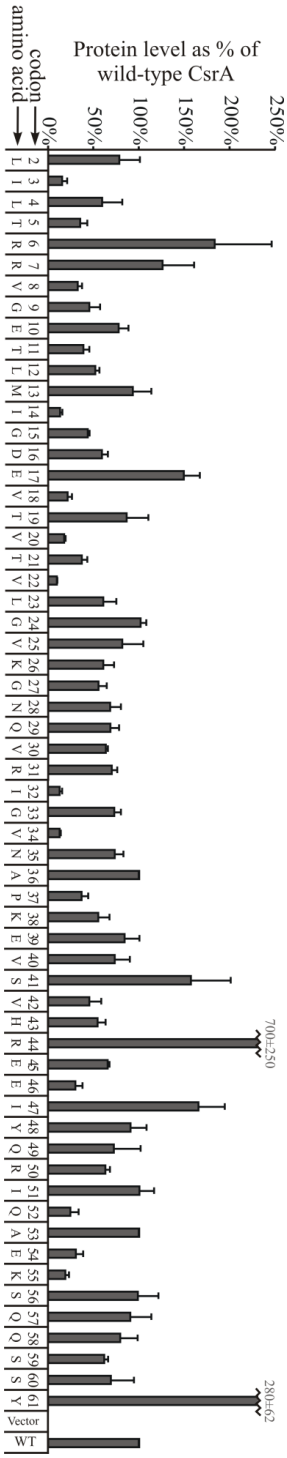

Protein level as % of wild-type CsrA
0%
50%
100%
150%
200%
250%
amino acid
codon
700±250
280±62
Vector
WT

Figure 2-3

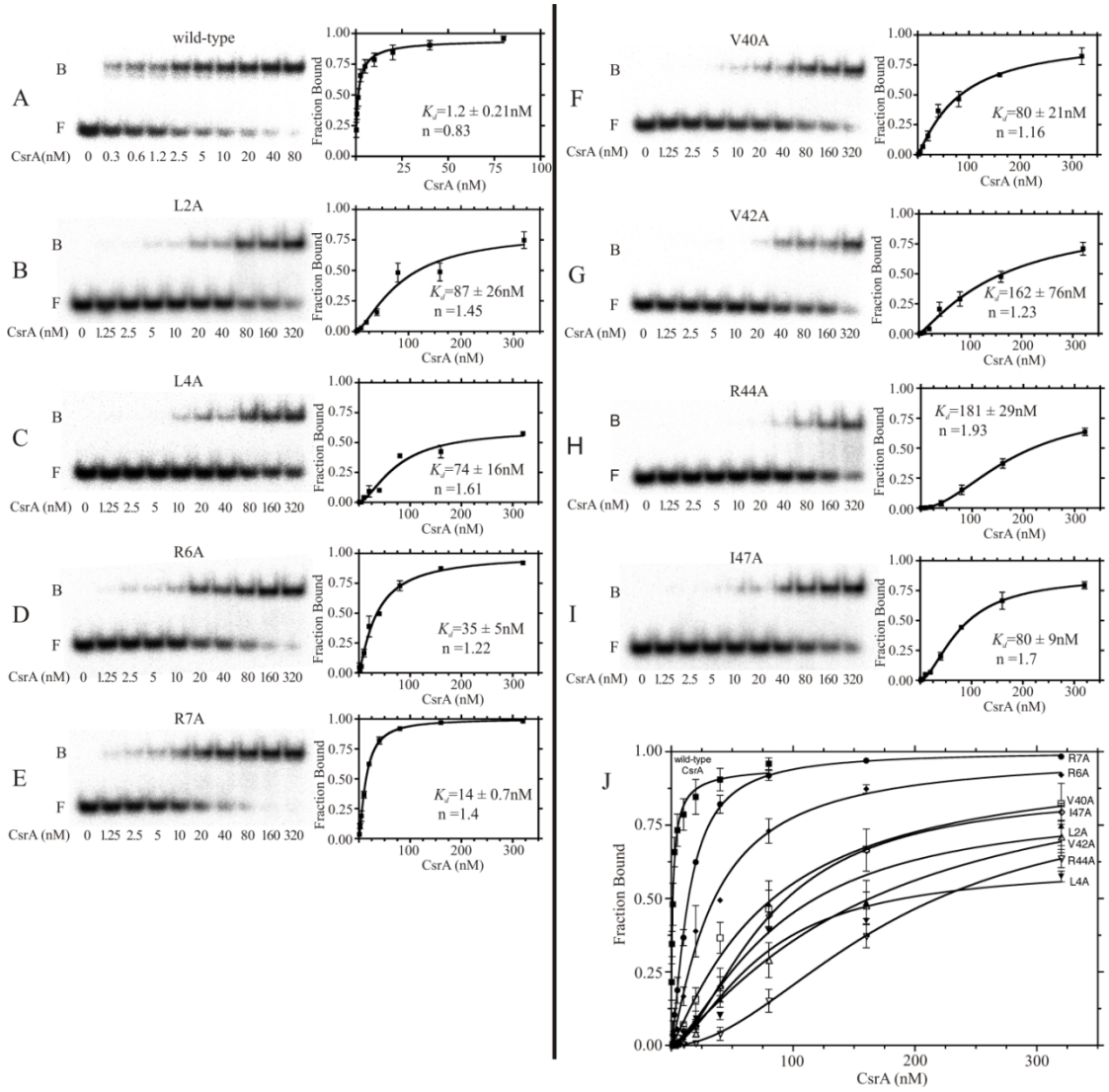

A
wild-type
B
F
CsrA(nM) 0 0.3 0.6 1.2 2.5 5 10 20 40 80
Fraction Bound
CsrA (nM)
K_d=1.2 ± 0.21nM
n =0.83
B
L2A
CsrA (nM) 0 1.25 2.5 5 10 20 40 80 160 320
K_d=87 ± 26nM
n =1.45
C
L4A
K_d=74 ± 16nM
n =1.61
D
R6A
K_d=35 ± 5nM
n =1.22
E
R7A
K_d=14 ± 0.7nM
n =1.4
F
V40A
K_d=80 ± 21nM
n =1.16
G
V42A
K_d=162 ± 76nM
n =1.23
H
R44A
K_d=181 ± 29nM
n =1.93
I
I47A
K_d=80 ± 9nM
n =1.7
J
wild-type CsrA
R7A
R6A
V40A
I47A
L2A
V42A
R44A
L4A

Figure 2-4

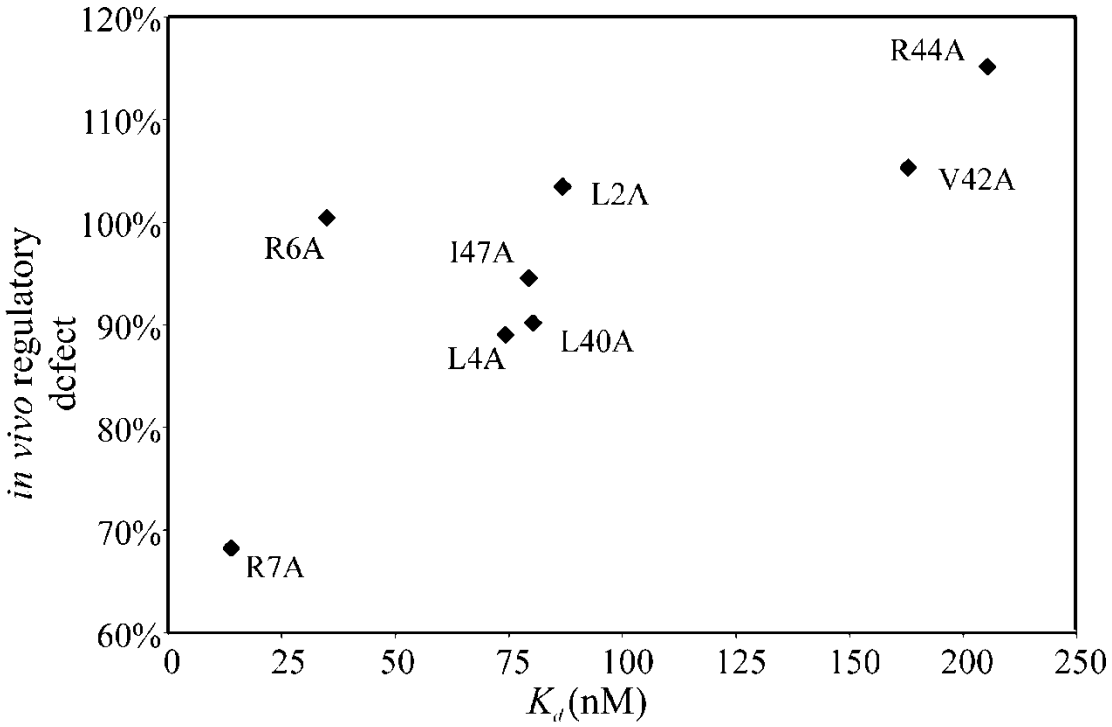

120%
110%
100%
90%
80%
70%
60%
in vivo regulatory defect
0
25
50
75
100
125
150
200
250
K_d(nM)
R44A
V42A
L2A
R6A
I47A
L4A
L40A
R7A

Figure 2-5

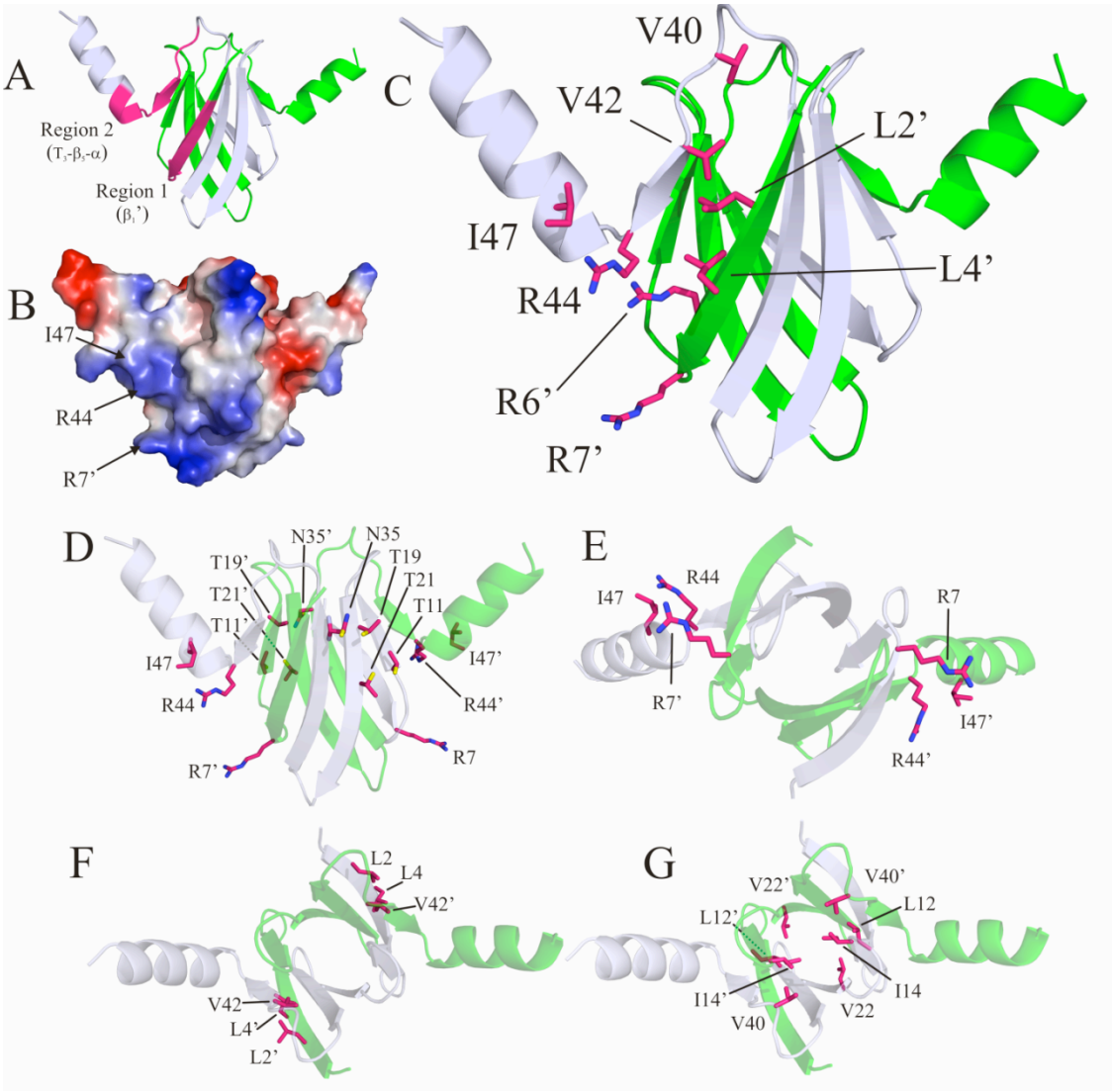
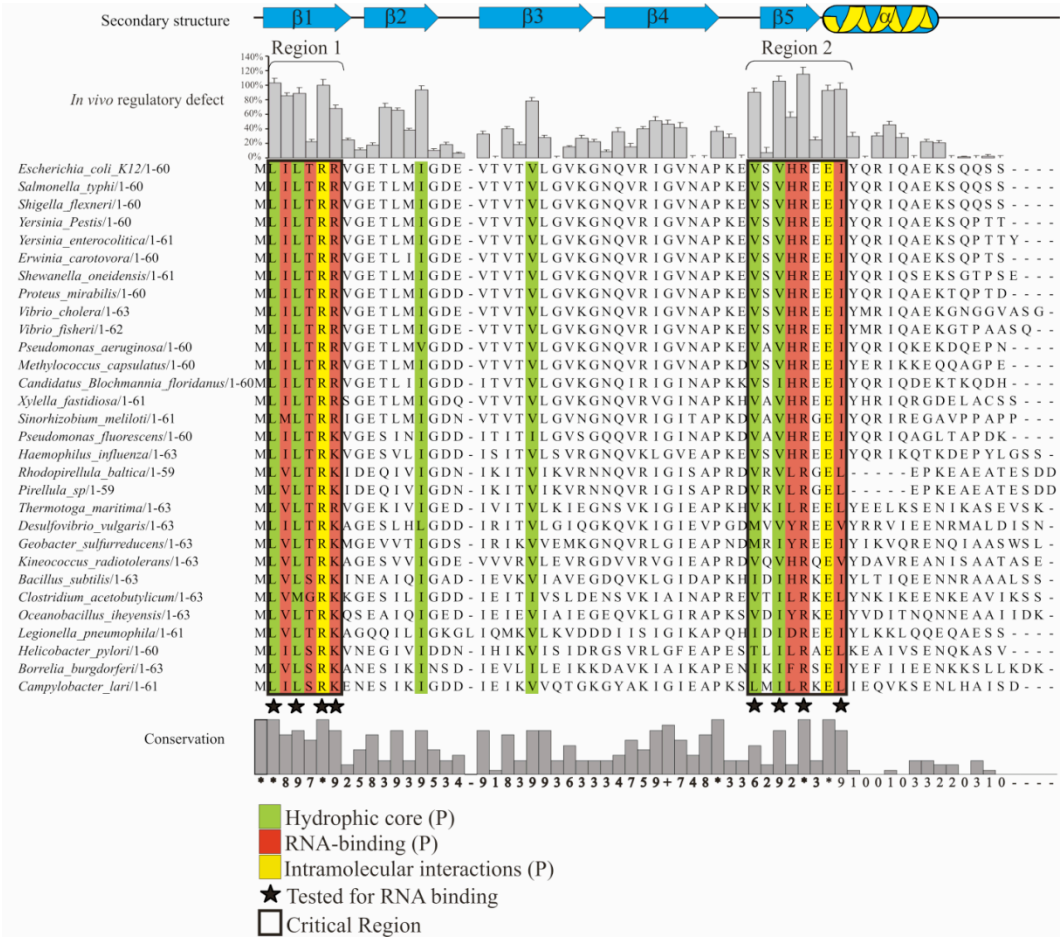

A
Region 2
(T3-β5-α)
Region 1
(β1')
B
I47
R44
R7'
C
V40
V42
L2'
I47
L4'
R44
R6'
R7'
D
N35'
N35
T19'
T19
T21'
T21
T11'
T11
I47
I47'
R44
R44'
R7'
R7
E
I47
R44
R7
R7'
I47'
R44'
F
L2
L4
V42'
V42
L4'
L2'
G
V22'
V40'
L12'
L12
I14'
I14
V40
V22

Figure 2-6


Secondary structure
β1
β2
β3
β4
β5
α
Region 1
Region 2
In vivo regulatory defect
Escherichia_coli_K12/1-60
Salmonella_typhi/1-60
Shigella_flexneri/1-60
Yersinia_Pestis/1-60
Yersinia_enterocolitica/1-61
Erwinia_carotovora/1-60
Shewanella_oneidensis/1-61
Proteus_mirabilis/1-60
Vibrio_cholera/1-63
Vibrio_fisheri/1-62
Pseudomonas_aeruginosa/1-60
Methylococcus_capsulatus/1-60
Candidatus_Blochmannia_floridanus/1-60
Xylella_fastidiosa/1-61
Sinorhizobium_meliloti/1-61
Pseudomonas_fluorescens/1-60
Haemophilus_influenza/1-63
Rhodopirellula_baltica/1-59
Pirellula_sp/1-59
Thermotoga_maritima/1-63
Desulfovibrio_vulgaris/1-63
Geobacter_sulfurreducens/1-63
Kineococcus_radiotolerans/1-63
Bacillus_subtilis/1-63
Clostridium_acetobutylicum/1-63
Oceanobacillus_iheyensis/1-63
Legionella_pneumophila/1-61
Helicobacter_pylori/1-60
Borrelia_burgdorferi/1-63
Campylobacter_lari/1-61
Conservation
Hydrophic core (P)
RNA-binding (P)
Intramolecular interactions (P)
Tested for RNA binding
Critical Region

Chapter 3. Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression.

Jeffrey Mercante[1], Adrianne N. Edwards[1], Ashok K. Dubey[2,4], Paul Babitzke[2], and Tony Romeo[1,3*]

[1]Department of Microbiology and Immunology, Emory University School of Medicine, 3153 Rollins Research Center, 1510 Clifton Road N.E., Atlanta, GA 30322, USA.

[2]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA.

[3]Department of Microbiology and Cell Science. P.O. Box 110700. University of Florida. Gainesville, FL 32611-0700, USA.

[4]Tata Chemicals Innovation Center, Anmol Pride, Baner Road, Pune 411045, India.

*Corresponding author

Tony Romeo. Present address is the University of Florida. Email: tromeo@ufl.edu

Phone: 352-392-5923 Fax: 352-392-5922

This chapter consists of a manuscript submitted for publication. The manuscript was written by Jeffrey Mercante, Paul Babitzke and Tony Romeo. Boundary analyses were performed by Ashok K. Dubey and a gel shift experiment was performed by Adrianne N. Edwards. All remaining experiments were performed by Jeffrey Mercante.

SUMMARY

The present study established that a single CsrA homodimer contains two independent RNA-binding surfaces that can bind either one or two high affinity SELEX-derived RNA targets. To investigate if both binding surfaces on a single protein are required for RNA-binding and regulation, we constructed a CsrA heterodimer (HD-CsrA) consisting of one wild-type surface and one that is defective for RNA interaction. In comparison to WT-CsrA protein, which bound either one or two independent RNA targets with high avidity, the HD-CsrA only bound to a single RNA and exhibited half the affinity. To test the RNA requirements for allowing a single CsrA dimer to interact with both targets of a dual site RNA, electrophoretic mobility shift assays (EMSA) were conducted where WT or HD-CsrA were combined with a series of synthetic RNA oligonucleotides containing one or two high affinity target sites set at varying distances from each other. It was determined that 18 nucleotides was the optimal distance between two sites, however, CsrA could simultaneously bind targets as close as 10 nt and as far apart as 63 nt. The spacing of 10 nt was found to be too short a distance, though, for stable association of CsrA at both targets. WT and HD-CsrA were also tested for their abilities to regulate expression from a native *E. coli* target through *in vitro* transcription-translations assays. WT-CsrA repressed expression from the 5'-untranslated leader of the *glgCAP* mRNA 14-fold more efficiently than HD-CsrA. When one target site in the leader, upstream from the SD, was deleted, repression by WT-CsrA dropped to 7-fold while HD-CsrA regulation did not change. This confirmed that dual-site binding is a biologically relevant mechanism for gene repression by CsrA and both binding surfaces are needed for full regulatory activity.

ABSTRACT

The global regulatory protein CsrA binds to the 5'-untranslated leader of target transcripts and alters their translation and/or stability. CsrA is a symmetrical homodimer containing two identical RNA-binding surfaces. Gel shift assays with model RNA substrates now show that CsrA can bind simultaneously at two target sites within a transcript (bridging or dual site binding). An intersite distance of ~18 nucleotides (nt) was optimal, although bridging occurred with an intersite distance of 10 to ≥ 63 nt. The close 10-nt spacing reduced the stability of dual site binding, as competition for one site by a second CsrA dimer readily occurred. Both RNA- binding surfaces of a single CsrA protein were essential for efficient *in vitro* repression of a *glgC'-'lacZ* translational fusion that contains four CsrA target sites within the untranslated leader. Heterodimeric CsrA (HD-CsrA) containing a single R44A replacement, which was defective for binding at its mutant surface but bound RNA normally at its wild-type (WT) surface, was ~14-fold less effective at repression than homodimeric WT-CsrA. Furthermore, deletion of a CsrA target site of *glgC* that lies upstream from the Shine-Dalgarno sequence did not affect regulation by HD-CsrA but decreased regulation by WT-CsrA, confirming a regulatory role of dual site binding. Finally, we propose a mechanism whereby a globular ribonucleoprotein complex is formed between CsrA and its noncoding RNA antagonist, CsrB. Because many target sites of CsrB are located closer together than is optimal for bridging, binding to non-adjacent sites should be energetically favored, causing multiple CsrA dimers to tether CsrB into the observed globular form rather than an extended CsrA-CsrB complex.

INTRODUCTION

RNA-binding proteins participate in a myriad of cellular processes and are necessary for all known forms of life to exist (Burd and Dreyfuss, 1994; Abaza and Gebauer, 2008). The abundance of these proteins mirrors their diverse functions, yet only a relatively small number of known RNA-binding motifs are distributed throughout this large family (Abaza and Gebauer, 2008; Bolognani and Perrone-Bizzozero, 2008; Glisovic, Bachorik *et al.*, 2008). The disparity is due to modularity in protein design and multimerization of RNA binding motifs. Multiple binding sites within the same protein can increase target specificity and RNA-binding avidity (Chmiel, Rio *et al.*, 2006). Both bacterial and eukaryotic RNA-binding proteins with multiple RNA binding motifs are known (Nogueira and Springer, 2000; Lunde, Moore *et al.*, 2007; Finn, Tate *et al.*, 2008; Babitzke, Baker *et al.*, 2009). For example, the eukaryotic proteins Dicer and U2AF35 each contain 3 and 4 different RNA-binding motifs, respectively, while Vigilin has 14 copies of the same motif (Lunde, Moore *et al.*, 2007; Glisovic, Bachorik *et al.*, 2008); the bacterial transcriptional terminator NusA contains an S1 and dual K-Homology domains (KH) (Gopal, Haire *et al.*, 2001; Beuth, Pennell *et al.*, 2005) while the circular, homo-undecameric *trp* RNA-binding attenuation protein (TRAP) is composed of 11 identical surfaces capable of binding (G/U)AG triplet repeats (Antson, Dodson *et al.*, 1999).

The molecular interactions between RNA and RNA-binding proteins are guided by the nature of the protein motifs that are present; however, the spacing of RNA-binding motifs and their target sequences is also essential for tight protein-RNA association and proper gene regulation. For example, TRAP was unable to bind RNA when the distance between trinucleotide repeats was changed from the optimum of 2 nt to either 1, 3 or 4 nt

(Babitzke, Bear *et al.*, 1995). Conversely, deletion of the linker between RRM2 and RRM3 in the mammalian CUG-binding protein (CUG-BP) severely compromised rUrG interaction, presumably by inhibiting cooperativity between two regions of the protein (Mori, Sasagawa *et al.*, 2008). The present study addresses questions about the spacing of RNA target sites and the biological necessity of multiple RNA-binding surfaces in the *E. coli* RNA-binding protein CsrA.

CsrA (and its orthologs RsmA, RsmE) is a global regulatory protein that controls carbon metabolism (Liu, Yang *et al.*, 1995; Sabnis, Yang *et al.*, 1995), biofilm formation (Jackson, Suzuki *et al.*, 2002; Wang, Dubey *et al.*, 2005), motility (Ang, Horng *et al.*, 2001; Wei, Brun-Zinkernagel *et al.*, 2001), peptide uptake (Dubey, Baker *et al.*, 2003), quorum-sensing (Lenz, Miller *et al.*, 2005) and virulence functions (Heurlier, Williams *et al.*, 2004; Fortune, Suyemoto *et al.*, 2006; Chao, Wei *et al.*, 2008; Mulcahy, O'Callaghan *et al.*, 2008) of many eubacteria. Regulation is accomplished by specific binding of CsrA to conserved sequences in the 5'-untranslated leader of target mRNAs, leading to effects on their translation and/or stability (Wei, Brun-Zinkernagel *et al.*, 2001; Wang, Dubey *et al.*, 2005). In addition to its mRNA targets, CsrA interacts with small noncoding RNAs, CsrB (Liu, Gui *et al.*, 1997) and CsrC (Weilbacher, Suzuki *et al.*, 2003) in *E. coli*, that sequester and antagonize this protein. These sRNAs are predicted to form multiple stem-loops containing conserved sequences resembling CsrA target sites in mRNA.

Studies of CsrA have culminated in the recent determination of its 3D-structure. Crystallography (Rife, Schwarzenbacher *et al.*, 2005) and solution NMR studies (Gutierrez, Li *et al.*, 2005; Heeb, Kuehne *et al.*, 2006) revealed that CsrA represents a new class of RNA-binding protein containing a unique RNA-binding motif. It is

assembled via the interdigitation of 5 b-strands of two identical polypeptides, forming a bilaterally symmetrical protein that is stabilized by a hydrophobic core and extensive hydrogen bonding. Two identical RNA-binding surfaces are located on opposite sides of the protein, each of which is composed primarily of the N-terminal b-strand (b_1) of one polypeptide and the parallel b_5 strand of the second polypeptide. These surfaces contain several amino acid residues that are needed for optimal RNA binding and regulation, the most important of which was R44 (Mercante, Suzuki *et al.*, 2006). NMR studies of a CsrA ortholog (RsmE) complexed with target RNAs confirmed the protein-RNA contacts that were implied by alanine substitution studies (Schubert, Lapouge *et al.*, 2007). Thus, as in TRAP (Babitzke, Stults *et al.*, 1994; Hopcroft, Manfredo *et al.*, 2004) and MBNL1 (He, Dang *et al.*, 2009), an RNA binding surface is formed by the interface of two interacting polypeptide subunits.

Electrophoretic mobility shift assays (EMSA), RNA footprinting and other studies have shown that multiple CsrA proteins can bind to a single mRNA or sRNA transcript that contains multiple binding sites (Baker, Morozov *et al.*, 2002; Wang, Dubey *et al.*, 2005). In conjunction with the evidence that CsrA contains two RNA-binding surfaces, we hypothesized that a single CsrA dimer might bridge two target sites on the same RNA (Mercante, Suzuki *et al.*, 2006). The mirrored orientation of the dual CsrA binding surfaces combined with the flexibility of single-stranded RNA should dictate the positioning of the second RNA target site after initial contact is made. Furthermore, the local concentration of a secondary target should be greatly increased upon CsrA binding to a primary site (Rippe, 2001), thus increasing the probability of a secondary intramolecular interaction, which should increase binding avidity.

Here, we demonstrate that a CsrA dimer can indeed bind two target sites on a single RNA simultaneously, resulting in a compact "bridge complex", whose stability depended on the distance separating the RNA target sites. Our findings further revealed that both RNA-binding surfaces of CsrA must function simultaneously for efficient post-transcriptional regulation, and in *glgC* repression, they facilitate bridging from a high-affinity stem-loop in the untranslated leader to a downstream inhibitory site that overlaps the SD sequence. A model for CsrA binding to RNAs with multiple target sites is presented and its implications for regulation of mRNA translation and CsrA sequestration by regulatory sRNAs are discussed.

RESULTS

WT-CsrA dimer binds two RNA oligonucleotides simultaneously

We previously demonstrated that a CsrA dimer contains two identical surfaces that are critical for RNA binding and regulation of gene expression (Mercante, Suzuki *et al.*, 2006). Subsequently, 3D-NMR analysis revealed that a CsrA ortholog can bind two RNA oligonucleotides in solution (Schubert, Lapouge *et al.*, 2007). To further examine CsrA binding properties, the WT-CsrA protein was tested for binding to a high affinity, SELEX-derived oligonucleotide (Dubey, Baker *et al.*, 2005), referred to as RNA-1A (Table 3-1). In this EMSA experiment, both a major and a minor complex were formed (Fig.1a). The addition of increasing amounts of specific unlabeled RNA competitor to the binding reaction with WT-CsrA resulted in the formation of the faster mobility minor complex at the expense of the slower moving major complex (Fig. 3-1b, lanes 2-6). Because this exchange was due to the addition of the RNA oligonucleotide to the complex, we concluded that the faster moving species was composed of one CsrA protein

bound to two independent RNAs and that the slower moving species was composed of one CsrA dimer bound to one RNA. The increased mobility of the larger complex (1 CsrA:2 RNA) must have been due to the greater influence of the negative charge conferred by the binding of a second RNA relative to its effect on the Stokes radius.

RNA binding by a CsrA heterodimer (HD-CsrA) containing mutant and wild type RNA-binding surfaces

To better understand how CsrA interacts with RNAs containing multiple target sites we prepared a heterodimeric CsrA protein that is composed of one wild-type polypeptide and one his-tagged polypeptide containing an alanine substitution (R44A). This substitution reduces the affinity of the homodimer mutant protein for a high-affinity RNA target by ~180-fold (Mercante, Suzuki *et al.*, 2006).

To test whether both RNA-binding surfaces of HD-CsrA were functional, increasing amounts of unlabeled RNA-1A were added to a binding reaction containing HD-CsrA. Unlike the reaction with WT-CsrA, the HD-CsrA protein never exhibited a second shift (Fig. 3-1b, lanes 7-10), confirming that it was unable to bind two RNAs simultaneously under our reactions conditions. The mobility of the HD-CsrA/RNA complex was slightly faster than the WT-CsrA complex on native polyacrylamide, presumably due to the R44A substitution, which results in a net increase in negative charge. HD-CsrA bound to RNA-1A with approximately one third of the binding affinity (K_d = 3.4 nM ± 0.4) of WT-CsrA protein (K_d = 1.25 nM ± 0.3) (Fig. 3-2).

Binding of the CsrA dimer to a series of RNAs containing dual target sites separated by varying distances

Most RNAs bound by CsrA (RsmA) contain multiple target sites that are related to the SELEX-derived consensus sequence "RUACARGGAUGU" where the underlined nucleotides were 100% conserved and were present in the loop of a hairpin structure (Liu, Gui *et al.*, 1997; Baker, Morozov *et al.*, 2002; Dubey, Baker *et al.*, 2003; Weilbacher, Suzuki *et al.*, 2003; Valverde, Lindell *et al.*, 2004; Burrowes, Abbas *et al.*, 2005; Dubey, Baker *et al.*, 2005; Wang, Dubey *et al.*, 2005; Lapouge, Sineva *et al.*, 2007; Sorger-Domenigg, Sonnleitner *et al.*, 2007; Yakhnin, Pandit *et al.*, 2007). In native RNA molecules, sequential CsrA target sites may be as close as 6 or 7 nt, as predicted in CsrC (Weilbacher, Suzuki *et al.*, 2003) and CsrB (Liu, Gui *et al.*, 1997), respectively, or they may be more than 50 nt apart, as determined in the 5'-leader of the *pgaABCD* message (Wang, Dubey *et al.*, 2005). Nevertheless, the way in which the respective RNA binding sites of a CsrA dimer interact with paired target sites has not been determined for any of these RNAs. Furthermore, we have proposed that the symmetrical structure of CsrA, with two identical binding surfaces on opposite sides of the dimer, appears to account for the observation that CsrB, which contains ~18-22 target sequences, forms a globular complex with ~9-10 CsrA dimers (Mercante, Suzuki *et al.*, 2006). The implication is that all or most of the CsrA dimers of this complex should form noncovalent bridges between two target sites on the same RNA.

To test the hypothesis that a CsrA dimer bridges two binding sites in target RNAs, and to probe the minimum and maximum intersite distances at which this occurs, we designed a series of RNA oligonucleotides containing dual high-affinity target sites set at

varying distances from each other (Table 3-1: RNA10-2BS, RNA14-2BS, RNA18-2BS, RNA28-2BS, RNA43-2BS and RNA63-2BS). These RNAs were examined by EMSA for the binding of WT-CsrA or HD-CsrA (Fig. 3-3a-f). A second series of RNAs were prepared, identical to the first series, except that the 3'-high-affinity binding sites were altered (i.e., the central GGA was changed to AAA or CCC; see Table 3-1 for an explanation of the use of CCC or AAA) to prevent CsrA binding (Table 3-1: RNA10-1BS, RNA14-1BS, RNA18-1BS, RNA28-1BS, RNA43-1BS and RNA63-1BS). The single-site RNAs were expected to form a 1:1 complex with HD-CsrA, providing a series of standard shifts to which the other complexes can be compared (Fig. 3-3a-f, complex 6). Complexes formed by two CsrA dimers bound to a single RNA should migrate considerably slower than the standards. Complexes that travel even faster than the 1:1 standards should be more compact, consistent with dual site bridging by CsrA. The design of all oligonucleotides except RNA10-2BS and RNA10-1BS included 16 nt RNA hairpins with stems of 6 nucleotide pairs, which was previously established to be a high affinity target for CsrA (Mercante, Suzuki *et al.*, 2006; Baker, Eory *et al.*, 2007). To decrease the distance between CsrA target sites even further, we constructed RNA10-2BS and RNA10-1BS with 12 nucleotide hairpins (4 nucleotide pair stems). The *in silico* analysis (Zuker, 1989; SantaLucia, 1998) of smaller RNA stems did not predict the formation of stable hairpin structures at a reaction temperature of 37 ^{o}C.

To characterize the target sites of the model RNAs, we initially confirmed that WT-CsrA bound with equal affinity to the target sites at the 5' and 3'-termini of dual target RNAs by synthesizing each stem loop independently (Table 3-1: RNA-1A and RNA-1B) (data not shown). We also established that target sites that were designed to

have negligible affinity for CsrA (Table 3-1: RNA-2A and RNA-2B) showed no binding up to 160 nM CsrA (data not shown).

All reactions containing HD-CsrA with a single-target RNA revealed a single complex, which was concluded to be composed of the WT RNA-binding surface of HD-CsrA bound to the high affinity stem-loop of the RNA (a 1:1 standard complex, labeled "6" in Fig. 3-3a, lanes 9 and 10; 3b-f, lanes 8 and 9). Strikingly, most of the dual site RNAs with intersite distances from 10 to 43 nt exhibited a WT-CsrA:RNA band (complex 1) of greater mobility (more compact) than the 1:1 standard complexes formed by HD-CsrA and the single-site RNAs (complex 6). This finding suggested that WT-CsrA can bridge closely adjoining RNA target sites in addition to distantly separated ones. At low HD-CsrA concentrations, every binding reaction with a dual high affinity target RNA (Fig. 3-3a-f, lane 4) exhibited a fast moving shift (complex 3) similar to the 1:1 complex 6. As the concentration of HD-CsrA was increased, a complex with considerably slower mobility (Fig. 3-3a-f, lane 5, complex 4) was formed at the expense of complex 3, implying that it is composed of 2:1 CsrA:RNA. This complex was expected because the affinity of the mutated binding surface found on the HD-CsrA for a high-affinity target site is very low (≥180 nM) and should not compete effectively with the WT surface of a second HD-CsrA molecule for RNA-binding (Mercante, Suzuki *et al.*, 2006). The concentration of free HD-CsrA required to fully compete for the second high-affinity site after the initial binding event was between 2.5 nM and 62.5 nM for every RNA (complex 4) except for RNA18-2BS. Thus, the second high affinity target site on RNA18-2BS appears to have a tighter association with the initially bound HD-CsrA dimer compared to the other RNAs, resulting in a complex that apparently hindered

competition by the second HD-CsrA dimer (Fig. 3-3c, lane 5, complex 4). Notably, the mobility of complex 3 (HD-CsrA bound to a dual site RNA) was not consistently equal to that of complex 6 (HD-CsrA bound to single-site RNA) for every pair of synthetic oligonucleotides, as might have been predicted (compare Fig. 3-3a, lane 4 with 3d, lane 4). It is known that while the R44A mutation severely reduces the binding affinity of a CsrA homodimer double mutant for RNA (~180 nM), it does not eliminate it (Mercante, Suzuki *et al.*, 2006). Therefore, this variation may be due to the residual affinity between the mutant RNA binding surface of CsrA and the second high-affinity binding site of the RNA (e.g., Fig. 3-3d lane 4).

Unlike the HD-CsrA, addition of the higher concentration of WT-CsrA to RNAs with two high affinity binding sites only led to putative 2:1 CsrA:RNA complexes with the two shortest RNAs, RNA10-2BS and RNA14-2BS (Fig. 3-3a, b, lanes 2 and 3, complex 2). At an intersite distance of 10 nt, WT-CsrA appears to be only weakly associated to one of the two target sites to which it is bound, allowing a second WT-CsrA to successfully compete for binding at one site. This observation suggests that dual site RNAs with intersite distances ≥ 18 nt (RNA18-2BS, RNA28-2BS, RNA43-2BS and RNA63-2BS) rapidly and tightly associate with a single WT-CsrA protein at both GGA targets, preventing competition by a second CsrA protein for binding to either target site. We conclude that RNAs with target sites separated by only 10 nt do not form stable bridge complexes with both RNA binding surfaces of CsrA, presumably due to steric restrictions. Curiously, the smallest synthetic RNA tested (RNA10-2BS) exhibited a distinct shifted species (Fig. 3-3a, lane 2, top band) that is not consistent with complexes 1 through 6. This may represent an alternate protein-RNA conformation or more likely, a

complex of higher stoichiometry (2:2 CsrA:RNA). Recall from Fig. 3-1 that addition of a small RNA to a CsrA-RNA complex can cause increased gel mobility when the relative effect of the altered charge exceeds that of Stokes radius.

At target separation distances ≥ 18 nt there was apparently no steric restriction imposed by the protein or RNA for dual-site binding. Therefore, 18 nt may be near the optimum spacing distance between RNA targets. As seen in figure 3-3c, the WT-CsrA/RNA18-2BS shift (lanes 2 and 3, complex 1) exhibited the same apparent mobility as the HD-CsrA/RNA18-2BS (lanes 4 and 5, complex 3) and WT-CsrA/RNA18-1BS (lanes 6 and 7, complex 5). These three complexes exhibited greater mobility than the 1:1 complex formed by HD-CsrA and RNA18-1BS (lanes 8 and 9, complex 6), which contained a single high-affinity site. The implication of this observation is that the high affinity target sites of RNA18-2BS, as well as the high and low affinity targets of RNA18-1BS, are optimally spaced and situated for dual binding. Thus, near-optimal spacing may counteract the effect of the R44A mutation in the HD-CsrA/RNA18-2BS complex, as well as the effect of the low affinity target site on RNA18-1BS as it binds WT-CsrA. Since the mobility of these shifts resembles that of the WT-dual site reaction (Fig. 3-3c, lanes 2 and 3, complex 1), these complexes probably represent bridge complexes. Thus, ideal spacing of two binding sites may compensate for relatively weak protein-RNA interactions with these molecules. However, the combined effects of an R44A protein substitution and a low affinity RNA target site in the HD-CsrA/RNA18-1BS combination (Fig. 3-3c, lanes 8 and 9, complex 6) apparently prevented the formation of a bridge complex despite near optimal spacing.

At intersite distances greater than 28 nt the WT-CsrA:dual site RNA complexes began to smear in a native polyacrylamide gel and the exact composition of the species was unclear (Fig. 3-3e and f, lanes 2 and 3, complex 1). Smearing may be due to the extensive linker size of RNA43-2BS and RNA63-2BS (30 and 50 nt respectively), which allows multiple degrees of freedom in the single-stranded intersite region when both target sites are bound or by transient dissociation of one interaction, resulting in an upward smearing due to loss of the compact structure.

Theoretically, when WT-CsrA is combined with a single-target RNA (i.e., RNA14-1BS), two simple complexes may form, a 1:1 protein-RNA and a 1:2 protein-RNA combination. The presence of only one species (Fig. 3-3b-f, lanes 6 and 7, complex 5) in all but one of these reactions suggests that although WT-CsrA has the capacity to bind two oligonucleotides, binding of one RNA may sterically inhibit the binding of a second RNA at the free surface. An exception to this observation was RNA10-1BS (Fig. 3-3a, lane 7), which formed a second shift at low protein concentrations that migrated slower than the apparent 1:1 HD-CsrA:RNA complex (Fig. 3-3a, lanes 9 and 10). We have not determined the basis of the latter observation but suspect that the small size of RNA10-1BS decreases bridging and allows binding of two RNAs to the same protein. In this case the RNA oligonucleotide appears to be of sufficient size such that the complex Stokes radius overcomes the influence of increased negative charge, resulting in a slower migrating shift.

Stoichiometry of CsrA:RNA complexes

The experiments described above suggested that fast moving complexes formed between WT-CsrA and RNAs with two high-affinity binding sites differed in conformation but not in stoichiometry from the 1:1 complexes formed between HD-CsrA and single-site RNAs. To test this hypothesis, the protein:RNA stoichiometries of four representative CsrA-RNA complexes were examined by EMSA. As described in the Materials and Methods, [γ-^{32}P]-labeled RNAs in gels were detected by phosphorimager analysis, followed by transfer of CsrA protein to a PVDF membrane and subsequent immunoblotting using a polyclonal antibody for CsrA (Gudapaty, Suzuki *et al.*, 2001). The stoichiometries of WT-CsrA and HD-CsrA complexes with RNA28-2BS and RNA28-1BS were chosen for this analysis because they yielded unambiguous shifts. Under the chosen conditions, the HD-CsrA:RNA28-1BS reaction yielded only a single shift (Fig. 3-4, lane 1), which was concluded to represent the 1:1 complex of HD-CsrA dimer bound to RNA, and its RNA:protein ratio was assigned a value of 1.0. The stoichiometries of the other shifted complexes were determined by comparison to this one (Fig. 3-4). The stoichiometry of the slow-moving complex formed by HD-CsrA:RNA28-2BS (Fig. 3-4, lane 3) was determined to be 1.93 ± 0.1, confirming that it contains two HD-CsrA dimers, each of which is bound to a high-affinity site of this RNA. The fast-moving WT-CsrA:RNA28-1BS complex (Fig. 3-4, lane 2) was present at a protein-RNA ratio of 1.12 ± 0.2. This species traveled significantly faster than HD-CsrA:RNA28-1BS (lane 1). This difference was likely due to an additional weak association between the low affinity RNA target site from RNA28-1BS and the free RNA binding surface of the WT-CsrA protein to which it was tethered. The fastest migrating complex, WT-CsrA:RNA28-2BS (Fig. 3-4, lane 4), had a protein-RNA ratio of 0.96 ± 0.1, consistent

with one WT-CsrA protein bound to a single RNA. This experiment demonstrated conclusively that although WT-CsrA/RNA28-2BS exhibits greater mobility, it nevertheless contains the same ratio of protein to RNA as the 1:1 HD-CsrA/RNA28-1BS complex. These results confirmed the interpretations from gel shifts (Fig. 3-3) that WT-CsrA bridges both high affinity target sites of RNA28-2BS forming a compact structure that migrates faster than HD-CsrA/RNA28-1BS. By extrapolation, the fast-moving complexes formed by the other dual site RNAs and WT-CsrA are also concluded to represent bridge complexes (Fig. 3-3 complex 1).

Both CsrA RNA-binding surfaces are necessary for efficient regulation

Having established parameters by which WT-CsrA protein can bind simultaneously to two target sites on a single RNA, we next explored the biological relevance of this arrangement. The presence of two or more RNA target sites within a CsrA-regulated transcript could cause tighter binding of a single CsrA protein, the formation of an RNA secondary structure such as a loop, and/or the recognition of a low affinity target by tethering of CsrA at a nearby high affinity site. Each of these consequences may result in increased translational inhibition with respect to a CsrA protein with a single binding surface. To test whether both CsrA binding surfaces of a dimer were necessary for efficient gene regulation, we performed S30-directed *in vitro* coupled transcription-translation (Romeo and Preiss, 1989) experiments using either WT-CsrA or HD-CsrA protein (Fig. 3-5). The 5'-untranslated leader of *glgCAP* contains two previously characterized CsrA target sites; one that overlaps the SD sequence and one that is situated 28 nt upstream (Baker, Morozov *et al.*, 2002). Supercoiled plasmids

containing a wild-type *glgC*'-'*lacZ* translational fusion (pCZ3-3) or the same construct with a 3 nt deletion of the upstream CsrA target site (*glgC*ΔGGA'-'*lacZ,* pCSB25) (Baker, Morozov *et al.*, 2002), were used as DNA templates for the reactions. An analysis of the expression data at 50% relative GlgC-LacZ polypeptide synthesis revealed that WT-CsrA was ~14 fold more efficient at repressing expression from the wild-type *glgC* template ($IC_{50} \cong 22$ nM) compared to the HD-CsrA ($IC_{50} \cong 323$ nM) (Fig. 3-5a). When this value is corrected for the number of RNA-binding surfaces per dimer, repression by the WT-CsrA was ~7-fold greater than HD-CsrA. When the upstream CsrA binding site of the *glgC*'-*lacZ* fusion was deleted from the template, repression by HD-CsrA was unaffected ($IC_{50} \cong 310$ nM). In contrast, translational repression by the WT-CsrA protein decreased to ~7-fold ($IC_{50} \cong 44$ nM) compared to the HD-CsrA, or ~3.5-fold when adjusted for the number of binding surfaces per dimer (Fig. 3-5b). These results indicated that both RNA-binding surfaces of CsrA and both CsrA binding sites of the *glgCAP* 5'-leader are required for maximal regulation of this operon.

Additional, uncharacterized CsrA target sites in the *glgCAP* 5'-leader RNA

HD-CsrA was expected to repress protein expression at similar levels from the WT and mutant *glgC*'-'*lacZ* templates because binding to the upstream CsrA target site should not block translation, and both templates contained the same CsrA binding site overlapping the SD sequence. However, a greater difference in IC_{50} might have been anticipated for WT-CsrA with these templates if the upstream binding site of *glgC* mRNA was primarily responsible for differential repression by the two proteins. A possible explanation was that the WT *glgCAP* leader contained additional CsrA binding

sites, which could mediate bridging of WT-CsrA to the SD sequence in the mutant transcript. Therefore, we searched for additional CsrA target sites in the *glgC* leader. While previous footprinting results identified only two CsrA binding sites in the *glgCAP* leader transcript (Fig. 3-6b, BS2 and BS4), a position-weighted matrix (pwm) search tool identified a total of four potential CsrA binding sites, including the two identified by footprinting (Baker, Morozov *et al.*, 2002; Baker, Eory *et al.*, 2007). As protection at BS1 and BS3 (Fig. 3-6b) was not observed by footprinting, a 3'-boundary analysis was performed to search for evidence of these additional CsrA binding sites. The information derived from a boundary experiment differs from the information obtained from a footprint analysis. Whereas a footprint reveals the nucleotides that are protected from enzymatic or chemical probes, 3'-boundary experiments reveal the 3'-nucleotides that are required for binding. Thus, the two techniques are complementary. After RNAs were 5'-end-labeled, base hydrolyzed and incubated in the presence of 1 μM CsrA, samples were fractionated through a native polyacrylamide gel followed by autoradiography. Four diffuse bands were observed corresponding to CsrA-*glg* leader RNA complexes and a single band containing unbound RNA. The RNA from each of these bands was gel purified and subsequently fractionated side by side under denaturing conditions. The cutoffs for the 3'-boundary analysis were relatively sharp and revealed the 3'-boundaries of CsrA binding sites at positions C86, U97, G113 and G128 (Fig. 3-6a and b). Each of these binding sites (BS1-BS4) corresponds to those identified by RNA footprinting and/or the pwm. Thus, boundary analysis confirmed that two CsrA target sequences (BS1 and BS3) lie on either side of the originally characterized upstream site within a stem loop (BS2), for a total of four target sites. The two new sites also are reasonable matches

to the SELEX-derived consensus of 5'-RUACARGGAUGU-3' (Dubey, Baker *et al.*, 2005) and an RNA secondary structure algorithm predicted that BS1 may form a stem loop structure (Fig. 3-6c) (Markham and Zuker, 2005; Markham and Zuker, 2008). The presence of two additional sites strongly suggests a redundant binding mechanism, wherein the deletion of a high affinity CsrA target site, which was identified by footprinting, was partially compensated by one or both of the nearby sites that remained in the mutant *glgC* transcript, reducing the overall impact of the GGA deletion on CsrA-mediated translational repression.

DISCUSSION

CsrA represents a novel class of RNA-binding global regulatory proteins that possess two binding surfaces located on opposite sides of a symmetrical dimer (Gutierrez, Li *et al.*, 2005; Mercante, Suzuki *et al.*, 2006; Lapouge, Sineva *et al.*, 2007). The amino acids that mediate RNA binding have been thoroughly defined, the most important being R44 (Heeb, Kuehne *et al.*, 2006; Mercante, Suzuki *et al.*, 2006). The RNA target to which CsrA binds has also been well characterized, consisting of a semi-conserved sequence surrounding a nearly invariant "GGA" motif, which is often centered in the single-stranded loop of a hairpin (Baker, Morozov *et al.*, 2002; Valverde, Lindell *et al.*, 2004; Dubey, Baker *et al.*, 2005). The present study examined the higher order binding properties of the CsrA dimer. RNA target-site positioning was found to be important for the stability of the resulting ribonucleoprotein complexes, and the presence of both RNA-binding surfaces of CsrA was determined to be important for its function as a translational repressor in a well-defined *in vitro* model system. Thus, our findings

further define the complex manner in which CsrA interacts with RNA targets and ultimately regulates gene expression.

By comparing recombinant CsrA proteins containing one (HD-CsrA) or two (WT-CsrA) functional RNA-binding surfaces, we demonstrated that a WT-CsrA dimer can bind to either one or two RNA molecules. Furthermore, CsrA can form a bridge complex wherein one protein is bound to two target sites of an RNA when they are located as close together as 10 nt or as distant as 63 nt. An intersite distance of 18 nt may provide near optimal spacing between target sites because this distance appeared to compensate for defects in either a secondary RNA target site or a CsrA binding surface, but not both. A spacing of < 18 nt was detrimental for tight bridging, as binding to one of the target sites was easily displaced by the addition of excess CsrA to form a tripartite complex containing two CsrA dimers and one RNA molecule. Conversely, RNAs with intersite distances of ≥ 18 nt formed stable bridge complexes with WT-CsrA, and neither of the protein-bound target sites could be displaced by free CsrA. We note that the RNAs used in these studies were designed to contain CsrA binding sites within single-stranded loops of hairpins, the preferred recognition site (Dubey, Baker *et al.*, 2005). It is possible that spacing constraints may differ in unstructured or alternatively structured RNAs. Presumably, the sequence recognition and spacing constraints of CsrA have evolutionary implications, wherein gene expression may be fine-tuned by altering the sequence of a CsrA target or by increasing or decreasing the distance between binding sites in the 5'-leader of a transcript.

Tethering of CsrA at one RNA target should lead to an increase in the local concentration of secondary sites in the RNA, favoring formation of a bridge complex

(Rippe, 2001). It follows that such binding may create an RNA repression loop or other secondary structure. Research by Rippe (2001) supporting such a model calculated that the optimum interaction distance between two proteins bound to the same RNA is 15-20 nt, and the local protein concentration increases to approximately 1 mM in this case. As this distance deviates from the optimum, the apparent local protein concentration decreases. While this cited work was based on protein-protein association, its conclusions can be applied to the present study, where the second RNA binding site in a dual target RNA is equivalent to a second protein. Thus, we estimate that given an RNA containing two neighboring target sites separated by 15-20 nt, initial binding of CsrA at one site will produce a local concentration on the order of 1 mM for the remaining target site, facilitating low affinity binding interactions. This principle was illustrated by the apparent formation of bridge complexes between HD-CsrA and a dual site RNA with 18 nt spacing (Fig. 3-3c, d, lane 4) or between WT-CsrA and RNA with a single high-affinity binding site (Fig. 3-3c, d, lanes 6 and 7).

Based on previous studies (Gudapaty, Suzuki *et al.*, 2001), we calculated the CsrA binding surface concentration in the cytoplasm to be ~20 μM (data not shown), which is nearly two orders of magnitude below the theoretical local concentration of CsrA when it is bound to a dual-site transcript. Consequently, CsrA binding at one site almost certainly leads to a cooperative interaction at an adjacent site, as opposed to contact by a second free CsrA dimer, under physiological conditions. Indeed, cooperative CsrA-RNA interaction has been observed previously (Dubey, Baker *et al.*, 2003; Weilbacher, Suzuki *et al.*, 2003; Wang, Dubey *et al.*, 2005). Note that our calculation for CsrA surfaces does not distinguish between surfaces bound to RNA

(CsrB/C and mRNA) or RNA-free surfaces. The actual free concentration is likely much lower (sub-micromolar), because most target transcripts identified to date are bound with K_d (apparent equilibrium binding constant) values in the range of 1-40 nM, making bridging interactions even more certain (Baker, Morozov *et al.*, 2002; Dubey, Baker *et al.*, 2003; Weilbacher, Suzuki *et al.*, 2003; Wang, Dubey *et al.*, 2005).

The observed spacing requirement for optimal binding may explain a long-standing observation that the CsrA-CsrB complex is globular in shape (Liu, Gui *et al.*, 1997). *E. coli* CsrB contains as many as 22 possible CsrA target sites with an average spacing of 12.25 nt, and all but 2 pairs of sites are separated by fewer than 18 nt (Table 3-3). This short average spacing distance suggests that a single CsrA protein would not tightly associate with most adjacent target sites. Instead, it is more likely that CsrA usually bridges non-adjacent sites (i.e., sites separated by ≥ 18 nt), giving rise to an energetically stable globular complex. A review of other small regulatory RNAs in the CsrA/RsmA circuitry of related bacteria revealed a similar pattern of short intersite distances (Table 3-3). Consequently, *V. cholerae* CsrB, *P. fluorescens* CHA0 RsmX and RsmZ, and *P. aeruginosa* PrrB would be expected to adopt a globular form when bound by CsrA/RsmA. We previously found that turnover of *E. coli* CsrB and CsrC RNAs requires a non-nucleolytic specificity factor, CsrD, in conjunction with the endonuclease RNase E (Suzuki, Babitzke *et al.*, 2006). Whether the formation of a globular complex with CsrA affects turnover of CsrB through the CsrD-RNaseE pathway or other biological properties of this complex remains an open question. In contrast, CsrA/RsmA-regulated mRNAs contain a higher proportion of target sites separated by ≥ 18 nt

compared to the noncoding regulatory RNAs (Table 3-3), strongly suggesting that CsrA often binds to adjoining target sites in CsrA-regulated mRNAs.

Our studies provide the first experimental demonstration of the function of dual RNA-binding sites of CsrA in regulation. The *glgCAP* 5'-leader contains two previously characterized CsrA target sites, one overlapping the SD sequence and the other 28 nt upstream in an RNA hairpin (Baker, Morozov *et al.*, 2002). Two additional target sites were identified in the present study, which are located on either side of the original upstream hairpin. We found that while the HD-CsrA protein displayed one third the affinity for a single-site oligoribonucleotide target compared to the WT-CsrA (Fig. 3-2), it was ~14-fold less efficient (7-fold based on molar equivalence of RNA binding surfaces) in repressing a *glgC'-'lacZ* reporter fusion (Fig. 3-5). When the GGA sequence of the upsteam RNA target site (Fig. 3-6, BS2) was deleted, regulation by HD-CsrA was unaffected, while the relative difference in regulation between the WT-CsrA and HD-CsrA proteins decreased to 7-fold (~3.5-fold based on equal numbers of binding surfaces) (Fig. 3-5). This finding supports a model of repression wherein two RNA target sites mediate translational repression by interacting simultaneously with one CsrA dimer (Fig 7). Apparently, binding of HD-CsrA only inhibits translation if it blocks the SD sequence, but not if it binds to an upstream stem-loop. For that reason, deletion of one upstream site negatively impacted WT-CsrA activity but did not affect regulation by HD-CsrA. WT-CsrA exhibited a modest, yet still significant, two-fold decrease in activity after deletion of the original stem-loop target (BS2) in the *glgCAP* leader. Deletion of the conserved GGA at BS2 was previously shown to severely decrease, but not abolish, CsrA binding affinity (K_d) for this leader from 39 nM to 150 nM (Baker, Morozov *et al.*, 2002);

therefore, the relatively modest effect with WT-CsrA might be partially explained by residual affinity for the deleted upstream site. The presence of two additional upstream targets (BS1 and BS3), including one that could form an RNA hairpin (Fig. 3-6c), likely compensates for the loss of the apparent high-affinity binding site, and might account for the entire effect. Given the distances between adjacent binding sites in the *glgCAP* 5'-leader (Table 3-3), bridging by CsrA is likely restricted to BS1 and BS3, BS1 and BS4, or BS2 and BS4.

While binding of CsrA at two sites was needed for complete repression in the *glgC* studies, our data suggest that it was not necessarily the occlusion of both target sites by one protein that is essential. Rather, efficient interaction with a low affinity target that directly overlaps the SD may require the tethering of CsrA to a nearby high affinity target within a stem-loop to increase the local concentration of this protein. Note that while primary sequence is critical for RNA recognition by CsrA, localization of a target sequence within the loop of a hairpin, as observed in the *glgC* leader, substantially improves binding to RNA (Dubey, Baker *et al.*, 2005). Thus, the apparent way in which CsrA regulates *glgC* expression may reflect a general repression mechanism that involves initial binding to a high affinity primary RNA target site (Fig. 3-7a), in the case of *glgC* regulation, this is the upstream hairpin (BS2) that was observed in previous footprinting studies or the alternate upstream hairpin (BS1) identified by boundary analysis (Baker, Morozov *et al.*, 2002). A tethered CsrA protein would then recognize and bind to the possibly lower affinity target that overlaps the SD sequence (Fig. 3-7b, c). An important evolutionary implication of this mechanism is that the SD sequence need not be altered

extensively to accommodate weak CsrA binding, which might otherwise affect intrinsic translational capacity of the translation initiation region.

Most negatively-regulated transcripts studied to date contain at least one CsrA target site in the 5'-leader mRNA in addition to a site that overlaps the SD (Table 3-3) (Dubey, Baker *et al.*, 2003; Wang, Dubey *et al.*, 2005; Yakhnin, Pandit *et al.*, 2007; Jonas, Edwards *et al.*, 2008). While our model in figure 3-7 describes the mechanism for an mRNA with two CsrA binding sites, several mRNA targets contain four or more binding sites, including *glgCAP* (Table 3-3). The presence of four CsrA binding sites in the *glgCAP* leader suggests that two CsrA dimers could be bound to any given transcript. While BS3 does not overlap the SD sequence, it is part of the ribosome binding site. Thus, the presence of one CsrA dimer bound to BS1 and BS3 and a second dimer bound to BS2 and BS4 should further reduce ribosome binding, and hence GlgC synthesis.

Because tethering increases the probability of molecular interaction and thus proper regulation, it is not surprising that examples can be found in both eukaryotes and prokaryotes. For instance, translational initiation in eukaryotes involves a large, complex structure which includes the poly(A)-binding protein (PABP), that is tethered to the mRNA 5'-leader and interacts with the poly(A) tail and other regulatory proteins found at the 3'-untranslated region (UTR), such as the GAIT complex (Abaza and Gebauer, 2008). In addition, the bacterial endoribonuclease RNaseE may specifically recognize monophosphates and tether itself to the 5'-termini of transcripts (and possibly to the 3'-poly(A) tail via bacterial PABP) while it cleaves internal targets (Coburn and Mackie, 1999; Kushner, 2002). Additional examples of tethering include the classic prokaryotic transcriptional regulators CI (λ phage repressor), GalR (galactose repressor), LacI

(lactose repressor) and AraC (arabinose activator/repressor), which bind operator sequences near DNA promoters. Tethering promotes protein self-association, bringing non-contiguous stretches of double stranded DNA together, creating repression loops in the case of *gal* (Irani, Orosz *et al.*, 1983; Majumdar and Adhya, 1984) and *lac* (Kramer, Niemoller *et al.*, 1987), activation loops for *ara* (Martin, Huo *et al.*, 1986; Lee and Schleif, 1989) and a combination lytic repressing/lysogenic activating loop for λ (Dodd, Shearwin *et al.*, 2004; Svenningsen, Costantino *et al.*, 2005). Similar to these transcriptional regulators, CsrA may alter the structure of the 5'-leader of a message by tethering distant sites, which by itself could affect ribosome access and translational capacity. However, we suspect the primary regulatory capability of CsrA is based on RNA-binding and direct occlusion of the translation initiation region of repressed genes. In support of this interpretation, we recently demonstrated that CsrA binds and modestly regulates *hfq* mRNA translation using only a single CsrA target site that overlaps the *hfq* SD sequence (Baker, Eory *et al.*, 2007). A role of higher order RNA structural alterations in CsrA-mediated activation processes, such as stabilization of *flhDC* mRNA (Wei, Brun-Zinkernagel *et al.*, 2001), is also possible, but at the present time, no molecular model exists for activation.

We should note that while a looping mechanism may be common to both CsrA and transcriptional regulators, the persistence length (measure of relative polymer stiffness, also known as Kuhn length) of double stranded DNA (~150 bp) dictates that it is approximately 20-fold less flexible than a similar length of single-stranded DNA or RNA (Rippe, 2001). Therefore, CsrA can bridge targets that are much closer together

than those of a dsDNA-binding protein; minimum of 90 bp for LacI (Kramer, Niemoller *et al.*, 1987) and 2.4 kb for CI (Dodd, Shearwin *et al.*, 2004) vs. ≤ 10 nt for CsrA.

As mentioned above, CsrA inhibits *hfq* expression by binding to a single RNA target site that blocks access to the SD sequence (Baker, Eory *et al.*, 2007). Is it possible that another RNA could be bound to the second CsrA binding surface that should be available in this case? Because CsrA contains two independent RNA-binding surfaces, a single CsrA dimer might bind two different messages or small RNAs in a manner reminiscent of Hfq itself (Schubert, Lapouge *et al.*, 2007). Hfq facilitates RNA-RNA interactions through two independent RNA-binding surfaces located on opposite sides of the hexamer; for instance, Hfq facilitates the interaction of DsrA sRNA with the *rpoS* message (Majdalani, Cunning *et al.*, 1998; Mikulecky, Kaw *et al.*, 2004). Although CsrA can bind simultaneously to two independent oligoribonucleotides, further studies are required to know whether such intermolecular bridging has any biological relevance.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

All strains and plasmids used in this study are listed in Table 3-2. For routine culture of bacteria, strains were grown in Luria-Bertani (LB) medium at 37 ^{o}C with shaking (Miller, 1972). Antibiotics were added to media as needed at the following concentrations: ampicillin, 100 μg ml^{-1}; kanamycin, 100 μg ml^{-1}; chloramphenicol, 25 μg ml^{-1}; tetracycline, 10 μg ml^{-1}; spectinomycin, 35 μg ml^{-1}; streptomycin, 35 μg ml^{-1}.

Recombinant DNA techniques

Plasmid pJMCN4 that expresses a wild type CsrA-Intein/Chitin Binding Domain (CBD) fusion monomer was constructed as follows: DNA encoding the exact *csrA* coding region, minus the termination codon, was cloned into the CBD-fusion plasmid pTWIN1 (NEB, Ipswich, MA) creating the CsrA-CBD fusion construct pTC14 (K. Suzuki, unpublished). The XbaI to PstI DNA fragment from pTC14 containing the CsrA-CBD fusion was subcloned into the arabinose-inducible expression vector pBAD33 (Guzman, Belin *et al.*, 1995) creating plasmid pBATC6 and the insert was then confirmed by sequencing (SeqWright DNA Technology, Houston, TX). This plasmid was then converted to Str/Spc^R by cloning a ClaI to NotI DNA fragment from pAH144 (Haldimann and Wanner, 2001) into the ScaI site of pBATC6, creating plasmid pJMCN4. This plasmid was designed such that the CBD was removed from the CsrA-CBD fusion by intein-mediated cleavage.

Protein purification

WT and HD-CsrA proteins were purified from strain TRCF7789B$^-$C$^-$ harboring both plasmids pCSRH6-19 and pJMCN4 or pCSRR44A and pJMCN4 as follows: 20 L of culture was grown overnight at 37^{o}C in LB medium with shaking and arabinose (0.2% final concentration) was added 3 hours prior to cell harvest. The cells were concentrated by centrifugation, resuspended in NEB wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) at 1 g per ml and lysed by sonication or French press. The resulting lysate was cleared by centrifugation and combined with 3 ml of Ni-NTA (NEB, Ipswich, MA) for 1 hr to allow binding of CsrA-His_6 to the nickel-NTA beads. The lysate-Ni-NTA mix was applied to a gravity flow column and the beads were washed with 30 ml of wash buffer containing 20 mM imidazole. CsrA protein was eluted with 10

ml wash buffer containing 250 mM imidazole and then dialyzed against at total of 8L of CBD Buffer #2 (20 mM Tris-HCl pH 7.0, 500 mM NaCl, 1 mM EDTA, NEB Impact Kit). The dialyzed protein was then passed by gravity flow over a column containing chitin beads (NEB, S6651S) that were previously equilibrated with Buffer #2. 10 ml of Buffer #3 (20 mM Tris-HCl pH 8.5, 500 mM NaCl, 1 mM EDTA, 40 mM DTT) was passed over the column, flow was stopped and the cleavage reaction was allowed to continue overnight at 4^{o}C. Flow was continued and the eluant was collected as one fraction. The column was then washed twice with 5 ml Buffer #2 and this was combined with the initial eluate. The pooled eluate was dialyzed against a total of 8L of NEB wash buffer without imidazole. Imidazole was then added to the dialyzed protein to 10 mM and 0.25ml of Ni-NTA slurry was added, incubated at 4^{o}C for 1 hour with gentle shaking and the beads were then applied to a gravity flow column. The beads were washed with wash buffer containing 10 mM imidazole and the HD-CsrA or WT-CsrA were eluted with 5 ml of wash buffer containing 250 mM imidazole. Protein size and purity was assessed by SDS-PAGE (15%) and Coomassie blue staining, which revealed two closely migrating protein bands of equal staining intensity and of the sizes predicted for a WT CsrA monomer (6,855 Da) and a site-directed alanine-substituted CsrA monomer with a C-terminal hexa-histidine tag (7,593 Da) (data not shown). Protein concentration was measured using the bicinchroninic acid (BCA) assay (Smith, Krohn *et al.*, 1985) (Pierce Biotech, Rockford, IL).

RNA Electrophoretic Gel Mobility Shift Assay

RNA gel mobility shift assays were carried out according to previously published procedures (Wang, Dubey *et al.*, 2005; Mercante, Suzuki *et al.*, 2006). Protein

calculations were based on the mass of the CsrA-His_6/CsrA WT dimer (14,516 Da) or the CsrAR44A-His_6/CsrA heterodimer (14,448 Da). RNA probes were designed based on the high-affinity consensus CsrA binding sequence determined by SELEX analysis (Wang, Dubey *et al.*, 2005; Mercante, Suzuki *et al.*, 2006). RNA probes were either synthesized by Integrated DNA Technologies (Coralville, IA) or synthesized *in vitro* with the Megashortscript Kit (Ambion, Austin, TX) and 5'-end-labelled using T4 polynucleotide kinase and [γ-^{32}P]-ATP. RNA radiolabeling, gel shift experiments, detection, quantitation and calculations were conducted according to previously published procedures (Yakhnin, Trimble *et al.*, 2000; Mercante, Suzuki *et al.*, 2006). All EMSAs were performed with [γ-^{32}P]-labeled RNA at 0.06 nM final concentration, unless stated otherwise. EMSAs were conducted at least twice for each set of RNA probes to confirm banding patterns and/or K_d measurements.

Protein-RNA Stoichiometry Measurements

RNA28-2BS and RNA28-1BS (Table 3-1) were 5'-end-labeled with [γ-^{32}P]-ATP and EMSAs were conducted as previously detailed using WT-CsrA and HD-CsrA (Dubey, Baker *et al.*, 2005; Wang, Dubey *et al.*, 2005; Mercante, Suzuki *et al.*, 2006). After electrophoresis, gels were wrapped in plastic and exposed to a flexible Kodak storage phosphor plate (Rochester, NY) for 1 hour, radioactive bands were then visualized with a GE LifeSciences Phosphoimager (Piscataway, NJ). CsrA protein was transferred from the previously visualized gels to a Biorad Immunoblot PVDF membrane (Hercules, CA) by electroblotting in 1X TBE buffer for 45 minutes at 4^{o}C. Blots were blocked with Superblock T20 (cat.# 37536, Pierce Biotech, Rockford, IL) for 1 hour at room temperature according the manufacturer's recommendations. Blots were probed

with a CsrA-specific polyclonal primary antibody (Gudapaty, Suzuki *et al.*, 2001) at 1:10,000 in blocking reagent overnight at 4^{o}C with gentle shaking, they were then washed 12 times with 30 ml of Tris-Buffered Saline with Tween (0.05%) (cat # 28360, Pierce Biotech, Rockford, IL) then probed with anti-rabbit IgG secondary antibody conjugated to HRP (cat.# 31460, Peirce Biotech, Rockford, IL) diluted in blocking buffer at 8 ng/ml for 1 hr at room temperature with gentle shaking. The blot was washed again 12 times with Tris-Buffered Saline with Tween and developed with SuperSignal West Femto chemiluminescent substrate (cat.# 34095, Pierce Biotech, Rockford, IL) according to the manufacturer's recommendations. Blots were then placed in a plastic development folder and visualized using a Biorad Versadoc for chemiluminescent detection. Radiolabeled RNA bands and chemiluminescent protein bands were physically matched up for size and migration distance and then quantified with ImageQuant (GE Life Sciences, Piscataway, NJ) and Quantity One Software (Biorad, Hercules, CA) respectively. The HD-CsrA/RNA10-1BS standard complex was assigned a stoichiometry of 1.0 because this combination of protein and RNA was specifically designed to only allow for a 1:1 ratio; all other protein:RNA ratios were compared relative to this standard complex. Stoichiometry values represent the average of triplicate experiments.

In vitro **coupled transcription-translation**

Effects of WT-CsrA or HD-CsrA proteins on *glgC'-'lacZ* (pCZ3-3) and *glgC'*ΔGGA-*'lacZ* (pCSB25) expression were examined by S30-driven transcription-translation using an S30 extract from a *csrA* mutant strain (TRBW3414), as previously described (Romeo and Preiss, 1989; Liu, Gui *et al.*, 1997). Reaction products were separated by 7.5% SDS-PAGE and radiolabeled proteins were detected by fluorography

using sodium salicylate (Chamberlain, 1979). ^{35}S-methionine incorporation was measured by densitometry as well as by dissolving excised gel slices in hydrogen peroxide and liquid scintillation counting (Romeo and Preiss, 1989).

Boundary Analysis

A 3'-boundary analysis was carried out by modifying previously published procedures (Bevilacqua, George *et al.*, 1998; Dubey, Baker *et al.*, 2003). Plasmid pAltC4 contains portions of the WT *glgCAP* operon leader and *glgC* coding regions (+46 to +179 relative to the start of transcription) cloned downstream of a T7 RNA polymerase promoter (Liu and Romeo, 1997) RNA was synthesized in vitro using the Ambion MEGAscript kit and plasmid pAltC4 that had been linearized with *Bam*H1 as template. Gel-purified RNA was dephosphorylated with calf intestinal alkaline phosphatase and subsequently 5'-end-labeled using [g-^{32}P] ATP and polynucleotide kinase. Labeled RNA was gel purified, suspended in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and renatured by heating to 85 °C followed by slow cooling to room temperature. To generate 5'-labeled alkaline hydrolysis ladders, 60-µl RNA samples (200 nM) were incubated for 5 min at 95 °C in alkaline hydrolysis buffer (100 mM $NaHCO_3/Na_2CO_3$, pH 9.0, 2 mM EDTA) and then recovered by ethanol precipitation. Hydrolyzed RNAs were mixed with 1 µM CsrA (20 µl reaction volume) in binding buffer (10 mM Tris-HCl, pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10% glycerol, 20 mM DTT). Reaction mixtures were incubated for 30 min at 37 °C to allow CsrA-*glgC* RNA complex formation. Samples were fractionated through 10% native polyacrylamide gels. Bound and unbound transcripts were visualized by autoradiography, excised and eluted from the gel. RNAs were ethanol precipitated and fractionated through 10% denaturing polyacrylamide gels. RNase T1

and alkali digestion ladders of the same transcript were prepared as described previously and used as molecular size standards (Bevilacqua and Bevilacqua, 1998).

ACKNOWLEDGEMENTS

We thank Kazushi Suzuki for construction of the CsrA-intein gene fusion. These studies were supported by the NIH (R01 GM059969).

REFERENCES

Abaza, I. and F. Gebauer (2008). "Trading translation with RNA-binding proteins." RNA **14**(3): 404-9.

Ang, S., Y. T. Horng, J. C. Shu, P. C. Soo, J. H. Liu, W. C. Yi, H. C. Lai, K. T. Luh, S. W. Ho and S. Swift (2001). "The role of RsmA in the regulation of swarming motility in *Serratia marcescens*." J Biomed Sci **8**(2): 160-9.

Antson, A. A., E. J. Dodson, G. Dodson, R. B. Greaves, X. Chen and P. Gollnick (1999). "Structure of the *trp* RNA-binding attenuation protein, TRAP, bound to RNA." Nature **401**(6750): 235-42.

Babitzke, P., C. S. Baker and T. Romeo (2009). "Regulation of Translation Initiation by RNA Binding Proteins." Annu Rev Microbiol.

Babitzke, P., D. G. Bear and C. Yanofsky (1995). "TRAP, the *trp* RNA-binding attenuation protein of *Bacillus subtilis*, is a toroid-shaped molecule that binds transcripts containing GAG or UAG repeats separated by two nucleotides." Proc Natl Acad Sci U S A **92**(17): 7916-20.

Babitzke, P., J. T. Stults, S. J. Shire and C. Yanofsky (1994). "TRAP, the *trp* RNA-binding attenuation protein of *Bacillus subtilis*, is a multisubunit complex that appears to recognize G/UAG repeats in the *trpEDCFBA* and *trpG* transcripts." J Biol Chem **269**(24): 16597-604.

Baker, C. S., L. A. Eory, H. Yakhnin, J. Mercante, T. Romeo and P. Babitzke (2007). "CsrA inhibits translation initiation of *Escherichia coli hfq* by binding to a single site overlapping the Shine-Dalgarno sequence." J Bacteriol **189**(15): 5472-81.

Baker, C. S., I. Morozov, K. Suzuki, T. Romeo and P. Babitzke (2002). "CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*." Mol Microbiol **44**(6): 1599-610.

Beuth, B., S. Pennell, K. B. Arnvig, S. R. Martin and I. A. Taylor (2005). "Structure of a *Mycobacterium tuberculosis* NusA-RNA complex." EMBO J **24**(20): 3576-87.

Bevilacqua, J. M. and P. C. Bevilacqua (1998). "Thermodynamic analysis of an RNA combinatorial library contained in a short hairpin." Biochemistry **37**(45): 15877-84.

Bevilacqua, P. C., C. X. George, C. E. Samuel and T. R. Cech (1998). "Binding of the protein kinase PKR to RNAs with secondary structure defects: role of the tandem A-G mismatch and noncontiguous helixes." Biochemistry **37**(18): 6303-16.

Bolognani, F. and N. I. Perrone-Bizzozero (2008). "RNA-protein interactions and control of mRNA stability in neurons." J Neurosci Res **86**(3): 481-9.

Burd, C. G. and G. Dreyfuss (1994). "Conserved structures and diversity of functions of RNA-binding proteins." Science **265**(5172): 615-21.

Burrowes, E., A. Abbas, A. O'Neill, C. Adams and F. O'Gara (2005). "Characterisation of the regulatory RNA RsmB from *Pseudomonas aeruginosa* PAO1." Res Microbiol **156**(1): 7-16.

Chamberlain, J. P. (1979). "Fluorographic detection of radioactivity in polyacrylamide gels with the water-soluble fluor, sodium salicylate." Anal Biochem **98**(1): 132-5.

Chao, N. X., K. Wei, Q. Chen, Q. L. Meng, D. J. Tang, Y. Q. He, G. T. Lu, B. L. Jiang, X. X. Liang, J. X. Feng, B. Chen and J. L. Tang (2008). "The *rsmA*-like gene *rsmA*(Xcc) of *Xanthomonas campestris* pv. *campestris* is involved in the control of various cellular processes, including pathogenesis." Mol Plant Microbe Interact **21**(4): 411-23.

Chmiel, N. H., D. C. Rio and J. A. Doudna (2006). "Distinct contributions of KH domains to substrate binding affinity of Drosophila P-element somatic inhibitor protein." Rna **12**(2): 283-91.

Coburn, G. A. and G. A. Mackie (1999). "Degradation of mRNA in *Escherichia coli*: an old problem with some new twists." Prog Nucleic Acid Res Mol Biol **62**: 55-108.

Dodd, I. B., K. E. Shearwin, A. J. Perkins, T. Burr, A. Hochschild and J. B. Egan (2004). "Cooperativity in long-range gene regulation by the lambda CI repressor." Genes Dev **18**(3): 344-54.

Dubey, A. K., C. S. Baker, T. Romeo and P. Babitzke (2005). "RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction." Rna **11**(10): 1579-87.

Dubey, A. K., C. S. Baker, K. Suzuki, A. D. Jones, P. Pandit, T. Romeo and P. Babitzke (2003). "CsrA regulates translation of the *Escherichia coli* carbon starvation gene,

cstA, by blocking ribosome access to the *cstA* transcript." J Bacteriol **185**(15): 4450-60.

Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer and A. Bateman (2008). "The Pfam protein families database." Nucleic Acids Res **36**(Database issue): D281-8.

Fortune, D. R., M. Suyemoto and C. Altier (2006). "Identification of CsrC and characterization of its role in epithelial cell invasion in *Salmonella enterica* serovar Typhimurium." Infect Immun **74**(1): 331-9.

Glisovic, T., J. L. Bachorik, J. Yong and G. Dreyfuss (2008). "RNA-binding proteins and post-transcriptional gene regulation." FEBS Lett **582**(14): 1977-86.

Gopal, B., L. F. Haire, S. J. Gamblin, E. J. Dodson, A. N. Lane, K. G. Papavinasasundaram, M. J. Colston and G. Dodson (2001). "Crystal structure of the transcription elongation/anti-termination factor NusA from *Mycobacterium tuberculosis* at 1.7 A resolution." J Mol Biol **314**(5): 1087-95.

Gudapaty, S., K. Suzuki, X. Wang, P. Babitzke and T. Romeo (2001). "Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*." J Bacteriol **183**(20): 6017-27.

Gutierrez, P., Y. Li, M. J. Osborne, E. Pomerantseva, Q. Liu and K. Gehring (2005). "Solution structure of the carbon storage regulator protein CsrA from *Escherichia coli*." J Bacteriol **187**(10): 3496-501.

Guzman, L. M., D. Belin, M. J. Carson and J. Beckwith (1995). "Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter." J Bacteriol **177**(14): 4121-30.

Haldimann, A. and B. L. Wanner (2001). "Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria." J Bacteriol **183**(21): 6384-93.

He, F., W. Dang, C. Abe, K. Tsuda, M. Inoue, S. Watanabe, N. Kobayashi, T. Kigawa, T. Matsuda, T. Yabuki, M. Aoki, E. Seki, T. Harada, Y. Tomabechi, T. Terada, M. Shirouzu, A. Tanaka, P. Guntert, Y. Muto and S. Yokoyama (2009). "Solution structure of the RNA binding domain in the human muscleblind-like protein 2." Protein Sci **18**(1): 80-91.

Heeb, S., C. Blumer and D. Haas (2002). "Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0." J Bacteriol **184**(4): 1046-56.

Heeb, S., S. A. Kuehne, M. Bycroft, S. Crivii, M. D. Allen, D. Haas, M. Camara and P. Williams (2006). "Functional analysis of the post-transcriptional regulator RsmA reveals a novel RNA-binding site." J Mol Biol **355**(5): 1026-36.

Heurlier, K., F. Williams, S. Heeb, C. Dormond, G. Pessi, D. Singer, M. Camara, P. Williams and D. Haas (2004). "Positive control of swarming, rhamnolipid synthesis, and lipase production by the posttranscriptional RsmA/RsmZ system in *Pseudomonas aeruginosa* PAO1." J Bacteriol **186**(10): 2936-45.

Hopcroft, N. H., A. Manfredo, A. L. Wendt, A. M. Brzozowski, P. Gollnick and A. A. Antson (2004). "The interaction of RNA with TRAP: the role of triplet repeats and separating spacer nucleotides." J Mol Biol **338**(1): 43-53.

Irani, M. H., L. Orosz and S. Adhya (1983). "A control element within a structural gene: the *gal* operon of *Escherichia coli*." Cell **32**(3): 783-8.

Jackson, D. W., K. Suzuki, L. Oakford, J. W. Simecka, M. E. Hart and T. Romeo (2002). "Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*." J Bacteriol **184**(1): 290-301.

Jonas, K., A. N. Edwards, R. Simm, T. Romeo, U. Romling and O. Melefors (2008). "The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins." Mol Microbiol **70**(1): 236-57.

Kay, E., C. Dubuis and D. Haas (2005). "Three small RNAs jointly ensure secondary metabolism and biocontrol in *Pseudomonas fluorescens* CHA0." Proc Natl Acad Sci U S A **102**(47): 17136-41.

Kramer, H., M. Niemoller, M. Amouyal, B. Revet, B. von Wilcken-Bergmann and B. Muller-Hill (1987). "*lac* repressor forms loops with linear DNA carrying two suitably spaced *lac* operators." EMBO J **6**(5): 1481-91.

Kushner, S. R. (2002). "mRNA decay in *Escherichia coli* comes of age." J Bacteriol **184**(17): 4658-65; discussion 4657.

Lapouge, K., E. Sineva, M. Lindell, K. Starke, C. S. Baker, P. Babitzke and D. Haas (2007). "Mechanism of *hcnA* mRNA recognition in the Gac/Rsm signal transduction pathway of *Pseudomonas fluorescens*." Mol Microbiol **66**(2): 341-56.

Lee, D. H. and R. F. Schleif (1989). "In vivo DNA loops in *araCBAD*: size limits and helical repeat." Proc Natl Acad Sci U S A **86**(2): 476-80.

Lenz, D. H., M. B. Miller, J. Zhu, R. V. Kulkarni and B. L. Bassler (2005). "CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*." Mol Microbiol **58**(4): 1186-202.

Liu, M. Y., G. Gui, B. Wei, J. F. Preston, 3rd, L. Oakford, U. Yuksel, D. P. Giedroc and T. Romeo (1997). "The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*." J Biol Chem **272**(28): 17502-10.

Liu, M. Y. and T. Romeo (1997). "The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein." J Bacteriol **179**(14): 4639-42.

Liu, M. Y., H. Yang and T. Romeo (1995). "The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability." J Bacteriol **177**(10): 2663-72.

Lunde, B. M., C. Moore and G. Varani (2007). "RNA-binding proteins: modular design for efficient function." Nat Rev Mol Cell Biol **8**(6): 479-90.

Majdalani, N., C. Cunning, D. Sledjeski, T. Elliott and S. Gottesman (1998). "DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription." Proc Natl Acad Sci U S A **95**(21): 12462-7.

Majumdar, A. and S. Adhya (1984). "Demonstration of two operator elements in *gal*: in vitro repressor binding studies." Proc Natl Acad Sci U S A **81**(19): 6100-4.

Markham, N. R. and M. Zuker (2005). "DINAMelt web server for nucleic acid melting prediction." Nucleic Acids Res **33**(Web Server issue): W577-81.

Markham, N. R. and M. Zuker (2008). "UNAFold: software for nucleic acid folding and hybridization." Methods Mol Biol **453**: 3-31.

Martin, K., L. Huo and R. F. Schleif (1986). "The DNA loop model for ara repression: AraC protein occupies the proposed loop sites in vivo and repression-negative mutations lie in these same sites." Proc Natl Acad Sci U S A **83**(11): 3654-8.

Mercante, J., K. Suzuki, X. Cheng, P. Babitzke and T. Romeo (2006). "Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance." J Biol Chem **281**(42): 31832-42.

Mikulecky, P. J., M. K. Kaw, C. C. Brescia, J. C. Takach, D. D. Sledjeski and A. L. Feig (2004). "*Escherichia coli* Hfq has distinct interaction surfaces for DsrA, *rpoS* and poly(A) RNAs." Nat Struct Mol Biol **11**(12): 1206-14.

Miller, J. H. (1972). Experiments in molecular genetics. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Mori, D., N. Sasagawa, Y. Kino and S. Ishiura (2008). "Quantitative analysis of CUG-BP1 binding to RNA repeats." J Biochem **143**(3): 377-83.

Mulcahy, H., J. O'Callaghan, E. P. O'Grady, M. D. Macia, N. Borrell, C. Gomez, P. G. Casey, C. Hill, C. Adams, C. G. Gahan, A. Oliver and F. O'Gara (2008). "*Pseudomonas aeruginosa* RsmA plays an important role during murine infection by influencing colonization, virulence, persistence, and pulmonary inflammation." Infect Immun **76**(2): 632-8.

Nogueira, T. and M. Springer (2000). "Post-transcriptional control by global regulators of gene expression in bacteria." Curr Opin Microbiol **3**(2): 154-8.

Rife, C., R. Schwarzenbacher, D. McMullan, P. Abdubek, E. Ambing, H. Axelrod, T. Biorac, J. M. Canaves, H. J. Chiu, A. M. Deacon, M. DiDonato, M. A. Elsliger, A. Godzik, C. Grittini, S. K. Grzechnik, J. Hale, E. Hampton, G. W. Han, J.

Haugen, M. Hornsby, L. Jaroszewski, H. E. Klock, E. Koesema, A. Kreusch, P. Kuhn, S. A. Lesley, M. D. Miller, K. Moy, E. Nigoghossian, J. Paulsen, K. Quijano, R. Reyes, E. Sims, G. Spraggon, R. C. Stevens, H. van den Bedem, J. Velasquez, J. Vincent, A. White, G. Wolf, Q. Xu, K. O. Hodgson, J. Wooley and I. A. Wilson (2005). "Crystal structure of the global regulatory protein CsrA from *Pseudomonas putida* at 2.05 A resolution reveals a new fold." Proteins **61**(2): 449-53.

Rippe, K. (2001). "Making contacts on a nucleic acid polymer." Trends Biochem Sci **26**(12): 733-40.

Romeo, T., M. Gong, M. Y. Liu and A. M. Brun-Zinkernagel (1993). "Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties." J Bacteriol **175**(15): 4744-55.

Romeo, T. and J. Preiss (1989). "Genetic regulation of glycogen biosynthesis in *Escherichia coli*: in vitro effects of cyclic AMP and guanosine 5'-diphosphate 3'-diphosphate and analysis of in vivo transcripts." J Bacteriol **171**(5): 2773-82.

Sabnis, N. A., H. Yang and T. Romeo (1995). "Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene *csrA*." J Biol Chem **270**(49): 29096-104.

SantaLucia, J., Jr. (1998). "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics." Proc Natl Acad Sci U S A **95**(4): 1460-5.

Schubert, M., K. Lapouge, O. Duss, F. C. Oberstrass, I. Jelesarov, D. Haas and F. H. Allain (2007). "Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA." Nat Struct Mol Biol **14**(9): 807-13.

Silhavy, T. J., M. L. Berman, L. W. Enquist and Cold Spring Harbor Laboratory. (1984). Experiments with gene fusions. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Smith, P. K., R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson and D. C. Klenk (1985). "Measurement of protein using bicinchoninic acid." Anal Biochem **150**(1): 76-85.

Sorger-Domenigg, T., E. Sonnleitner, V. R. Kaberdin and U. Blasi (2007). "Distinct and overlapping binding sites of *Pseudomonas aeruginosa* Hfq and RsmA proteins on the non-coding RNA RsmY." Biochem Biophys Res Commun **352**(3): 769-73.

Suzuki, K., P. Babitzke, S. R. Kushner and T. Romeo (2006). "Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E." Genes Dev **20**(18): 2605-17.

Svenningsen, S. L., N. Costantino, D. L. Court and S. Adhya (2005). "On the role of Cro in lambda prophage induction." Proc Natl Acad Sci U S A **102**(12): 4465-9.

Valverde, C., M. Lindell, E. G. Wagner and D. Haas (2004). "A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of *Pseudomonas fluorescens*." J Biol Chem **279**(24): 25066-74.

Wang, X., A. K. Dubey, K. Suzuki, C. S. Baker, P. Babitzke and T. Romeo (2005). "CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a

biofilm polysaccharide adhesin of *Escherichia coli*." Mol Microbiol **56**(6): 1648-63.

Wei, B. L., A. M. Brun-Zinkernagel, J. W. Simecka, B. M. Pruss, P. Babitzke and T. Romeo (2001). "Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*." Mol Microbiol **40**(1): 245-56.

Weilbacher, T., K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke and T. Romeo (2003). "A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*." Mol Microbiol **48**(3): 657-70.

Yakhnin, A. V., J. J. Trimble, C. R. Chiaro and P. Babitzke (2000). "Effects of mutations in the L-tryptophan binding pocket of the Trp RNA-binding attenuation protein of *Bacillus subtilis*." J Biol Chem **275**(6): 4519-24.

Yakhnin, H., P. Pandit, T. J. Petty, C. S. Baker, T. Romeo and P. Babitzke (2007). "CsrA of *Bacillus subtilis* regulates translation initiation of the gene encoding the flagellin protein (*hag*) by blocking ribosome binding." Mol Microbiol **64**(6): 1605-20.

Zuker, M. (1989). "On finding all suboptimal foldings of an RNA molecule." Science **244**(4900): 48-52.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-15.

Table 3-1. Synthetic RNA oligonucleotides used in this study. [a]

Name	Intersite Distance	Linker size (N)	Sequence
RNA-1A			5'-GGCACAAGGAUGUGCC-3'
RNA-1B			5'-CGCACAAGGAUGUGCG-3'
RNA-2A			5'-GGCACAACCCUGUGCC-3'
RNA-2B			5'-GGCACAAAAAUGUGCC-3'
RNA10-2BS	10	1	5'-CACAAGGAUGUG (N) GACAAGGAUGUC-3'
RNA10-1BS	10	1	5'-CACAAGGAUGUG (N) GACAAAAAUGUC-3'
RNA14-2BS	14	1	5'-GGCACAAGGAUGUGCC (N) CGCACAAGGAUGUGCG-3'
RNA14-1BS	14	1	5'-GGCACAAGGAUGUGCC (N) CGCACAACCCUGUGCG-3'
RNA18-2BS	18	5	5'-GGCACAAGGAUGUGCC (N) CGCACAAGGAUGUGCG-3'
RNA18-1BS	18	5	5'-GGCACAAGGAUGUGCC (N) CGCACAAAAAUGUGCG-3'
RNA28-2BS	28	15	5'-GGCACAAGGAUGUGCC (N) CGCACAAGGAUGUGCG-3'
RNA28-1BS	28	15	5'-GGCACAAGGAUGUGCC (N) CGCACAAAAAUGUGCG-3'
RNA43-2BS	43	30	5'-GGCACAAGGAUGUGCC (N) CGCACAAGGAUGUGCG-3'
RNA43-1BS	43	30	5'-GGCACAAGGAUGUGCC (N) CGCACAAAAAUGUGCG-3'
RNA63-2BS	63	50	5'-GGCACAAGGAUGUGCC (N) CGCACAAGGAUGUGCG-3'
RNA63-1BS	63	50	5'-GGCACAAGGAUGUGCC (N) CGCACAACCCUGUGCG-3'

[a] Unless otherwise noted, the linker (N) was composed of a poly (A) tract. The intersite distance is measured in ribonucleotides starting on either side of the invariant "GGA" motif. Arrows under each sequence indicate annealing strands of a stem in a likely stem-loop structure as modeled by MFOLD (Zuker, 1989; Zuker, 2003). The conserved GGA nucleotides of high affinity binding targets are highlighted in gray. Oligoribonucleotide nomenclature reflects both the intersite distance and the number of functional target sites (ie.,intersite distance of 10 nt with 2 binding sites = RNA10-2BS). RNAs were supplied by Integrated DNA Technologies (Coralville, IA) or synthesized using the Megashortscript Kit (Ambion, Austin, TX). During the course of experiments, GGA replacements were changed from CCC to AAA when it was revealed that a CCC trinucleotide could potentially facilitate unwanted RNA oligonucleotide secondary structures as predicted by MFOLD (Zuker, 1989; Zuker, 2003).

Table 3-2. Bacterial strains and plasmids used in this study**.**

Strain or plasmid	Description or relevant genotype	Source or reference
E. coli K-12 Strains		
TRBW3414	*csrA::kan, rpoS*	(Romeo, Gong *et al.*, 1993)
TRCF7789B$^-$C$^-$	CF7789 *csrA*::*kan* Δ*csrB::cam* Δ*csrC::tet*	(Mercante, Suzuki *et al.*, 2006)
Plasmids		
pJMCN4	*aadA* gene from pAH144 cloned into pBATC6; Str^R/Spc^R	This Study
pCZ3-3	Φ*glgC*'-'*lacZ* in pMLB1034, Amp^R	(Romeo and Preiss, 1989)
pCSB25	pCZ3-3 with upstream ΔGGA	(Baker, Morozov *et al.*, 2002)
pAltC4	T7-driven *glgCAP* RNA expression construct (+46 to +179 relative to the start of transcription)	(Liu and Romeo, 1997)
pMLB1034	*'laZ* translational fusion vector; Amp^R	(Silhavy, Berman *et al.*, 1984)
pCSRH6-19	pKK223-3 with C-terminal His_6-tag CsrA; Amp^R	(Liu, Gui *et al.*, 1997)
pAH144	CRIM plasmid, *aadA;* Str^R/Spc^R	(Haldimann and Wanner, 2001)
pTWIN1	Chitin Binding Domain (CBD) fusion plasmid	NEB
pTC14	CsrA-CBD fusion in pTWIN1	K. Suzuki
pBAD33	Arabinose inducible vector; Cm^R	(Guzman, Belin *et al.*, 1995)
pBATC6	CsrA-CBD fragment from pTC14 cloned into pBAD33	This Study
pCSRR44A	*csrA* R44A C-terminal His_6 tag, derived from pCSRH6-19	(Mercante, Suzuki *et al.*, 2006)

Table 3-3. RNA targets of CsrA (or its orthologs) and predicted or experimentally determined spacing[a].

RNA target[a]	Source[b]	Target sites predicted or determined	Average distance between adjacent sites (nt)	Pairs of sites separated by <18 nt	Pairs of sites separated by ≥18 nt	Target overlap with SD?	Reference
CsrB	*E. coli*	~21	12.25	18	2	N/A	(Liu, Gui *et al.*, 1997)
CsrC	*E. coli*	~9	11.4	7	1	N/A	(Weilbacher, Suzuki *et al.*, 2003)
CsrB	*V. cholerae*	~25	11.6	21	3	N/A	(Lenz, Miller *et al.*, 2005)
RsmX	*P. fluorescens*	~5	10	4	0	N/A	(Kay, Dubuis *et al.*, 2005)
RsmZ	*P. fluorescens*	~7	9.8	6	0	N/A	(Heeb, Blumer *et al.*, 2002)
RsmY	*P. fluorescens*	~6	12.8	3	2	N/A	(Valverde, Lindell *et al.*, 2004)
PrrB	*P. aeruginosa*	~6	11	5	0	N/A	(Burrowes, Abbas *et al.*, 2005)
hag	*B. subtilis*	2	26	0	1	Yes	(Yakhnin, Pandit *et al.*, 2007)
ycdT	*E. coli*	2	29	0	1	No[c]	(Jonas, Edwards *et al.*, 2008)
ydeH	*E. coli*	2	11	0	1	Yes[c]	(Jonas, Edwards *et al.*, 2008)
hcnA	*P. fluorescens*	5	15.5	2	2	Yes	(Lapouge, Sineva *et al.*, 2007)
glgC	*E. coli*	4	11.3	3	0	Yes	(Baker, Morozov *et al.*, 2002)
cstA	*E. coli*	4	14	2	1	Yes	(Dubey, Baker *et al.*, 2003)
pgaA	*E.coli*	6	29.4	3	2	Yes	(Wang, Dubey *et al.*, 2005)

[a] sRNAs that bind to CsrA are capitalized (CsrB – PrrB); mRNA leaders are italicized.
[b] Abbreviations: *E.*, *Escherichia*; *V.*, *Vibrio*; *P.*, *Pseudomonas*.
[c] Putative mRNA target site, not tested experimentally. Note that *ycdT* lacks a predicted target site at SD, but contains one overlapping the putative initiation codon, which should also inhibit translation.

FIGURE LEGEND

Fig. 3-1. Electrophoretic Mobility Shift Assay (EMSA) demonstrating that WT-CsrA contains 2 independent RNA binding surfaces while the HD-CsrA contains only one. a) EMSA with WT-CsrA protein showing the presence of a "major" and "minor" shifted complex (arrows), as well as "free RNA", suggesting that CsrA can bind 2 independent RNAs. b) Competition experiment with increasing concentrations of specific unlabeled competitor RNA. Note that addition of competitor RNA caused a redistribution of the WT-CsrA complex to the lower "minor" position, while HD-CsrA failed to exhibit a secondary complex with additional competitor,

Fig. 3-2. EMSA demonstration that HD-CsrA binds to a single-site RNA with approximately one third the avidity of WT-CsrA. a,b) EMSA of WT-CsrA and HD-CsrA proteins and their corresponding binding curves. The avidity of WT-CsrA for RNA, as reflected in the apparent equilibrium binding constant (K_d = 1.3 nM), is greater than that of the HD-CsrA (K_d = 3.4 nM) for the same labeled RNA oligonucleotide. Note that the shorter development time and contrast adjustment in this image vs. Fig. 3-1 deemphasized the "minor complexes". Graphpad Prism version 3 was used for graphical representations and calculations

Fig. 3-3. EMSA analysis of the optimal CsrA binding site spacing using a series of synthetic RNA oligonucleotides. EMSAs were performed with the RNAs listed in Table 3-1. Two concentrations of WT-CsrA or HD-CsrA protein were employed for each series of binding reactions, the first reaction in each binding set (i.e., gel "a" lane 2) contains a low concentration (2.5 nM) of the protein listed, while the reaction to the right of it (i.e., gel "a" lane 3) contains a high concentration (62.5 nM) of CsrA protein. RNAs

containing either one or two high affinity binding sites are also shown. Important shifted species are numbered, described in the results, and correspond to the structural interpretations in the boxed area to the right. The intersite distance and linker size (N) for each RNA oligo used is measured in nucleotides. Experiments were performed at least twice to confirm banding patterns and migration distances.

Fig. 3-4. Stoichiometric measurements of CsrA:RNA ratios within shifted complexes. EMSA experiments were conducted similar to figure 3-3 with WT-CsrA or HD-CsrA in combination with either RNA28-2BS or RNA28-1BS. After detection of labeled RNA by phosporimager analysis (left panel, "RNA-EMSA"), protein was transferred to a PVDF membrane by electroblotting and probed for CsrA protein (right panel, "Western Analysis") using a CsrA-specific polyclonal antibody (Gudapaty, Suzuki *et al.*, 2001); the exact protocol is outlined in the Materials and Methods. RNA and protein bands were quantified using Quantity One software (Biorad, Hercules, CA) and the HD-CsrA:RNA28-1BS (complex 1) ratio was set to 1.0 (explained in materials and methods and results). All other protein:RNA complexes were compared to HD-CsrA:RNA28-1BS for determination of their relative CsrA:RNA ratios as follows: WT-CsrA:RNA28-1BS (complex 2) = 1.12 ± 0.2; HD-CsrA:RNA28-2BS (complex 3) = 1.93 ± 0.1; WT-CsrA:RNA28-2BS (complex 4) = 0.96 ± 0.1. Stoichiometric analyses were performed in triplicate and representative EMSA and western blots are shown.

Fig. 3-5. Effects of WT-CsrA or HD-CsrA on the expression of *glgC'-'lacZ* or *glgC*ΔGGA'-*'lacZ* translational fusions in S-30 coupled transcription-translation reactions. a) Reactions contained 2 μg of supercoiled plasmid pCZ3-3 (*glgC'-'lacZ*) and either WT-CsrA (top panel) or HD-CsrA (bottom panel) at the concentrations indicated;

right panel graphically displays quantified results. b) Similar reactions were conducted with 2 μg of supercoiled plasmid pCSB25 (*glgC*ΔGGA'-*'lacZ*). The concentration of CsrA that leads to a 50% inhibition of gene expression (IC_{50}) was calculated based on a simple point-to-point curve and a Hill-Slope (4PL) model, both of which gave similar values. Graphpad Prism version 3 was used for graphical representations and IC_{50} calculations. *In vitro* transcription-translation assays were conducted at least twice for each set, with similar results.

Fig. 3-6. Boundary Analysis revealing additional uncharacterized CsrA target sites in the *glgCAP* 5'-leader. A) Limited alkaline hydrolysis ladders of *glg* RNA incubated with 1 μM CsrA. CsrA-RNA complexes were separated from unbound RNA on a native gel and subsequently fractionated through a 10% denaturing gel (shown). Lanes corresponding to distinct bound complexes (B1, B2, B3 or B4) and unbound (U) RNA are shown. Lanes corresponding to limited base hydrolysis (OH) and RNase T1 digestion (T1) ladders are indicated. Positions of the boundaries are marked with arrows on the right. Numbering on the left is from the start of *glgCAP* transcription. b) The *glgCAP* leader transcript contains four CsrA binding sites (BS1-BS4). Positions of the 3' boundaries (B1-B4), as well as the previously identified CsrA footprints are shown (Baker, Morozov *et al.*, 2002). Positions of the *glgC* SD sequence and translation initiation codon (Met) are also shown. c) UNAfold model (an algorithm based on Mfold) (Smith, Krohn *et al.*, 1985) of RNA structure that includes BS1 within the single-stranded region of a hairpin loop. Note that hairpins containing BS1 and BS2 will not form simultaneously because they share common stem sequences. ΔG of this structure and a comparison of BS1 with the

SELEX-derived CsrA consensus (Heeb, Blumer *et al.*, 2002; Kay, Dubuis *et al.*, 2005) sequence is shown (Dubey, Baker *et al.*, 2005).

Fig. 3-7. A model for CsrA regulation by binding to the 5' leader of repressed transcripts, based on results from RNA gel shifts and *glgC* S30 transcription-translation. a) CsrA initially binds at a high affinity to a target site, commonly located within a hairpin structure that lies proximal to the SD. b) After initial binding, the increased local concentration of CsrA allows the free RNA-binding surface to interact with the downstream low affinity target site overlapping the SD sequence. c) Binding of the low affinity target site sequesters the SD sequence, thereby blocking ribosome loading and decreasing translational initiation. SD = Shine Dalgarno; AUG = methionine start codon.

Fig. 3-1

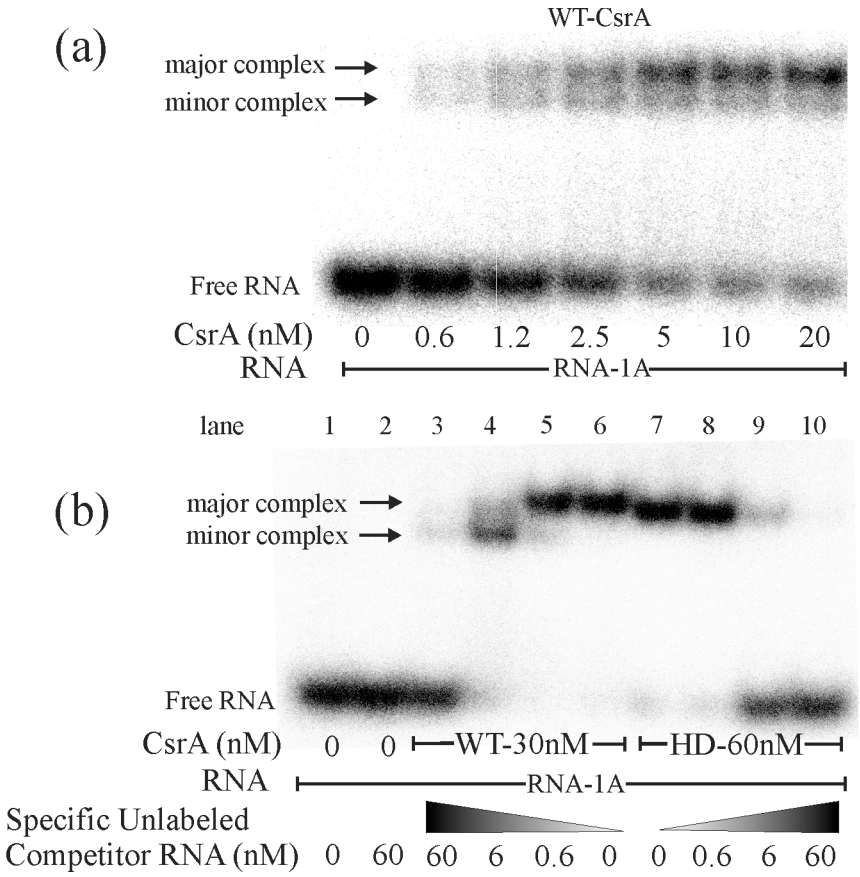

(a)
WT-CsrA
major complex
minor complex
Free RNA
CsrA (nM) 0 0.6 1.2 2.5 5 10 20
RNA
RNA-1A
lane 1 2 3 4 5 6 7 8 9 10
(b)
major complex
minor complex
Free RNA
CsrA (nM) 0 0
WT-30nM
HD-60nM
RNA
RNA-1A
Specific Unlabeled
Competitor RNA (nM) 0 60 60 6 0.6 0 0 0.6 6 60

Fig. 3-2

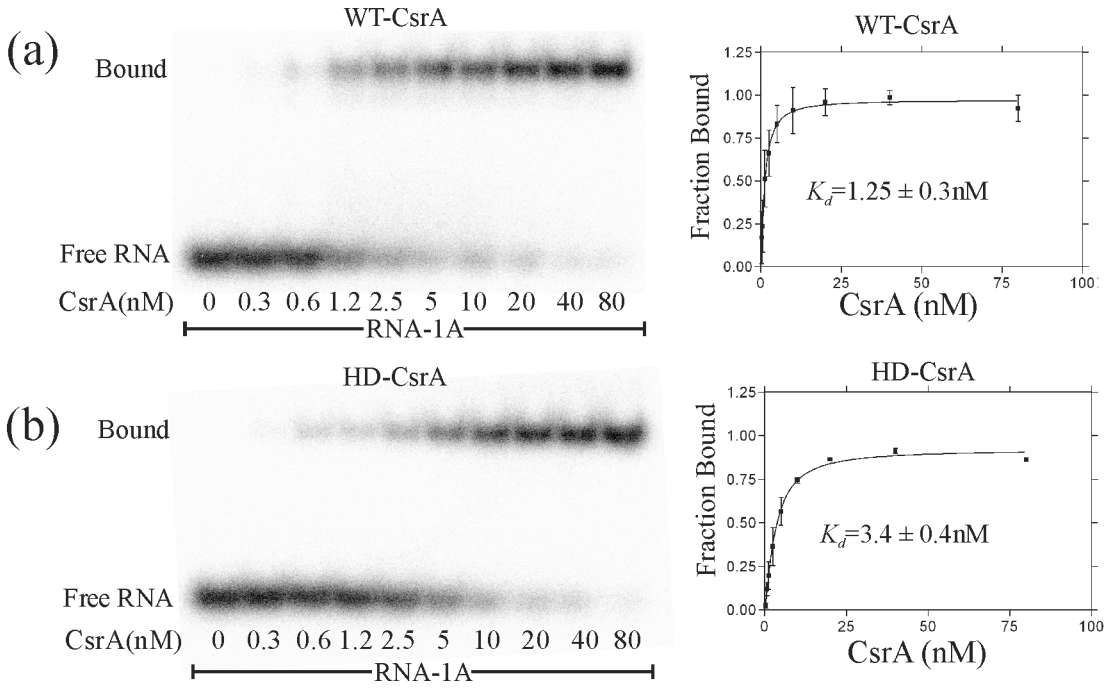

(a)
WT-CsrA
Bound
Free RNA
CsrA(nM) 0 0.3 0.6 1.2 2.5 5 10 20 40 80
RNA-1A
WT-CsrA
Fraction Bound
K_d=1.25 ± 0.3nM
CsrA (nM)
(b)
HD-CsrA
Bound
Free RNA
CsrA(nM) 0 0.3 0.6 1.2 2.5 5 10 20 40 80
RNA-1A
HD-CsrA
Fraction Bound
K_d=3.4 ± 0.4nM
CsrA (nM)

Fig. 3-3a,b,c

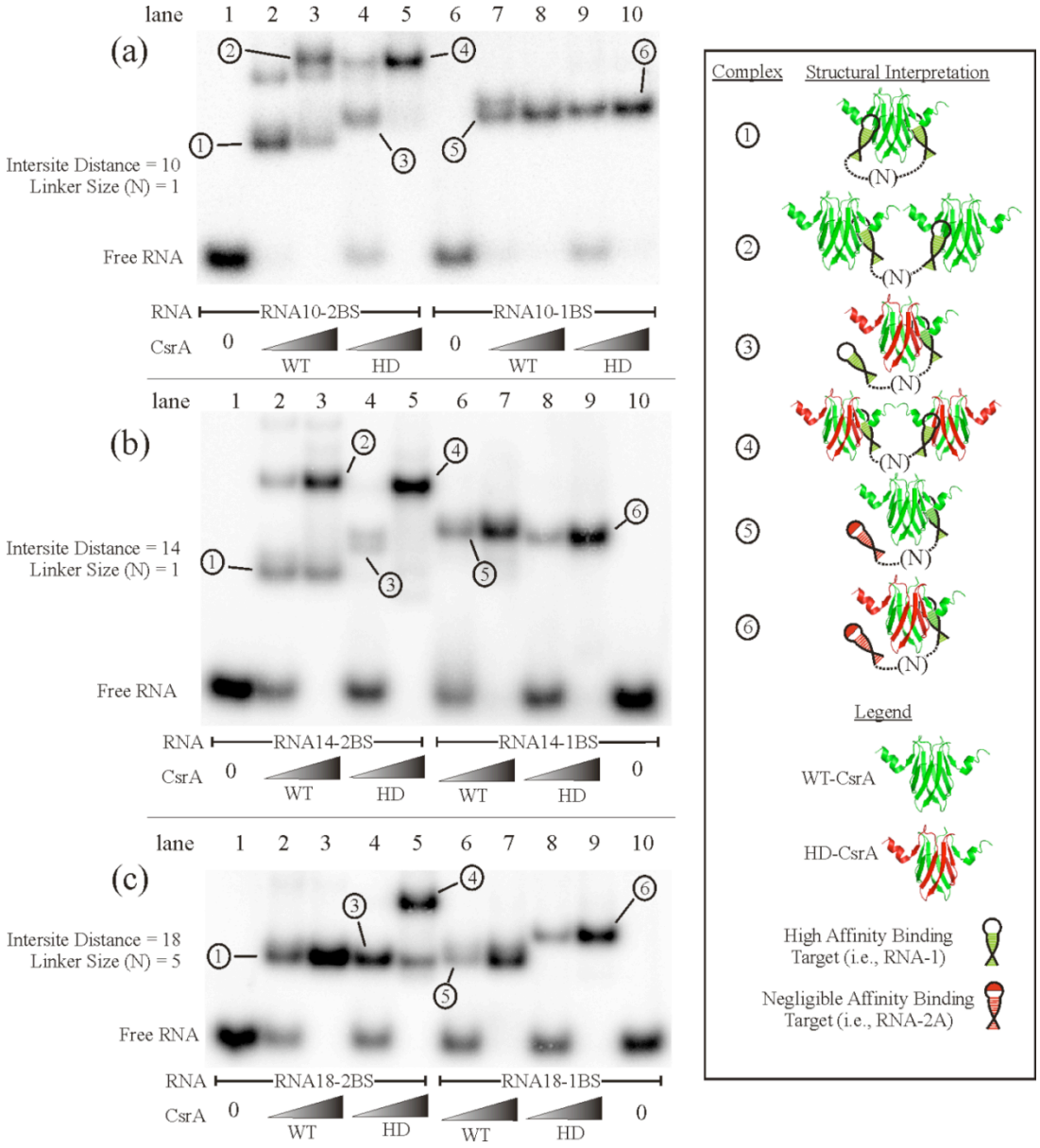

lane 1 2 3 4 5 6 7 8 9 10
(a)
Intersite Distance = 10
Linker Size (N) = 1
Free RNA
RNA RNA10-2BS RNA10-1BS
CsrA 0 WT HD 0 WT HD
lane 1 2 3 4 5 6 7 8 9 10
(b)
Intersite Distance = 14
Linker Size (N) = 1
Free RNA
RNA RNA14-2BS RNA14-1BS
CsrA 0 WT HD WT HD 0
lane 1 2 3 4 5 6 7 8 9 10
(c)
Intersite Distance = 18
Linker Size (N) = 5
Free RNA
RNA RNA18-2BS RNA18-1BS
CsrA 0 WT HD WT HD 0
Complex
Structural Interpretation
(N)
Legend
WT-CsrA
HD-CsrA
High Affinity Binding Target (i.e., RNA-1)
Negligible Affinity Binding Target (i.e., RNA-2A)

Fig. 3-3d,e,f

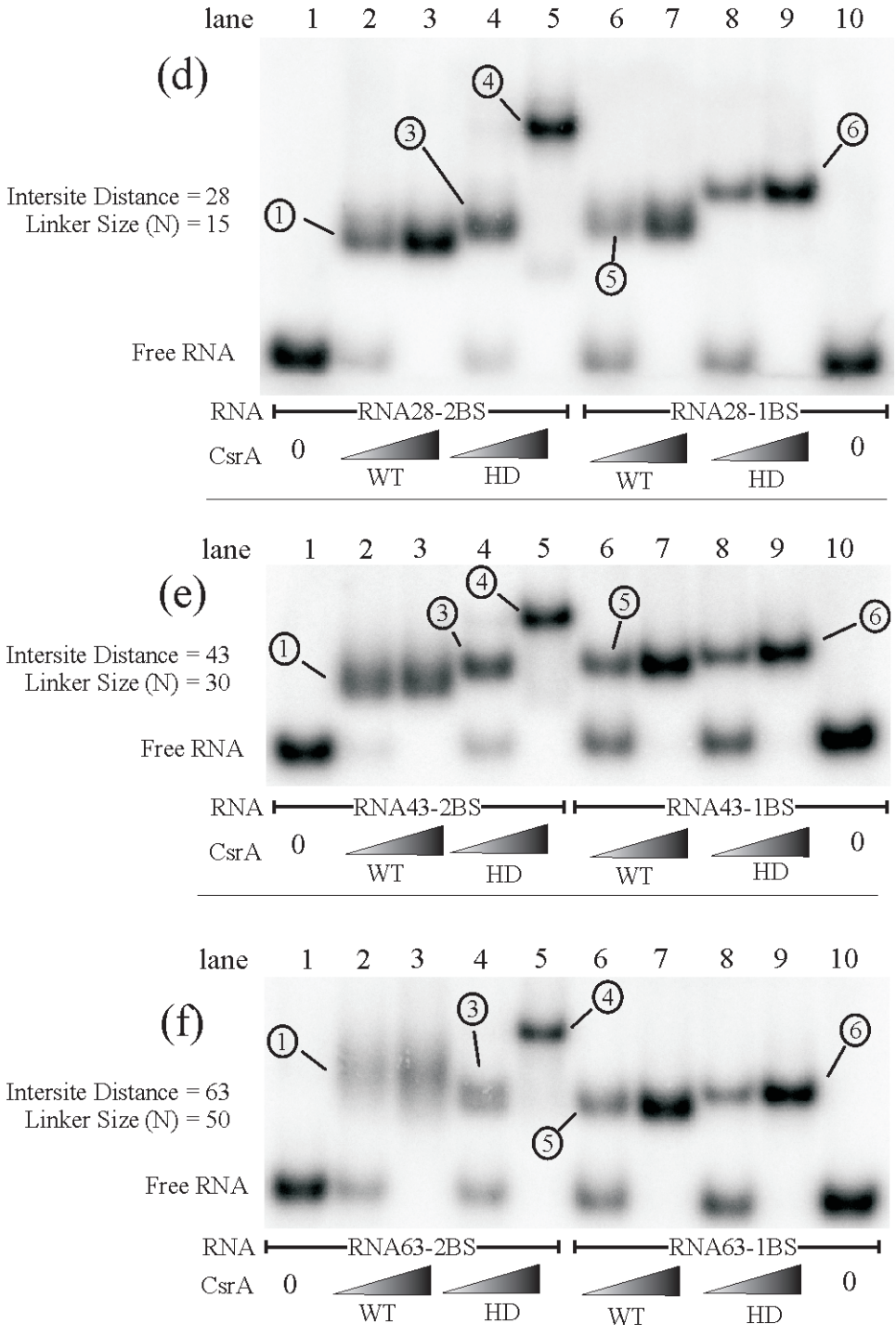

lane 1 2 3 4 5 6 7 8 9 10
(d)
Intersite Distance = 28
Linker Size (N) = 15
Free RNA
RNA RNA28-2BS RNA28-1BS
CsrA 0 WT HD WT HD 0
lane 1 2 3 4 5 6 7 8 9 10
(e)
Intersite Distance = 43
Linker Size (N) = 30
Free RNA
RNA RNA43-2BS RNA43-1BS
CsrA 0 WT HD WT HD 0
lane 1 2 3 4 5 6 7 8 9 10
(f)
Intersite Distance = 63
Linker Size (N) = 50
Free RNA
RNA RNA63-2BS RNA63-1BS
CsrA 0 WT HD WT HD 0

Fig. 3-4

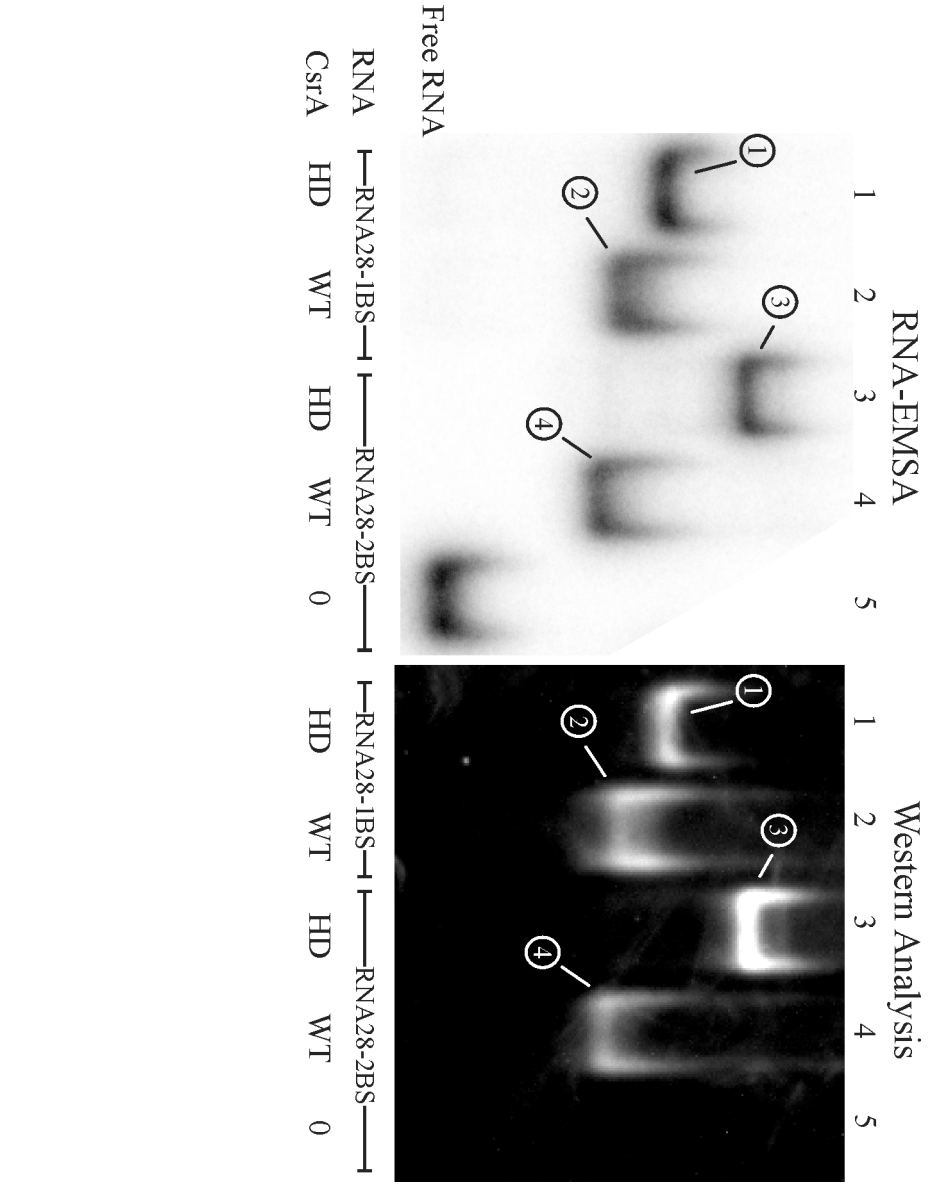

RNA-EMSA
1 2 3 4 5
Free RNA
1
2
3
4
Western Analysis
1 2 3 4 5
1
2
3
4
RNA
RNA28-1BS
RNA28-2BS
RNA28-1BS
RNA28-2BS
CsrA HD WT HD WT 0 HD WT HD WT 0

Fig. 3-5

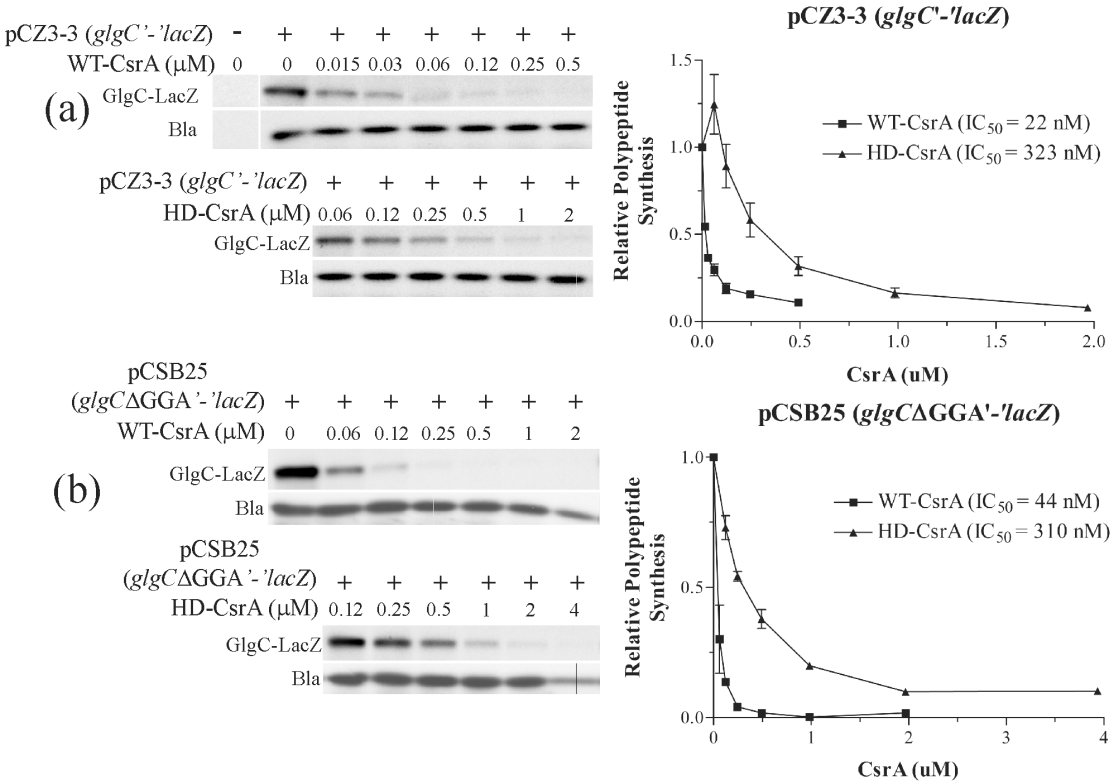

pCZ3-3 (glgC'-'lacZ) - + + + + + + +
WT-CsrA (μM) 0 0 0.015 0.03 0.06 0.12 0.25 0.5
(a)
GlgC-LacZ
Bla
pCZ3-3 (glgC'-'lacZ) + + + + + + +
HD-CsrA (μM) 0.06 0.12 0.25 0.5 1 2
GlgC-LacZ
Bla
pCZ3-3 (glgC'-'lacZ)
Relative Polypeptide Synthesis
WT-CsrA (IC50 = 22 nM)
HD-CsrA (IC50 = 323 nM)
CsrA (uM)
pCSB25 (glgCΔGGA'-'lacZ) + + + + + + +
WT-CsrA (μM) 0 0.06 0.12 0.25 0.5 1 2
(b)
GlgC-LacZ
Bla
pCSB25 (glgCΔGGA'-'lacZ) + + + + + + +
HD-CsrA (μM) 0.12 0.25 0.5 1 2 4
GlgC-LacZ
Bla
pCSB25 (glgCΔGGA'-'lacZ)
Relative Polypeptide Synthesis
WT-CsrA (IC50 = 44 nM)
HD-CsrA (IC50 = 310 nM)
CsrA (uM)

Fig. 3-6

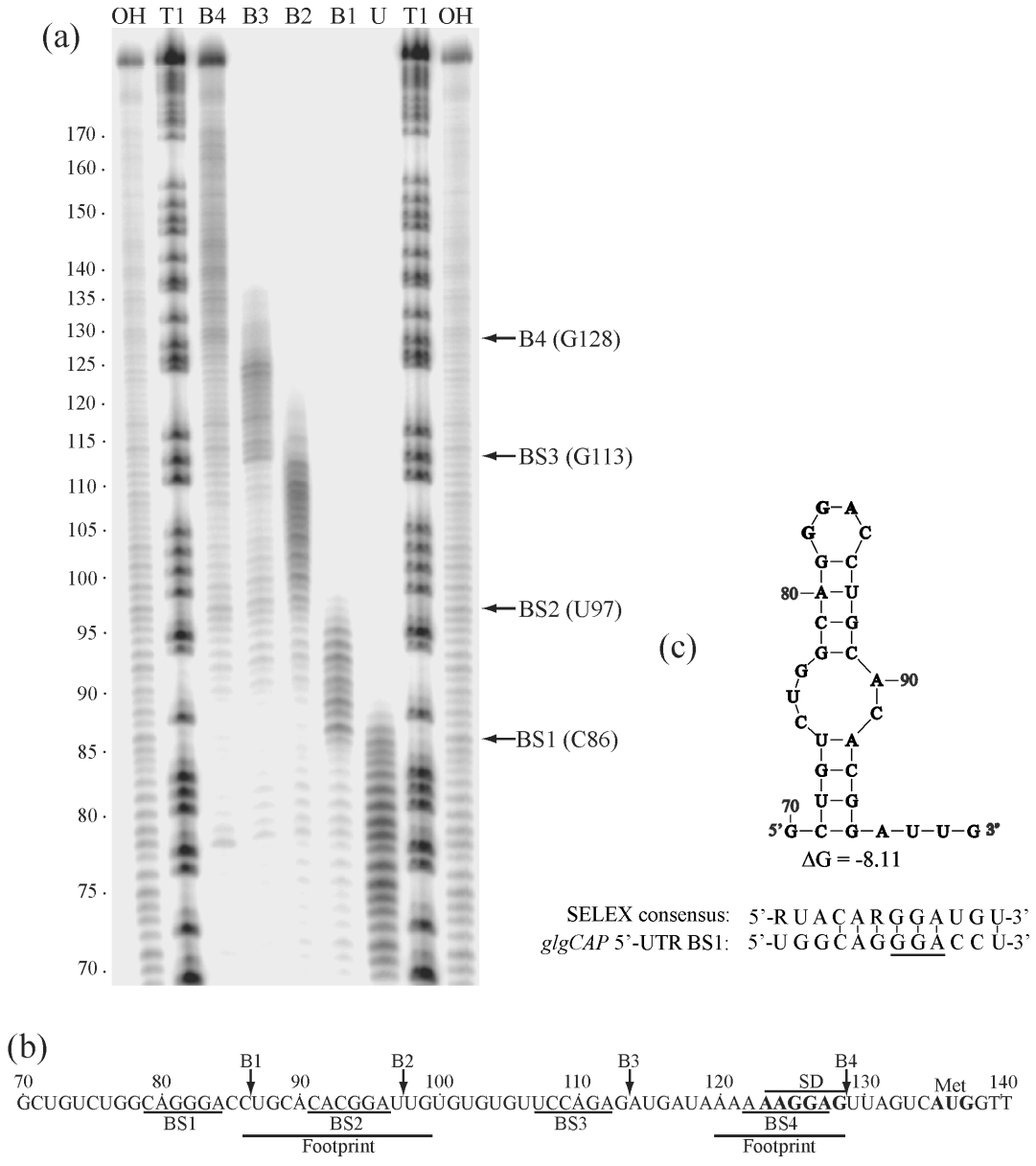
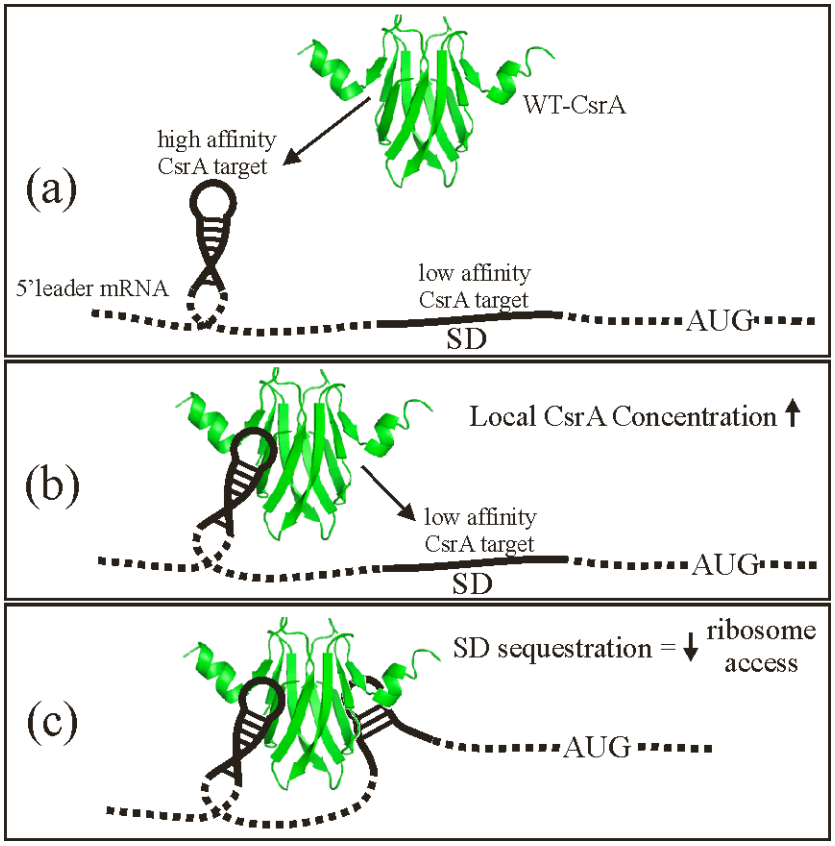

(a)
OH T1 B4 B3 B2 B1 U T1 OH
170
160
150
140
135
130
125
120
115
110
105
100
95
90
85
80
75
70
B4 (G128)
BS3 (G113)
BS2 (U97)
BS1 (C86)
(c)
80
90
70
5'G–C–G–A–U–U–G 3'
ΔG = -8.11
SELEX consensus: 5'-R U A C A R G G A U G U-3'
glgCAP 5'-UTR BS1: 5'-U G G C A G G G A C C U-3'
(b)
B1
B2
B3
B4
70 80 90 100 110 120 SD 130 Met 140
GCUGUCUGGCAGGGACCUGCACACGGAUUGUGUGUUCCAGAGAUGAUAAAAAAGGAGUUAGUCAUGGTT
BS1
BS2
Footprint
BS3
BS4
Footprint

Fig. 3-7


WT-CsrA
high affinity
CsrA target
(a)
5'leader mRNA
low affinity
CsrA target
SD
AUG
(b)
Local CsrA Concentration
low affinity
CsrA target
SD
AUG
(c)
SD sequestration = ribosome access
AUG

Chapter 4. Discussion

CsrA bilateral symmetry and amino acid conservation

Recent progress in X-ray (Heeb, Blumer *et al.,* 2002; Rife, Schwarzenbacher *et al.,* 2005) and 3D-NMR (Gutierrez, Li *et al.,* 2005) solution studies described CsrA as a new class of protein fold. The work presented here built upon this novel structure to clearly define the protein regions which are critical for CsrA-RNA interaction; the contribution of every individual amino acid side chain to proper gene regulation was examined along with the RNA-binding capacity of 8 different CsrA mutant proteins (Mercante, Suzuki *et al.,* 2006). Our studies revealed two protein sub-domains, composed of highly conserved residues, which are essential for both these functions. We consequently established that CsrA regulates gene expression primarily through RNA-binding. Because CsrA is a symmetrical homodimer, any single mutation made to the *csrA* gene results in two corresponding amino acid changes; while basic, this realization was fundamental to our understanding of CsrA structure and function. The present alanine mutagenesis defined a cleft formed between β_1 and β_5' near R44 as a likely RNA binding region or subdomain (Mercante, Suzuki *et al.,* 2006), therefore, the opposite symmetrical face of CsrA also functioned as an RNA-binding surface. This structural prediction was confirmed by subsequent 3D-NMR solution studies of RsmE complexed with 2 independent RNAs (Schubert, Lapouge *et al.,* 2007).

The modular design of RNA-binding proteins allows inclusion of several different binding motifs or multiple copies of the same motif within a protein (e.g., Dicer, vigilin, and Hfq; Brennan and Link, 2007; Lunde, Moore *et al.*, 2007; Glisovic, Bachorik *et al.*, 2008), which increases their binding avidities and specificities. Although CsrA contains

two RNA-binding surfaces, it does not conform to this traditional definition of modularity because these two subdomains are an integral part of the entire protein structure, and are probably inseparable from the hydrophobic core. While it is tempting to speculate that a larger protein could encode several "CsrA domains", it would be difficult to envision because while the CsrA C-termini are solvent-exposed, the N-termini of both polypeptides are buried in hydrophobic pockets and are relatively inaccessible (Gutierrez, Li *et al.,* 2005; Rife, Schwarzenbacher *et al.,* 2005; Mercante, Suzuki *et al.,* 2006). The first four amino acids of the vast majority of CsrA orthologs are conserved (>90%) hydrophobic residues (Fig. 6-2 and data not shown), strongly indicating that the N-terminus must be stably maintained (Clamp, Cuff *et al.,* 2004). Thus, CsrA modularity should only be compatible at the C-terminus, where orthologs are found to diverge greatly in both length and amino acid sequence. It is possible, though by no means certain, that these variable regions may encode as yet unidentified and possibly species-specific functions. Evidence for potential modularity and accessory functions comes from CsrA-like orthologs in *Blastopirellula marina* (Amann, Ferriera *et al.,* 2006) and *Lawsonia intracellularis* (Kaur, Zhang *et al.,* 2005) which contain a probable sensory transduction histidine kinase and a FliW domain, respectively, at the C-terminus. Whether the functionality of either of these apparent fusion proteins involves CsrA-like RNA-binding remains to be determined.

Consistent with an established RNA-binding role, all CsrA orthologs appear to retain the same or similar residues at the RNA-binding surface (Mercante, Suzuki *et al.,* 2006). Indeed, many orthologs can functionally complement an *E. coli csrA* mutant (Cui, Chatterjee *et al.,* 1995; Liu, Cui *et al.,* 1998; Liaw, Lai *et al.,* 2003; Lenz, Miller *et al.,*

2005; Krin, Derzelle *et al.,* 2008; Sahr, Bruggemann *et al.,* 2009), suggesting a common post-transcriptional mechanism retained across different species. However, there are two examples of orthologs that are unable to complement an *E. coli* mutant: CsrA of *Bacillus subtilis* (Yakhnin, Pandit *et al., 2007)* and *Helicobacter pylori* (Barnard, Loughlin *et al.,* 2004). While the 6 N-terminal amino acids of *H. pylori* region 1 (Fig. 2-6) are highly similar to that of *E. coli* (with compatible substitutions R7K and T5S), residues in region 2 near the α-helix are less conserved. Assuming that the *H. pylori* CsrA protein is expressed in *E. coli*, two significant non-equivalent substitutions (V40T and H42L) may account for the incompatibility. It would be interesting to know if these changes affect RNA-binding and, if so, whether the *H. pylori* CsrA target sequences have changed to accommodate this difference in amino acid composition. In contrast, a comparison of *B. subtilis* and *E. coli* CsrA show no major region 1 or region 2 differences, with strictly conserved substitutions such as those listed for *H. pylori* region 1 and similar region 2 changes (V40I and V42I). However, compared to *E. coli*, the *B. subtilis* CsrA contains a T11A substitution that our current studies demonstrated was detrimental to *E. coli* CsrA-mediated regulation of *glgC'-*, *pgaA'-* and *flhDC'-lacZ* expression (Fig. 2-1; Mercante, Suzuki *et al.,* 2006). Curiously, the *E. coli* T11 side chain is solvent exposed, not found at the RNA-binding interface, and 3D modeling does not show it to associate with other important residues (J. Mercante, unpublished results). Thus, it is possible that T11 serves some value in *E. coli*, such as an interface for protein-protein interaction, which has no equivalent in *B. subtilis,* yet is necessary for protein function. However, further studies will be required to resolve this question.

It is interesting to note that while the majority of the ~300 eubacterial species or strains predicted to express CsrA-like proteins have a single ortholog, at least 27 have multiple CsrA paralogs (J. Mercante, unpublished data; Finn, Tate *et al.,* 2008). In fact, some species carry 5 (*Pseudomonas syringae* pv. tomato, *Pseudomonas putida* W619, *Legionella pneumophila*, strains Corby and Paris), 7 (*Blastopirellula marina* DSM 36457) and even 9 (*Planctomyces maris* DSM 8797) copies of *csrA*-like genes (Finn, Tate *et al.,* 2008). Why a bacterium would need so many copies of this gene is not fully understood, as other species, including *S. enterica* (Altier, Suyemoto *et al.,* 2000; Lawhon, Frye *et al.,* 2003), *V. cholerae* (Lenz, Miller *et al.,* 2005), *E. carotovora* (Cui, Chatterjee *et al.,* 1995) and *P. aeruginosa* (Mulcahy, O'Callaghan *et al.,* 2006; Mulcahy, O'Callaghan *et al.,* 2008), in which CsrA dramatically affects growth and cell invasion, quorum sensing, extracellular enzyme production, and virulence factor expression, respectively, encode only a single ortholog. CsrA paralogs within a single species are more often non-identical (i.e., the amino acid identity of 5 CsrA paralogs in *L. pneumophila* strain Corby differ from each other by 20-75%; J. Mercante, unpublished data; Finn, Tate *et al.,* 2008), therefore it is possible that they are non-redundant, perhaps regulating distinct RNA targets. Also likely is that CsrA paralogs bind some of the same RNA targets but with different affinities, or they may be differentially expressed via environmental, temporal or spacial cues. Studies of the *P. fluorescens* CsrA paralogs RsmA and RsmE (Reimmann, Valverde *et al.,* 2005) support these later two hypotheses; the proteins are ~ 70 % identical and cooperate to fully regulate expression of the *hcnA* gene required for hydrogen cyanide synthesis. However, RsmA, which is expressed consistently throughout the cell cycle, binds the RsmY sRNA with an apparent K_d ~550

nM, while RsmE, which is synthesized mainly in stationary phase, has a much higher affinity for the same sRNA (K_d ~138 nM).

The majority of *csrA*-like genes are found in the γ-proteobacteria, yet many α-, δ- and ε-proteobacteria also contain highly conserved orthologs; CsrA-like proteins are also predicted to be present in representatives of the Actinobacteria, Planctomycetes, Spirochaetes, Firmicutes and Thermotogae (J. Mercante, unpublished data). Clues to how *csrA* may have been distributed to such a wide range of eubacteria in different environments comes from two complementary sources. First, *Legionella pneumophila* serogroup 1 Philadelphia 1 strains contain a horizontally acquired ~ 65-kb pathogencity island, LpPI-1, that encodes genes for several virulence factors including a putative type IV secretion system, a multidrug resistance protein (YttB) and a CsrA ortholog, among others (Brassinga, Hiltz *et al.,* 2003). LpPI-1 is flanked on both sides by phage-related genes needed for transposition as well as by tRNA genes that are likely the targets for pathogenecity island integration. Second, a double-stranded DNA phage of Pseudomonads, F116, encodes a CsrA-like protein whose product is ~ 55% identical to *E. coli* CsrA and which is conserved for the critical RNA-binding residues in regions 1 and 2 (Finn, Tate *et al.,* 2008). Thus, while CsrA-like genes are concentrated in the proteobacteria, it is possible that it was seeded throughout the entire eubacterial domain through horizontal gene transfer. However, a more thorough examination of CsrA phylogeny is required to ascertain the origin of *csrA* or its exact method of dissimination throughout bacterial families.

A study of SirA orthologs (UvrY in *E. coli*) suggests that this gene may have originally been a member of a flagellar operon and thus one of its conserved functions is

to regulate motility (Goodier and Ahmer, 2001). Additionally, one consistent observation made in almost every species containing Csr circuitry is that CsrA affects cellular motility (Ang, Horng *et al.,* 2001; Fettes, Forsbach-Birk *et al.,* 2001; Lawhon, Frye *et al.,* 2003; Liaw, Lai *et al.,* 2003; Barnard, Loughlin *et al.,* 2004; Heurlier, Williams *et al.,* 2004; Heroven, Bohme *et al.,* 2008); while the basis for this phenotype has not been investigated in most cases, it appears that CsrA acts directly by targeting expression of flagellar transcription factors such as *flhDC* (Wei, Brun-Zinkernagel *et al.,* 2001) or structural components like FliC (Lawhon, Frye *et al.,* 2003). And an examination of the Csr circuit in species where both CsrA and SirA orthologs have been described suggests their activities are usually linked through the action of CsrB/C or a similar sRNA (J. Mercante, unpublished data). The question then becomes whether CsrA was also part of an ancient flagellar regulatory element or if it was co-opted to regulate expression of existing motility circuits. A brief search of the CsrA-containing species for which we have genomic information reveals that while *csrA* is found within or close to the flagellar operons of *L. intracellularis* (Kaur, Zhang *et al.,* 2005), *B. subtilis* (Yakhnin, Pandit *et al.*, 2007), *Borrelia burgdorferi* and *Clostridium botulinum*, it is more often not located at flagellar loci (J. Mercante, unpublished data; National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov; Finn, Tate *et al.*, 2008). In addition, while a CsrA-like protein has been identified in ~300 species, SirA orthologs are found in ~500 different eubacteria (Finn, Tate *et al.,* 2008). Thus, while it is possible that CsrA was originally a flagellar gene, it is more probable that CsrA was adapted independently in many cases to regulate motility. The mechanism for this adaptation may have naturally evolved from the capacity of CsrA to bind small, structured RNA target sites; and at least

in *E. coli* and *S. enterica*, the extensively structured *flhDC* 5'-untranslated leader may have provided ample binding possibilities.

CsrA molecular geometry and target site binding

The current studies are the first to experimentally examine a role for the dual RNA-binding surface arrangement of CsrA. We dissected the higher-order RNA requirements that allow CsrA to bind two targets simultaneously within a single message; dual target sequences separated by fewer than 18 nt can bound by a single CsrA protein, however, CsrA interaction at one of the two target sites is not stably maintained. These findings may help explain why a CsrA-CsrB complex is globular, due to CsrA tethering of non-adjacent binding targets. We also determined that a transcript with multiple binding sites is regulated more efficiently by a WT-CsrA protein than a heterodimer protein containing only a single RNA-binding surface. These results are likely applicable to other CsrA-regulated genes in *E. coli* and related eubacteria.

An examination of the handful of messages shown to be directly bound by CsrA (Liu and Romeo, 1997; Baker, Morozov *et al.*, 2002; Dubey, Baker *et al.*, 2003; Wang, Dubey *et al.*, 2005; Jonas, Tomenius *et al.*, 2006; Baker, Eory *et al.*, 2007; Yakhnin, Pandit *et al.*, 2007) reveals that most transcripts contain 2 or more target sites (Table 3-3). This fact is quite compelling when considering the RNA-binding properties of CsrA that were revealed in the present studies. With the knowledge that RNA targets must be spaced appropriately for dual-binding, we can now reexamine known CsrA-bound transcripts to better describe how coordinated binding might occur and affect regulation. The *glgCAP* 5'-untranslated leader contains 4 CsrA target sites, one of which overlaps the SD sequence; 2 new binding sites were characterized in this series of studies (Fig. 3-

6b; Baker, Morozov *et al.,* 2002). The distance between site BS1 and BS2 = 9 nt, BS2 and BS3 = 13 nt, and BS3 and BS4 = 12 nt; therefore, no pairs of adjacent sites on this transcript are separated by a sufficient distance as would allow a single CsrA dimer to bind both targets simultaneously. With the requirement that BS4, which overlaps the SD, must be bound in order to occlude ribosome loading, then it is probable that in regulation, a single CsrA homodimer binds either BS1 and BS4 or BS2 and BS4. BS1 and BS2 also form potential stem-loop structures which share common stem nucleotides; thus, the presentation of one binding site (ie., BS2) in a hairpin loop would preclude the other binding site (ie., BS1) from also being displayed in the single-stranded region of a hairpin structure (J. Mercante unpublished results). Current boundary analysis identified BS3 as a CsrA target, therefore it is feasible that a single CsrA protein can bind both BS1 and BS3, which is the only dual-binding combination that is compatible with BS3. Alternatively, BS3 could be bound independently of any other binding target, as was demonstrated for the *cstA* transcript (Dubey, Baker *et al.,* 2003). Thus, there may exist a CsrA concentration-dependent population of CsrA-*glgCAP* complexes that encompass all possible binding site configurations. However, additional experiments will be required to confirm this hypothesis.

The binding site layout for *cstA*, which encodes a putative peptide transporter, also warrants further examination; this is the only CsrA-repressed transcript known to possess CsrA target sites within the open reading frame as well as overlapping the SD sequence (Dubey, Baker *et al.,* 2003). Two targets are located 9 nucleotides from each other within the untranslated leader (BS1 and BS2) while two additional sites separated by 12 nucleotides (BS3 and BS4) are located 20 nucleotides downstream. Based on

target site positioning in combination with previous nucleolytic probing for exposed single-stranded RNA regions (Dubey, Baker *et al.,* 2003), it is probable that at least two different CsrA-*cstA* configurations may coexist: one where CsrA is bound at both the farthest upstream site (BS1) and one of the ORF targets (BS3 or BS4), and a second combination where CsrA is bound at the second upstream site (BS2) and one additional downstream ORF site. It would not be unimaginable that both complex configurations could form concurrently or that a secondary loop structure forms after CsrA binding to *cstA* that includes the initiation codon; loop formation in this region is supported by previous studies showing increased RNaseT2 cleavage at the translational start site (Dubey, Baker *et al.,* 2003). Loop formation within *cstA* is also consistent with our current work showing that a single CsrA protein binds two target sites within a single oligonucleotide, thus creating a compact, fast migrating complex as visualized by native PAGE.

We also demonstrated here that one CsrA RNA-binding surface could be occupied independently of the other in an RNA concentration-dependent manner (Fig. 3-1 and 3-2). This observation could help explain the existence of target transcripts that contain only one CsrA binding site, such as *hfq* (Baker, Eory *et al*., 2007). And a recent study suggests single-site messages may not be uncommon (Brencic and Lory, 2009); Brencic *et.al*., (2009) recovered RNAs bound to a his-tagged, column-purified CsrA protein. Among the directly bound transcripts identified were two hypothetical proteins, each with a single putative CsrA target site, however, direct CsrA binding to these targets has not been confirmed. Thus, for those genes that contain single target sites with sufficient affinity for CsrA, another higher affinity binding target may not be needed for

full regulation. Indeed, some messages may have evolved only a single site so as limit CsrA influence.

Outlook

Many relevant questions remain regarding CsrA structure, binding and regulation. For example, what is the importance of conserved CsrA residue side chains that do not participate in RNA-binding, and are not likely needed to stabilize the dimer? How does RNA target site abundance (on a single transcript) impact gene expression? Do more target sites equate to tighter regulatory control? Does CsrA promote stem-loop formation when bound to a target, and/or does it destabilize hairpin formation when bound? Do CsrA-CsrA interactions play a role in regulation *in vivo*? And, a broad question that has eluded us for many years, how is positive regulation accomplished? Does CsrA bind to the *flhDC* 5'-untranslated leader and destabilize a repression loop or possibly stabilize a structure that activates translation?

An avenue for understanding regulation as it pertains to CsrA and RNA structure could be to take a reductionist approach based on minimizing the very large number of variables, especially in terms of RNA structure. While it is always prudent to verify the biological relevance of a system by using native proteins and RNAs, we may learn more about the intricacies of the Csr system when the components are individually defined and studied. Future studies into CsrA function will undoubtedly help clarify how *E. coli* and related species use the Csr circuit to fine-tune gene expression in response to their rapidly changing environments.

REFERENCES

Altier, C., M. Suyemoto and S. D. Lawhon (2000). "Regulation of *Salmonella enterica* serovar Typhimurium invasion genes by *csrA*." Infect Immun **68**(12): 6790-7.

Amann, R., S. Ferriera, J. Johnson, S. Kravitz, A. Halpern, K. Remington, K. Beeson, B. Tran, Y.-H. Rogers, R. Friedman and J. C. Venter (2006). *Blastopirellula marina* DSM 3645 1099463000737, whole genome shotgun sequence, .

Ang, S., Y. T. Horng, J. C. Shu, P. C. Soo, J. H. Liu, W. C. Yi, H. C. Lai, K. T. Luh, S. W. Ho and S. Swift (2001). "The role of RsmA in the regulation of swarming motility in *Serratia marcescens*." J Biomed Sci **8**(2): 160-9.

Baker, C. S., L. A. Eory, H. Yakhnin, J. Mercante, T. Romeo and P. Babitzke (2007). "CsrA inhibits translation initiation of *Escherichia coli hfq* by binding to a single site overlapping the Shine-Dalgarno sequence." J Bacteriol **189**(15): 5472-81.

Baker, C. S., I. Morozov, K. Suzuki, T. Romeo and P. Babitzke (2002). "CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*." Mol Microbiol **44**(6): 1599-610.

Barnard, F. M., M. F. Loughlin, H. P. Fainberg, M. P. Messenger, D. W. Ussery, P. Williams and P. J. Jenks (2004). "Global regulation of virulence and the stress response by CsrA in the highly adapted human gastric pathogen *Helicobacter pylori*." Mol Microbiol **51**(1): 15-32.

Brassinga, A. K., M. F. Hiltz, G. R. Sisson, M. G. Morash, N. Hill, E. Garduno, P. H. Edelstein, R. A. Garduno and P. S. Hoffman (2003). "A 65-kilobase pathogenicity

island is unique to Philadelphia-1 strains of *Legionella pneumophila*." J Bacteriol **185**(15): 4630-7.

Brencic, A. and S. Lory (2009). "Determination of the regulon and identification of novel mRNA targets of *Pseudomonas aeruginosa* RsmA." Mol Microbiol **72**(3): 612-32.

Brennan, R. G. and T. M. Link (2007). "Hfq structure, function and ligand binding." Curr Opin Microbiol **10**(2): 125-33.

Clamp, M., J. Cuff, S. M. Searle and G. J. Barton (2004). "The Jalview Java alignment editor." Bioinformatics **20**(3): 426-7.

Cui, Y., A. Chatterjee, Y. Liu, C. K. Dumenyo and A. K. Chatterjee (1995). "Identification of a global repressor gene, *rsmA*, of *Erwinia carotovora* subsp. *carotovora* that controls extracellular enzymes, N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting *Erwinia* spp." J Bacteriol **177**(17): 5108-15.

Dubey, A. K., C. S. Baker, K. Suzuki, A. D. Jones, P. Pandit, T. Romeo and P. Babitzke (2003). "CsrA regulates translation of the *Escherichia coli* carbon starvation gene, *cstA*, by blocking ribosome access to the *cstA* transcript." J Bacteriol **185**(15): 4450-60.

Fettes, P. S., V. Forsbach-Birk, D. Lynch and R. Marre (2001). "Overexpresssion of a *Legionella pneumophila* homologue of the *E. coli* regulator *csrA* affects cell size, flagellation, and pigmentation." Int J Med Microbiol **291**(5): 353-60.

Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer and A. Bateman (2008). "The Pfam protein families database." Nucleic Acids Res **36**(Database issue): D281-8.

Glisovic, T., J. L. Bachorik, J. Yong and G. Dreyfuss (2008). "RNA-binding proteins and post-transcriptional gene regulation." FEBS Lett **582**(14): 1977-86.

Goodier, R. I. and B. M. Ahmer (2001). "SirA orthologs affect both motility and virulence." J Bacteriol **183**(7): 2249-58.

Gutierrez, P., Y. Li, M. J. Osborne, E. Pomerantseva, Q. Liu and K. Gehring (2005). "Solution structure of the carbon storage regulator protein CsrA from *Escherichia coli*." J Bacteriol **187**(10): 3496-501.

Heeb, S., C. Blumer and D. Haas (2002). "Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0." J Bacteriol **184**(4): 1046-56.

Heroven, A. K., K. Bohme, M. Rohde and P. Dersch (2008). "A Csr-type regulatory system, including small non-coding RNAs, regulates the global virulence regulator RovA of *Yersinia pseudotuberculosis* through RovM." Mol Microbiol **68**(5): 1179-95.

Heurlier, K., F. Williams, S. Heeb, C. Dormond, G. Pessi, D. Singer, M. Camara, P. Williams and D. Haas (2004). "Positive control of swarming, rhamnolipid synthesis, and lipase production by the posttranscriptional RsmA/RsmZ system in *Pseudomonas aeruginosa* PAO1." J Bacteriol **186**(10): 2936-45.

Jonas, K., H. Tomenius, U. Romling, D. Georgellis and O. Melefors (2006). "Identification of YhdA as a regulator of the *Escherichia coli* carbon storage regulation system." FEMS Microbiol Lett **264**(2): 232-7.

Kaur, K., Q. Zhang, D. Beckler, S. Munir, L. Li, K. Kinsley, L. Herron, A. Peterson, B. May, S. Singh, C. Gebhart and V. Kapur (2005). "The complete genome sequence of *Lawsonia intracellularis*: the causative agent of proliferative enteropathy
".

Krin, E., S. Derzelle, K. Bedard, M. Adib-Conquy, E. Turlin, P. Lenormand, M. F. Hullo, I. Bonne, N. Chakroun, C. Lacroix and A. Danchin (2008). "Regulatory role of UvrY in adaptation of *Photorhabdus luminescens* growth inside the insect." Environ Microbiol **10**(5): 1118-34.

Lawhon, S. D., J. G. Frye, M. Suyemoto, S. Porwollik, M. McClelland and C. Altier (2003). "Global regulation by CsrA in *Salmonella typhimurium*." Mol Microbiol **48**(6): 1633-45.

Lenz, D. H., M. B. Miller, J. Zhu, R. V. Kulkarni and B. L. Bassler (2005). "CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*." Mol Microbiol **58**(4): 1186-202.

Liaw, S. J., H. C. Lai, S. W. Ho, K. T. Luh and W. B. Wang (2003). "Role of RsmA in the regulation of swarming motility and virulence factor expression in *Proteus mirabilis*." J Med Microbiol **52**(Pt 1): 19-28.

Liu, M. Y. and T. Romeo (1997). "The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein." J Bacteriol **179**(14): 4639-42.

Liu, Y., Y. Cui, A. Mukherjee and A. K. Chatterjee (1998). "Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. *carotovora* that controls production of extracellular enzymes and secondary metabolites." Mol Microbiol **29**(1): 219-34.

Lunde, B. M., C. Moore and G. Varani (2007). "RNA-binding proteins: modular design for efficient function." Nat Rev Mol Cell Biol **8**(6): 479-90.

Mercante, J., K. Suzuki, X. Cheng, P. Babitzke and T. Romeo (2006). "Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance." J Biol Chem **281**(42): 31832-42.

Mulcahy, H., J. O'Callaghan, E. P. O'Grady, C. Adams and F. O'Gara (2006). "The posttranscriptional regulator RsmA plays a role in the interaction between *Pseudomonas aeruginosa* and human airway epithelial cells by positively regulating the type III secretion system." Infect Immun **74**(5): 3012-5.

Mulcahy, H., J. O'Callaghan, E. P. O'Grady, M. D. Macia, N. Borrell, C. Gomez, P. G. Casey, C. Hill, C. Adams, C. G. Gahan, A. Oliver and F. O'Gara (2008). "*Pseudomonas aeruginosa* RsmA plays an important role during murine infection by influencing colonization, virulence, persistence, and pulmonary inflammation." Infect Immun **76**(2): 632-8.

National Center for Biotechnology Information, U. S. N. L. o. M. Entrez Genomes.

Reimmann, C., C. Valverde, E. Kay and D. Haas (2005). "Posttranscriptional repression of GacS/GacA-controlled genes by the RNA-binding protein RsmE acting together with RsmA in the biocontrol strain *Pseudomonas fluorescens* CHA0." J Bacteriol **187**(1): 276-85.

Rife, C., R. Schwarzenbacher, D. McMullan, P. Abdubek, E. Ambing, H. Axelrod, T. Biorac, J. M. Canaves, H. J. Chiu, A. M. Deacon, M. DiDonato, M. A. Elsliger, A. Godzik, C. Grittini, S. K. Grzechnik, J. Hale, E. Hampton, G. W. Han, J. Haugen, M. Hornsby, L. Jaroszewski, H. E. Klock, E. Koesema, A. Kreusch, P. Kuhn, S. A. Lesley, M. D. Miller, K. Moy, E. Nigoghossian, J. Paulsen, K. Quijano, R. Reyes, E. Sims, G. Spraggon, R. C. Stevens, H. van den Bedem, J. Velasquez, J. Vincent, A. White, G. Wolf, Q. Xu, K. O. Hodgson, J. Wooley and I. A. Wilson (2005). "Crystal structure of the global regulatory protein CsrA from *Pseudomonas putida* at 2.05 A resolution reveals a new fold." Proteins **61**(2): 449-53.

Sahr, T., H. Bruggemann, M. Jules, M. Lomma, C. Albert-Weissenberger, C. Cazalet and C. Buchrieser (2009). "Two small ncRNAs jointly govern virulence and transmission in *Legionella pneumophila*." Mol Microbiol.

Schubert, M., K. Lapouge, O. Duss, F. C. Oberstrass, I. Jelesarov, D. Haas and F. H. Allain (2007). "Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA." Nat Struct Mol Biol **14**(9): 807-13.

Wang, X., A. K. Dubey, K. Suzuki, C. S. Baker, P. Babitzke and T. Romeo (2005). "CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*." Mol Microbiol **56**(6): 1648-63.

Wei, B. L., A. M. Brun-Zinkernagel, J. W. Simecka, B. M. Pruss, P. Babitzke and T. Romeo (2001). "Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*." Mol Microbiol **40**(1): 245-56.

Yakhnin, H., P. Pandit, T. J. Petty, C. S. Baker, T. Romeo and P. Babitzke (2007). "CsrA of *Bacillus subtilis* regulates translation initiation of the gene encoding the flagellin protein (*hag*) by blocking ribosome binding." Mol Microbiol **64**(6): 1605-20.

Appendix. Additional Publications

CsrA Inhibits Translation Initiation of *Escherichia coli* *hfq* by Binding to a Single Site Overlapping the Shine-Dalgarno Sequence

Carol S. Baker[1], Lél A. Eöry[1,2,4], Helen Yakhnin[1], Jeffrey Mercante[3], Tony Romeo[3], and Paul Babitzke[1]*

[1]*Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*

[2]*School of Biochemistry and Microbiology, University of Leeds, Leeds, UK*

[3]*Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia 30322, USA*

Current address: [4]*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK*

***Corresponding author:** Paul Babitzke**,** Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802

Tel. (814)865-0002; Fax (814)863-7024; E-mail pxb28@psu.edu

This manuscript was published in *J Bacteriol*. 2007 Aug; 189(15): 5472-81. Jeffrey Mercante conducted the real-time RT-PCR experiments in this study.

ABSTRACT

Csr (carbon storage regulation) of *Escherichia coli* is a global regulatory system that consists of CsrA, a homodimeric RNA binding protein, two non-coding sRNAs (CsrB and CsrC) that function as CsrA antagonists by sequestering this protein, and CsrD, a specificity factor that targets CsrB and CsrC for degradation by RNase E. CsrA inhibits translation initiation of *glgC*, *cstA*, and *pgaA* by binding to their leader transcripts and preventing ribosome binding. Translation inhibition is thought to contribute to the observed mRNA destabilization. Each of the previously known target transcripts contains multiple CsrA binding sites. A position-specific weight matrix search program was developed using known CsrA binding sites in mRNA. This search tool identified a potential CsrA binding site that overlaps the Shine-Dalgarno sequence of *hfq*, a gene encoding an RNA chaperone that mediates sRNA-mRNA interaction. This putative CsrA binding site matched the SELEX-derived binding site consensus sequence in 8 out of 12 positions. Results from gel mobility shift and footprint assays demonstrated that CsrA binds specifically to this site in the *hfq* leader transcript. Toeprint and cell-free translation results indicated that bound CsrA inhibits Hfq synthesis by competitively blocking ribosome binding. Disruption of *csrA* caused elevated expression of an *hfq'-'lacZ* translational fusion, while overexpression of *csrA* inhibited expression of this fusion. We also found that *hfq* mRNA is stabilized upon entry into stationary phase growth by a CsrA-independent mechanism. The interaction of CsrA with *hfq* mRNA is the first example of a CsrA-regulated gene that contains only one CsrA binding site.

INTRODUCTION

Bacteria have evolved several regulatory strategies that ensure their survival in response to changes in their growth environment. The Csr (carbon storage regulation) and homologous Rsm (repressor of secondary metabolites) global regulatory systems of several eubacterial species control numerous genes and processes post-transcriptionally. Csr systems consist of at least one RNA binding protein that either activates or represses expression of target mRNAs, as well as one or more small non-coding regulatory RNAs (sRNAs) that contain multiple CsrA binding sites. The sRNAs function as antagonists of the RNA binding protein(s) via protein sequestration (reviewed in 1, 26). The Csr system of *Escherichia coli* is involved in the repression of several stationary phase processes and the activation of some exponential phase functions. Four major components of Csr in this organism include the homodimeric RNA binding protein CsrA, two sRNA antagonists of CsrA (CsrB and CsrC), and CsrD, a protein that specifically targets both sRNAs for degradation by RNase E (18, 35, 45). CsrA represses gluconeogenesis, glycogen metabolism, peptide transport and biofilm formation (9, 16, 27, 28, 42, 48), while it activates glycolysis, acetate metabolism, and flagellum biosynthesis (28, 43, 44). CsrB and CsrC sequester CsrA and prevent its interaction with mRNA targets. Multiple imperfect repeat sequences in these regulatory RNAs function as CsrA binding sites, such that each sRNA is capable of sequestering several CsrA dimers (14, 18, 45).

CsrA negatively regulates expression of the glycogen biosynthetic gene, *glgC*, by binding to four sites in the untranslated leader of the *glgCAP* operon transcript, one of which overlaps the *glgC* Shine-Dalgarno (SD) sequence (2, unpublished results). CsrA binding to the *glgCAP* leader transcript inhibits GlgC synthesis by blocking ribosome binding. Presumably, CsrA-mediated inhibition of *glgC* translation is responsible for the accelerated rate of *glgCAP* mRNA decay (19). CsrA also represses translation of *cstA*, a carbon starvation-induced gene thought to be involved in peptide transport (9, 31), as well as the first gene in the *pgaABCD* operon, a cluster of genes

that are required for the synthesis of the polysaccharide adhesin poly-β-1,6-*N*-acetyl-D-glucosamine (PGA) that participates in biofilm formation (42). CsrA binds to four sites in the *cstA* transcript and to six sites in the *pga* operon leader transcript. In each case, one of the CsrA binding sites overlaps the cognate SD sequence. Translational repression of these genes proceeds by a mechanism that is similar to the mechanism identified for *glgC* (9, 42). Considerable sequence variation exists among the known *E. coli* CsrA binding sites; however, GGA is a highly conserved sequence element, which is often present in the loop of short RNA hairpins. Systematic evolution of ligands by exponential enrichment (SELEX) was used to isolate high-affinity CsrA ligands (10). The high-affinity RNA ligands contained a single CsrA binding site with a consensus sequence of RUACARGGAUGU, with the underlined residues being 100% conserved. In each case the GGA motif was present in the loop of a short predicted hairpin (10).

A bioinformatics approach was used to search the *E. coli* genomic database for genes containing potential CsrA binding sites. A potential CsrA binding site was identified that overlaps the *hfq* SD sequence, suggesting that CsrA might regulate translation initiation of this gene. *E. coli* Hfq is a toroid-shaped homohexamer that was discovered as a protein required for *in vitro* transcription of bacteriophage Qβ RNA (12, 29). Hfq is present in a wide range of bacterial species and its role in global control of gene expression is readily apparent as it impacts numerous physiological processes such as virulence, bacteriocin production and nitrogen fixation (40). Numerous studies have established that Hfq functions as an RNA chaperone in promoting sRNA-mRNA base-pairing (reviewed in 13, 34). For example, it is well established that Hfq activates translation of σS, the stationary phase sigma factor, by promoting base-pairing of two sRNAs to the *rpoS* leader transcript. Base-pairing of either sRNA disrupts an inhibitory RNA structure in the *rpoS* leader such that translation is stimulated (13, 23, 34).

We confirmed that CsrA binds to the site in *hfq* identified *in silico*. As this CsrA binding site overlaps the *hfq* SD sequence, bound CsrA inhibits translation initiation of *hfq* by blocking

ribosome binding. The interaction of CsrA with the *hfq* transcript described here is unique as this is the first example of a CsrA-regulated mRNA that contains only a single CsrA binding site. Because Hfq mediates many intermolecular sRNA-mRNA interactions in the cell, these findings imply that CsrA has a substantially greater influence on global regulatory networks than was previously recognized.

MATERIALS AND METHODS

Bacterial strains and plasmids. Plasmid pCSB52 contains the WT *hfq* leader and the first 55 nt of the *hfq* coding region (+1 to +124 relative to the start of of $P3_{hfq}$ promoter transcription) (37) cloned into the pTZ18U polylinker (United States Biochemical Corp.). pCSB60 contains an *hfq'-'lacZ* translational fusion consisting of the $P3_{hfq}$ promoter and leader region, as well as the first 18 codons of *hfq* (-66 to +124 relative to the start of $P3_{hfq}$ transcription), cloned in frame with the *lacZ* gene of pMLB1034 (32). Three nucleotide substitutions in the *hfq* leader just upstream of the SD sequence and within the CsrA binding site (A51T:T52G:A53C) were introduced into pCSB60 using the QuikChange II protocol (Stratagene), resulting in plasmid pCSB62. *E. coli* strains used for β-galactosidase assays were constructed to create single-copy chromosomal gene insertions of *hfq'-'lacZ* translational fusions into the λ *att* site using the λInCh protocol as described previously (5). Strains PLB785 and PLB786 contain the *hfq'-'lacZ* fusion from pCSB60 integrated into the chromosome of strains CF7789 (MG1655 Δ*lacI-Z* [*Mlu*I]) and TR1-5CF7789 (CF7789 *csrA::kan*), respectively. Strains PLB923 and PLB924 contain the *hfq'-'lacZ* fusion from pCSB62 integrated into the chromosomes of CF7789 and TR1-5CF7789, respectively. Plasmid pCRA16 (36) contains *csrA* cloned into pBR322 (4). Strain PLB786 was transformed with pBR322 and pCRA16 to generate strains PLB789 and PLB793, respectively. Plasmid pYH109 was generated by replacing the *B. subtilis trp* operon sequence in pYH28 (30) with a PCR fragment containing the *hfq* leader and amino

terminal coding sequence (+1 to +179 relative to $P3_{hfq}$ transcription), resulting in an *hfq'-'gfp* translational fusion (37th *hfq* codon fused in frame with *gfp*). The *E. coli smpB* gene was cloned into the pET28A+ polylinker (Novagen) to create pETB. Unless otherwise indicated, all strains were grown at 37°C in Lennox LB medium. When appropriate, growth media were supplemented with antibiotics to the following concentrations: ampicillin (100 μg/ml), kanamycin (50 μg/ml), and tetracycline (20 μg/ml).

Gel mobility shift assay. Quantitative gel mobility shift assays followed a previously published procedure (46). *E. coli* CsrA-H6 protein was purified as described previously (10). RNA was synthesized *in vitro* using the MEGAscript kit (Ambion) and linearized pCSB52 as template. Gel-purified RNA was 5' end-labeled with [γ-32P]ATP as described previously (46). RNA suspended in TE buffer was renatured by heating to 80°C followed by slow cooling to room temperature. Binding reactions (10 μl) contained 10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 100 mM KCl, 32.5 ng of yeast RNA, 7.5% glycerol, 20 mM dithiothreitol (DTT), 4U of RNase inhibitor (Promega), 0.5 nM *hfq* leader RNA, purified CsrA-H6 (various concentrations) and 0.1 mg/ml xylene cyanol. Competition assays also contained unlabeled RNA competitor (see text for details). Reaction mixtures were incubated for 30 min at 37°C to allow CsrA–RNA complex formation. Samples were then fractionated on native 8% polyacrylamide gels. Radioactive bands were visualized with a phosphorimager. Free and bound RNA species were quantified using IMAGEQUANT (Molecular Dynamics), and the apparent equilibrium binding constant (K_d) of CsrA-*hfq* RNA interaction was calculated as described previously (46).

Toeprint assay. Toeprint assays were performed by modifying published procedures (2, 15). *hfq* leader transcripts used in this analysis were synthesized using the MEGAscript kit and

linearized pCSB52 as template. Gel-purified *hfq* leader RNA (500 nM) in TE was renatured and hybridized to a 5' end-labeled DNA oligonucleotide (500 nM) in TE that was complementary to the 3' end of the transcript. Hybridization was accomplished by heating the mixture to 80°C followed by slow cooling to room temperature. Toeprint assay mixtures contained various concentrations of CsrA-H6 and/or 260 nM 30S ribosomal subunits and 5 μM $tRNA^{fMet}$. *E. coli* ribosomes were purified as described previously (25). Purified 30S ribosomal subunits were obtained by denaturing 70S ribosomes, followed by purification through a sucrose gradient. 30S subunit fractions were pooled and stored at -80°C in 10 mM Tris-HCl, pH 7.5, 60 mM NH_4OAc, 6 mM 2-mercaptoethanol, 10 mM $MgCl_2$ and 10% glycerol. Previously frozen 30S ribosomal subunits were thawed, activated by incubation at 37°C for 15 min, and kept on ice until addition to toeprint reaction mixtures. Toeprint reactions (20 μl) contained 2 μl of the hybridization mixture, 375 μM each dNTP and 10 mM DTT in toeprint buffer (10 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 60 mM NH_4OAc, 6 mM 2-mercaptoethanol). Mixtures containing CsrA were incubated for 30 min at 37°C to allow CsrA-mRNA complex formation. 30S ribosomal subunit toeprint reactions were performed by incubating RNA, 30S ribosomal subunits and $tRNA^{fMet}$ in toeprint buffer as described previously (15). Following the addition of 3 U of avian myeloblastoma virus reverse transcriptase (Roche), the reaction mixture was further incubated for 15 min at 37°C. Reactions were terminated by the addition of 12 μl of stop solution (70 mM EDTA, 85% formamide, 0.1x TBE, 0.025% xylene cyanol and 0.025% bromophenol blue). Samples were heated at 95°C for 5 min prior to fractionation through standard 6% polyacrylamide sequencing gels. Sequencing reactions were performed using pCSB52 as the template and the same end-labeled DNA oligonucleotide as a primer. Radioactive bands were visualized with a phosphorimager.

RNA Footprint assay. Preparation of 5' end-labeled *hfq* leader transcripts was as described for the gel shift analysis. Titrations of RNAse T1 (Roche) and RNAse T2 (Sigma) were performed to identify the amount of enzyme in which ~90% of the transcripts were full length to minimize multiple cleavages in any one transcript. Binding reactions (10 μl) containing various concentrations of CsrA-H6 and 50 nM *hfq* RNA were otherwise identical to those described for the gel shift assay. After the initial binding of CsrA-H6, either RNase T1 (0.025 U) or RNase T2 (0.03 U) was added to the reactions, and incubation was continued for 15 min at 37°C. Reactions were terminated by the addition of 10 μl of Gel Loading Buffer II (Ambion) and kept on ice. Partial alkaline hydrolysis and RNase T1 digestion ladders of each transcript were prepared as described previously (3). Samples were fractionated through standard 6% polyacrylamide sequencing gels. Radioactive bands were visualized with a phosphorimager.

β-galactosidase assays. Bacterial cultures growing in liquid medium at 37°C were monitored using a Klett-Summerson colorimeter (No. 52 green filter). Culture samples (4 ml) were harvested at various times, washed once with 10 mM Tris-HCl, pH 7.5, and frozen as cell pellets at -20°C. Cell extracts were prepared by suspending frozen cell pellets in 0.5 ml of BugBuster protein extraction reagent (Novagen), followed by incubation at 37°C in an air shaker. After 30 min, 0.3 ml of Z buffer (24) containing 0.2 mg/ml lysozyme was added to each sample and incubation was continued for 30 min at 37°C in an air shaker. Cell debris was removed by centrifugation at 4°C. β-galactosidase assays were performed using the cell extracts (2, 24). Protein concentrations were determined by the Bio-Rad protein assay with BSA as a standard.

mRNA abundance and mRNA half-life assays. Bacterial cultures were mixed with 2 volumes of RNAprotect Bacterial Reagent (Qiagen) and incubated at room temperature for 5 min.

Cells were then harvested and total RNA was prepared using the Masterpure RNA Purification Kit (Epicentre) and treated with DNase I according to the manufacturer's recommendations. RNA was quantified by measuring the absorbance at 260 and 280 nm.

To measure *hfq* transcript levels, strains MG1655 (wild type) and TR1-5MG1655 (*csrA::kan*) were grown at 37°C in Lennox LB medium to exponential phase (OD_{600} = 0.4) or to early stationary phase (OD_{600} = 4.0). Total RNA was purified and the steady state level of *hfq* mRNA was determined by real-time quantitative reverse transcription-polymerase chain reaction (rt-qRT-PCR) using the primer pair hfq-Fw (5' AAGCACGCGATTTCTACTGTTG 3') and hfq-Rv (5' CCACCGGCGTTGTTACTGT 3') and the probe Hfq-6FAM-BHQ1 (5' CCCGTCTCGCCCGGTTTCTCA 3'), which was 5' end-labeled with 6-carboxyfluorescein (6FAM) and 3' end-labeled with Black Hole Quencher-1 (BHQ-1). rt- qRT-PCR was performed using the iScript one-step RT-PCR Kit for Probes (Biorad) with a Biorad iCycler IQ real-time system. The conditions used for RT-PCR were as follows: 50°C for 10 min; 95°C for 5 min; and 40 cycles of 95°C for 15 sec and 65°C for 30 sec. Unless otherwise noted, all primers and probes were used at a final concentration of 200 nM. Real-time measurements were taken during the 65°C step. Reactions were performed in triplicate in two independent experiments, each time with 100 ng of template RNA. A set of reactions lacking reverse transcriptase was performed for each RNA sample as a control for DNA contamination. For normalization of *hfq* RNA levels, rt-qRT-PCR reactions were performed with each sample for 16S rRNA quantitation using the primer pair 16S-Fw (5' CGTGTTGTGAAATGTTGGGTTAA 3') and 16S-Rv (5' CCGCTGGCAACAAAGGATA) and the probe 16S-6FAM-BHQ1 (5' TCCCGCAACGAGCGCAACC 3'). The reaction conditions for rt-qRT-PCR of 16S rRNA were identical to *hfq* except that 1 ng of RNA template was used for each reaction. The $2^{-\Delta\Delta CT}$ method was used to calculate relative *hfq* RNA levels, which allowed for the use of a 16S rRNA control (20).

For *hfq* mRNA half-life studies, strains MG1655 (wild type) and TR1-5MG1655 (*csrA::kan*) were grown as described above. Cells were harvested at various times following the addition of rifampicin (200 ug/ml final concentration) and total RNA was purified as described above using RNAprotect. rt-qRT-PCR was performed as described above for steady state *hfq* RNA determination. The percentage of RNA remaining through the time course was determined by calculating the difference in cycle threshold (ΔC_T) and the subsequent fold difference compared to the 0 min time point after controlling for 16S rRNA levels.

RNA directed cell-free translation. Cell-free translation reactions followed previously published procedures (11, 47). Transcripts for this analysis were synthesized using the Ambion MEGAscript kit. *hfq'-'gfp* and *smpB* transcripts were synthesized using linearized pYH109 and pETB as templates, respectively. *bla* was transcribed from a 1020 nt PCR fragment containing a T7 promoter. CsrA-deficient *E. coli* S-30 extract was prepared from TR1-5 MG1655 (*csrA::kan*) according to published procedures (47). The S-30 extract was preincubated with RNase-free DNase I for 15 min at 37°C to remove chromosomal DNA and to allow time for *E. coli* RNases to degrade endogenous mRNA. Reaction mixtures (24 µl) contained 60 mM Tris-HEPES, pH 7.5, 60 mM NH_4Cl, 5 mM to 15 mM $MgCl_2$ (determined empirically for each transcript), 12 mM KCl, 0.5 mM EGTA, 5 mM DTT, 2 mM ATP, 0.6 mM GTP, 0.08 mM calcium folinate, 4 mg/ml of aprotinin, 4 mg/ml of leupeptin, 4 mg/ml of pepstatin A, 4 µl S-30 extract (12 mg of total protein), 800 U/ml of DNase I, 500 U/ml of RNasin, 10 mM phosphoenolpyruvate, 35 U of pyruvate kinase, 0.4 mg/ml of *E. coli* tRNA, 0.04 µg/ml of mRNA, 10 µCi [^{35}S] methionine, 0.5 mM of each of the other amino acids and 0.8 mM spermidine. Reaction mixtures were incubated for 45 min at 37°C and terminated by adding 6 µl of stop buffer (125 mM Tris-HCl, pH 6.8, 5% SDS, 25% glycerol, 2% 2-mercaptoethanol and 12.5 mg/ml of bromophenol blue). Samples were

heated at 95°C for 5 min prior to fractionation by 14% SDS-PAGE. Radioactive bands were visualized with a phosphorimager and quantified using IMAGEQUANT.

RESULTS

Identification of *hfq* as a potential CsrA-regulated gene. A genome search program was developed to identify potential CsrA binding sites by exploring the properties of known CsrA binding motifs. A total of fourteen CsrA binding sites were previously identified in the leader regions of *glgC*, *cstA* and *pgaA* (2, 9, 42, unpublished results). These sequences were aligned using ClustalW (7) and a position weight matrix (pwm) was calculated from the alignment using the MATCH^{TM} tool (17). The pwm was then used as a scoring function to identify potential CsrA binding sites within the *E. coli* genomic database. The scores were assigned from 0 to 100 % according to the minimum and the maximum score calculated from the pwm. The details of the pwm will be published elsewhere. This program predicted the presence of CsrA binding sites in 278 genes with scores of 96.5% or above, including the three genes that were used in generating the pwm (*cstA, glgC* and *pgaA*). A CsrA binding site with a score of 96.8% was identified that overlaps the *hfq* SD sequence. This predicted sequence also conformed to the SELEX-derived consensus sequence in 8 out of 12 positions (Fig. 1). However, unlike the CsrA binding sites identified by SELEX, secondary structural predictions using MFOLD (49) indicated that the GGA motif within this putative CsrA binding site was not present in the loop of a hairpin. The pwm also identified three potential sites with scores between 81.8 and 83.5%; however, since all three of these sites overlapped the predicted site with a score of 96.8% and none of them contained an appropriately positioned GGA motif, it appeared that *hfq* contained one likely CsrA binding site. Because all of the known CsrA-controlled genes contained four to six CsrA binding

sites, experiments were carried out to determine whether *hfq* contained a single CsrA binding site and whether CsrA could bind to this site in *hfq* and regulate its expression.

CsrA binds to the predicted site in *hfq*. *hfq* transcription is driven by three promoters just upstream of its coding sequence. The *hfq* transcript originating from the SD sequence proximal P3 promoter ($P3_{hfq}$) contains a 68 nt untranslated leader (37). To characterize the interaction of CsrA with *hfq* RNA, quantitative gel mobility shift assays were performed with an *hfq* transcript containing nucleotides +1 to +124 relative to the start of $P3_{hfq}$ transcription. Since quantitative gel mobility shift assays using native CsrA or C-terminal His-tagged CsrA (CsrA-H6) did not show any significant difference in binding affinities for target transcripts (data not shown), CsrA-H6 was used in all *in vitro* assays and is referred to as CsrA from here on. CsrA bound to this *hfq* transcript as a distinct band in native gels between 4 and 512 nM CsrA (Fig. 2A). A complete shift was observed at 128 nM CsrA and no additional shifted species of higher molecular weights were observed as the concentration of CsrA was increased further, suggesting that CsrA binds to a single site in the *hfq* leader transcript. A nonlinear least-squares analysis of these data yielded an estimated K_d value of 38 ± 13 nM CsrA. For comparison, the affinity of CsrA for *glgC* (4 CsrA binding sites), *cstA* (4 CsrA binding sites) and *pgaA* (6 CsrA binding sites) was 39 nM, 40 nM and 22 nM, respectively (2, 9, 42).

The specificity of the CsrA-*hfq* RNA interaction was investigated by performing competition experiments with specific (*hfq* leader and a SELEX-derived ligand) and non-specific (*Bacillus subtilis trp* leader) unlabeled RNA competitors (Fig. 2B). Unlabeled *hfq* and SELEX transcripts were effective competitors, whereas the *B. subtilis trp* leader RNA (*trpL*) did not compete with CsrA-*hfq* RNA interaction. These results establish that CsrA binds specifically to *hfq* RNA.

CsrA-*hfq* RNA footprint experiments were conducted to identify the position of bound CsrA in the *hfq* transcript. Single strand-specific RNases were used as probes for these studies. As the concentration of CsrA was increased from 0 to 2 μM, protection of several nucleotides from RNase T1 (G specific) and RNase T2 (A preference) cleavage was observed (Fig. 3A). CsrA protected G49, G57 and G58 from RNase T1 cleavage, as well as, residues A51 through A61 from RNase T2 cleavage. Importantly, the entire sequence overlapping the *hfq* SD sequence identified *in silico* was protected from RNase cleavage. The composite footprint indicates that CsrA protects one RNA segment extending from G49 through A61 (Fig. 3B). Previous RNA structure mapping identified two stable RNA hairpins in the *hfq* transcript (h1 and h2) (41). The presence of these hairpins was confirmed; residues corresponding to h1 and h2 were protected from ribonuclease cleavage in the absence of bound CsrA (Fig. 3). Bound CsrA also resulted in increased RNase T2 cleavage of C20 and C42. These two residues are located within the 5' side bulge of h1 and just downstream of h1, respectively. CsrA-dependent protection was also observed for residues A76 (RNase T2), G77 (RNase T1) and G78 (RNase T1). These residues are present within the stem of a predicted RNA hairpin containing a GNRA tetraloop ($\Delta G = -4.0$ kcal/mol) (22, 49). Thus, it appears that bound CsrA stabilizes this structure (Fig. 3).

Bound CsrA inhibits translation initation of *hfq*. Primer extension inhibition (toeprint) experiments were performed to determine whether CsrA was capable of competing with 30S ribosomal subunits for binding to the *hfq* transcript. The presence of bound CsrA or 30S ribosomal subunits should stop primer extension by reverse transcriptase, resulting in a toeprint band near the 3' boundary of the bound macromolecule. Stable secondary structures are also capable of inhibiting extension by reverse transcriptase, resulting in a toeprint band near the 3' end of the RNA hairpin. The toeprint results are presented in figure 4 and summarized in figure 3B. The presence of CsrA resulted in toeprints at positions A43, A82 and A91. The toeprint at

A91 provides additional evidence for the hairpin containing the GNRA tetraloop as this structure ends at A89 (Fig. 3B). The toeprint at A43 likely corresponds to the base of h1 as this RNA hairpin ends at G39. The origin of the A82 toeprint is unclear as it is not near the 3' end of a stable hairpin and is 21 nt downstream from the 3' end of the CsrA footprint (Fig. 3 and 4). Since a toeprint corresponding to the 3' boundary of bound CsrA was not observed, it appears that reverse transcriptase is effective at displacing CsrA when bound to a single site.

Toeprint assays were also performed to identify the positions of bound 30S ribosomal subunits. A prominent $tRNA^{fMet}$-dependent 30S ribosomal subunit toeprint band was observed 15 nt down from the A of the AUG initiation codon, which is the same distance from the translation initiation codon that was previously observed for several mRNAs (e.g. 2, 9, 42, 47). The second 30S ribosomal subunit-dependent toeprint at G72 was not expected and its origin is unknown. Toeprint experiments were also carried out to determine whether bound CsrA could inhibit ribosome binding. Importantly, when CsrA was bound to the *hfq* transcript prior to the addition of 30S ribosomal subunits and $tRNA^{fMet}$, each of the CsrA-dependent toeprint bands were observed, whereas the 30S ribosomal subunit toeprint bands were eliminated (Fig. 3B and 4). Thus, our toeprinting results demonstrate that bound CsrA prevents ribosome binding to the *hfq* transcript, suggesting that CsrA could capable of preventing translation initiation and Hfq synthesis.

Since our *in vitro* binding studies demonstrated that bound CsrA blocks ribosome binding, RNA-directed cell-free translation experiments were conducted to determine whether CsrA could inhibit Hfq synthesis (Fig. 5). Our initial attempt to examine the effect of CsrA on Hfq synthesis was problematic as we observed multiple bands, suggesting that hexameric Hfq was not completely denatured (data not shown). Because we previously found that using *gfp* translational fusions circumvented similar problems in cell-free translation experiments with *Bacillus subtilis* S-30 extracts, an *hfq'-'gfp* translational fusion transcript was tested in the *E. coli* S-30 extract. In this case, well-behaved Hfq-Gfp fusion polypeptides were produced that migrated

as a doublet. Importantly, addition of increasing concentrations of CsrA to the cell-free translation system led to a corresponding decrease in the Hfq-Gfp synthesis (Fig. 5). Similar cell-free translation experiments were carried out using *smpB* and *bla* transcripts as negative controls. The *smpB* transcript contained a SD sequence derived from pET28A+, whereas the *bla* transcript contained its natural SD sequence. Slight CsrA-dependent translation inhibition was observed for the negative controls, although the level of inhibition was far less than for Hfq-Gfp (Fig 5). These results, in conjunction with the *in vitro* binding studies, demonstrate that CsrA inhibits translation initiation of *hfq* by blocking ribosome access to the *hfq* transcript.

As Hfq was previously shown to repress its own translation (41), it was of interest to determine whether CsrA- and Hfq-mediated translation control were additive. Results from cell-free translation experiments confirmed that Hfq represses its own translation (Fig. 5C and data not shown). Moreover, an additive effect of CsrA and Hfq on translation inhibition was observed at protein concentrations ≥ 0.8 μM (Fig. 5C). This latter result is somewhat surprising as one of the Hfq binding sites overlaps the single CsrA binding site.

CsrA inhibits *hfq* expression. CsrA-dependent regulation of *hfq* was examined *in vivo* using an *hfq'-'lacZ* translational fusion whose expression was driven by $P3_{hfq}$ (37). This fusion was integrated into the lambda *att* site of the *E. coli* chromosome in single copy and expression was examined throughout the growth curve in both wild-type and *csrA* mutant backgrounds. When compared to the wild type strain, a small but reproducible increase in expression (~30%) was observed in the *csrA* mutant during stationary phase growth (Fig. 6A). Expression was also examined when cells were grown in LB plus 1% glucose because we previously found that expression of a *cstA'-'lacZ* fusion in wild type and mutant strains differed to a greater extent in this glycolytic growth condition (9). A somewhat higher increase in expression (~50%) was

observed in the *csrA* mutant from late exponential to stationary phase growth when cells were grown in LB plus 1% glucose (Fig. 6C). Introduction of *csrA* on a plasmid complemented the *csrA* mutant defect, resulting in a twofold reduction of *hfq'-'lacZ* expression beginning in late exponential phase (Fig. 6B).

We attempted to examine the influence of a mutant CsrA binding site on *hfq* expression. Because the critical GGA motif in this binding site is part of the *hfq* SD sequence, three CsrA binding site residues located just upstream from the *hfq* SD sequence were altered (A51T:T52G:A53C; see Fig. 3B). Expression from this mutant fusion was reduced ~5-fold in both wild type and *csrA* mutant strains (Fig. 6C), suggesting that sequence alterations this close to the SD sequence had deleterious effects on translation initiation. Furthermore, it is apparent that these substitutions did not eliminate CsrA-dependent inhibition of *hfq* expression, suggesting that a more substantial change to the CsrA binding site would be needed to prevent CsrA-*hfq* RNA interaction. While the reason for reduced translation of the binding site mutant is not known, RNA secondary structure predictions using MFOLD (49) suggest that RNA structural rearrangements are not the cause.

Previous studies established that the mRNA of several CsrA-repressed genes were stabilized in *csrA* mutant strains (1, 26). The steady-state level of *hfq* mRNA was determined by rt-qRT-PCR in wild type and *csrA* mutant strains in exponential and early stationary phase growth. The relative abundance of *hfq* mRNA was 2-fold and 1.7-fold higher in the *csrA* mutant strain during exponential and early stationary phase growth, respectively. The increased level of *hfq* mRNA in the *csrA* mutant strain could have been caused by increased transcription or a reduction in the mRNA decay rate. Results from mRNA half-life experiments indicated that CsrA does not affect the stability of *hfq* mRNA, suggesting that CsrA indirectly influences *hfq* transcription (Fig. 7). While the half-life of *hfq* mRNA was similar in wild type and *csrA* mutant strains, we found that *hfq* transcripts were dramatically stabilized in early stationary phase growth

(Fig. 7). Thus, it appears that mRNA stabilization contributes to increased *hfq* expression in stationary phase by a CsrA-independent mechanism.

DISCUSSION

CsrA is a global regulatory RNA binding protein that represses or activates gene expression post-transcriptionally. Bound CsrA inhibits expression of several genes by binding to multiple sites within target transcripts, one of which overlaps the cognate SD sequence, thereby blocking ribosome binding. Inhibition of translation is thought to contribute to the observed accelerated rate of mRNA decay (1, 26). In the case of gene activation, bound CsrA can stabilize target transcripts, although the mechanism of message stabilization is not known (43). Two sRNA antagonists of CsrA, CsrB and CsrC, contain multiple CsrA binding sites and function by sequestering this protein (18, 45). Expression of *csrA*, *csrB* and *csrC* increases as the culture approaches stationary phase (14, 45). The BarA/UvrY two-component signal transduction system activates transcription of *csrB* and *csrC* (36). Although the signal for this activation is not known, BarA signaling appears to be pH dependent (21). Interestingly, CsrA indirectly activates synthesis of both of the sRNAs via the response regulator UvrY, resulting in an autoregulatory circuit for CsrA, CsrB and CsrC (36, 45). A fourth Csr component, CsrD protein, was recently shown to be a specificity factor that targets CsrB and CsrC for degradation by RNase E (35). As CsrA acts downstream of transcriptional regulation and generally affects gene expression in the 1.5- to 10-fold range (e.g. 9, 19, 27, 42, 43), it appears that CsrA functions in a fashion similar to a "governor" on a motor by reducing expression of some genes and increasing expression of others, rather than as an on-off switch.

A pwm search tool identified a potential CsrA binding site that overlaps the *hfq* SD sequence. The finding that this single site was similar to the SELEX-derived binding site

consensus (Fig. 1), led us to investigate CsrA-dependent regulation of this gene. *hfq* is located in the *amiB-mutL-miaA-hfq-hflX* superoperon, which contains both $E\sigma^{32}$- and $E\sigma^{70}$-specific promoters (37, 38). Transcription of this operon is driven by at least five promoters during exponential growth ($P_{mutL,}$ $P_{miaA,}$ $P1_{hfq,}$ $P2_{hfq,}$ and $P3_{hfq}$). Our studies focused on the 68 nt mRNA leader originating from the SD sequence-proximal promoter, $P3_{hfq}$. Our gel shift (Fig. 2) and footprint (Fig. 3) results demonstrate that CsrA binds to the single site identified *in silico*. Moreover, the toeprint (Fig. 4) and cell-free translation (Fig. 7) results establish that bound CsrA inhibits Hfq synthesis by competitively blocking ribosome binding. While *hfq* mRNA contains only a single CsrA binding site, the affinity of CsrA-*hfq* RNA interaction (K_d = 38 nM) is comparable to the affinity that CsrA has for mRNAs containing 4 to 6 binding sites (K_d = 22 to 40 nM). Despite the high affinity for *hfq* RNA, CsrA-mediated regulation was only 1.5 to 2-fold under our growth conditions (LB and LB + glucose) (Fig. 5). This level of regulation is comparable to CsrA-dependent regulation of *cstA* expression in LB; however, regulation of *cstA* was ~5-fold in LB + glucose (9). Thus, it is possible that growth conditions for optimum CsrA-dependent regulation of *hfq* were not achieved.

The finding that translational repression did not alter the stability of *hfq* mRNA constitutes the first example in which CsrA-mediated translational repression did not lead to accelerated mRNA decay (Fig. 6). The observation that CsrA did not influence the stability of *hfq* mRNA, combined with the finding that the steady-state level of *hfq* transcripts was elevated in *csrA* mutant strains, suggests that CsrA has a negative effect on *hfq* transcription as well as translation. While CsrA caused only a small reduction in expression of the *hfq'-'lacZ* fusion used in this study, it is important to note that our fusion contained only one of five known *hfq* promoters, $P3_{hfq}$. Since it is reasonable to assume that CsrA is capable of repressing translation initiation of transcripts derived from any of the *hfq* promoters, it appears likely that CsrA indirectly represses transcription from (at least) one of the other *hfq* promoters. Thus, the *in vivo*

effect of CsrA on translation, as determined by the *hfq'-'lacZ* reporter, and the apparent indirect effect of CsrA on *hfq* mRNA transcription, as determined by rt-qRT-PCR, likely contribute to the overall effect of CsrA on Hfq levels.

All previously characterized CsrA target mRNAs contain four or more CsrA binding sites (2, 9, 18, 42, 45, unpublished results). As CsrA is not a general repressor of translation, the finding that CsrA is capable of inhibiting translation of an mRNA containing a single CsrA binding site that overlaps its cognate SD sequence raises the question of how CsrA distinguishes one SD sequence from another? RNA secondary structure does not appear to be the only answer, as the majority of the known CsrA binding sites in mRNA targets, including the single binding site in *hfq*, do not contain their GGA motif in the loop of an RNA hairpin. Thus, additional conserved RNA sequence elements of binding sites must contribute to binding specificity.

Hfq is known to inhibit its own translation by binding to two sites within the *hfq* leader and initially translated region (41). Site A is located just upstream of the h1 hairpin, while binding site B overlaps its SD sequence (Fig. 3B). Both CsrA and Hfq are effective at inhibiting formation of a translation initiation complex. As Hfq-mediated autoregulation was reported to be about twofold, it is apparent that the level of CsrA-dependent and Hfq-dependent inhibition of *hfq* translation is similar. Moreover, it appears that CsrA-dependent and Hfq-dependent inhibition of Hfq synthesis is additive despite the fact that the CsrA binding site and Hfq binding site B overlap (Fig. 5C).

Additional studies further establish an interrelationship between CsrA and Hfq. A global analysis of protein-protein interactions in *E. coli* using Hfq as the "bait" protein identified a stable interaction with CsrA (6). The intracellular levels of Hfq hexamers and CsrA dimers were determined to be approximately 9,000 and 16,000 molecules, respectively (14, 39). The mRNA signal intensity from a transcription profile of cells grown in minimal medium showed that the

relative *hfq* transcript level (11884) was comparable to that of *csrA* (7783) (8). An additional study reported that Hfq stabilized RsmY RNA, a CsrB homolog in *Pseudomonas aeruginosa* (33). Moreover, this study suggesteds that RsmA, a CsrA homolog, and Hfq could bind concurrently to RsmY, a known antagonist of RsmA.

Csr and homologous Rsm systems have been identified in a wide variety of bacterial species (1, 26). Depending on the particular organism, this global regulatory system controls a variety of cellular processes and behaviors (e.g., RpoS stress signaling, quorum sensing, biofilm development, motility and chemotaxis, central carbon flux and pathogenesis). The finding that the Csr circuitry is interconnected with other global regulatory networks suggests that Csr governs cellular behavior and physiology on a scale that is not yet fully understood.

ACKNOWLEDGEMENTS

The authors thank Gisela Storz for Hfq, Ken Keiler for pETB, Emine Koc for technical assistance in the preparation of *E. coli* 30S ribosomal subunits and Isaac Magaña for computer support.

This work was supported by Public Health Service grant GM59969 from the National Institutes of Health. Lél Eöry was supported in part by a fellowship from the Worldwide University Network (WUN).

REFERENCES

1. **Babitzke, P. and T. Romeo.** 2007. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr. Opin. Microbiol. **10**:156-163.

2. **Baker, C. S., I. Morozov, K. Suzuki, T. Romeo, and P. Babitzke.** 2002. CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. Mol. Microbiol. **44**:1599-1610.

3. **Bevilacqua, J. M., and P. C. Bevilacqua.** 1998. Thermodynamic analysis of an RNA combinatorial library contained in a short hairpin. Biochemistry **37**:15877-15884.

4. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95-113.

5. **Boyd, D., D. S. Weiss, J. C. Chen, and J. Beckwith.** 2000. Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system. J. Bacteriol. **182**:842-847.

6. **Butland, G., J. M. Peregrín-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattie, N. Krogan, M. Davey, J. Parkinson, J. Greenblatt, and A. Emili.** 2005. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 433:531-537.

7. **Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson.** 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. **31**:3497-3500.

8. **Corbin, R.W., O. Paliy, F. Yang, J. Shabanowitz, M. Platt, C. E. Lyons, Jr., K. Root, J. McAuliffe, M. I. Jordan, S. Kustu, E. Soupene, and D. F. Hunt.** 2003. Toward a protein profile of *Escherichia coli*: comparison to its transcription profile. Proc. Natl. Acad. Sci. USA **100**:9232-9237.

9. **Dubey, A. K., C. S. Baker, K. Suzuki, A. D. Jones, P. Pandit, T. Romeo, and P. Babitzke.** 2003. CsrA regulates translation of the *Escherichia coli* carbon starvation gene, *cstA*, by blocking ribosome access to the *cstA* transcript. J. Bacteriol. **185**:4450-4460.

10. **Dubey, A. K., C. S. Baker, T. Romeo, and P. Babitzke.** 2005. RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. RNA **11**:1579-1587.

11. **Du, H. and P. Babitzke.** 1998. *trp* RNA-binding attenuation protein-mediated long distance RNA refolding regulates translation of *trpE* in *Bacillus subtilis*. J. Biol. Chem. **273**:20494-20503.

12. **Franze de Fernandez, M. T., L. Eoyang, and J. T. August.** 1968. Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. Nature **219**:588-590.

13. **Gottesman, S.** 2004. The small RNA regulators of *Escherichia coli*: roles and mechanisms. Annu. Rev. Microbiol. **58**:303-328.

14. **Gudapaty, S., K. Suzuki, X. Wang, P. Babitzke, and T. Romeo.** 2001. Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*. J. Bacteriol. **183**:6017-6027.

15. **Hartz, D., D. S. McPheeters, R. Traut, and L. Gold.** 1988. Extension inhibition analysis of translation initiation complexes. Methods Enzymol. **164**:419-425.

16. **Jackson, D.W., K. Suzuki, L. Oakford, J.W. Simecka, M.E. Hart, and T. Romeo.** 2002. Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. J. Bacteriol. **184**:290-301.

17. **Kel, A. E., E. Gößling, I. Reuter, E. Cheremushkin, , O.V. Kel-Margoulis, and E. Wigender.** 2003. MATCH^{TM}: a tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res. **31**:3576-3579.

18. **Liu, M. Y., G. Gui, B. Wei, J. F. Preston, 3rd, L. Oakford, U. Yuksel, D. P. Giedroc, and T. Romeo.** 1997. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. J. Biol. Chem. **272**:17502-17510.

19. **Liu, M. Y., H. Yang, and T. Romeo.** 1995. The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. J. Bacteriol. **177**:2663-2672.

20. **Livak, K. J., and T. D. Schmittgen.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta T}$ method. Methods **25**:402-408.

21. **Mondragon, V., B. Franco, K. Jonas, K. Suzuki, T. Romeo, Ö. Melefors, and D. Georgellis.** 2006. pH dependent activation of the BarA-UvrY two-component system in *Escherichia coli*. J. Bacteriol. **188**:8303-8306.

22. **Moody, E. M., J. C. Feerrar, and P. C. Bevilacqua.** 2004. Evidence that folding of an RNA tetraloop hairpin is less cooperative than its DNA counterpart. Biochemistry **43**:7992-7998.

23. **Muffler, A., D. Fischer, and R. Hengge-Aronis,** 1996. The RNA-binding protein HF-1 plays a global regulatory role which is largely, but not exclusively, due to its role in expression of the sigma(s) subunit of RNA polymerase in *Escherichia coli*. J. Bacteriol. **179**:297-300.

24. **Platt, T., B. Müller-Hill, and J. H. Miller.** 1972. Assay of b972. Assay of üller-Hill, a*In* , J. H. Miller (ed.). Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

25. **Remold-O'Donnell, E., and R. E. Thach**. 1970. A new method for the purification of initiation factor F2 in high yield, and an estimation of stoichiometry in the binding reaction. J. Biol. Chem. **245**:5737-5742.

26. **Romeo, T.** 1998. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. Mol. Microbiol. **29**:1321-1330.

27. **Romeo, T., M. Gong, M. Y. Liu, and A.-M. Brun-Zinkernagel.** 1993. Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. J. Bacteriol. **175**:4744-4755.

28. **Sabnis, N. A., H. Yang, and T. Romeo.** 1995. Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene *csrA*. J. Biol. Chem. **270**:29096-29104.

29. **Sauter, C., J. Basquin, and D. Suck.** 2003. Sm-like proteins in eubacteria: the crystal structure of the Hfq protein from *Escherichia coli.* Nucleic Acids Res. **32**:4091-4098.

30. **Schaak, J. E., H. Yakhnin, P. C. Bevilacqua, and P. Babitzke**. 2003. A Mg2+-dependent RNA tertiary structure forms in the *Bacillus subtilis trp* operon leader transcript and appears to interfere with *trpE* translation control by inhibiting TRAP binding. J. Mol. Biol. **332**:555-574.

31. **Schultz, J. E. and A. Matin**. 1991. Molecular and functional characterization of a carbon starvation gene of *Escherichia coli*. J. Mol. Biol. **218**:129-140.

32. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with Gene Fusions. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

33. **Sorger-Domenigg, T., E. Sonnleitner, V. R. Kaberdin, and U. Bläsi**. 2007. Distinct and overlapping binding sites of *Pseudomonas aeruginosa* Hfq and RsmA proteins on the non-coding RNA RsmY. Biochem. Biophys. Res. Commun. **352**:769-773.

34. **Storz, G., J. A. Opdyke, and A. Zhang**. 2004. Controlling mRNA stability and translation with small, noncoding RNAs. Curr. Opin. Microbiol. **7**:140-144.

35. **Suzuki K., P. Babitzke, S. R. Kushner, and T. Romeo.** 2006. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. Genes Dev. **20**:2605-2617.

36. **Suzuki, K., X. Wang, T. Weilbacher, A. K. Pernestig, Ö. Melefors, D. Georgellis, P. Babitzke, and T. Romeo.** 2002. Regulatory circuitry of the CsrA/CsrB and BarA/UvrY systems of *Escherichia coli*. J. Bacteriol. **184**:5130-5140.

37. **Tsui, H. C., G. Feng, and M. E. Winkler.** 1996. Transcription of the *mutL* repair*, miaA* tRNA modification, *hfq* pleiotropic regulator, and *hflA* region protease genes of *Escherichia coli* K-12 from clustered Esigma32-specific promoters during heat shock J. Bacteriol. **178**:5719-5731.

38. **Tsui, H. C., H. C. Leung, and M. E. Winkler.** 1994. Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. Mol. Microbiol. **13**:35-49.

39. **Valentin-Hansen, P, M. Eriksen, and C. Udesen**. 2004. The bacterial Sm-like protein Hfq: a key player in RNA transactions. Mol. Microbiol. **51**:1525-1533.

40. **Vasil'eva, I. M., and M. B. Garber**. 2002. The regulatory role of the Hfq protein in bacterial cells. Mol. Biol. **36**:785-791.

41. **Vecerek, B., I. Moll, and U. Bläsi.** 2005. Translational autocontrol of the *Escherichia coli* *hfq* RNA chaperone gene. RNA **11**:976-984.

42. **Wang, X., A. K. Dubey, K. Suzuki, C. S. Baker, P. Babitzke, and T. Romeo**. 2005. CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. Mol. Microbiol. **56**:1648-1663.

43. **Wei, B.L., A. M. Brun-Zinkernagel, J. W. Simecka, B. M. Prüß, P. Babitzke, and T. Romeo.** 2001. Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*. Mol. Microbiol. **40**:245-256.

44. **Wei, B.L., S. Shin, D. LaPorte, A. J. Wolfe, and T. Romeo.** 2000. Global regulatory mutations in *csrA* and *rpoS* cause severe central carbon stress in *Escherichia coli* in the presence of acetate. J. Bacteriol. **182**:1632-1640.

45. **Weilbacher, T., K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke, and T. Romeo.** 2003. A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. Mol. Microbiol. **48**:657-670.

46. **Yakhnin, A. V., J. J. Trimble, C. R. Chiaro, and P. Babitzke.** 2000. Effects of mutations in the L-tryptophan binding pocket of the *trp* RNA-binding attenuation protein of *Bacillus subtilis*. J. Biol. Chem. **275**:4519-4524.

47. **Yakhnin, H., H. Zhang, A. V. Yakhnin, and P. Babitzke**. 2004. The *trp* RNA-binding attenuation protein of *Bacillus subtilis* regulates translation of the tryptophan transport gene *trpP* (*yhaG*) by blocking ribosome binding. J. Bacteriol. **186**:278-286.

48. **Yang, H., M. Y. Liu, and T. Romeo.** 1996. Coordinate genetic regulation of glycogen catabolism and biosynthesis in *Escherichia coli* via the CsrA gene product. J. Bacteriol. **178**:1012-1017.

49. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31**:3406-3415.

FIGURE LEGEND

FIG. 1. Predicted CsrA binding site overlapping the *hfq* SD sequence. The SELEX-derived CsrA binding site consensus sequence is shown above the predicted CsrA binding site in *hfq* mRNA. Vertical lines mark the residues in the predicted site that match those in the consensus. Positions of the *hfq* SD sequence and translation initiation codon (Met) are shown.

FIG. 2. Gel mobility shift anlaysis of CsrA-*hfq* RNA interaction. 5'-end-labeled *hfq* RNA was incubated with the concentration of CsrA shown at the bottom of each lane. Gel sift assays were performed in the absence (panel A) or presence (panel B) of various unlabeled competitor RNAs. The concentration of each competitor RNA is shown at the bottom of each lane in panel B. The positions of bound (B) and free (F) *hfq* RNA are shown at the left of each gel. (A) Determination of the equilibrium binding constant of CsrA-*hfq* RNA interaction. The simple binding curve for this data is shown at the right. (B) Competition assay for CsrA-*hfq* RNA interaction to establish binding specificity. Lanes corresponding to competition with specific (*hfq* and SELEX) and non-specific (*trpL*) RNAs are indicated.

FIG. 3. CsrA-*hfq* RNA footprint analysis. (A) *hfq* RNA was treated with RNase T1 or RNase T2 in the absence or presence of CsrA. The concentration of CsrA used is indicated at the top of each lane. Partial alkaline hydrolysis (OH) and RNase T1 digestion (T1) ladders, as well as control (C) lanes in the absence of RNase treatment are shown. The RNase T1 ladder was generated under denaturing conditions so that every G residue in the transcript could be visualized. Residues in which RNase cleavage was reduced (–) or enhanced (+) in the presence of CsrA are marked. The positions of the CsrA footprint (FP), the *hfq* SD sequence and the translation initiation codon (AUG) are shown. Two RNA segments corresponding to RNA secondary structures (h1 and h2) that were previously identified are shown (41). Numbering at the left of each gel is from the start of *hfq* transcription. (B) Summary of the *hfq* footprint (Fig. 3A) and toeprint (Fig. 4) results. The composite CsrA footprint shows the residues in which cleavage was reduced (–) or enhanced (+) by the presence of bound CsrA. Residues corresponding to the CsrA-dependent and 30S ribosomal subunit (Rib) toeprints are marked with arrow heads. An additional 30S ribosomal subunit-dependent toeprint is marked (*). The positions of the *hfq* SD sequence and translation initiation codon (Met) are indicated. Inverted horizontal arrows identify the residues corresponding to h1, h2 and a short RNA hairpin containing a GNRA tetraloop. Vertical arrows identify a triple nucleotide substitution introduced into the CsrA binding site. Numbering is from the start of *hfq* transcription.

FIG. 4. CsrA and 30S ribosomal subunit toeprint analysis of *hfq* RNA. The concentration of CsrA used in each reaction, as well as the absence (–) or presence (+) of $tRNA^{fMet}$ and 30S ribosomal subunits (30S Rib) is shown at the top of each lane. CsrA was added prior

to 30S ribosomal subunits when both were present in the same reaction. Arrows identify CsrA-dependent and 30S ribosomal subunit (Rib) toeprint bands. An additional 30S ribosomal subunit-dependent toeprint is marked (*). Positions of the *hfq* SD sequence and the translation initiation codon (AUG) are shown. The RNA segment corresponding to h1 is also shown. Sequencing lanes to reveal G, U, A or C residues are marked. Numbering is from the start of *hfq* transcription.

FIG. 5. Effect of CsrA and Hfq on in vitro translation of *hfq′-′gfp* mRNA. The *E. coli* S-30 extract was prepared from a CsrA-deficient strain. (A) Reactions were carried out with the concentration of CsrA indicated at the top of each lane in the absence (–) or presence (+) of *hfq′-′gfp* or control (*smpB* or *bla*) transcripts. Hfq-GFP, SmpB and Bla translation products were analyzed by SDS-PAGE. (B) Relative levels of Hfq-GFP, SmpB and Bla polypeptide synthesis as a function of CsrA concentration. All of the bands shown in panel A were used for quantifying the effect of CsrA on protein synthesis. Symbols are: Hfq-GFP, solid circles; Bla, open circles; SmpB, solid squares. The level of polypeptide synthesis in the absence of CsrA was set to 1.0 for each transcript. (C) Reactions were carried out with the concentration of CsrA and/or Hfq indicated at the top of each lane in the absence (–) or presence (+) of *hfq′-′gfp* transcripts. Hfq-GFP products were analyzed by SDS-PAGE.

FIG. 6. CsrA-dependent regulation of an *hfq′-′lacZ* translational fusion. Cells were harvested at various times throughout growth and assayed for β-galactosidase activity.

Growth media was LB (A and B) or LB supplemented with 1% glucose (C). Growth curves for each strain in (A), (B) or (C) were essentially identical. Time is hours of cell growth. These experiments were conducted at least three times with similar results. Results from representative experiments are shown. (A) Symbols for β--galactosidase activity: PLB785 (wild type), solid circles; PLB786 (*csrA::kan*), open circles. Symbols for growth: PLB785, x. (B) Symbols for β--galactosidase activity: ((*csrA::kan* / pCRA16 [*csrA*+]), solid circles; PLB789 (*csrA::kan* / pBR322), open circles. Symbols for growth: PLB789, x. (C) Symbols for β-galactosidase activity: (*csrA::kan*), open circles; PLB923 (wild type with mutant *hfq'-'lacZ* fusion), solid squares; PLB924 (*csrA::kan* with mutant *hfq'-'lacZ* fusion), open squares.

FIG. 7. Effect of growth phase and CsrA on *hfq* mRNA stability. *hfq* mRNA half-lives were determined in wild type and *csrA* mutant strains during exponential and early stationary phase growth. The relative levels of mRNA remaining at 0, 1, 2, 4, 8, 15 and 32 min after the addition of rifampicin was determined by rt-qRT-PCR. The mRNA level corresponding to each 0 min time point was set to 100. The mRNA half-life for each strain and growth phase is shown next to the corresponding symbol. Symbols are: MG1655 (wild type) in exponential phase (exp), solid circles; TR1-MG1655 (*csrA::kan*) in exponential phase, open circles; MG1655 in early stationary phase (stat), solid squares; TR1-MG1655 in early stationary phase, open squares.

Figure 1

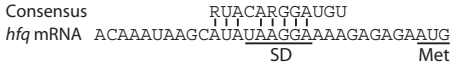

Consensus
RUACARGGAUGU
hfq mRNA
ACAAAUAAGCAUAUAAGGAAAAGAGAGAAUG
SD
Met

Figure 2

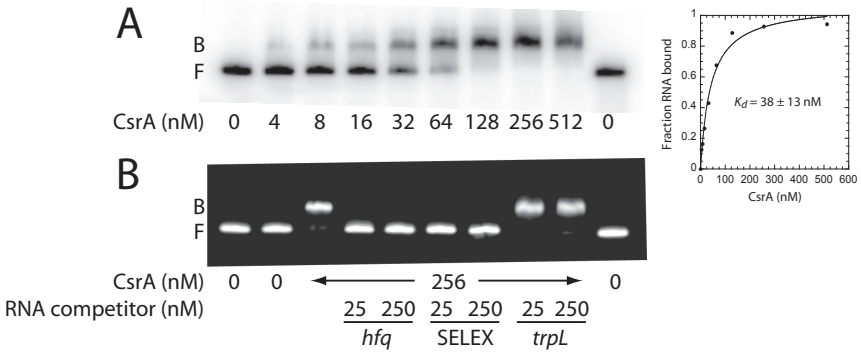

A
B
F
CsrA (nM) 0 4 8 16 32 64 128 256 512 0
Fraction RNA bound
K_d = 38 ± 13 nM
CsrA (nM)
B
B
F
CsrA (nM) 0 0 256 0
RNA competitor (nM) 25 250 25 250 25 250
hfq
SELEX
trpL

Figure 3

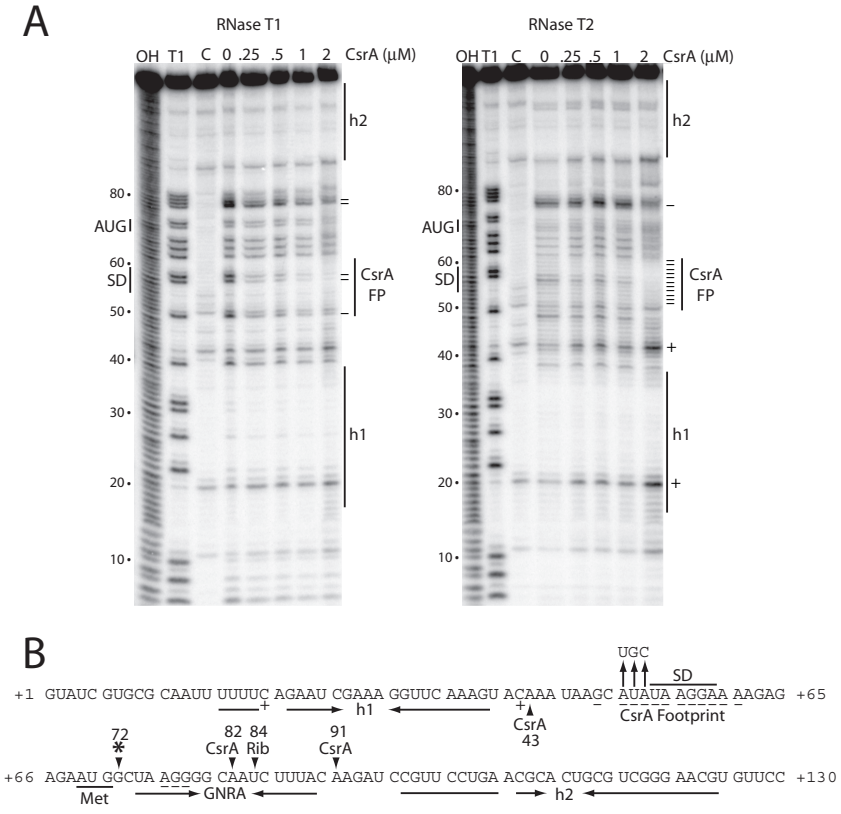

A
RNase T1
RNase T2
OH T1 C 0 .25 .5 1 2 CsrA (μM)
h2
h1
CsrA FP
AUG
SD
80
60
50
40
30
20
10
B
+1 GUAUC GUGCG CAAUU UUUUC AGAAU CGAAA GGUUC AAAGU ACAAA UAAGC AUAUA AGGAA AAGAG +65
h1
CsrA 43
UGC
SD
CsrA Footprint
72
82 84
CsrA Rib
91
CsrA
+66 AGAAU GGCUA AGGGG CAAUC UUUAC AAGAU CCGUU CCUGA ACGCA CUGCG UCGGG AACGU GUUCC +130
Met
GNRA
h2

Figure 4

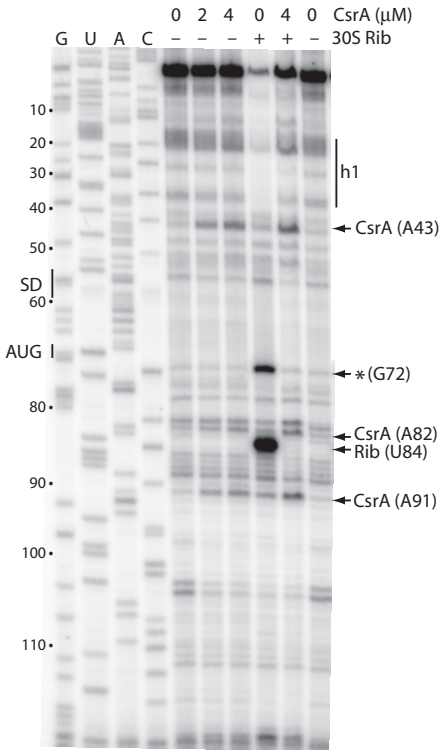

0 2 4 0 4 0 CsrA (μM)
G U A C – – – + + – 30S Rib
10
20
30
40
50
SD
60
AUG
80
90
100
110
h1
CsrA (A43)
*(G72)
CsrA (A82)
Rib (U84)
CsrA (A91)

Figure 5

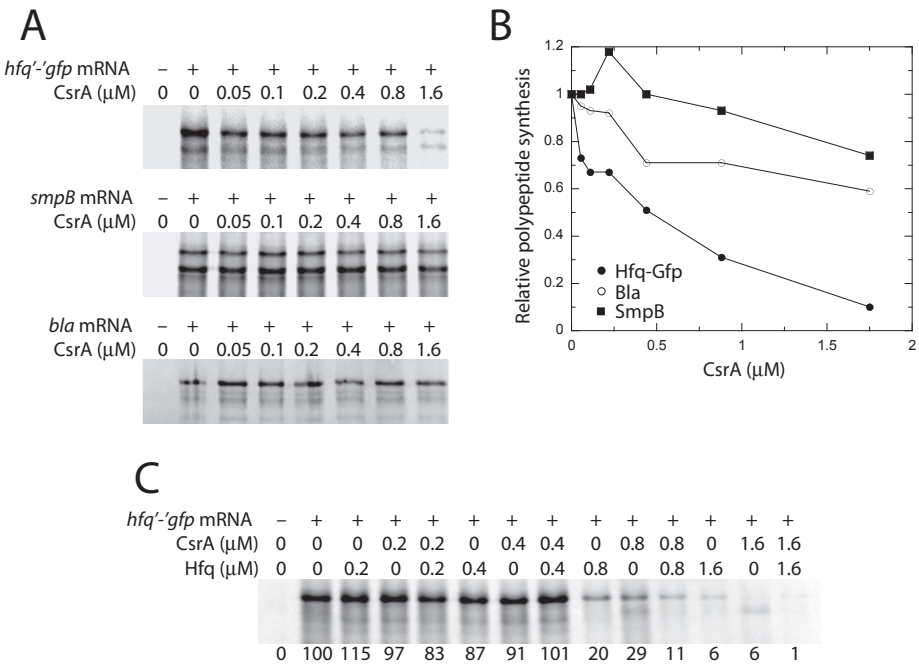

A
hfq'-'gfp mRNA – + + + + + + +
CsrA (μM) 0 0 0.05 0.1 0.2 0.4 0.8 1.6
smpB mRNA – + + + + + + +
CsrA (μM) 0 0 0.05 0.1 0.2 0.4 0.8 1.6
bla mRNA – + + + + + + +
CsrA (μM) 0 0 0.05 0.1 0.2 0.4 0.8 1.6
B
Relative polypeptide synthesis
Hfq-Gfp
Bla
SmpB
CsrA (μM)
C
hfq'-'gfp mRNA – + + + + + + + + + + + + +
CsrA (μM) 0 0 0 0.2 0.2 0 0.4 0.4 0 0.8 0.8 0 1.6 1.6
Hfq (μM) 0 0 0.2 0 0.2 0.4 0 0.4 0.8 0 0.8 1.6 0 1.6
0 100 115 97 83 87 91 101 20 29 11 6 6 1

Figure 6

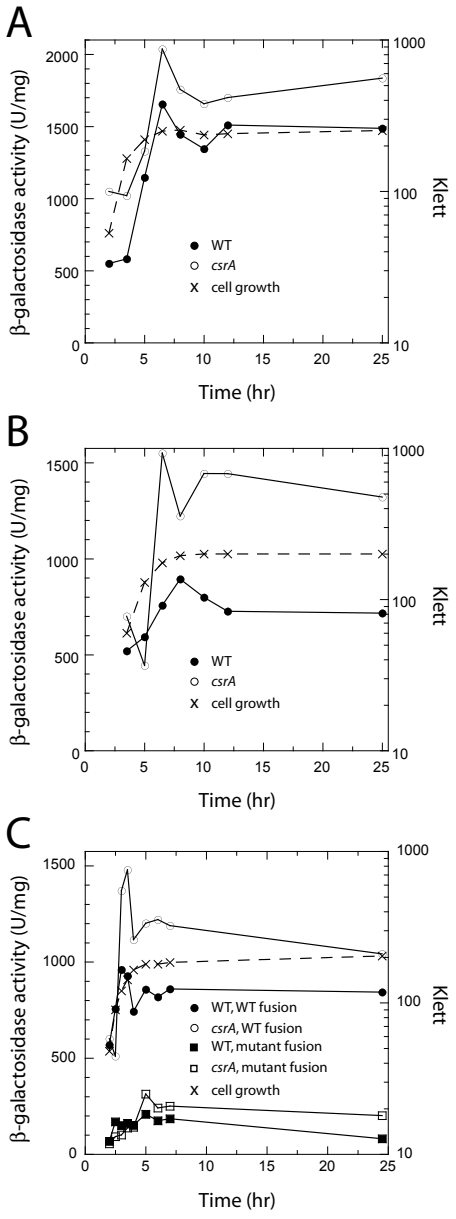
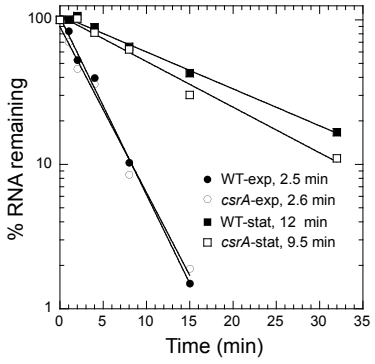

A
β-galactosidase activity (U/mg)
Klett
Time (hr)
WT
csrA
cell growth
B
β-galactosidase activity (U/mg)
Klett
Time (hr)
WT
csrA
cell growth
C
β-galactosidase activity (U/mg)
Klett
Time (hr)
WT, WT fusion
csrA, WT fusion
WT, mutant fusion
csrA, mutant fusion
cell growth

Figure 7


100
10
1
% RNA remaining
0
5
10
15
20
25
30
35
Time (min)
WT-exp, 2.5 min
csrA-exp, 2.6 min
WT-stat, 12 min
csrA-stat, 9.5 min

The evolution of contact-dependent inhibition in non-growing populations of *Escherichia coli*

Marc Lemonnier[1,2,3], Bruce R Levin[3]*, Tony Romeo[4], Kim Garner[3], María-Rosario Baquero[5], Jeff Mercante[4], Emmanuel Lemichez[2], Fernando Baquero[5] and Jesús Blázquez[6]

[1]*Department of Molecular Microbiology, Centro de Investigaciones Biológicas CSIC*

28040 Madrid, Spain

[2]*INSERM*

Faculté de Médecine, U627, 06107 Nice, France

[3]*Department of Biology, Emory University*

Atlanta, GA 30322, USA

[4]*Department of Microbiology and Immunology, Emory University School of Medicine*

Atlanta, GA 30322, USA

[5]*Department of Microbiology, Ramón y Cajal Hospital*

28034 Madrid, Spain

[6]*Departamento de Biotecnología Microbiana, Centro Nacional de Biotecnología CSIC*

28049 Madrid, Spain

Author for correspondence (blevin@emory.edu)

This manuscript was published in Proc Biol Sci. 2008 Jan 7; 275(1630):3-10. Jeffrey Mercante constructed bacterial mutants and conducted glycogen accumulation experiments in this study.

Abstract

In the course of liquid culture, serial passage experiments with *Escherichia coli* K-12 bearing a mutator gene deletion (Δ*mut*S) we observed the evolution of strains that appeared to kill or inhibit the growth of the bacteria from where they were derived, their ancestors. We demonstrate that this inhibition occurs after the cells stop growing and requires physical contact between the evolved and ancestral bacteria. Thereby, it is referred to as stationary phase contact-dependent inhibition (SCDI). The evolution of this antagonistic relationship is not anticipated from existing theory and experiments of competition in mass (liquid) culture. Nevertheless, it occurred in the same way (parallel evolution) in the eight independent serial transfer cultures, through different single base substitutions in a gene in the glycogen synthesis pathway, *glgC*. We demonstrate that the observed mutations in *glgC*, which codes for ADP-glucose pyrophosphorylase, are responsible for both the ability of the evolved bacteria to inhibit or kill their ancestors and their immunity to that inhibition or killing. We present evidence that without additional evolution, mutator genes, or known mutations in *glgC*, other strains of *E. coli* K-12 are also capable of SCDI or sensitive to this inhibition. We interpret this, in part, as support for the generality of SCDI and also as suggesting that the *glgC* mutations responsible for the SCDI, which evolved in our experiments, may suppress the action of one or more genes responsible for the sensitivity of *E. coli* to SCDI. Using numerical solutions to a mathematical model and *in vitro* experiments, we explore the population dynamics of SCDI and postulate the conditions responsible for its evolution in mass culture. We conclude with a brief discussion of the potential ecological significance of SCDI and its possible utility for the development of antimicrobial agents, which unlike existing antibiotics, can kill or inhibit the growth of bacteria that are not growing.

1. Introduction

When bacteria compete in liquid culture, the evolution of mutants that replicate faster on existing or by-product resources made available by the metabolism or death of other members of the community is anticipated and has been observed (Helling *et al*. 1987; Lenski *et al*. 1991; Zambrano & Kolter 1996; Farrell & Finkel 2003). Not expected in these liquid (mass) culture conditions is the evolution of antagonistic relationships, mechanisms that enable bacteria to kill or inhibit the growth of competitors, like the production of bacteriocins or antibiotics. In the absence of spatial heterogeneity, all members of the community gain equally from the resources made available by the death or growth inhibition of other members. As a result, individual bacteria producing these allelopathic agents and engendering the cost of their production would not gain advantage that would enable them to increase in frequency when they are rare (Chao & Levin 1981; Levin 1988; Kerr *et al*. 2002). Nevertheless, in broth cultures of *Escherichia coli* we observed the rapid evolution of mutants that appeared to inhibit the growth or kill the bacteria from where they were derived—their ancestors. This inhibition does not occur until these bacteria are at stationary phase, a time when they are refractory to other toxic agents like antibiotics (Bigger 1944; McDermott 1958; Levin & Rozen 2006).

In this report, we examine the properties of this evolved antagonistic interaction, determine its genetic basis, explore its generality in other strains of *E. coli* and with the aid of a mathematical model explore the population dynamics of this process and the conditions responsible for its evolution. We conclude with a brief discussion of the potential ecological implications and practical applications of this phenomenon.

2. Material and methods

In the following we describe some of the materials and methods used in this investigation. More information about these methods can be found in the electronic supplementary material.

(a) Culture medium

Unless noted, bacterial strains were grown at 37°C in Luria–Bertani (LB) media (Miller 1972) or Kornberg (KB) media (1.1% K_2HPO_4, 0.85% KH_2PO_4, 0.6% yeast extract and 1% glucose).

(b) Sampling methods

Estimates of the densities of the bacterial cultures and the relative frequencies of the different populations were obtained by serial dilution and plating on LB agar, LB agar with streptomycin ($100\,\mu g\,ml^{-1}$), ampicillin ($100\,\mu g\,ml^{-1}$), chloramphenicol ($25\,\mu g\,ml^{-1}$) or tetracycline ($10\,\mu g\,ml^{-1}$). Tetrazolium arabinose (TA; Levin *et al*. 1977) agar or McConkey agar (DIFCO 1984) containing 1% arabinose was also used to estimate the relative frequencies of strains that differed by an arabinose marker. Unless otherwise noted, liquid cultures were in 50 ml Erlenmeyer flask containing LB (10 ml) and were maintained at 37°C with shaking at between 150 and 250*g*.

(c) Bacteria

AB1157. thr-1 leu-6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1 proA2 his-4 argE3 str-31 tsx-33 supE44 rec$^+$*F-*.

AB1157ΔmutS::spc gyrA (a derivative of AB1157 provided by Ivan Matic). For convenience, we designate this most prominently used strain as ABM. We also used spontaneous Ara$^+$ revertants of this strain for mixed culture experiments.

MC1061. F$^-$ *araD139 Δ*(*ara-leu*)*7696 galE15 galK16 Δ*(*lac*)*X74 rpsL* (*Strr*) *hsdR2* (*rK− mK+*) *mcrA mcrB1*.

MG1655. *ilvG rfb-50 rph*$_-$ λ^--F-.

C600. F-, *thi-1*, *thr-1*, *leuB6*, *lacY1*, *tonA21*, *supE44*, λ^-.

Escherichia coli B. No mutations known *CGSC#:* 5713.

(d) Mixed culture experiments

Unless otherwise noted in the electronic supplementary material, strains in the mixed culture experiments were distinguished by an arabinose marker, pink Ara$^+$ and deep red Ara$^-$ colonies on the TA agar. For this, we used spontaneous Ara$^+$ mutants of the ancestral and evolved ABM. Reciprocal Ara$^+$ evolved and Ara$^-$ ancestral and Ara$^-$ evolved and Ara$^+$ ancestral mixed culture experiments were performed to control for the effects of the Ara$^-$ marker. None were observed.

(e) Fluorescence-activated cell sorting analysis

Discosoma red (DsRed)-labelled *E. coli* were obtained by transformation of ABM cells with plasmid pMW211; a colE1 derivative that expresses a DsRed variant protein and confers resistance to ampicillin (a kind gift from Dr Arthur Altenhoefer, University of Wuerzburg). Green fluorescent protein (GFP)-labelled *E. coli* were obtained following transformation of ABM cells with the multicopy plasmid pGFP (Amp^R), (a kind gift from S. Méresse, centre d'Immunologie de Marseille-Luminy). Flow cytometry was carried out using a Becton Dickinson fluorescence-activated cell sorting (FACS) Vintage SE instrument with a 70 µm sorting nozzle at low pressure (12 psi). GFP and DsRed were excited using a 488 nm blue laser and detected using 530/30 and 585/42 nm filters, respectively. DsRed-labelled ancestral ABM and GFP-labelled evolved ABM *E. coli* were prepared by overnight growth in LB broth supplemented with ampicillin (75 µg ml^{-1}). Mixed cultures (10 ml) initiated at a 10 : 1 ancestor-to-evolved ratio were incubated in 500 ml

culture flasks at 37°C with shaking at 250*g* for the times indicated (figure 3). Evolved-bound ancestor cells and free ancestor cells were isolated by sorting approximately 30 000 particles of each cell population into separate tubes containing phosphate-buffered saline (0.3 ml). Serial dilutions of the samples were plated on McConkey agar plates containing arabinose to discriminate between Ara^- (evolved) and Ara^+ (ancestors) bacteria. For the control experiment shown in figure 3*b*,*d*, the GFP-labelled ancestors were Ara^-, to distinguish them from the DsRed-labelled Ara^+ ancestors. The viability of DsRed-ancestral cells was determined by counting colony-forming units per particle sorted from each gated population.

3. Results

(a) The phenomenon

The evolved antagonistic interaction reported here was an unanticipated (serendipitous) outcome of *in vitro* evolution experiments. Eight independent cultures of a mutator strain of *E. coli* K-12 AB1157Δ*mutS::spc* (Bregeon *et al*. 1999; here referred to as ABM) were maintained in 50 ml flasks by daily transfer (100 µl–10 ml) in LB broth. After 62 passages (approx. 412 generations), 50 µl of overnight cultures derived from single colonies from 'evolved' cultures were mixed in 10 ml LB with 50µl of ABM and grown together for 24 hours. For all eight cultures, the density (CFUs) of the ancestral cells after 24 hours of growth was less than 10^{-3} that of the evolved. This rapid decline in the density of the ancestral cells did not occur when these bacteria were in single clone culture, and in mixed culture it only occurred after the populations stopped growing (figure 1*a*,*b*). The details of the methods for this and the other experiments reported here and more information about the phenomenon can be found in the §2 and in the electronic supplementary material.

This precipitous decline in the density of the ancestral bacteria also occurs when ancestral cells are mixed at stationary phase with evolved bacteria from any of the eight independent serial transfer cultures or after stationary phase when they are grown together. This inhibition or killing does not occur when evolved cells from independent cultures are mixed with each other (data not shown). Moreover, this apparent killing of the ancestral cells by evolved is frequency and density dependent; it does not occur when the frequency of the evolved cells in the mixture is low or when the total density of bacteria in sterile filtrates of stationary phase cells is less than approximately 5×10^7 (figure 2).

(b) Killing or inhibition?

Because the preceding results are based on colony counts, it is possible that in the liquid cultures being sampled the ancestral cells were viable, but their ability to form colonies was inhibited by the evolved cells upon entry into the stationary phase of growth. To ascertain whether this is the case or if the ancestral cells are in fact killed, we tested for the selective inclusion of propidium iodide (PI), which does not occur in viable cells (Boulos *et al*. 1999). The ancestral ABM cells do not selectively include PI in the mixed stationary phase cultures with evolved (see figure S9 and more detailed information in the electronic supplementary material). Thus, we cannot conclude that the ancestral bacteria are killed by the evolved, but may well be converted to a viable but not culturable (VBNC) state (see §4). For this reason, we shall refer to this phenomenon as inhibition rather than killing.

(c) Stationary phase contact-dependent inhibition

In theory (the results of mathematical modelling studies), the observed frequency- and density-dependent inhibition of ancestral bacteria by evolved can be attributed to either the evolved bacteria releasing extracellular toxins, allelopathic agents that inhibit the ancestral bacteria, or

inhibition through physical contact with the evolved bacteria (Levin 1988; electronic supplementary material, figure S13). Two lines of evidence support the hypothesis that the observed inhibition cannot be attributed to extracellular toxins: (i) the rate of decline in the density of ancestral cells in filtrates of stationary phase cultures of evolved cells or mixtures of evolved and ancestral cells is no different than that in filtrates of their own cells (electronic supplementary material, figure S1) and (ii) the density of ancestral cells does not decline at this accelerated rate when they are separated from the evolved bacteria by 0.45 μm filters in 'U-tube' experiments (electronic supplementary material, figure S2).

The above experiments, however, do not formally rule out inhibition by a highly labile, secreted allelopathic agent and are therefore only suggestive of a contact-dependent mechanism. Direct evidence that this inhibition involves contact of ancestral bacteria with the evolved comes from FACS experiments (Aoki *et al*. 2005). Evolved ABM were labelled with the GFP and the ancestral ABM with the DsRed fluorescent protein and grown together in LB. Three different bacterial populations were observed: free green evolved cells; free red ancestral cells; and aggregates of evolved and ancestral cells (figure 3*a*). This aggregation also occurred in control cultures with mixtures of DsRed- and GFP-labelled ancestral cells (figure 3*b*). The viability of the free ancestral cells was compared (from CFU data) with that of the clustered, evolved-bound ancestors following FACS of mixed cultures of evolved and ancestral bacteria. The results indicated that the evolved-bound ancestors declined at a faster rate than their unbound counterparts following the onset of stationary phase contact-dependent inhibition (SCDI; figure 3*c*). A lesser but significant decline in the density of recovered free ancestral cells was also observed in these experiments. We attribute this to their release (dissociation) from the aggregates with evolved as a result of agitation during sampling. Consistent with this interpretation is the observation that there was no loss of viability in control cultures with mixtures of GFP- and DsRed-labelled ancestral cells (figure 3*d*).

(d) SCDI in other *E. coli* strains

To begin to explore the generality of the SCDI phenomenon, we grew mixtures of overnight LB cultures of the evolved and the ancestral ABM with five different stocks of *E. coli* K-12 and a strain of *E. coli* B. (i) The evolved ABM inhibited the AB1157 *mutS*+ from where they were derived and two of its derivatives (AB1157-D and JC5129) and an *E. coli* K-12 strain, MC1061, which has a different genetic ancestry than AB1167. (ii) The ancestral ABM as well as AB1157 *mutS*+ are inhibited at stationary phase when they are mixed with the *E. coli* K-12 strains MG1655 and C600 and a wild-type *E. coli* B (electronic supplementary material, figures S13 and S14). In all of these cases, inhibition does not occur until after the bacteria are at stationary phase nor does it occur when the inhibited strain is in sterile supernatants of the challenging *E. coli*. Finally, the replication of the evolved ABM is not inhibited when it is mixed with either MG1655, C600 or *E. coli* B, despite the ability of these bacteria to inhibit its unevolved ancestor.

(e) The genetic basis of SCDI

As a first attempt to identify the gene(s) responsible for the observed SCDI, we took a genetic approach that was based on the assumption that the alleles which evolved in the serial passage, evolution experiments with ABM would be dominant (details in the electronic supplementary material). We found that in this evolved ABM, SCDI can be attributed to single base substitution mutations in *glgC*, which encodes ADP-glucose pyrophosphorylase, a regulatory enzyme that catalyses the first reaction of bacterial glycogen synthesis (Ballicora *et al*. 2003). The details of the procedures used to identify this gene and demonstrate that it is responsible for both the immunity of the evolved strains to killing and their ability to kill the ancestral strain are presented in the electronic supplementary material. In summary: (i) a shotgun library of genomic DNA from an evolved 62nd transfer clone was constructed with a mini-F plasmid which was used to transform the ancestral ABM. (ii) Independent transformants were pooled, mixed with evolved

cells and passaged twice to enrich for bacteria carrying potential genes for immunity to SCDI. (iii) The mini-F plasmid pCDI1 of the pSCDI-refractory clone obtained contained the *glgCAP* operon for glycogen synthesis which bore a *glgC* allele that differed from the ancestral *glgC* by a C to T transition in the 17th codon (proline to serine; figure 4*a*). (iv) This mutation, *glgC17*, was cloned in a multicopy pBR322 plasmid, pMLM141. (v) Both the pMLM141 transformants of the ancestral AB1157 Δ*mutS*, and the original evolved strain produced high concentrations of glycogen as measured by iodine staining, while the ancestral cell line and the evolved strains deleted for the chromosomal *glgC17* produced little or no glycogen (figure 4*b*). (vi) Ancestral cell transformants bearing pMLM141 were not inhibited by the evolved strain while those bearing pBR322 without the *glgC17* were (figure 4*c*). (vii) Ancestral cells carrying pMLM141 inhibited those bearing pBR322 without *glgC17* as well as evolved cells that were deleted for *glgC17* (figure 4*d*).

Our results indicate that evolution of SCDI in the eight serial passage cultures occurred independently but was convergent phenotypically as well as genetically. Although the extent varied among the eight independently evolved SCDI strains, all overproduced glycogen relative to the ancestral strain as well as to an *E. coli* K-12 MG1655 control (data not shown). Furthermore, all eight evolved SCDI strains bore single non-synonymous base substitutions in *glgC*; two in the 17th codon and one each in the 14th, 66th, 125th, 318th, 330th and 336th codons (electronic supplementary material, table 1).

A biochemical clue to the reason for the high glycogen phenotype of these mutants is provided by the finding that the G336D *glgC* mutation increases enzymatic activity and alters the allosteric behaviour of ADP-glucose pyrophosphorylase (Ballicora *et al*. 2003). This overproduction of glycogen appears to be necessary for the inhibition of ancestral cells. AB1157 *glgC17* mutants deleted for *glgA* (glycogen synthase), that no longer synthesize glycogen, do not cause SCDI when mixed with ancestral ABM (electronic supplementary material, figure S8). On the other

hand, inactivation of *glgP* had no impact on SCDI. This gene encodes glycogen phosphorylase, the major exolytic enzyme of glycogen catabolism Alonso-Casajús *et al*. 2006, which removes glucosyl units from the non-reducing ends of the polymer (electronic supplementary material, figure S8). This suggests that glycogen catabolism may not be required for SCDI.

While it is clear that *glgC* mutations are responsible for evolved SCDI in ABM, our results suggest that these mutations are most probably suppressors that compensate for defects in ABM. AB1157 and MC1061 and possibly other strains may have similar or identical defects that make them susceptible to SCDI. The DNA sequences of *glgC* in MG1655 and the ancestral ABM as well as AB1157 are identical (data not shown). Thus, mutations in *glgC* are not necessary for SCDI by MG1655. The genetic basis for these defects awaits further elucidation before any plausible mechanistic model for SCDI can be proposed.

(f) Population and evolutionary dynamics of SCDI

As noted in §1, the evolution of an antagonistic interaction such as SCDI in liquid (mass) culture is inconsistent with the proposition that mechanisms to kill or inhibit the growth of competitors will not evolve under these conditions (Chao & Levin 1981; Levin 1988; Frank 1994; Kerr *et al*. 2002). In accordance with this theory and these experiments, if there is a cost in the fitness of the evolved SCDI strain associated with its ability to inhibit the growth or kill the ancestral strain, when rare SCDI mutants should not be able to increase to frequencies where they would be detected. To illustrate this and other properties of the population dynamics of SCDI, we used numerical solutions to the differential equations of a simple mathematical model, a computer simulation. In this model we consider a serial transfer, liquid culture populations of the sort used in our evolution experiments with 1/100 dilutions into fresh medium occurring every 24 hours. We assume a mass action process so that interaction between the ancestral and evolved bacteria occurs at a rate proportional to the product of their densities. In this model, ancestral cells are

killed instantly upon contact with evolved and this contact-dependent killing does not start until 16 hours after the populations starts to grow, when the bacteria when the culture would be at stationary phase. The simulated populations are initiated with only ancestral cells. The evolved SCDI cells are produced at random by recurrent mutation. For more details about this simulation and the values of the parameters used in these illustrations, see the electronic supplementary material.

If SCDI engenders a 1% cost in the exponential growth rate, evolved cells continue to be present, due to recurrent mutation, but their population does not increase in frequency, despite the SCDI advantage (figure 5A). If the ability to inhibit ancestral cells engenders no cost or benefit to the density- and frequency-independent fitness of the evolved bacteria, then as a consequence of SCDI the density of the evolved population increases at a rate which itself increases with the density of the evolved cells (figure 5B). However, if SCDI is the only advantage possessed by the evolved bacteria, their rate of ascent will be low and they would probably not have been detected, much less dominated the population by the 62nd transfer (as we observed). On the other hand, if in addition to SCDI the ancestral bacteria had a higher rate of mortality than the evolved, the frequency of the evolved would increase due to the combination of this density-independent fitness advantage and the density/frequency-dependent advantage resulting from SCDI (figure 5C). As can be seen in the electronic supplementary material, in single clone culture the evolved bacteria in fact die at a lower rate than the ancestral strain (electronic supplementary material; figure 4).

4. Discussion

The observation that initiated this investigation was serendipitous. Based on existing theory and observations (Chao & Levin 1981; Levin 1988; Frank 1994; Kerr *et al*. 2002), the evolution in

mass (liquid) culture of bacteria that inhibit the growth or kill the cells from where they are derived, their ancestors, would not have been anticipated or sought. However, in retrospect, the observation that contact-dependent inhibition (CDI), SCDI, evolved independently by different mutations in the same gene in eight cultures of a mutator strain of *E. coli* is an excellent and novel, but not at all surprising, example of parallel evolution. For other examples of parallel changes occurring in independent experimental populations of bacteria derived from the same ancestor lineage, see Cunningham *et al*. (1997), Wichman *et al*. (1999, 2000, 2005), Cooper *et al*. (2003), Crozat *et al*. (2005), Sachs & Bull (2005), Woods *et al*. (2006) and Bantinaki *et al*. (2007).

Owing to a yet unknown mutation or a pleiotropic effect of a known mutation in the course of its long existence in laboratory culture, the *mut*S+ AB1157 strain of *E. coli* K-12 may have become sensitive to CDI at stationary phase by at least some other strains of *E. coli* including other K-12 derivatives (MG1655, C600 as well as the *glgC* mutants that evolved in these experiments). This sensitivity was maintained in the mutator construct of AB1157, Δ*mutS::spc* (Bregeon *et al*. 1999) used in the evolution experiments where this phenomenon was first observed. In the course of these experiments, missense mutations were generated in *glgC* that both suppressed the sensitivity of these bacteria to CDI during the stationary phase these serial transfer experiments and permitted these bacteria to inhibit the ancestral strain. Although we have not formally demonstrated it, by moving the *glgC* mutant genes responsible for SCDI to separate AB1157 backgrounds, our results and parsimony suggest that these *glgC* mutations also provided ABM with a density- and frequency-independent fitness advantage by reducing their rate of mortality. As a consequence of the latter, these *glgC* mutations were able to increase in frequency when they were rare and achieve densities where they had the additional advantage of inhibiting/killing the dominant population with wild-type *glgC* loci. The primary, if not exclusive, role of the *mutS* gene in this process was to increase the rate at which the suppressing *glgC* mutations occurred,

which in addition to providing the variation needed for this evolution, reduced the likelihood of their loss in the bottlenecks associated with serial passage (Levin *et al*. 2000). Presumably SCDI would eventually evolve if *mutS*+ strains of AB1157 or other strains with this sensitivity were maintained long enough in serial transfer cultures of this type.

In some ways, the SCDI that evolved in these experiments is similar to the growth advantage in stationary-phase (GASP) phenomenon studied by Kolter and colleagues (Zambrano & Kolter 1996; Farrell & Finkel 2003); it too is manifest at stationary phase and requires high pH (greater than 8.5; electronic supplementary material, figure S5). SCDI is, however, clearly different from GASP both functionally and genetically. (i) The decline in plating efficiency of ancestral cells when mixed with evolved (figure 1*b*) occurs at a rate nearly ten times as great as the highest reported for GASP (Zambrano & Kolter 1996). (ii) Contrary to what would be anticipated by the scavenging mechanism (differential use of resources made available from dead cells) postulated for GASP, there was little or no growth of evolved cells when they were inoculated at low densities into filtrates of stationary phase ancestral cells or filtrates of sonicated ancestral cells (data not shown). (iii) The *lrp* and *rpoS* mutations that account for GASP are not present in the evolved ABM, nor in their ancestors ABM and AB1157 (data not shown).

The underlying mechanisms by which cells are inhibited in the SCDI reported here and the CDI reported by Aoki *et al*. (2005) have not yet been elucidated. However, it is clear that these two inhibitory processes are functionally and genetically distinct. CDI of *E. coli* K-12 (MG1655) by a naturally occurring strain of *E. coli* from a urinary tract infection occurred when the bacteria were growing, rather than at stationary phase. Moreover, the *cdiA* and *cdiB* genes that were found to be responsible for CDI by the wild *E. coli* are not present in *E. coli* K-12, wherein SCDI evolved in the present study.

In this report we use the term ‘inhibition’ rather than ‘killing’ in our description of the observed SCDI because the results of the PI test suggested that the ancestral bacteria are VBNC, when mixed with the evolved cells at stationary phase. It should be noted, however, that the concept of VBNC bacteria is controversial; whether these cells are dormant or are progressively undergoing cell death is unclear (Nystrom 2003; Oliver 2005). What is clear from our studies is that the decline in the density of the inhibited strain in these cultures cannot be attributed to their failure to form colonies on agar. Their densities continue to decline in successive serial passages (electronic supplementary material, figure S10).

The mechanism by which the *glgC* mutations that evolved in these experiments convert the ancestral strain into an inhibitor requires further research. Excessive glycogen synthesis, the apparent phenotype derived from the *glgC* mutants, could trigger the inhibitory function. Glycogen excess mutations of *E. coli* have been mapped to the *glgC* gene encoding ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis (Ballicora *et al*. 2003 and references therein). In particular, certain amino acid substitutions of GlgC are known to alter the allosteric behaviour of the protein and increase glycogen synthesis (Meyer *et al*. 1993; Wu & Preiss 1998). However, to the best of our knowledge there is no evidence that glycogen synthesis is correlated with any known bactericidal activity. In solution, glycogen is devoid of antimicrobial activity (data not shown). High glycogen accumulation might alter the cell structure, leading to effects involved in killing. Indeed, evolved hyperglycogenic strains are more susceptible than the ancestors to SDS 0.2% (data not shown), suggesting possible alterations of the bacterial cell envelope. We cannot discount at the present time that the excess glycogen synthesis may be related to oxidative or other envelope stress pathways. Hence, our report here of the SCDI phenomenon in *E. coli* populations sets the basis for future research towards the precise mechanism of SCDI in this and eventually other bacteria.

How important the CDI reported by Aoki *et al*. (2005) and that observed in this study are to the ecology of bacteria is unclear at this time but is certainly intriguing and worthy of further investigation. If in natural populations CDI and SCDI have effects of the magnitude observed in these laboratory studies, they could have a profound effect on the structure of bacterial communities. CDI could facilitate the invasion of strains into habitats occupied by other members of their species and lead to the elimination of these competitors. From a medical perspective CDI may well point to as yet unexploited chinks in the bacterial armour. By elucidating the mechanisms responsible for the sensitivity of *E. coli* K-12 to SCDI, new targets for antimicrobial agents could be identified. Most if not all classic and contemporary antibiotics have little effect on bacteria that are not replicating (Bigger 1944; McDermott 1958; Levin & Rozen 2006). In practice (Eagle 1952; McDermott 1958; Tuomanen 1986) as well as in theory (Levin 2004; Wiuff *et al*. 2005; Levin & Rozen 2006) tolerant, persistent and latent populations of non-replicating, antibiotic-refractory bacteria can extend the term of treatment as well as lead to treatment failure. The mechanism responsible for the SCDI reported here is particularly appealing because the inhibition or killing occurs when the bacteria are not replicating.

Acknowledgments

We are grateful to Roberto Kolter and Maria Zambrano not only for their useful comments, suggestions during the early phase of this project, but also for their inspiration. We also thank Luba Beylina, Nina Walker and Mariastella Tucker for their superb technical support, Julie Cazareth at the cell sorting facility from CNRS UMR-6097 and Grégoire Lauvau at INSERM E-344 (IPMC, Valbonne, France) for invaluable help and suggestions during the FACS experiments, Anne Doye (INSERM, U627) for excellent assistance with the confocal microscopy analysis and for preparing high-quality micrographs, Xin Wang for advice and assistance in constructing *glgA* and *glgP* deletions, Lauren Ancel, Mark Jensen, Mary Reynolds, Jeff Smith,

Mark Tanaka and Renata Zappala for their comments and suggestions. We are also grateful to Ivan Matic, Mary Berlyn and her colleagues at the Coli Genetic Stock Center for providing the strains of bacteria used in this investigation, and to Arthur Altenhoefer and Stéphane Méresse for constructs needed for the fluorescent proteins. We express a particular depth of gratitude for the intellectual and practical (space and facilities) generosity of Ramón Díaz Orejas (CIB, Madrid) in whose laboratory most of the molecular genetic experiments were performed. This research was supported by grants from the US Nation Institutes of Health, AI40662, GM33782 (B.R.L.), GM59969 (T.R.), CO3/14 of the Spanish Fondo de Investigaciones Sanitarias and SAF-2002-04649, BFU2005-03911/BMC (R. D. O.), BFU2004-00879 (J.B.) of the Spanish Ministerio de Educación y Ciencia, and a grant from the European Union (QLK2-CT-2001-873, and BIO2005-04278 and LSHM-CT-2005-518152 (F.B.)). M.L. was supported by a postdoctoral fellowship from La Ligue Nationale Contre le Cancer. M.R.B. was in part supported by the Alfonso X el Sabio University in Madrid.

References

1. **Alonso-Casaju´ s, N., Dauville´e, D., Viale, A. M., Mun˜ oz, F. J., Baroja-Ferna´ndez, E., Mora´n-Zorzano, M. T., Endallin, G., Ball, S. & Pozueta-Romero, J.** 2006 Glycogen phosphorylase, the product of the *glgP* gene catalyses glycogen breakdown by removing glucose units from the nonreducing ends in *Escherichia coli*. *J. Bacteriol*. 188, 5266–5272.
2. **Aoki, S. K., Pamma, R., Hernday, A. D., Bickham, J. E., Braaten, B. A. & Low, D. A.** 2005 Contact-dependent inhibition of growth in *Escherichia coli*. *Science* 309, 1245–1248.

3. **Ballicora, M. A., Iglesias, A. A. & Preiss, J.** 2003 ADPglucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. *Microbiol. Mol. Biol. Rev*. 67, 213–225.
4. **Bantinaki, E., Kassen, R., Knight, C. G., Robinson, Z., Spiers, A. J. & Rainey, P. B.** 2007 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* 176, 441–453.
5. **Bigger, J. W.** 1944 Treatment of staphylococcal infections with penicillin. *Lancet* 2, 497–500.
6. **Boulos, L., Prevost, M., Barbeau, B., Coallier, J. & Desjardins, R.** 1999 LIVE/DEAD BacLight: application of a new rapid staining method for direct enumeration of viable and total bacteria in drinking water. *J. Microbiol*. *Methods* 37,77–86.
7. **Bregeon, D., Matic, I., Radman, M. & Taddei, F.** 1999 Insufficient mismatch repair: genetic defects and down regulation. *J. Genet*. 78, 21–28.
8. **Chao, L. & Levin, B. R.** 1981 Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc. Natl Acad. Sci*. *USA* 78, 6324–6328.
9. **Cooper, T. F., Rozen, D. E. & Lenski, R. E.** 2003 Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli. Proc. Natl Acad. Sci. USA* 100, 1072–1077.
10. **Crozat, E., Philippe, N., Lenski, R. E., Geiselmann, J. & Schneider, D.** 2005 Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169, 523–532.
11. **Cunningham, C. W., Jeng, K., Husti, J., Badgett, M., Molineux, I. J., Hillis, D. M. & Bull, J. J.** 1997 Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. *Mol. Biol. Evol*. 14, 113–116.
12. DIFCO 1984 Difco manual: dehydrated culture media and reagents for microbiology. Detroit, MI: Difco Laboratories.

13. **Eagle, H.** 1952 Experimental approach to the problem of treatment failure with penicillin. *Am. J. Med*. 13, 389–399.
14. **Farrell, M. J. & Finkel, S. E.** 2003 The growth advantage in stationary-phase phenotype conferred by rpoS mutations is dependent on the pH and nutrient environment. *J. Bacteriol*. 185, 7044–7052.
15. **Frank, S. A.** 1994 Spatial polymorphism of bacteriocin and other allelopathic traits. *Evol. Ecol.* 8, 369–386.
16. **Helling, R. B., Vargas, C. N. & Adams, J.** 1987 Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* 116, 349–358.
17. **Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J.** 2002 Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature* 418, 171–174.
18. **Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C.** 1991 Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. *American Naturalist* 91, 1315–1341.
19. **Levin, B. R.** 1988 Frequency-dependent selection in bacterial populations. *Phil. Trans. R. Soc. B* 319, 459–472.
20. **Levin, B. R.** 2004 Microbiology. Noninherited resistance to antibiotics. *Science* 305, 1578–1579.
21. **Levin, B. R. & Rozen, D. E.** 2006 Non-inherited antibiotic resistance. *Nat. Rev. Microbiol*. 4, 556–562.
22. **Levin, B. R., Stewart, F. M. & Chao, L.** 1977 Resource—limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. *Am. Nat*. 977, 3–24.
23. **Levin, B. R., Perrot, V. & Walker, N.** 2000 Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* 154, 985–997.

24. **McDermott, W.** 1958 Microbial persistence. *Yale J. Biol. Med*. 30, 257–291.

25. **Meyer, C. R., Ghosh, P., Nadler, S. & Preiss, J.** 1993 Cloning, expression, and sequence of an allosteric mutant ADPglucose pyrophosphorylase from *Escherichia coli* B. *Arch. Biochem. Biophys*. 302, 64–71.

26. **Miller, J. H.** 1972 Experiments in molecular genetics. Cold Spring Harbor, NY: Cold Spring Harbor Press.

27. **Nystrom, T.** 2003 Conditional senescence in bacteria: death of the immortals. *Mol. Microbiol.* 48, 17–23

28. **Oliver, J. D.** 2005 The viable but nonculturable state in bacteria. J. Microbiol. 43, 93–100. Sachs, J. L. & Bull, J. J. 2005 Experimental evolution of conflict mediation between genomes. *Proc. Natl Acad. Sci. USA* 102, 390–395.

29. **Tuomanen, E.** 1986 Phenotypic tolerance: the search for beta-lactam antibiotics that kill nongrowing bacteria. *Rev. Infect. Dis*. (Suppl. 3), S279–S291.

30. **Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull, J. J.** 1999 Different trajectories of parallel evolution during viral adaptation. *Science* 285, 422–424.

31. **Wichman, H. A., Scott, L. A., Yarber, C. D. & Bull, J. J.** 2000 Experimental evolution recapitulates natural evolution. *Phil. Trans. R. Soc. B* 355, 1677–1684.

32. **Wichman, H. A., Millstein, J. & Bull, J. J.** 2005 Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* 170, 19–31

33. **Wiuff, C., Zappala, R. M., Regoes, R. R., Garner, K. N., Baquero, F. & Levin, B. R.** 2005 Phenotypic tolerance: antibiotic enrichment of noninherited resistance in bacterial populations. *Antimicrob. Agents Chemother*. 49, 1483–1494.

34. **Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A. & Lenski, R. E.** 2006 Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl Acad. Sci. USA* 103, 9107–9112.

35. **Wu, M. X. & Preiss, J.** 1998 The N-terminal region is important for the allosteric activation and inhibition of the *Escherichia coli* ADP-glucose pyrophosphorylase. *Arch. Biochem. Biophys*. 358, 182–188

36. **Zambrano, M. M. & Kolter, R.** 1996 GASPing for life in stationary phase. *Cell* 86, 181–184.

Figure Legend

Figure 1. Changes in the density of ancestral (Ara^+) and evolved (Ara^-) ABM in mixed and single clone cultures. (*a*) Mixed culture during exponential growth and early stationary phase. Solid diamonds, density of ancestral (Ara^+); solid square, evolved (Ara^-). (*b*) Mixed and single clone cultures of ancestral and evolved cells at stationary phase. Solid diamond and solid squares with solid lines are, respectively, the density of ancestral and evolved cells in mixed cultures. Open diamonds and open squares with broken line are, respectively, the density of ancestral and evolved cells in single clone culture. In this figure and other data presented, the evolved strain used was that isolated as a single colony from the 62nd transfer (approx. 412 generations) of one of the evolution experiments.

Figure 2. Changes in the ratio of evolved and ancestral ABM mixed at different initial frequencies or average densities. (*a*) Frequency dependence: ancestral (Ara^+) and evolved (Ara^-) cells from 24 hour stationary phase cultures were mixed at different ratios of ancestral and evolved cells and put into fresh LB (1/100). (*b*) Density dependence: mixtures of ancestral (Ara^-) and evolved (Ara^+) cells from 14 hour stationary phase were diluted at different initial densities in the sterile filtrate of a 14 hour mixture of ancestral and evolved cells. The average density is that calculated over the entire sampling period.

Figure 3. FACS and colony formation analyses of mixed cultures of evolved and ancestral ABM cells. (*a*) Ancestral cells constitutively expressing DsRed were mixed with evolved cells constitutively expressing GFP at a 10∶1 ancestor-to-evolved ratio and were analysed during 28 hours of mixed growth using flow cytometry. The relative GFP and DsRed fluorescence is shown on the *x*- and *y*-axes. The 'F' and 'A' windows enclose the *f*ree ancestors and *a*ggregated cell populations, respectively, that were gated during the subsequent FACS sorting (see below). The 'A' population contained at least one evolved cell and one or more ancestor cells per particle. (*b*) As in (*a*), except that ancestor cells constitutively expressing GFP were used in place of evolved cells. (*c*) 'A' and 'F' cell populations were isolated from mixed cultures of GFP-labelled evolved and DsRed-labelled ancestor cells at the indicated times of incubation using FACS sorting. The plating efficiency was scored as the number of CFUs per sorted particle for a given population. Free ancestor cells are shown as red bars and evolved-bound ancestor cells are shown as black bars. The right axis shows the frequency of ancestral cells in the mixed cultures during the sampling period. (*d*) As in (*c*), except that GFP-labelled ancestral cells were used in place of GFP-labelled evolved cells. Data are means±s.d., $n \geq 2$ (time 28 hours, $n=3$).

Figure 4. Genetic basis of SCDI in ABM *E. coli*. (*a*) Map of the pSCDI1 plasmid. The genes *glgC*, *glgA* and *glgP* encode the enzymes ADP-glucose pyrophosphorylase, glycogen synthase and glycogen phosphorylase, respectively. *orf* denotes an open reading frame that encodes a putative membrane protein with no significant homology to any known protein (data not shown). *glpD′* is the truncated 3′ portion of the gene for the glycerol-3-phosphate dehydrogenase. *ori2*, *repE* and *incC* are the *cis*-elements that are necessary for the mini-F type plasmid replication. *cat*, chloramphenicol acetyl transferase. The partial sequence below the *glgC* gene shows the base pair substitution leading to the Pro to Ser change in GlgC17. (*b*) Iodine staining of the pBR322-carrying ABM derivatives ancestor, evolved and evolved Δ*glgC17*, and of the ancestor carrying the *glgC17*-bearing multicopy plasmid pMLM141 (ancestor+*glgC17*). (*c*) *glgC17* protects from

SCDI. Plating efficiency of naive cells containing pMLM141 (triangles) or the pBR322 vector (squares) in mixed cultures with evolved cells (solid symbols) or in single cultures (empty symbols). (*d*) *glgC17* is responsible for SCDI. Plating efficiency of pBR322-containing ancestral Ara^+ cells was monitored in single cultures (diamonds) or in mixed cultures with Ara^- ancestral cells containing pBR322 (squares); evolved, evolved Δ*glgC17* cells containing pBR322 (hexagons); Ara^- ancestral cells containing pMLM141 (triangles); evolved cells containing pBR322 (circles).

Figure 5. Evolution of SCDI simulation results: change in the density of the evolved strain. (A) Evolved strain has a 1% lower growth rate than the ancestral strain. (B) Evolved and ancestral strains have the same fitness. (C) The ancestral strain has a 25% higher rate of mortality than the evolved. For more details about this model and the values of the parameters, see the electronic supplementary material.

Figure 1

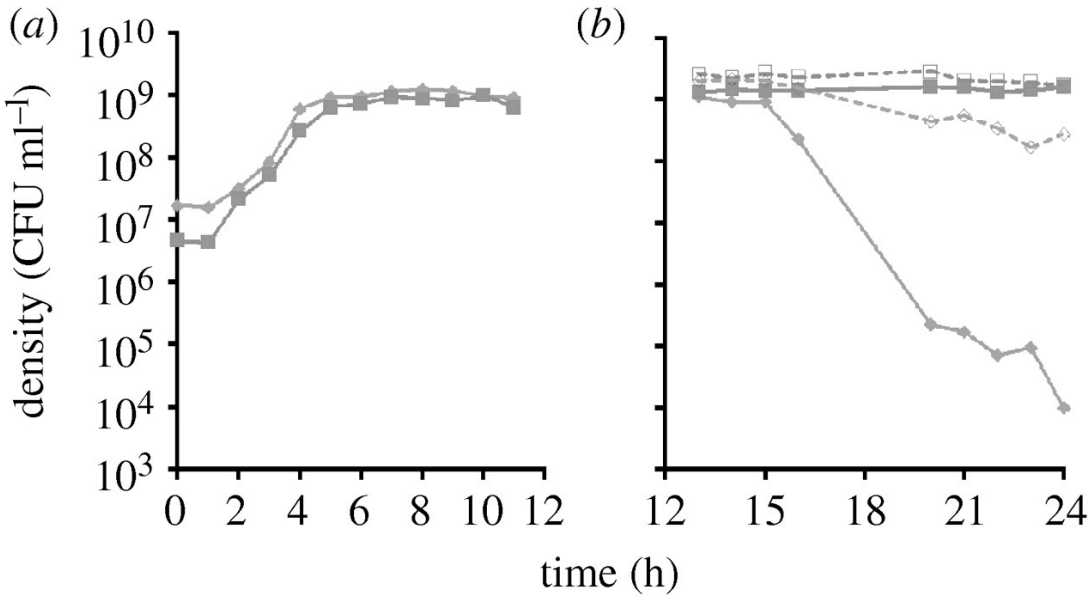

(a)
(b)
density (CFU ml−1)
10^10
10^9
10^8
10^7
10^6
10^5
10^4
10^3
0 2 4 6 8 10 12
12 15 18 21 24
time (h)

Figure 2

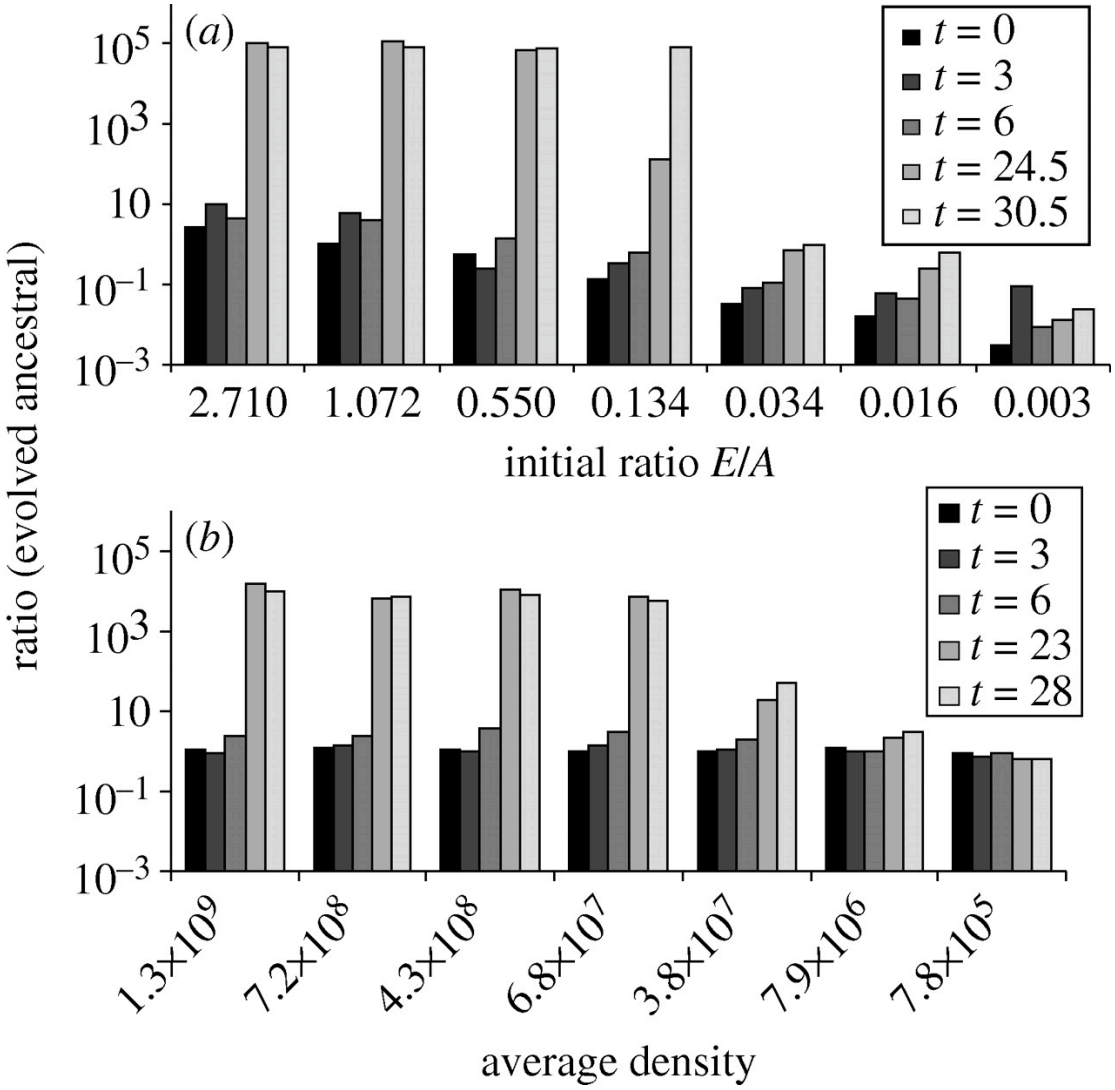

(a)
10^5
10^3
10
10^{-1}
10^{-3}
$t = 0$
$t = 3$
$t = 6$
$t = 24.5$
$t = 30.5$
2.710
1.072
0.550
0.134
0.034
0.016
0.003
initial ratio E/A
ratio (evolved ancestral)
(b)
$t = 0$
$t = 3$
$t = 6$
$t = 23$
$t = 28$
1.3×10^9
7.2×10^8
4.3×10^8
6.8×10^7
3.8×10^7
7.9×10^6
7.8×10^5
average density

Figure 3

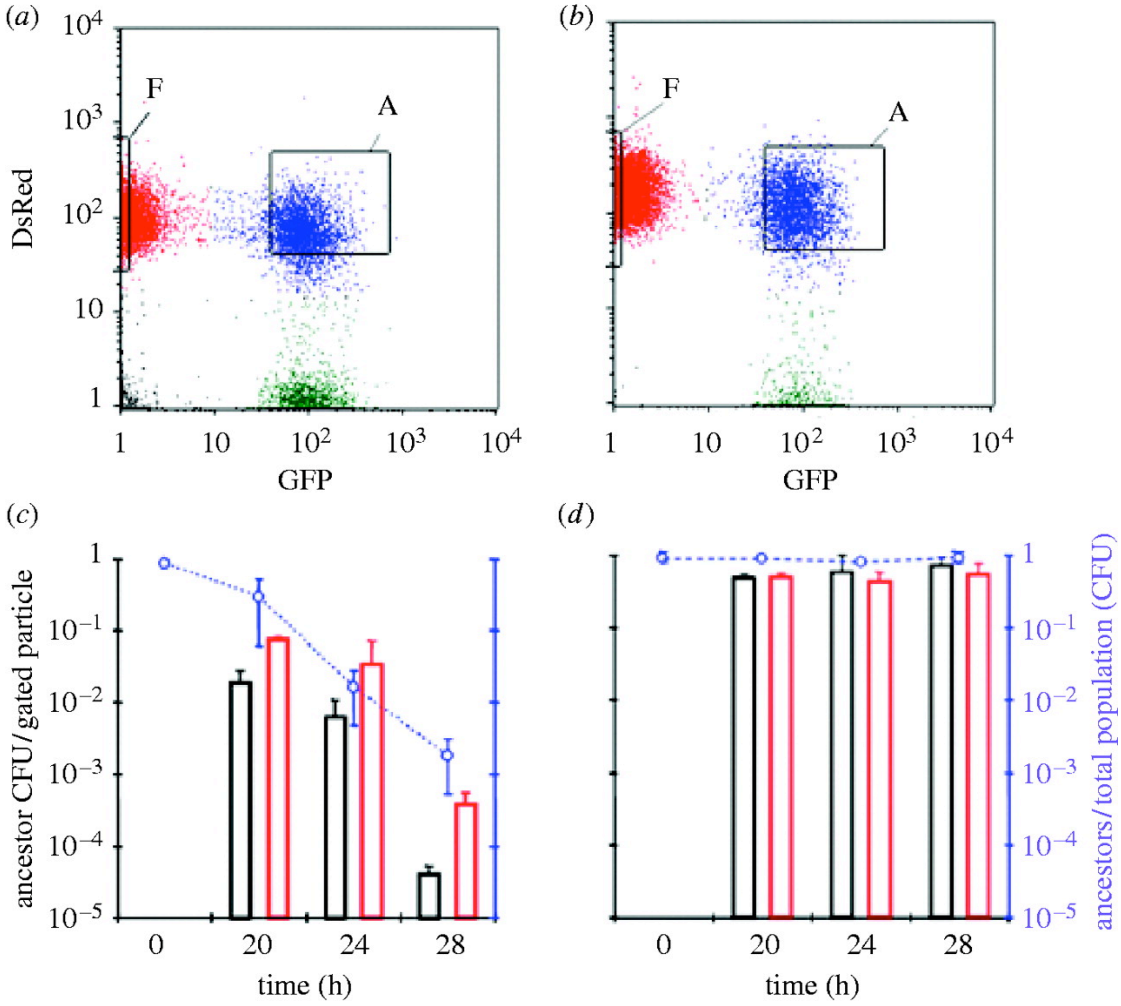

(a)
(b)
(c)
(d)
F
A
DsRed
GFP
ancestor CFU/gated particle
ancestors/total population (CFU)
time (h)
0
20
24
28
1
10
10^2
10^3
10^4
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}

Figure 4

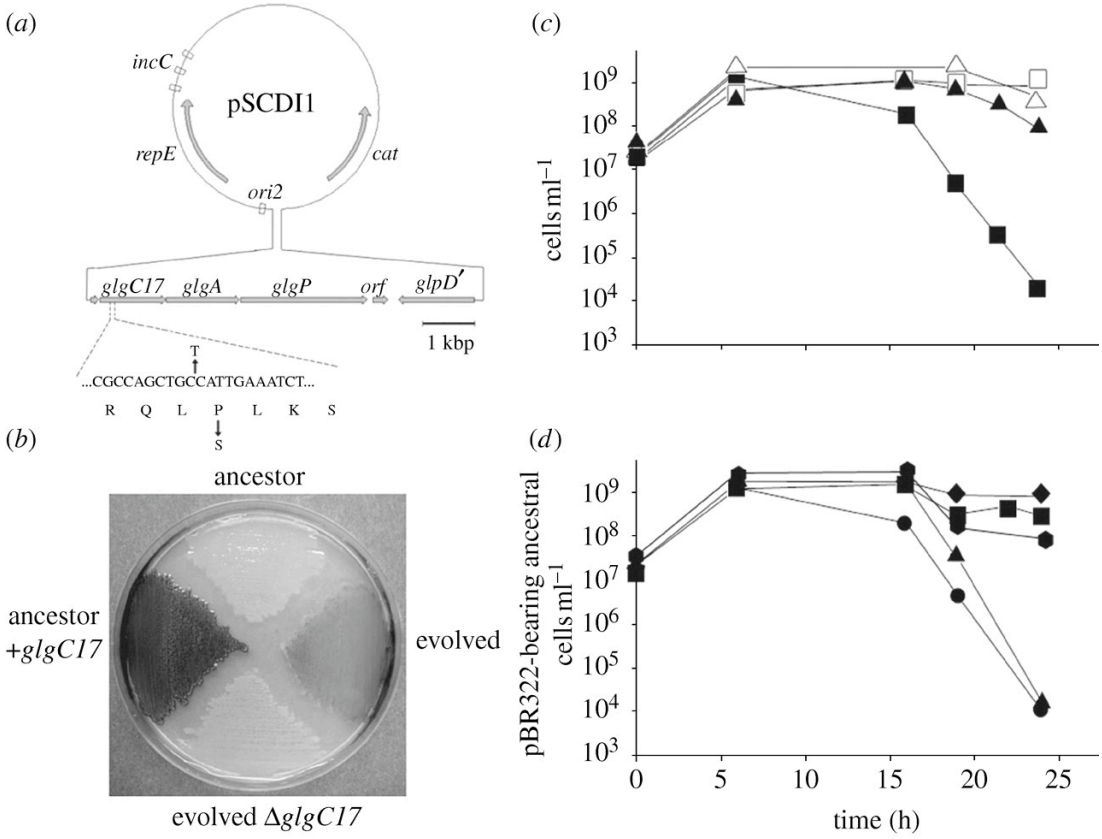
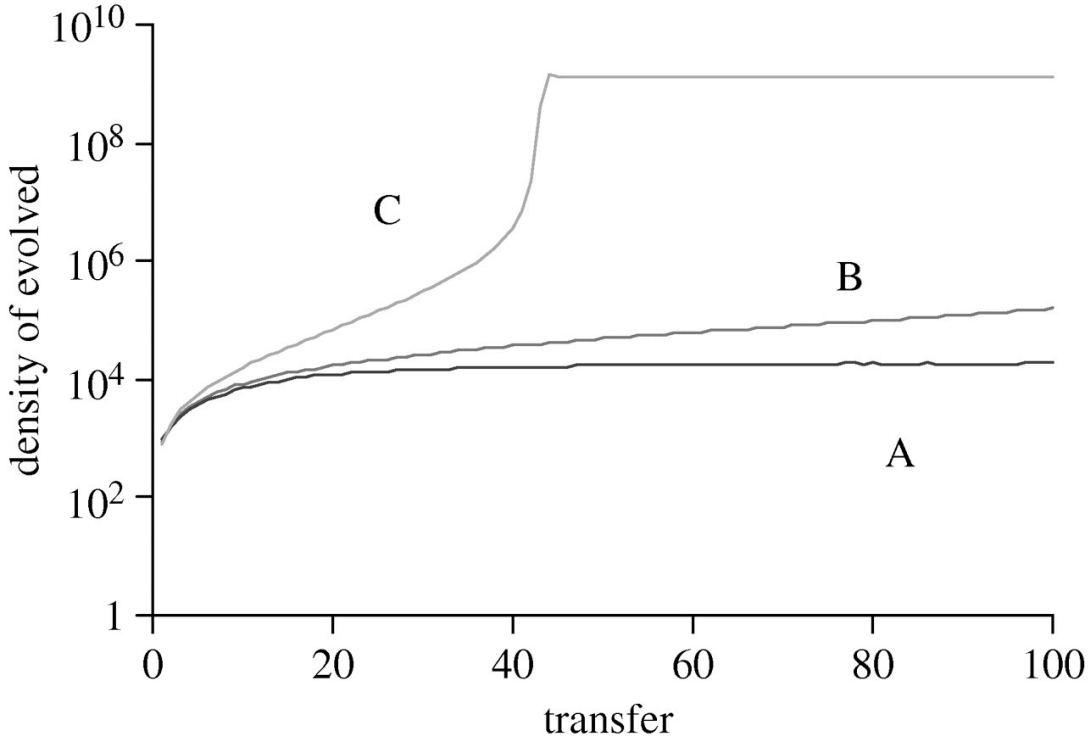

(a)
incC
pSCDI1
repE
cat
ori2
glgC17
glgA
glgP
orf
glpD′
1 kbp
T
...CGCCAGCTGCCATTGAAATCT...
R Q L P L K S
S
(b)
ancestor
ancestor
+glgC17
evolved
evolved ΔglgC17
(c)
cells ml−1
(d)
pBR322-bearing ancestral
cells ml−1
time (h)

Figure 5


10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
1
density of evolved
C
B
A
0
20
40
60
80
100
transfer

Supplemental Information

1- Additional Evidence for contact dependence:

In addition to the FACS experiments considered in our report, two other lines of evidence support the hypothesis that the inhibition of growth of the ancestral ABM strains requires physical contact with viable evolved *glgC* mutants. One is that the rates of decline in the estimated density of ancestral cells are effectively the same when they are in sterile filtrates or sonicated or heat killed suspensions of stationary phase cultures of the evolved cells or their own cells. The results of two of the experiments providing this evidence are presented in Figure S1. In these experiments stationary phase cultures of ancestral (A) or evolved (E) cells were incubated with shaking at 37oC in stationary phase filtrates or suspensions of heat-killed stationary phase cells. The densities of ancestral and evolved cells in these media were estimated from CFU after one and two days of incubation.

The second line of additional evidence for cell-cell contact being required for the observed inhibition comes from experiments with variant of the classic Davis U-tube. In this case the "U-tubes" were pairs of 10 ml plastic syringes connected either directly with silicon tubes or through pairs of 0.45 micron filter (Figure S2). The ancestral and evolved cultures were put into opposite syringes and mixed by forcing the plungers at least seven times during a 24 hour periods. In between mixing they were held vertically, shaken and plugged to allow for air to enter, The initial densities of ancestral (Ara+) and evolved (Ara-) cells in the filter separated

experimental, and control syringes were 9.8x108 and 7.4x108, respectively for an Evolved/Ancestral (E/A) ratio of 0.76. At 24 hours, in the experimental syringes the

mean number of bacteria from two samples taken from the syringes carrying the ancestral and evolved cells were, 4.1 x108 and 5.5 x108 respectively for an evolved to ancestral ratio of 1.3. In the control syringes where the cells were mixed this ratio was 300.6.

2- The SCDI Phenomenon and bacteria

(i) Variability: In the course of this investigation, we performed more than 150 mixed culture experiments similar to those reported in Figure 1 with the ancestral Ara- or Ara+ ABM bacteria and different separately evolved Ara+ and Ara- bacteria at high densities with evolved cells in frequencies in excess of 0.1. Quantitatively SCDI is a somewhat variable process; the extent to which the ancestral cells were inhibited by the evolved varied between experiments with the same pairs of ancestral ABM and evolved as well with Ara+ and Ara- ancestral cells with independently evolved ABM. In growing LB culture initiated with approximately equal frequencies of ancestral and evolved cells, the decline in the density of ancestral population varied from 200 fold to more than 5000 fold. Within a particular experiment, there is relatively little variability in the extent of inhibition of the ancestral strain. This can be seen in Figure S3, where we plot the estimated densities of evolved and ancestral bacteria at 5.5 hours (soon after cells stopped growing) and at 24 hours. In this experiment 100 μl of a mixture or single clones of overnight ancestral and evolved cells were put in eight flasks each containing 10 mls of LB.

(ii) Mortality Rates: The above Figure S3 also illustrates another property of the evolved bacteria. In single clone culture, as determined by CFU data the evolved bacteria die or enter a viable but not culturable state, VBNC, at lower rate than the

ancestral. This can be seen from experiments were single clone cultures were sampled soon after they stopped growing and after they had been at stationary phase for at least 15 hours see Figure S4. While these separate cultures vary in the extent of "mortality" (reduced number of CFU between time of onset of stationary phases), it is clear than on average the evolved cells die at a lower rate than the ancestral.

(iii) SCDI requires high pH: As noted in Figure 1 of our article, the onset of SCDI does not occur until mixed cultures have stopped growing for 12 or so hours. By 24 hours the pH of the unbuffered LB used in these experiments is on the order of 9.0 or higher. To ascertain if this high pH is necessary for SCDI we performed these mixed ancestral evolved ABM experiments in two buffered versions of LB. In these media, the ingredients of Luria-Bertani broth (per liter 10 mgs Tryptone, 5 mgs yeast extract and 5 grams of NaCl) were added to 60 millimolar solutions of a TRIS or BIS-TRIS. (2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol) buffers obtained from SigmaTM. The pH's of these buffered LB solutions were adjusted to 7.0 by the addition of HCl prior to autoclaving. The results obtained in the buffered LBs and their unbuffered control are depicted in Figure S5 . We interpret the absence of SCDI in the pH7.0 buffered medium as consistent with the hypothesis that high pH is a necessary condition for SCDI.

(iv) SCDI does not occur in continuous culture: Although it is clear that SCDI occurs when the cells stop growing, the data presented in our article and above do not rule the possibility that SCDI can also occur when the evolved and ancestral bacteria are together for an extended time at high densities but are still replicating. To test this hypothesis we mixed ancestral and evolved cells from chemostats with LB medium. For more information about the design of this "home made" chemostats see the appendix to (Chao

et al. 1977) and/or write to us for more details about the more modern versions of these continuous culture devices. To initiate the chemostats overnight cultures of an ancestral, ABM Ara+ and an evolved (62-2) Ara- were put into separate vessels and grown for 72 hours at dilution rates of ~ 0.55 per hour. The vessels were emptied and the evolved and ancestral cells were mixed and 10 mls of this mixture was returned to the vessels of their origin, chemostats 1 and 3. In chemostat 200ul of this mixture was added to LB and allowed to reach full density in the chemostat. As controls we used two batch cultures, one with just the same mixture used in the chemostats without dilution, (stationary), and one where 100 ul of this mixture of evolved and ancestral cells was added to 10 ml LB (exponential). The results of this chemostat experiment are presented in Figure S6.
We interpret the results of these experiments as support for the hypothesis that SCDI requires the bacteria to be non-growing and presumably in stationary phase. While SCDI occurred in the control stationary phase and initially exponential batch cultures, it did not occur in the chemostat. If anything, in these chemostats the evolved cells had a selective disadvantage relative to the ancestral. There is, however, a caveat to this interpretation, we cannot rule out a pH effect. Although the LB in the above chemostat was not buffered the pH remained between 8.0 and 8.5 which, about 1 pH unit lower than that in the batch control cultures at 24 hour.

3 - The genetic basis of SCDI and immunity to SCDI in AB1157Δ*mutS::spc* (ABM)

Plasmid constructions

pMLM02 is a partition-deficient mini-F carrying a chloramphenicol resistance gene (*cat*) that was derived from pMLM1 (Lemonnier et al. 2000) by deleting an *Eco47*III- *Hpa*I fragment (3063 bps) containing the *sopABC* operon. pCDI1, that was isolated in this work, derives from the insertion of a 7472 bps fragment (containing the *glgC17*, *glgA*, and *glgP* genes)

generated by partial *Sau3A* digestion of the evolved chromosome (see below) into the *BamH*I site of pMLM02. pMLM141, was obtained by inserting a 2158 bps *glgC17*-containing fragment generated by cleavage of pSCDI1 with *Xba*I and *Sph*I into a vector obtained by cleaving pBR322 (Balbas et al. 1988) with *NheI* and *Sph*I. pMLM143 (*cat*, *rpsl*+), the integration plasmid designed to knock-out *glgC* in the chromosome, was constructed as follows. A 3496 bps fragment containing the *glgC17* gene and its flanking regions was obtained by PCR amplification using the evolved ABM chromosome as DNA template and the pair of primers *glgC5*1 (5'-CAGCGACCAGGCATTACTATC-3') and glgC31 (5'- GTGCTCTCGCAGGTGAAGTTA-3'). The PCR product was digested with *Dra*I and *Pst*I, followed by end-filling with Klenow fragment of DNA polymerase, prior to being inserted into the *Bgl*II (end-repaired with Klenow) and *Nsi*I sites of the integration excision vector pLN135 (Cornet et al. 1996) to give pMLM142. The bulk of the *glgC17* gene was then removed by cutting pMLM142 with *Asc*I and *Xmn*I (Figure S6, A) followed by end-repair using Klenow polymerase and self- ligation of the vector fragment, to give pMLM143.

Construction of the Δ*glgC* strain MLM385. The multi-step procedure to substitute the Δ*glgC17* mutation present in pMLM143 for *glgC17* in the chromosome was performed essentially as described (Lemonnier et al. 2000). Briefly, pMLM143 (*cat*, *rpsl*+) was used to transform the streptomycin-resistant (SmR) evolved strain. Integration of pMLM143 into the chromosome was selected by plating cells on chloramphenicol-containing medium at 42°C. Excision of *glgC17* from the chromosome was selected by plating on medium containing streptomycin (the excision event being tightly linked to the loss of the pMLM143-borne *rpsL*+ allele that confers sensitivity to streptomycin). The deletion of *glgC17* in the chromosome was verified by PCR using the oligonucleotides glgC51 and glgC31 and by monitoring glycogen accumulation in vivo (Figure 3 in the article) using

iodine vapor staining (see below). The Δ*glgC17* strain was named MLM385.

Construction of the *E. coli* genomic library and isolation of the *SCDIR* clones. Genomic DNA was prepared from the evolved ABM strain. After partial digestion by the restriction enzyme *Sau3A*, DNA fragments between 6 and 8 kbp were isolated by agarose gel electrophoresis and ligated into the *BamH*I site of pMLM02. The random shotgun generated library was electroporated into the ancestral ABM strain using a Bio-Rad MicroPulser™. The electro-transformants were selected at 37ºC on LB-agar plates containing chloramphenicol (20 µg.ml-1). Approximately 4000 transformants were pooled, washed several times in fresh LB medium, diluted to ≈ 2.5 x 107 cells/ml and mixed with an equal density of evolved ABM cells. 10 ml of fresh LB medium were inoculated with this mixture and were cultured at 37ºC with aeration for 24 hours. Appropriate dilutions were then plated on LB-agar plates containing chloramphenicol. Approximately 4000 colonies were pooled and were submitted to a new round of competition with evolved cells as described above. The chloramphenicol-resistant colonies obtained after these two rounds were individually tested for resistance to SCDI in pairwise competition experiments with evolved ABM. Clones that showed significant resistance to SCDI were kept for subsequent analyses.

Assessment of glycogen accumulation by iodine staining. Staining with iodine vapor (Govons et al. 1969) was used to examine glycogen accumulation in colonies grown for 18 hours at 37ºC in Kornberg medium (1.1 % K2HP04, 0.85 % KH2P04, 0.6 % yeast extract, 1 % glucose).

Convergent evolution and variation in *glgC* in ABM. As noted in our article, SCDI evolved in eight separate serial transfer cultures. Clones of the eight evolved strains

isolated at the 62nd transfer were refractory to SCDI by the other evolved strains and all eight over produced glycogen as measured by the above iodine staining protocol. DNA sequence data of these eight evolved clones indicate that all have missense, base substitutions in *glgC* but at seven different sites. Only two mutations were identical, the proline to serine substitution at the 17th codon (Table 1). In other words, the evolution of SCDI in ABM was convergent in the sense that the same phenotype evolved through different mutations.

4- Glycogen production is not needed for immunity to SCDI but is needed for Inhibition

In an effort to ascertain the role of glycogen production in this contact-dependent inhibition using AB1157 *mutS*+ we constructed three strains with different combinations of *glg* alleles and tested their sensitivity to SCDI by evolved 62-2 and ability to inhibit the ancestral ABM in mixed culture experiments. The strains constructed for these tests were:

An – AB1157 *glgC*17 (a replacement of the *glgC* gene with the evolved *glgC17* from 62-2, ABM *glgC*17)

glgA - AB1157 *glgC*17 *glgA* (**An** with a knock out of *glgA*, glycogen synthase)

glgP – AB1157 *glgC*17Δ *glgP* (**An** with a knock out of *glgP*, glycogen phosphorylase)

The bacteria and methods employed for these constructions are described below.

Procedures for constructing An and the *glgA* and *glgP* knockouts.

Bacteria, bacteriophage, and growth conditions. Bacterial strains and bacteriophage are described in **Table 2**. Unless noted, bacterial strains were grown at 37 oC in Luria-

Bertani (LB) medium (Miller 1972) or Kornberg medium (KB) (1.1 % K2HP04, 0.85 % KH2P04, 0.6 % yeast extract, 1 % glucose). Media were solidified using 1.5% agar and were supplemented with antibiotics, as needed, at the following concentrations: ampicillin 100 µg ml-1; chloramphenicol 25 µg ml-1; and tetracycline 10 µg ml-1.

Strain construction. Strains AB18450C17GA1 (Δ*glgA*) and AB18450C17GP1 (Δ*glgP)* were constructed as follows. First, P1*vir* was used to transduce the tetracycline resistance (*tet*) gene from mapping strain CAG18450 (*zhf*-50::Tn*10*) (Singer et al. 1989)to strain AB62-2MS carrying the *glgC*17 allele. A transductant exhibiting both tetracycline resistance (TetR) and high glycogen accumulation, as determined by dark colony staining upon exposure to iodine vapor when grown overnight on KB agar plates, was isolated and designated AB62-2MS18450. The *glgC*17 allele and *zhf*-50::Tn*10* were then cotransduced to a *mutS* wild type strain of AB1157 to create strain AB18450-C17-1, by selecting for TetR and screening for high glycogen accumulation. The *glgA* and *glgP* genes of AB18450C17-1 and the *glgA* gene of MG1655 were replaced with a chloramphenicol (*cat*) expressing cassette by λ-Red mediated gene replacement (Datsenko & Wanner 2000) to create strains AB18450C17GA1, AB18450C17GP1 and MG1655GA12, respectively. The *cat* cassettes used to replace *glgA* and *glgP* were PCRamplified from the template plasmid pKD3 using primer sets (for *glgA*) AH1P1 (5'-aatgctacggaagttagggcataaacaggagcgataagtgtaggctggagctgcttc), AH2P2 (5'- 13 gctaagcgtgggcgatgaatatgtaaacggagcattcatatgaatatcctccttagt) and (for *glgP*) PH1P1 (5'-gcgaagtcgtaccgtgagctttactatcgcttgaaataggtgtaggctggagctgcttc), PH2P2 (5'-gatcgatatgccagatatgatcggcgtactctttgatagtcatatgaatatcctccttagt). Conditions for amplification of the *cat* cassette from pKD3 were as follows: 1 cycle at 94oC for 1 min.; 30 cycles of 94oC for 30 seconds, 65oC for 30 sec., 68oC for 90 sec.; 1 cycle at 68oC for 10 min. The PCR product containing the *cat* gene with terminal *glgA* or *glgP* homology

regions, was treated with DpnI, gel purified and introduced by electroporation into MG1655[pKD46] and AB18450C17-1[pKD46], which had been grown in LB containing arabinose (1 mM). Transformants were selected for chloramphenicol resistance (CamR) and screened for loss of glycogen accumulation. The targeted gene replacements were confirmed by PCR using the primers (for *glgA*) Aconfirm1 (5'-tagccacgggatgacccttaactc), Aconfirm2 (5'-tgggcacggtttgaacgtaacc) and (for *glgP*) Pconfirm1 (5'-ttcactgtggcggtttgtgc), Pconfirm2 (5'- ggatgccgaaaaaagtcattac). Confirmatory PCR conditions were as follows: 1 cycle at 94 oC for 1 min.; 30 cycles of 94 oC for 1 min., 53 oC for 1 min., 68 oC for 3 min.; 1 cycle at 68 oC for 10 min. A non-polar deletion of the *cat* cassette from strain MG1655-GA12 was constructed using FLP-mediated recombination at flanking FRT sites (Datsenko & Wanner 2000). This deletion was confirmed by PCR as described above. We thank Xin Wang who created MG1655-GA12 and designed the primers for amplification and confirmation of the pKD3 *cat* cassette replacement of *glgA* and *glgP*.

Mixed culture SCDI experiments with these knockouts

The above described An, *glgA* and *glgP* Ara- strains were mixed with evolved ABM Ara+ (62-2) and with the ancestral ABM Ara+ and put into LB 100 ul to 10 ml in 50ml flasks. The densities of these competing strains were estimated at 5 hours (when cell growth ceased) and 24 hours. The results of this experiment are presented in Figure S10. The glgA construct (AB1157 *glgA glgC*17) did not produce glycogen and was not inhibited by the evolved ABM *glgC*17. On the other hand when this strain was mixed with the ancestral ABM it did not cause SCDI. We interpret these results to indicate that glycogen production is not essential for the immunity of AB1157 to SCDI by the evolved *glgC* strains but is required for the inhibition of ancestral cells. On the other hand, the glycogen producing AB1157 *glgC*17 and AB1157 *glgC*17 *glgP* constructs did

inhibit the ancestral strain. While this suggests that glycogen catabolism is not required for SCDI in ABM cells, other minor enzymatic systems can also support glycogen turnover which may be sufficient to promote inhibition.

5-Cell viability analysis during SCDI using confocal microscopy

The SCDI phenomenon described here is based on the results of plating experiments the relative abilities of the mixed strains to form colonies on agar. Not clear from these result is whether the failure of the ancestral cells to form colonies in mixed culture with the evolved cells is due to contact-mediated killing or contact with these inhibiting bacteria inducing a viable but non-culturable (VBNC) state. To begin to address this question we performed a confocal microscopy analysis of mixed cultures of evolved and ancestral ABM cells after 28 hours of co-incubation at 37°C. As can be seen in Figure S9 panel a, ancestral cells expressing the DsRed protein were present at similar proportions and displayed similar fluorescence intensities irrespective of whether they were in mixed cultures with GFP-labelled evolved cells or mixed with GFP-labelled ancestral cells. In the former case the plating efficiency (CFU) of ancestor cells was less than 10-3 that of the evolved, as expected (data not shown). These observations were fully consistent with the results from the FACS experiment where no significant variation in the fluorescence of ancestral cells expressing the DsRed protein was observed, despite a drop of several orders of magnitude in their plating efficiency when mixed with evolved cells during 28 hours (Figure 3). These experiments suggested that the ancestral cells remained viable during SCDI in stationary phase cultures. To further address this hypothesis we performed a propidium iodide (PI) inclusion assay, which is used as a standard criterion to distinguish dead from viable bacterial cells. Bacteria that undergo membrane disruption and cell death appear as red fluorescing cells following PI inclusion (Boulos et al. 1999). Our results

indicated that death was not occurring in bacterial cultures experiencing SCDI. As shown in Figure S11, panel b, no PI inclusion could be detected in cells from either the ancestors/evolved or the ancestors(GFP)/ancestors mixed cultures. At the same time the cells were visualized by staining with 4 ,6-diamidino-2-phenylindole (DAPI), which yields blue fluorescing cells upon binding to DNA. In addition, green fluorescence was clearly detected in GFP-producing ancestral cells. When the mixed cultures were exposed to polymyxin B (an antimicrobial compound that binds to cell membranes and leads to rapid cell disruption (Lehtinen et al. 2006)) red fluorescence due to IP inclusion was clearly observed. Altogether, our results support the hypothesis that ancestral cells enter a viable but not culturable (VBNC) state upon contact-dependent inhibition in stationary-phase cultures. However, at this stage we cannot predict whether this VBNC state remains unaltered or rather evolves to rapid cell death upon dilution of the cultures in fresh medium and plating on agar. Indeed, this is a matter of ongoing debate in the field of VBNC research (Nystrom 2003; Oliver 2005)). In this respect, SCDI in stationary phase has inherent properties (like culturable (evolved) and non-culturable (ancestors) cells co-existing and being easily sorted and purified;

good knowledge of the parameters that control the phenomenon (cell frequency and density, pH) that we anticipate will make it a prominent model for VBNC studies. Particularly, it will be interesting to determine whether cell deterioration via protein carbonylation is specifically taking place in the ancestral cells during SCDI, as reported in the case of non-culturable cells in starving *E. coli* populations (Desnues et al. 2003).

Experimental procedures for the Confocal Microscopy

Confocal laser scanning microscopy was performed using a Zeiss 510 META Confocal instrument. Excitation wavelengths used to elicit red, blue or green fluorescence were 543, 405 and 488 nm, respectively. DsRed- and GFP-labelled *E. coli* cells were prepared as described in the general experimental procedures. For bacterial cell viability analyses, aliquots (500 µl) of 28

hours mixed cultures of evolved and ancestral ABM were centrifuged and the pellets were resuspended in PBS (10 ml). 1 ml aliquots of this resuspension were treated with propidium iodide (PI, 2.5 µg. ml-1) and DAPI (4 ,6-diamidino-2-phenylindole, 2 µg. ml-1). The samples were kept for 10 min at room temperature in the dark. The stained suspensions were then centrifuged and the pellets were resuspended in Moviol (10 µl) and mounted in coverslips. The samples were observed and photographed using confocal microscopy. Exposure of the bacterial cultures to polymyxin B was performed by adding 1.5 ml fresh LB supplemented with polymyxin B (5 µg. ml-1) to aliquots (500 µl) of the same 28 hours mixed cultures used above. The samples were incubated at 37°C for 15 min and then were centrifuged and resuspended in PBS (10 ml) prior to being treated with IP and DAPI as above.

SCDI is not due to plating efficiency differences:

If, as the preceding results suggest, SCDI does not operate through killing but rather via the induction of a viable but not culturable state; it is conceivable that this VBNC state is only manifest by the inability of these bacteria to form colonies on agar. Were they to remain in liquid LB, they would grow again. To ascertain whether this is the case, we did serial passage experiments with AB1157 and MG1655. These strains, rather than the evolved and ancestral ABM were used for this experiment for two reasons. First AB1157 had a marker *rpsL* (Str-r) that enable us, by selective plating, to measure its frequency when rare in cultures with MG1655 (Str-s), which was not the case for the ABM evolved and ancestral strains. Second, because of the high mutation rate in ABM rare mutants for a single allele are commonly produced by mutation from the dominant strains. In these experiment 100 µl of a mixture of MG1655 and AB1157 were introduced into 10 ml LB in 50 ml Erlenmeyer flasks. Samples were taken each day and 100 µl of the mixture was transferred to a fresh flask with 10 ml of LB. The densities of these two cell lines were estimated on Tetrazolium arabinose agar and/or streptomycin LB agar to detect AB1157

(*rpsL*) when it was rare. The results of one of these experiments are presented in Fig.S12. The "Expected line is that which would obtain if AB1157 declined by a factor of 100 (the dilution factor) at each transfer. If the inhibition was due to contact with MG1655 solely reducing the ability of AB1157 to form colonies, the relative frequency of AB1157 would not continue to decline in successive passages in liquid culture as it does.

6- Evidence for SCDI with other strains of *E. coli*

***E. coli K-12* MG1655, C600 and wild-type *E. coli* B as inhibitors of AB1157Δ*mutS::spc nal* (ABM) *and AB1157mutS+* :** Overnight Ara+ cultures of *E. coli* K-12 MG1655 (MGA+), C600 Ara+ (C600A+) and a wild-type strain of *E. coli* B (BWTA+), were mixed with Ara- AB1157*mutS*+ (ABA-), an Ara- ancestral ABM (AA-) or an evolved variant of this strain, EA-. For each mixed pair of overnight cultures 100 µl put into 10 ml LB and were incubated with shaking at 37o C. Samples were taken at 5 hours, when the cultures were no longer growing, and again at 24 hours and the estimated densities of Ara+ and Ara- cells estimated from CFU data. The results of one of these experiments are depicted in Figure S11. We interpret these results as evidence that when cultured with either *E.* coli K-12 MG1655, C600 or wild-type *E. coli* B, *the E. coli* K-12 strains AB1157*mutS*+ or ancestral ABM employed in the evolution experiments are subject to inhibition at stationary phase. These results also indicate that the evolved *glgC17* mutant of ABM is immune to inhibition by either *E. coli* K-12 MG1655 or *E. coli* B. Since there was no evidence for the stationary phase AB1157 dying at higher rates in sterile filtrates of their own cultures than in filtrates of stationary phase cultures of MG1655 or *E. coli* B we conclude that this inhibition requires cell-cell contact, i.e. is SCDI (data not shown).

MC1061 is subject to SCDI: Overnight LB cultures of *E. coli* K-12 strain MC1061 (Ara-) (MCA-) were mixed with MG1655 (MGA+) or the evolved ABM Ara+ (EA+) and 100µl of the mixtures were put into 10 ml LB. Samples were taken at 5 hours, when the cultures were no longer growing, and again at 24 hours. The results of one of these experiments is depicted in Figure S12 We interpret the results of this experiment as support for the hypothesis that *E. coli* K-12 strain MC1061 is sensitive to inhibition at stationary phase when mixed with either *E. coli* MG1655 or the evolved ABM. The extent of the inhibition is somewhat less when it is mixed with the evolved ABM strain than with MG1655. Filtrate experiments similar to those considered above, support the interpretation that this inhibition of MC1061 requires physical contact with these other bacteria, i.e. is contact dependent, SCDI (data not shown).

7 - Population and Evolutionary Dynamics of SCDI: Theoretical Considerations:

(i) The frequency- and density- dependence of SCDI is anticipated from a simple mathematical model. To model contact-dependent inhibition we consider a non-growing population of evolved and ancestral cells with densities E and A bacteria per ml, respectively. We assume: (i) the E and A bacteria and die at rates de and da h-1 respectively due to processes that have nothing to do with inhibition by a competing population, (ii) bacteria in these cultures randomly collide with each other at rates proportional to the product of their densities; a fraction of contacts with the evolved strain, the parameter τ, prevents the ancestral strain from replicating (kills them or induces a VBNC state). It should be noted, that this model of SCDI does not assume that the inhibition occurs through a single contact by pairs of individual cells or that cells remain in permanent contact, it is a mass process that may require many contacts by

groups of cells. It also does not assume that the inhibited cells are dead, but rather that they no longer replicate or can be detected by forming colonies, i.e. they can be viable but not culturable, VBNC. With these definitions and assumptions, the rates of change in the densities of the inhibited (dying) populations of ancestral and evolved cells are given by,

$dA/dt = - da\ A - \tau\ A\ E$

$dE/dt = -de\ E$

To illustrate the properties of this model, we use a numerical solution to these equations (a computer simulation). This and the other simulations used here were programmed in Berkeley MadonnaTM. Copies of this program, which is the same program used for the evolution simulations below can be downloaded from www.eclf.net. As can be seen in the below simulation (Figure S13), when the total density of bacteria (A+E) or the frequency of the inhibitor strain (E) are low SCDI has little effect on the number of ancestral cells recovered. It should be noted, that frequency- and densitydependence

would also be obtained if the inhibition of the ancestral cells was mediated by an extracellular toxin like a bacteriocin (Levin 1988).

(ii) Mathematical model and computer simulation of the evolution of SCDI

In this model we assume that within a transfer there are two phases. In Phase 1 there is no contact-dependent inhibition of the ancestral cells. The rates of change in the densities of the ancestral and evolved populations, A and E, respectively, are solely a function of their resource concentration dependent growth rates and a constant death rate. For the former we use a Monod function (Monod 1949), $va\Psi(R)$ and $ve\Psi(R)$, where $\Psi(R)= R/(R+k)$, where R is the concentration of the limiting resource, va and ve are the maximum hourly growth rates of the ancestral and evolved cells respectively and k μg is the concentration of the resource where the growth rate is half of its maximum value, the

Monod constant. As in (Stewart & Levin 1973) we assume, the resource is taken up at rates proportional to the densities of the bacterial populations, their replication rates (the Monod functions) and a parameter, e, the conversion efficiency; the production of a new cell requires e μg of resource. With these definitions and assumption, the rates of change in resource concentration and density of bacteria during Phase 1 is

$dR/dt = - \Psi(R) [A*va +E*ve]e$

$dA/dt = A*va*\Psi(R) – da*A$

$dE/dt = E*ve* \Psi(R)- de*E$

where da and de are, respectively, the death rates for the ancestral and evolved bacteria. When the time within a transfer exceeds a defined period, 16 hours in our simulations, Phase 2 commences. Phase 2 is the situation described above the populations are no longer growing but continue to "die" at the density independent rates, da and de, and the ancestral cells die (or enter a VBNC state) at a rate proportional to the product of the densities of the ancestral and evolved cells and an SCDI rate constant, τ ml x cell per hour. With these assumptions and definitions the rates of change in the density of the bacteria in Phase 2 are given by,

$dA/dt = - da *A - \tau* A*E$

$dE/dt = -de* E$

In these simulations the population is maintained by serial passage with a transfer every 24 hours. At each transfer, cells surviving the previous cycle are diluted by a factor of 0.01, and R is set equal to 1000 and the growth cycle starts again. In the runs made in Figure 5, the population was initially composed of 107 ancestral cells and no evolved bacteria. The evolved bacteria arose by recurrent mutation at rate μ per cell per generation from the ancestral population which we simulated by a Monte Carlo process. At each finite time interval Δt a pseudo random number x ($0 \leq x \leq 1$) from a rectangular distribution. When $x < N* \Delta t* \Psi(R)* \mu$, a single mutant cell was added to the E

population. We did not consider recurrent mutation to the A state from E. The parameter values used for the runs made in Figure 5 in our article are presented in Table 2.

References

1. **Balbas, P., Soberon, X., Bolivar, F. & Rodriguez, R. L**. 1988 The plasmid, pBR322. *Biotechnology* 10, 5-41.
2. **Boulos, L., Prevost, M., Barbeau, B., Coallier, J. & Des*jardins, R*.** *19*99 LIVE/DEAD BacLight : application of a new rapid staining method fo*r direct enume*ration of viable and total bacteria in drinking water. *J Microbiol Methods* 37, 77-86.
3. **Chao, L., Levin, B. R. & Stewart, F. M**. 1977 A com*plex community in a* simple habitat: an experimental study with bacteria and phage. *Ecology* 58, 369-378.
4. **Cornet, F., Louarn, J., Patte, J. & Louarn, J. M**. 1996 Restriction of the activity of the recombination site dif to a small zone of the *Escherichia coli* chromosome. *Genes Dev* 10, 1152-61.
5. **Datsenko, K. A. & Wanner, B. L.** *2000 One-step in*activation of *chromosom*al genes in *Escherichia coli* K-12 using P*CR products. Proc Natl Acad Sci U S A 9*7, 6640-5.
6. **Desnues, B., Cuny, C., Gregori, G., Dukan, S., Aguilaniu, H. & Nystrom, T.** 2003 Differential oxidative damage and expression of stress defence regulons in culturable and non-culturable *Escherichia coli* cells. *EMBO Rep* 4, 400-4.
7. **Govons, S., Vinopal, R., Ingraham, J. & Preiss, J.** 1969 Isolation of mutants of *Escherichia coli* B altered in their ability to synthesize glycogen. *J Bacteriol* 97, 970-2.
8. **Lehtinen, J., Jarvinen, S., Virta, M. & Lilius, E. M.** 2006 Real-time monitoring of antimicrobial activity with the multiparameter microplate assay. *J Microbiol Methods* 66, 381-9.

9. **Lemonnier, M., Bouet, J. Y., Libante, V. & Lane, D.** 2000 Disruption of the F plasmid partition complex in vivo by partition protein SopA. *Mol Microbiol* 38, 493-505.
10. **Levin, B. R.** 1988 Frequency-dependent selection in bacterial populations. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 319, 459-72.
11. **Miller, J. H.** 1972 *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Press.
12. **Monod, J.** 1949 The growth of bacterial cultures. *Annual Review of Microbiology* 3, 371-394.
13. **Nystrom, T.** 2003 Conditional senescence in bacteria: death of the immortals. *Mol Microbiol* 48, 17-23.
14. **Oliver, J. D.** 2005 The viable but nonculturable state in bacteria. *J Microbiol* 43 Spec No, 93-100.
15. **Singer, M., Baker, T. A., Schnitzler, G., Deischel, S. M., Goel, M., Dove, W., Jaacks, K. J., Grossman, A. D., Erickson, J. W. & Gross, C. A.** 1989 A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. *Microbiol Rev* 53, 1-24.
16. **Stewart, F. M. & Levin, B. R.** 1973 Resource partitioning and the outcome of interspecific competition: a model and some general considerations. *American Naturalist* 107, 171-198.

Table 1. *glgC* sequence variation in the evolved ABM strains from the 62nd transfer

Evolved strain	Codon change	Site and amino-acid change
62.1	CGC /CAC	R14H
62.2	CCA/TCA	P17S
62.3	GGC/GAC	G336D
62.4	ATC/TTC	I66F
62.5	GAC/TAC	D125Y
62.6	CTT/ATT	L330I
62.7	CCA/TCA	P17S
62.8	TTC/CTC	F318L

Table 2. Bacterial strains, plasmids, phage and growth conditions.

Strain, plasmid or phage	Description or genotype	Source or reference
E. coli K12 Strains		
MG1655	$F^- \lambda^-$	Michael Cashel
CAG18450	MG1655 *zhf*-50::Tn*10*	Singer, et.al., 1989(Singer et al. 1989)
ABM62-2MS	AB1157 *glgC*17 *mutS*	This study
MG1655-GA12	MG1655 Δ*glgA*	X. Wang, unpublished
ABM18450	AB1157 *zhf*-50::Tn*10*	This study
AB1157	Standard Lab strain	E. A. Adelberg (CGSC)
ABM62-2MS18450	AB1157 *glgC*17 *mutS zhf*-50::Tn*10*	This study
AB18450C17-1 Designated **An**	AB1157 *glgC*17 *zhf*-50::Tn*10*	This study
AB18450C17GA1 Designated **glgA**	AB1157 *glgC*17 *zhf*-50::Tn*10* Δ*glgA::cat*	This study
AB18450C17GP1 Designated **glgP**	AB1157 *glgC*17 *zhf*-50::Tn*10* Δ*glgP::cat*	This study
Phage		
P1*vir*	Strictly lytic P1	Carol Gross
Plasmids		
pKD46	For arabinose induction of λ Red system	Datsenko and Wanner, 2000(Datsenko & Wanner 2000)
pKD3	Contains the *cat* gene	Datsenko and Wanner, 2000(Datsenko & Wanner 2000)

Table 3. Parameter values used for the simulations presented in Figure 5*

Parameter	Definition	A	B	C
va	Maximum growth rate A	1.0	1.0	1.0
ve	Maximum growth rate E	**0.99**	1.0	1.0
e	Conversion efficiency	$5x10^{-7}$	$5x10^{-7}$	$5x10^{-7}$
k	Monod Constant	0.25	0.25	0.25
da	Death rate A	0.02	0.02	**0.025**
de	Death rate E	0.02	0.02	0.02
τ	SCDI killing constant	10^{-11}	10^{-11}	10^{-11}

*** These parameter values were chosen to illustrate the process rather than estimated experimentally, although the population growth and resource uptake and use parameters are in a realistic range.**

Figure Legend

Figure S1. Survival of stationary phase ancestral and evolved cells in sterile filtrates or cultures of heat killed stationary phase ancestral or evolved cells. A-A ancestral in filtrates or suspensions of dead ancestral cells; A-E ancestral cells in filtrates or suspensions of dead evolved cells; E-A evolved cells in filtrates or suspensions of dead ancestral cells; E-E evolved cells in filtrates of suspensions of heat-killed evolved! cells. (A) Sterile filtrates, (B) Suspensions of heat-killed (CFU <104 viable) cells.

Figure S2: Make-shift “U-tube” used to test for contact-dependence The initial densities of ancestral (Ara+) and evolved (Ara-) cells in the filter separated experimental, and control syringes were 9.8x108 and 7.4x108, respectively for an Evolved/Ancestral (E/A) ratio of 0.76. At 24 hours, in the experimental syringes the mean number of bacteria from two samples taken from the syringes carrying the ancestral and evolved cells were, 4.1 x108 and 5.5 x108 respectively for an evolved to ancestral ratio of 1.3. In the control syringes where the cells were mixed this ratio was 300.6.

Figure S3. Change in the density of ancestral and evolved bacteria in mixed and single clone cultures between 5.5 hours (blue) and 24 hours (maroon). ME and MA are, respectively, the estimated densities of the evolved and ancestral cells in mixed culture. SE and SA are respectively the estimated densities of the evolved and ancestral cells in single clone culture. Means and standard errors of eight cultures of each type.

Figure S4. Ratio of later to early stationary phase densities of ancestral and evolved ABM(62-2) in single clone culture; AM and AP, ancestral Ara- and Ara+; EM and EP, evolved Ara- and Ara+. The numbers designate different cultures with different sampling times 1 and 2: 7 and 25 hours; 3: 13 and 23 hours; 4: 5.5 and 24 hours.

Figure S5. Ratio of evolved to ancestral ABM at 6 and 24 hours (blue and maroon, respectively) in mixed cultures of unbuffered Luria Bertani broth, PM, and Bis-Tris and Tris buffered LB, BPM and TPM respectively.

Figure S6. Ancestral ABM (62-2) Ara+ in chemostat and batch culture (insert) with evolved ABM(62-2) Ara-, Ratio of evolved to ancestral cells. In the batch culture the blue is the ratio at 5.5 hours and the maroon at 24 hours. In the stationary phase batch culture the mixture of cells used to start the chemostats were mixed without supplemental LB, in the exponential, 100µl of the mixture was added to 10 ml LB.

Figure S7. Deletion of *glgC17* in the chromosome of the evolved cells. (A) map of the *glgC* region before and after the knock-out of *glgC17*. The restriction enzymes used to delete a 1032 bps fragment of coding sequences inside *glgC* were AscI (A) and XmnI (X). (B) PCR amplification of chromosomal DNA from evolved (left) and evolved Δ*glgC17* cells (right) using the oligonucleotides glgC51 and glgC31 (see plasmid construction). The size of the expected PCR fragments were 3496 and 2465 bps, respectively. The leftmost lane is a DNA ladder.

Figure S8 The constructed glg variants of AB1157 *glgc17* Ara- (glgA, glgP and An) in mixed cultures with the evolved Ara+ (E) and ancestral Ara+ (A) ABM. Ratio of the Ara constructs with the Ara+ ancestral and evolved ABM strains at 5 hours (blue) and 24 hours (maroon).

Figure S9. Cell viability analysis of SCDI using confocal microscopy. a) Fluorescent emission of DsRed-labelled ancestral ABM (AR) mixed with either GFP-labelled evolved (EG) or FP labelled ancestral cells (AG) at a 1:1 initial ratio and grown together

for 28 hours at 37°C. b) The integrity of bacteria in 28 hours mixture of GFP-labelled ancestral cells (AG) with unlabelled evolved (E) or ancestral (A) cells was monitored using propidium iodide (PI). In addition, bacterial cells were visualized through staining of their nucleoids by using DAPI. The lower micrographs show the cell samples after treatment with the membrane damaging compound polymyxin B.

Figure S10. Observed (blue) and Expected (fusia) changes in the relative frequency of AB1157 in successive 1:100 serial transfers in mixed LB culture with MG1655. The expected decline in the relative frequency of AB1157 was calculated under the assumption of 100 fold reduction in the density of AB1157 at each transfer.

Figure S11. SCDI with other strains; ratio of Ara+ to Ara- in mixed cultures at 5 and 24 hours. See the text for the strain designations.

Figure S12. MC1061 is subject to SCDI: Changes in ratio of Ara+ and Ara- between 5 and 24 hours; MG1655 (MGA+), or an Ara+ evolved ABM with MC1061 (MCA-).

Figure S13. Simulation results: Changes in ratio of Evolved and Ancestral cells at different initial total cell densities and different initial frequencies of evolved bacteria (e.g. 1E8- 1.0) is an initial cell density of 1 x 108 and an initial relative frequency of the evolved strain of 0.5. The parameter values in these runs, da=de=0.02 per hour and τ = 10-11 were chosen to illustrate the properties of this model and are not derived from actual estimates.

Figure S1

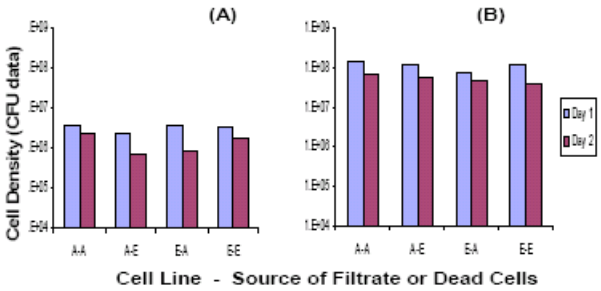

(A)
(B)
Cell Density (CFU data)
Day 1
Day 2
A-A
A-E
E-A
E-E
Cell Line - Source of Filtrate or Dead Cells

Figure S2

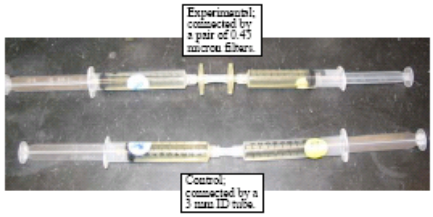

Experimental;
connected by
a pair of 0.45
micron filters.
Control;
connected by a
3 mm ID tube.

Figure S3

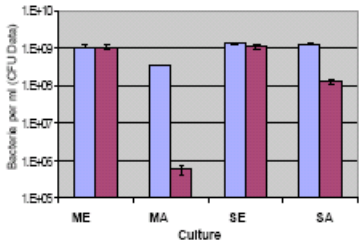

1.E+10
1.E+09
1.E+08
1.E+07
1.E+06
1.E+05
Bacteria per ml (CFU Data)
ME
MA
SE
SA
Culture

Figure S4

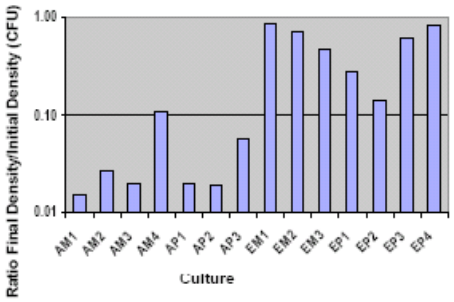

Ratio Final Density/Initial Density (CFU)
1.00
0.10
0.01
AM1
AM2
AM3
AM4
AP1
AP2
AP3
EM1
EM2
EM3
EP1
EP2
EP3
EP4
Culture

Figure S5

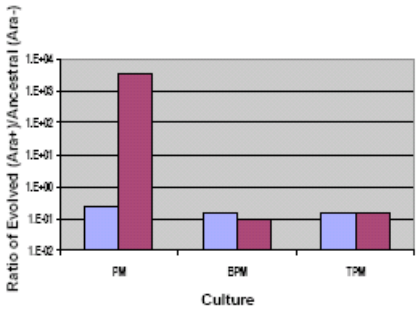

Ratio of Evolved (Ara+)/Ancestral (Ara-)
1.E+04
1.E+03
1.E+02
1.E+01
1.E+00
1.E-01
1.E-02
PM
BPM
TPM
Culture

Figure S6

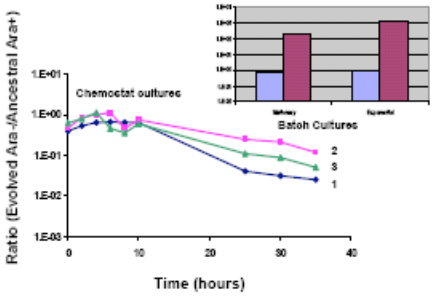

Chemostat cultures
Batch Cultures
Ratio (Evolved Ara-/Ancestral Ara+)
Time (hours)
1.E+01
1.E+00
1.E-01
1.E-02
1.E-03
0
10
20
30
40
2
3
1

Figure S7

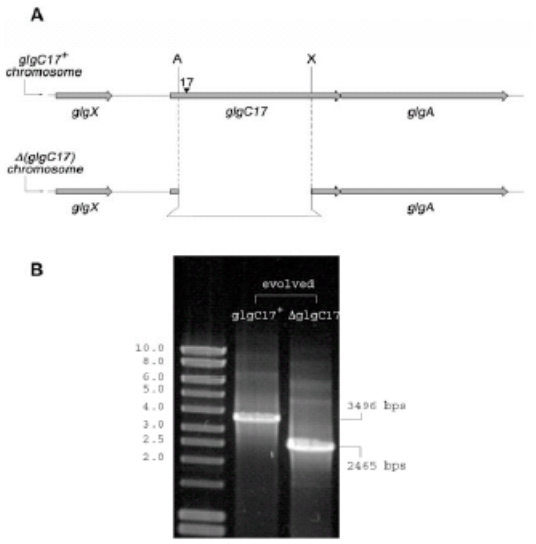

A
glgC17+ chromosome
A
X
17
glgX
glgC17
glgA
Δ(glgC17) chromosome
glgX
glgA
B
evolved
glgC17+ ΔglgC17
10.0
8.0
6.0
5.0
4.0
3.0
2.5
2.0
3496 bps
2465 bps

Figure S8

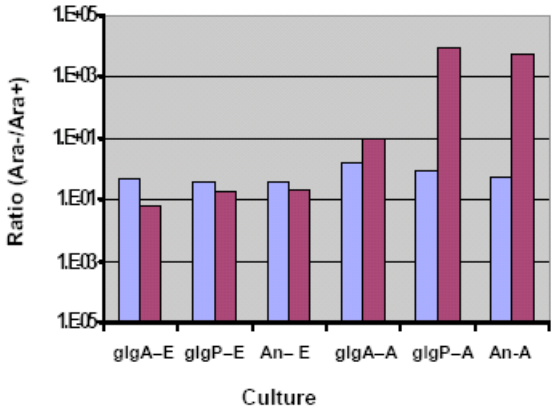
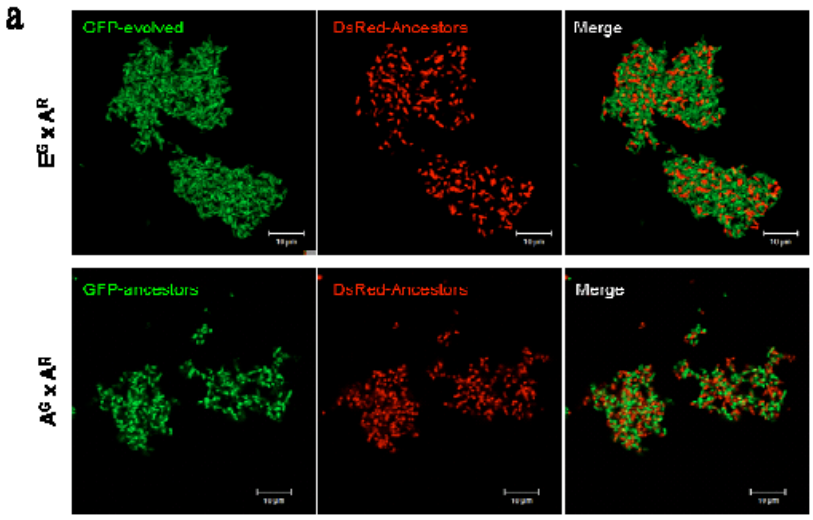

1.E+05
1.E+03
1.E+01
1.E01
1.E03
1.E05
Ratio (Ara-/Ara+)
glgA–E
glgP–E
An– E
glgA–A
glgP–A
An-A
Culture

Figure S9


a
GFP-evolved
DsRed-Ancestors
Merge
E^G x A^R
GFP-ancestors
DsRed-Ancestors
Merge
A^G x A^R

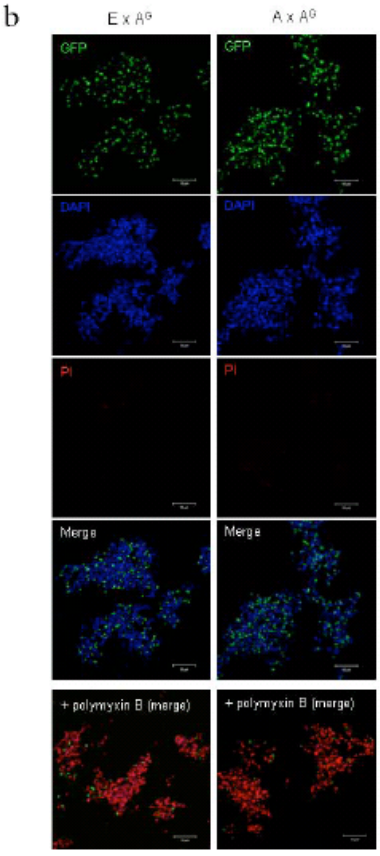

b
E x A^G
A x A^G
GFP
GFP
DAPI
DAPI
PI
PI
Merge
Merge
+ polymyxin B (merge)
+ polymyxin B (merge)

Figure S10

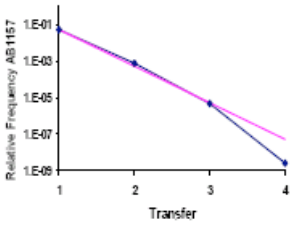

Relative Frequency AB1157
1.E-01
1.E-03
1.E-05
1.E-07
1.E-09
1
2
3
4
Transfer

Figure S11

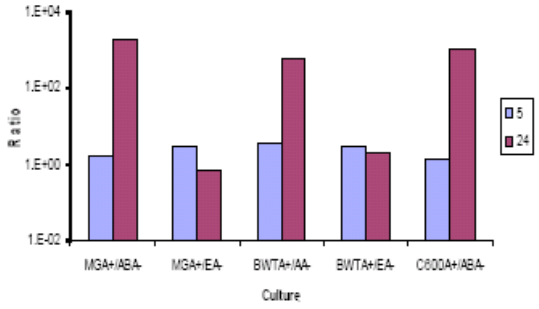

1.E+04
1.E+02
1.E+00
1.E-02
Ratio
5
24
MGA+/ABA-
MGA+/EA-
BWTA+/AA-
BWTA+/EA-
C600A+/ABA-
Culture

Figure S12

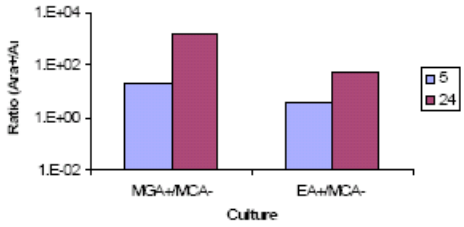
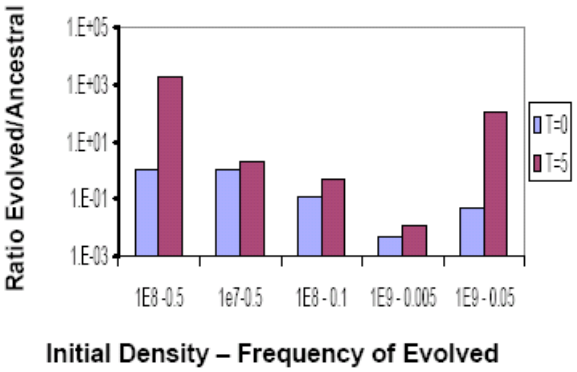


Figure S13


Ratio Evolved/Ancestral
1.E+05
1.E+03
1.E+01
1.E-01
1.E-03
1E8 -0.5
1e7-0.5
1E8 - 0.1
1E9 - 0.005
1E9 - 0.05
T=0
T=5
Initial Density – Frequency of Evolved