SCIENTIFIC REPORTS

OPEN

A Canonical Biophysical Model of the CsrA Global Regulator Suggests Flexible Regulator-Target Interactions

Received: 15 January 2018
Accepted: 4 June 2018
Published online: 02 July 2018

A. N. Leistra[1], G. Gelderman[1], S. W. Sowa[2], A. Moon-Walker[3], H. M. Salis[4] & L. M. Contreras[1]

Bacterial global post-transcriptional regulators execute hundreds of interactions with targets that display varying molecular features while retaining specificity. Herein, we develop, validate, and apply a biophysical, statistical thermodynamic model of canonical target mRNA interactions with the CsrA global post-transcriptional regulator to understand the molecular features that contribute to target regulation. Altogether, we model interactions of CsrA with a pool of 236 mRNA: 107 are experimentally regulated by CsrA and 129 are suspected interaction partners. Guided by current understanding of CsrA-mRNA interactions, we incorporate (i) mRNA nucleotide sequence, (ii) cooperativity of CsrA-mRNA binding, and (iii) minimization of mRNA structural changes to identify an ensemble of likely binding sites and their free energies. The regulatory impact of bound CsrA on mRNA translation is determined with the RBS calculator. Predicted regulation of 66 experimentally regulated mRNAs adheres to the principles of canonical CsrA-mRNA interactions; the remainder implies that other, diverse mechanisms may underlie CsrA-mRNA interaction and regulation. Importantly, results suggest that this global regulator may bind targets in multiple conformations, via flexible stretches of overlapping predicted binding sites. This novel observation expands the notion that CsrA always binds to its targets at specific consensus sequences.

Large-scale omics techniques have been applied with increasing frequency to the study of bacterial post-transcriptional global regulators (e.g., *E. coli* Hfq, ProQ, and CsrA), aiming to elucidate the scope of their targets and regulatory effects[1–6]. Results have established that these global regulators can act upon over hundreds of targets. For example, the Hfq global regulator has been implicated as a chaperone for nearly all characterized small RNA (sRNA)-messenger RNA (mRNA) interactions in *E. coli*[7,8]. Hfq-RNA interactions are characterized by Hfq binding U-rich sequences in the 3′ portions of sRNAs[3,9–11] and A-rich sequences in the 5′ untranslated regions (UTRs) of mRNAs[3,12,13]. These motifs are considered specific to the proximal and distal binding faces, respectively, of the Hfq hexamer[12,14]. However, interaction-specific variation is observed, in which sRNAs contact both faces, like ChiX and McaS[15,16]. The impact of such variation in binding on target control and network regulation is still unfolding.

For the case of CsrA, approximately 800 mRNA have been identified across multiple environmental conditions as potentially interacting with CsrA[2,5]: this total approaches 20% of the *E. coli* genome. Generally, the CsrA homodimer binds a target mRNA at two copies[17] of a consensus sequence (ANGGA)[18], preferentially located in the loop of a hairpin structure[18,19]. However substantial variation in these traits is observed. For example, among the 31 classical and well-characterized targets of CsrA (defined in Table 1), mRNAs like *pgaA* may hold to the general pattern[20], but *clpB*, *dps*, *patA*, and *purM* do not present the consensus ANGGA binding motif[5]. Similarly, *hfq* and *ycdT* present only a single copy of the consensus sequence[20–24] and *cstA*[21] presents footprinted binding sites outside of the typical stem loop structures that have been shown favorable to CsrA binding.

[1]McKetta Department of Chemical Engineering, University of Texas at Austin, 200 E. Dean Keeton St. Stop C0400, Austin, TX, 78712, USA. [2]Microbiology Graduate Program, University of Texas at Austin, 100 E. 24th St. Stop A6500, Austin, TX, 78712, USA. [3]Biological Sciences Program College of Natural Sciences, University of Texas at Austin, 120 Inner Campus Drive Stop G2500, Austin, TX, 78712, USA. [4]Department of Chemical Engineering, Pennsylvania State University, 210 Agricultural Engineering Building, University Park, PA, 16802, USA. Correspondence and requests for materials should be addressed to L.M.C. (email: lcontrer@che.utexas.edu)

mRNA Target Type	Foot-printed?	Direct Binding?	Regulation of translation?	Literature of mRNA association with CsrA	Number of mRNA modeled	Number of mRNA with Correctly-Predicted Regulation (rep or act)
Classical	Yes	Yes	Yes	Multiple Studies[20,21,23,25,39,42–47]	10	8
Well-characterized	No	Yes	Yes	Multiple Studies[5,6,24]	21	13
Functional	No	No	Yes	Sequence-Search model[40,41]	30	23
				Multi-Omics study[5]	28	17
				Co-IP[2], HITS-CLIP[5], or mRNA stability[4] study	18	5
Total					***107***	***66***

Table 1. mRNAs experimentally regulated by CsrA.

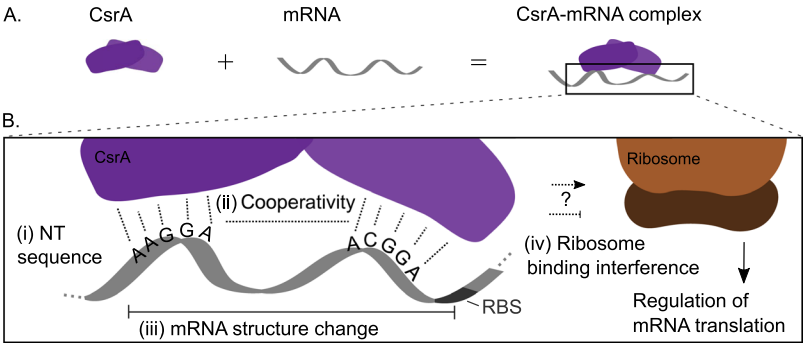

Figure 1. Biophysical considerations for modeling the CsrA post-transcriptional regulator. (**A**) The CsrA protein regulator binds mRNA, preferentially with a stoichiometry of one CsrA homodimer to two binding sites within an mRNA. (**B**) Zooming in on the CsrA-mRNA interaction highlights the biophysical factors that affect it. CsrA preferentially binds (i) ANGGA nucleotide sequences (ii) spaced approximately 10–55 nucleotides apart to support cooperative dual-site binding. The (iii) structure of the mRNA determines how accessible these sequences may be and the energetic cost of CsrA-mRNA binding. Lastly, CsrA bound to the mRNA may (iv) interfere with ribosome binding to regulate mRNA translation.

The question thus arises of how post-transcriptional global regulators execute hundreds of interactions with targets of variable sequence and structural features while retaining specificity. Although mutation followed by biochemical footprinting, gel shift, and reporter assays are typically used to assess how such features may contribute to regulator-target interactions[25–27], these are low throughput. For this reason, omics techniques and, in particular, co-immunoprecipitation studies have proven helpful for establishing pools of a post-transcriptional regulator's potential targets[3,5,28]. These studies, however, cannot assess how specific molecular features of a target mRNA may influence its recognition and control by the regulator.

Thermodynamic modeling approaches offer higher-throughput opportunities to test the impact of specific molecular features hypothesized to be important on regulator-target interactions. For instance, thermodynamic models have been used to characterize the energetics of single post-transcriptional regulator-target, e.g., sRNA-mRNA, interactions and predict energetically likely mRNA targets from bacterial genomes[29]. These models have also begun to incorporate estimation of a regulator's effect on RNA target expression[30–32] and to predict local RNA accessibility patterns that indicate potential areas for RNA interactions[33]. Recent advances in thermodynamic modeling of translation initiation[34–37] have further expanded the ability to calculate a regulator's impact on the expression of its mRNA targets by bridging molecular interactions with cellular translation.

In this work, we develop, validate, and apply a biophysical, statistical thermodynamic model of CsrA-mRNA targets to investigate how the molecular features of a target mRNA may impact its binding and regulation by CsrA. We employ the *E. coli* CsrA protein (Fig. 1A) as a model post-transcriptional regulator given that its binding and regulation of several mRNA targets has been well-characterized. Moreover, we investigate this system because, as described above, its well-characterized mRNA targets demonstrate variance in the sequence and structural features typically considered as hallmarks of CsrA regulatory interactions. This experimentally-demonstrated flexibility in mRNA recognition[17,19,38] raises questions as to how molecular diversity of target interactions could enable diverse control schemes to support the large proposed scope and known complexity of the CsrA target network[2,5,39].

Previous models of CsrA-mRNA interactions have focused on genome-wide identification of potential mRNAs bound by CsrA with variations of the ANGGA consensus binding site sequence[18,40,41]. Specifically, work by McKee *et al.* identified *E. coli* genes with a [A/C/U]A[A/G/U]GGA[A/G/U][A/C/U] version of the CsrA binding motif within a 22 nucleotide window upstream of their translation start site[40]. Similarly, Kulkarni *et al.* published an algorithm that identified genes containing an A(N)GGA sequence (where N is any or a gap nucleotide) in a window 30 nucleotides upstream to 5 nucleotides downstream of their translation start sites. Additionally, the

number and spacing of total available A(N)GGA-like binding sites were considered. It was required that within a window from its transcription start site to the 30th nucleotide of coding sequence, the gene must include at least 3 total A(N)GGA sequences within 10–60 nucleotides of each other or 2A(N)GGA sequences, each with a degenerate GGA-like sequence within a 10 nucleotide distance[41].

Given the detailed, mechanistic characterizations of several CsrA-mRNA interactions, termed "classical" targets[20,21,23,25,39,42–47] (defined in Table 1), and the recent expansion of the well-characterized CsrA mRNA target repertoire[5,6], we expand the previous sequence-based approaches to build a model with greater molecular-level resolution. Specifically, we build a model of canonical CsrA-mRNA binding and regulation by incorporating the biophysical factors characteristic of these known regulatory interactions: mRNA nucleotide sequence[18], potential cooperativity of two-site binding[17,19], and minimization of mRNA structural changes. We use these factors to identify and estimate the free energies of an ensemble of CsrA-bound conformations of an mRNA of interest (Fig. 1B, part i–iii). Importantly, these calculations are performed in the context of translation initiation rate calculations[34,35] to capture the impact of bound CsrA on translation of the mRNA (Fig. 1B, part iv). In this way, our model is able to capture mRNAs expected to be activated, bound but not regulated, or repressed by CsrA binding outside of the Shine-Dalgarno region. Moreover, the current approach enables observation of the effects molecular variations in mRNA features can have on CsrA-mRNA regulation; these variations include mRNA structure and the number, location, spacing, and sequence of potential binding sites.

We employ this model to assess in detail molecular traits that contribute to CsrA regulation of experimentally-verified CsrA-regulated mRNAs that can be described by the canonical model. For this purpose, we analyze probable CsrA-target interactions captured within ensembles of CsrA-bound target conformations. Specifically, we model 236 mRNA: 107 are experimentally-confirmed CsrA-regulated mRNAs (includes the 31 classical and well-characterized targets) that are either repressed or activated in the presence of CsrA (Supplementary Table S1). We conduct detailed molecular analysis of resulting ensembles of CsrA-bound conformations for 66 of these mRNA targets given that their regulation is well-captured by our canonical biophysical model. The model is also applied to 129 mRNAs for which regulation by CsrA is suspected based on results of an integrated omics analysis of the Csr system[5] or prior literature evidence (summarized in Supplementary Table S1), but has not been experimentally confirmed. Importantly, results suggest that CsrA may bind targets in multiple conformations, allowing use of flexible "binding pockets" that contain a continuous stretch of potential binding sites; this is a novel observation that expands the notion that CsrA always binds to its targets rigidly at a specific five-nucleotide consensus sequence. As such, this work provides insights into molecular features that can be important in understanding ways by which a global post-transcriptional regulator can bind and affect a large pool of cellular mRNAs.

Results

Biophysics of CsrA-mRNA Interactions Guide Canonical Model of CsrA Binding and Regulation. To develop a thermodynamic model of CsrA-mRNA interaction, we considered four biophysical factors understood to govern CsrA binding and regulation, as inferred from studies of classical mRNA targets: consensus sequence (the best-studied aspect of CsrA-mRNA binding), cooperativity, structural change, and ribosome binding ability (Fig. 1B). As such, we term it the "canonical biophysical model". With respect to nucleotide sequence, we identified plausible five-nucleotide CsrA binding sites in an mRNA of interest based on sequence similarity to the optimal AAGGA binding site sequence previously identified experimentally[17,18]. To do this quantitatively, we derived and employed a position-specific mono-nucleotide free energy model (FEM) that quantifies the energetic contribution of each nucleotide of a five-nucleotide sequence to CsrA-mRNA binding (Fig. 2A) (Methods). Thus the likelihood of CsrA binding to a five-nucleotide sequence, quantified as $\Delta G_{site,}$ is the sum of the energetic contributions of each nucleotide. Five-nucleotide sequences with $\Delta G_{site} < 0$ are identified as potential CsrA binding sites.

To address cooperativity in CsrA binding, we quantified the likelihood of two sites within an mRNA being bound by a single CsrA homodimer after identifying all unique pairs of identified potential binding sites. In our model, the cooperativity term ($\Delta G_{cooperativity}$) creates a distance-dependent negative free energy for site pairs spaced 10 to 55 nucleotides apart, as this distance was observed to support cooperative binding[17,19] and has been used in a prior model to represent two-site binding[41] (Fig. 2B). The $\Delta G_{cooperativity}$ term has the greatest energetic contribution when the binding sites are 10 nucleotides apart and decreases to no contribution for sites greater than 55 nucleotides apart (Methods).

All of the unique combinations of two potential binding sites identified in an mRNA of interest constitute an ensemble of the possible conformations in which the CsrA can bind the mRNA at two sites. For each member of this ensemble, we quantified the energetic cost of unfolding the mRNA structure to accommodate CsrA binding (Fig. 2C). The change in free energy, $\Delta\Delta G_{unfold}$, between the unbound mRNA and the mRNA bound to CsrA (as predicted by the ViennaRNA RNA folding package[48]) is determined for each member of the ensemble (Methods).

The sum of the contributions of each individual term described above results in the ability to calculate the total free energy of CsrA-mRNA binding ($\Delta G_{mRNA:CsrA}$) for each CsrA-bound mRNA conformation (Fig. 2D):

$$\Delta G_{mRNA:CsrA} = \Delta G_{site1} + \Delta G_{site2} + \Delta G_{cooperativity} + \Delta\Delta G_{unfold} \quad (1)$$

In this model, each CsrA-bound mRNA conformation represents varying optimality of binding site sequence, cooperativity, and structural unfolding. We quantified the probability of CsrA binding to the mRNA of interest in a given conformation as the Boltzmann probability of the system existing in that particular energy state, described by $\Delta G_{mRNA:CsrA}$, compared to the energy states of all the other possible conformations (Methods).

To predict the effects of CsrA binding on mRNA translation, we selected the 15 most-likely (lowest energy $\Delta G_{mRNA:CsrA}$) CsrA-bound mRNA conformations and modeled the regulatory impact of CsrA on their translation

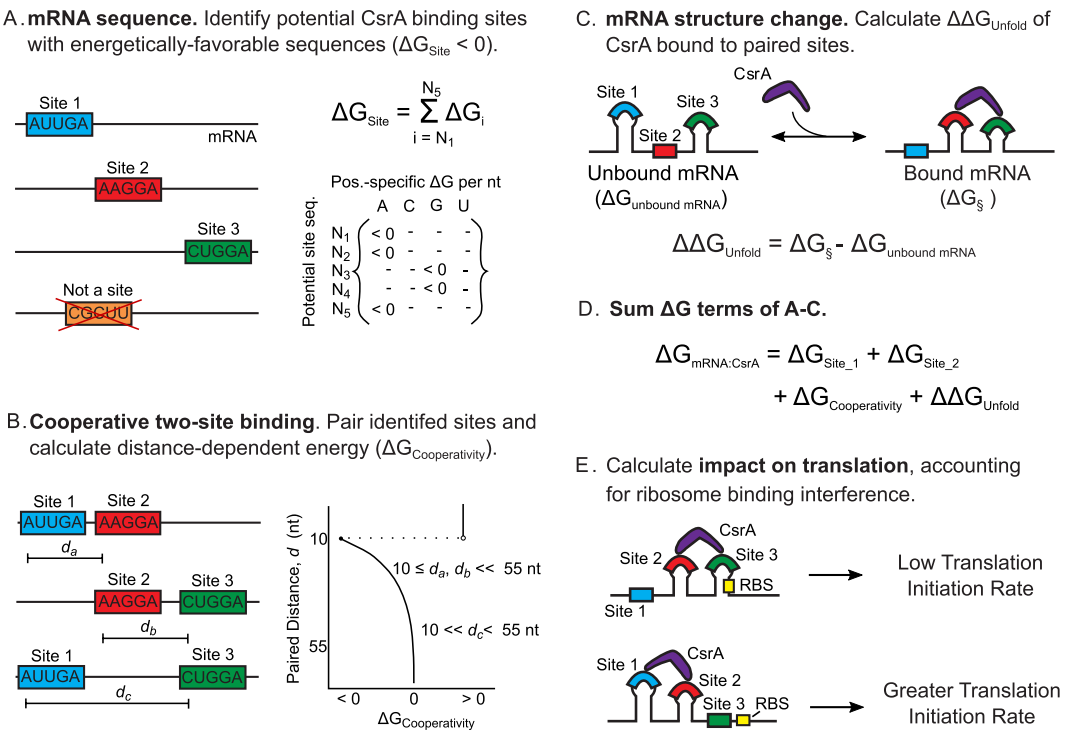

Figure 2. Biophysical, statistical thermodynamic model of CsrA-mRNA interaction and impact on translation. (**A**) Potential CsrA binding sites are identified within an mRNA sequence. A position weight matrix, which quantifies the energy of CsrA binding (ΔG_{Site}) to 5 nucleotide-long sequence segments, is used to identify favorable binding sites ($\Delta G_{Site} < 0$). (**B**) To account for cooperativity in CsrA binding to two sites in an mRNA, potential CsrA binding sites are paired and $\Delta G_{cooperativity}$ calculated as a function of their distance apart. Sites 10–55 nucleotides apart are rewarded ($\Delta G_{cooperativity} < 0$), and sites less than or equal to 9 nucleotides apart penalized. (**C**) The free energy change due to CsrA binding and altering mRNA structure ($\Delta\Delta G_{unfold} \geq 0$) is estimated for each pair of sites. (**D**) The ΔG terms calculated in panels A–C are summed for each pair of sites and (**E**) translation initiation modeled for the top 15 most energetically-favorable bound conformations to determine CsrA-mRNA regulation.

initiation rates, the rate-limiting step of translation, using the RBS Calculator[34,35] (Fig. 2E). We selected the top 15 conformations, as ranked by high to low Boltzman probability, because this portion represents the majority of the most likely conformations in the distribution of possible CsrA-mRNA interactions. As a point of reference, the 15th most-likely conformation is, on average, 6.0 ± 4.6 fold less likely than the most-likely bound conformation for an mRNA. The fold change in translation for each mRNA target was calculated for each of the 15 lowest energy CsrA-bound conformations relative to the unbound/reference state (Methods). Three possible outcomes result from this calculation: (1) repression of translation (e.g., if CsrA binding directly or indirectly blocks the RBS or start codon), (2) activation of translation (e.g., if CsrA binding leads to increased accessibility of the RBS or start codon), or (3) no impact on translation (e.g., if CsrA binding does not alter translation initiation rate relative to the unbound conformation). It is worth noting that the model-calculated fold change in translation and the type of regulation predicted for each of the top 15 energetically-favorable CsrA-bound conformations of an mRNA can vary in magnitude and in direction given that unique binding site pairs comprise each conformation. Overall, the complete model has only 8 free parameters; 5 parameters quantifying CsrA's binding affinity to mRNA sites and 3 parameters quantifying the free energy of dimeric CsrA cooperativity. Importantly, free model parameters were fit to experimental measurements prior to performing predictions on full-length mRNA sequences, and no attempt was made to optimize model parameter values.

Selection of 236 CsrA-mRNA Interactions. We identified a set of 236 mRNAs with known or suspected CsrA interaction and/or regulation and applied the canonical biophysical model to predict how CsrA regulates their translation rates. Specifically, 10 mRNA are classical CsrA targets whose binding sites have been footprinted *in vitro*, direct binding demonstrated *in vitro* or *in vivo*, and regulation demonstrated *in vivo* in prior studies[20,21,23,25,39,42–47]. 21 mRNAs represent well-characterized CsrA targets, whose direct binding has been demonstrated *in vivo* or *in vitro* (without footprinting data) and regulation demonstrated *in vivo*[5,6,24]. Lastly, 76 mRNAs are functional targets, as their regulation by CsrA has been shown *in vivo*, but their direct binding with CsrA has not been confirmed or footprinted. These three types of CsrA targets comprise a set of 107 mRNA that are regulated by CsrA experimentally (Tables 1 and S1).

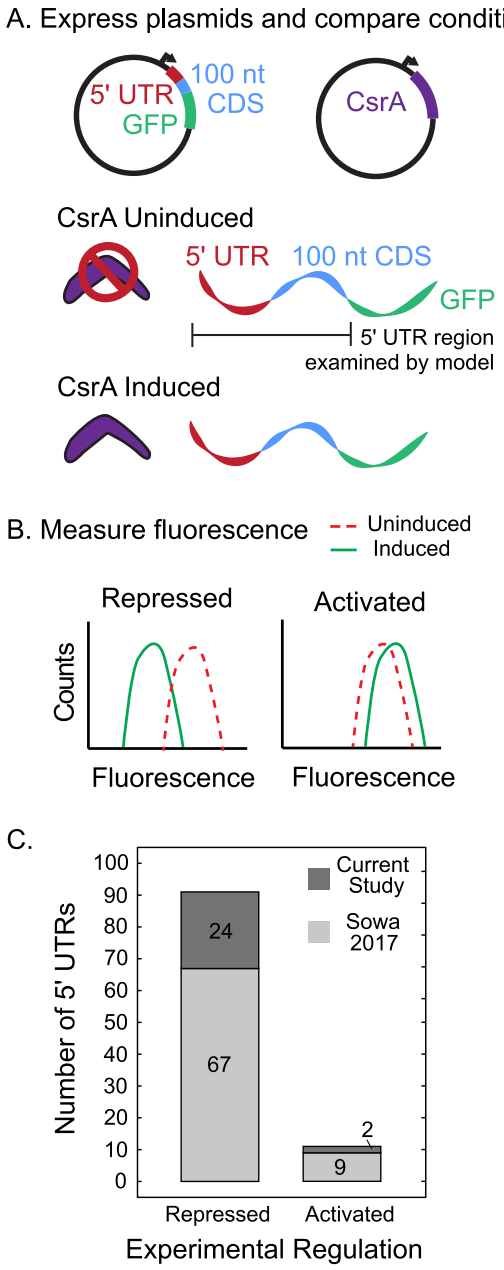

Figure 3. Fluorescent translational reporter assay determines 5′ UTRs repressed and activated in the presence of CsrA. (**A**) Schematic of the fluorescent translational reporter assay. A two plasmid system, one that contains an inducible CsrA and the other that constitutively expresses the 5′ UTR of an mRNA of interest fused to GFP, is used to determine the regulatory relationship of CsrA to the target 5′ UTR. The assay compares 5′ UTR-controlled GFP fluorescence under two conditions: CsrA present (induced condition) and CsrA absent (uninduced condition). (**B**) Significant down- or up-shift in fluorescence upon induction indicates CsrA regulates the 5′ UTR-GFP reporter. (**C**) Bars indicate number of 5′ UTRs which display repression (downshift, P-value < 0.1), or activation (upshift, P-value < 0.1). Bar shading indicates the study the reporter assay results are published in: Sowa *et al.* (lower, light gray) (ref.[5]) or the current work (upper, dark gray).

The majority of the well-characterized and functional CsrA targets were previously tested for CsrA regulation using an *in vivo* fluorescent translational reporter assay (Supplementary Table S1)[5]. This study performed an integrated omics analysis of the Csr system in *E. coli* and identified mRNAs likely to be regulated by CsrA; many were tested for CsrA regulation with the *in vivo* assay as follow-up. In the current work, additional mRNAs were tested for CsrA regulation using the same fluorescent translational reporter assay (Fig. 3A,B) (Supplementary Methods, Supplementary Table S2). It is worth acknowledging that for 23 functional targets showing activation or repression by CsrA in the current work, this is their first specific evidence of CsrA regulation. Figure 3 C summarizes results of the fluorescent translational reporter assay across both the initial (Sowa *et al.*)[5] and the current works. A total of 91 mRNAs showed repression and 11 showed activation by CsrA. It should be noted that the

Experimental Regulation	Model-Calculated Regulation				*Total*
	Repression	Activation	No Impact	Heterogeneous	
Repression	65	3	11	16	95
Activation	5	1	0	6	12
Total	*70*	*4*	*11*	*22*	*107*

Table 2. Experimentally-measured and model-calculated regulation of mRNAs.

remaining 5 mRNA targets of the 107 experimentally CsrA-regulated mRNAs are classical or well-characterized targets extensively studied in other works[24,39,45,46] (Supplementary Table S1).

The remaining 129 of the 236 modeled mRNAs were initially identified as potential CsrA-controlled mRNAs in a prior integrated omics study[5], other transcriptomics-based experimental studies[2,4], or sequence-search based computational studies[40,41] (more detail in Supplementary Table S1). We tested the vast majority of these mRNAs (124) for CsrA regulation by fluorescent translational reporter assay (Supplementary Table S2 or in ref.[5]). However, fluorescence results of the 5′ UTR-GFP constructs were inconclusive for various reasons, largely attributed to noise in the fluorescence data, inconsistencies amongst the biological replicas, or low signal indistinguishable from fluorescence background (Supplementary Methods). These are tabulated and regulation indicated as undetermined in Supplementary Table S1.

Canonical Biophysical Model Captures CsrA Regulation of 66 mRNA Targets. For each of the 236 mRNAs, we extracted its 5′ untranslated region and the first 100 nucleotides of protein coding sequence, and inputted this sequence into the canonical biophysical model to calculate: (i) the locations where CsrA binds, (ii) the predicted CsrA-bound mRNA structures, (iii) the calculated CsrA-mRNA binding free energies for all ensemble conformations (Supplementary Data 1), and (iv) the predicted changes in translation rate for the top (i.e. lowest-energy) fifteen ensemble conformations (Supplementary Table S3). For analysis, we then classified each mRNA as being repressed, activated, or not impacted by CsrA-mediated translation regulation. In most cases, the top fifteen ensemble conformations had similar modes of translation regulation; however, when the most probable ensemble conformations had different equally-likely regulatory modes, we classified these model predictions as heterogeneous (Methods).

To assess model performance, we used the subset of 107 mRNAs known to be regulated by CsrA (10 classical, 21 well-characterized, and 76 functional targets) (Tables 1 and S1). Across the set of 107 mRNAs with known experimental CsrA-mediated regulation, the canonical biophysical model correctly predicted the regulatory modes of 66 mRNAs: 65 of these mRNA have repressed expression, while one is activated (Table 2). It is unsurprising that the canonical biophysical model lacks robust capability for predicting activating CsrA-mRNA interactions given that the biophysical principles used to "train" the model represent the canonical understanding of CsrA interactions with its classical targets. Only one classical target is activated by CsrA, while nine are repressed. This pattern holds for the well-characterized targets (19 repressed and 2 activated) and in Fig. 3C; it is also supported by the understanding that CsrA typically binds in the Shine-Dalgarno region of an mRNA, due to similarities of the consensus CsrA binding site and Shine-Dalgarno sequences, and likely represses translation through direct occlusion of the RBS. The model was best able to correctly classify how CsrA regulates translation when the mRNA contained consensus or near-consensus CsrA binding sites, as in the classical and functional targets identified by sequence search studies (Table 1). Notably here, the model's formula for calculating ΔG_{site} was parameterized by using binding affinity measurements to consensus CsrA sites with only single nucleotide mutations[18]. Therefore, the model does not account for any non-additive energetic contributions to CsrA's binding affinity, which could play a role when mRNAs contain non-consensus CsrA sites.

Remarkably, for the 41 mRNAs where the model incorrectly predicted regulation, regulation of 22 mRNAs was predicted as heterogeneous: in 21 of 22 cases, regulation of at least 3 of the top 15 most-likely ensemble conformations was correct, but predicted regulation of the ensemble was heterogeneous overall. For 14 mRNAs, repression was missed (3 were predicted as activated and 11 as non-impacted) and for the remaining 5 mRNAs repression was falsely predicted (experimental data showed activation) (Table 2). These results indicate that rather than over-predicting repression, the model does not as readily capture activation mechanisms. More broadly, the 41 mRNAs not captured by the canonical biophysical model might be regulated by CsrA through mechanisms not yet fully described in the literature. Targets for which the model does not clearly predict one regulatory outcome (i.e. heterogeneous regulation) or predicts regulation that does not match experimental results constitute excellent candidates for further exploration of potentially new mechanisms of CsrA post-transcriptional regulation.

Comparison to Prior CsrA-mRNA Interaction Models. The canonical biophysical model is unique from prior models of CsrA-mRNA interactions in that it (i) calculates a predicted free energy of CsrA-mRNA binding that (ii) incorporates changes in predicted mRNA structure upon CsrA binding. It also (iii) estimates a regulatory outcome of CsrA binding on mRNA translation by calculating a change in estimated translation rate and (iv) predicts *an ensemble* of potential CsrA-bound mRNA conformations and their translation rates. Previous models focused on using consensus binding site sequences to predict high affinity interactions between CsrA and mRNAs across the genome. Importantly, when attempting to predict whether CsrA plays a role in repressing an mRNA's translation rate, the canonical biophysical model's accuracy is greater than published sequence-based models[40,41] (Supplementary Table S4). The difference in accuracy is mainly due to an increased number of correct repression predictions by the canonical biophysical model (65 mRNAs) compared to the McKee and Kulkarni

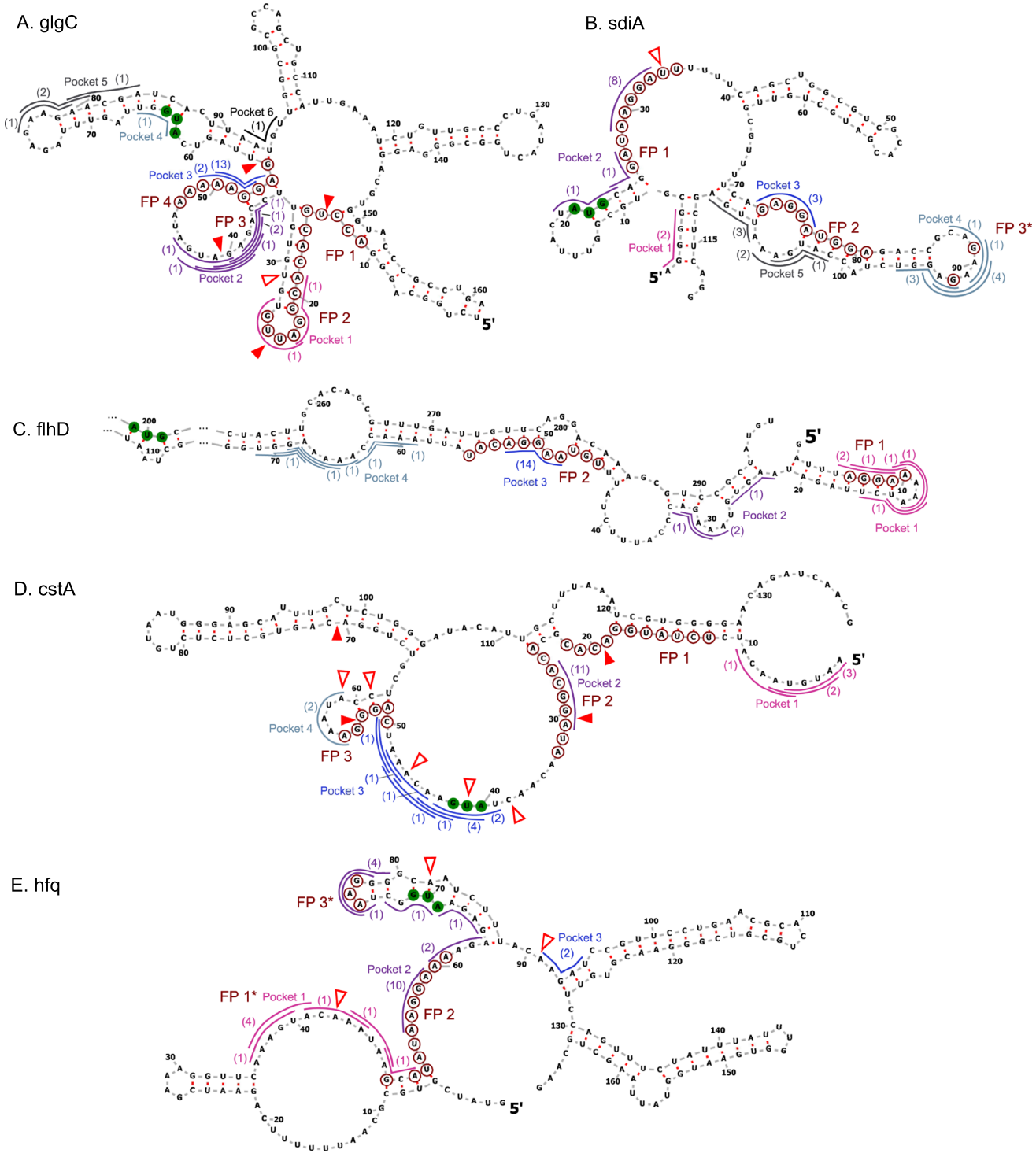

Figure 4. Model binding site predictions capture known CsrA-mRNA footprints. (**A**–**E**) Vienna-RNA predicted structures of eight canonical CsrA-mRNA targets (unbound by CsrA). Model-identified CsrA binding sites of the fifteen most-likely CsrA-bound conformations of each target are shown as colored outlines. Number in parentheses adjacent to the outline indicates the number of times that particular 5-nucleotide binding site is identified in the fifteen most-likely CsrA-bound mRNA conformations. Overlapping or adjacent mRNA binding sites form pockets, indicated by the color of the outline. Start codon nucleotides are shown with green fill. Nucleotides within experimentally-determined *in vitro* footprints are marked with red circular outlines. Red fill wedges indicate 3′ edges of binding sites identified by *in vitro* 3′ boundary analyses; red-outlined wedges indicate 3′ edges of binding sites identified by *in vitro* toeprint analyses. *In vitro*-identified binding sites, regardless of method, are labeled from 5′ to 3′, starting with "FP 1". *In vitro* binding assays were performed in, (**A**) refs[17,25]; (**B**) ref.[43]; (**C**) ref.[44]; (**D**) ref.[21]; and (**E**) ref.[23]. (**B** and **E**) FP 3 of *sdiA* and FP 1 and FP 3 of *hfq* are marked with an "*" to indicate weaker *in vitro* signals newly interpreted as binding sites. (**D**) Toeprint signals within pocket 3 of *cstA* did not indicate a clear binding site, and thus was not labelled with an "FP".

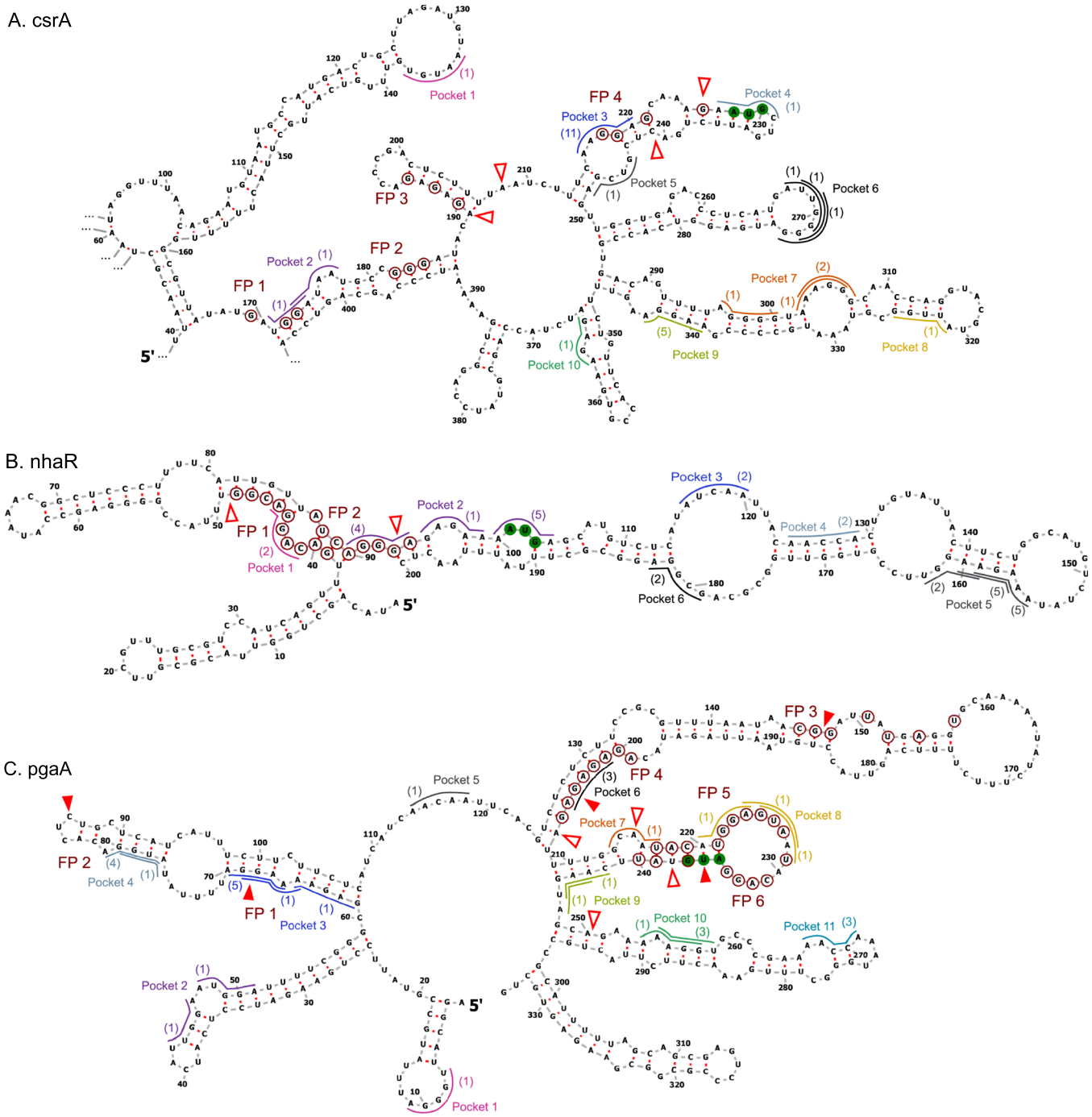

Figure 5. Model binding site predictions capture known CsrA-mRNA footprints. (**A**–**C**) Vienna-RNA predicted structures of eight canonical CsrA-mRNA targets (unbound by CsrA). Model-identified CsrA binding sites of the fifteen most-likely CsrA-bound conformations of each target are shown as colored outlines. Number in parentheses adjacent to the outline indicates the number of times that particular 5-nucleotide binding site is identified in the fifteen most-likely CsrA-bound mRNA conformations. Overlapping or adjacent mRNA binding sites form pockets, indicated by the color of the outline. Start codon nucleotides are shown with green fill. Nucleotides within experimentally-determined *in vitro* footprints are marked with red circular outlines. Red fill wedges indicate 3′ edges of binding sites identified by *in vitro* 3′ boundary analyses; red-outlined wedges indicate 3′ edges of binding sites identified by *in vitro* toeprint analyses. *In vitro*-identified binding sites, regardless of method, are labeled from 5′ to 3′, starting with "FP 1". *In vitro* binding assays were performed in, (**A**) ref.[47]; (**B**) ref.[42]; and (**C**) ref.[20].

sequence-search approaches (17 and 40 mRNAs, respectively) (Supplementary Table S4 and Supplementary Fig. S1). While the high stringency of the sequence-search approaches is useful for identifying potential targets with a low false positive rate on a genome scale (and avoiding heterogeneous predictions), the greater number of

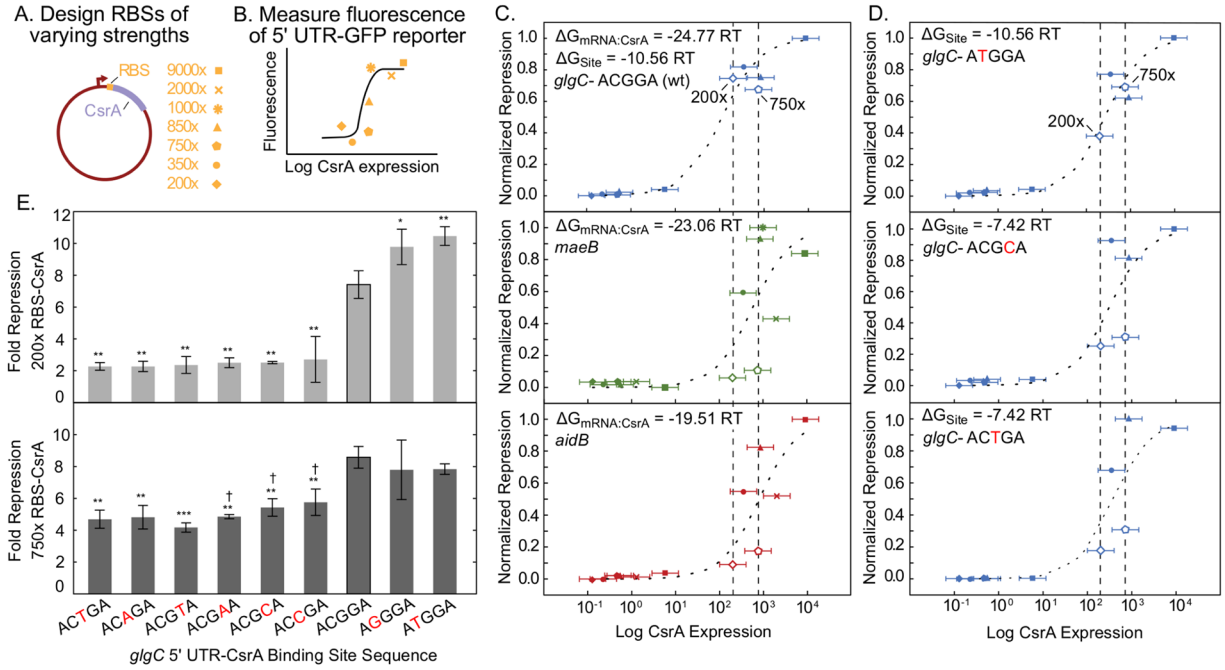

Figure 6. Relative *in vivo* CsrA-mRNA affinity correlates with calculated free energies. (**A**) Using the RBS calculator, multiple RBSs were designed to vary CsrA expression. These RBS-CsrA constructs were paired with target 5′ UTR-GFP reporter plasmids to titrate expression of the reporter and (**B**) fluorescence was measured. (**C**) Titration of *glgC*, *maeB*, and *aidB* 5′ UTRs. Symbols used for each RBS construct are indicated in (**A**). Normalized repression (minimum and maximum repression ratios scaled linearly from 0 to 1) is plotted as function of CsrA expression, as determined by the RBS calculator, on a log scale. Dotted lines represent fit to titration curve (Methods). Induced 200x and 750x RBS-CsrA constructs (open diamond and pentagon symbols, respectively) offer best resolution of differences between 5′ UTRs. 1000x and 2000x RBS constructs produced inconsistent results and were excluded from analysis in some cases. (**D**) Titration, as described for (**C**), of mutated *glgC* sequences. Induced 200x and 750x RBS-CsrA constructs offer best resolution of mutations' effects. (**E**) The fluorescent translational reporter assay performed with the 200x (upper panel) and 750x (lower panel) RBS-CsrA constructs for the wild type (ACGGA, black outlined bar) and eight mutant *glgC* 5′ UTR-GFP reporters. Asterisks indicate results of heteroscedastic one-tailed T-tests, comparing average fold repression (Uninduced/Induced) of mutants to wild type *glgC*: *P-value < 0.05; **P-value < 0.01; ***P-value < 0.001. (lower panel) Daggers indicate results of heteroscedastic one-tailed T-tests, comparing average fold repression (Uninduced/Induced) of ACTGA, ACAGA, ACGAA, ACGCA and ACCGA constructs to the ACGTA mutant: †P-value < 0.05.

repression predictions provided by the canonical biophysical model is useful for studying mRNAs expected to be canonically regulated by CsrA and for identifying those that could be regulated by alternative mechanisms.

Validation of Model Binding Site Predictions and Free Energy Calculations. *Model-identified Binding Sites Align with Known CsrA-mRNA Footprints.* After establishing that the model correctly predicted CsrA regulatory effects on 66 mRNAs, we assessed if the model captured specific experimentally-determined CsrA-mRNA binding sites. We aimed to validate model predictions by comparing model-identified binding sites with experimentally-determined CsrA footprints in eight classical CsrA targets: *glgC*, *sdiA*, *flhD*, *cstA*, *hfq*, *csrA*, *nhaR*, and *pgaA* (Figs 4 and 5). For most of the classical mRNA targets, the model-identified binding sites of the 15 lowest-energy CsrA-bound conformations align well with at least one of the known footprints (Figs 4A–E and 5A). For this subset of classical targets (Figs 4A–E and 5A), the model identifies binding sites within one of its known footprints at least 9 times among the top 15 ensemble conformations. It is important to note that (i) the same binding site can be identified multiple times, as long as it is paired with a second site that makes the binding site *pair* unique. Additionally, it is noteworthy that (ii) binding sites identified multiple times within the 15 most-likely CsrA-bound conformations of an mRNA adhere more strictly to the ANGGA consensus sequence. Model-predicted binding sites in the *nhaR* and *pgaA* mRNAs (Fig. 5B,C) present exceptions.

Given that overlapping binding sites are commonly identified by the model in the top fifteen CsrA-bound conformations of the classical mRNA targets, we grouped these identified neighboring binding sites into "pockets". A pocket is defined as a set of contiguous predicted binding sites that overlap, are adjacent, or have, at maximum, one nucleotide between them. Importantly, analysis of predicted binding site pockets within the classical CsrA-mRNA targets indicated that the model can capture CsrA-mRNA interactions of a wide range of *in vivo* affinities. For example, just upstream of the *glgC* Shine-Dalgarno sequence, binding site pocket 2, which contains 7 overlapping binding sites, aligns with a weaker, non-consensus sequence CsrA-mRNA binding site[17] (FP 3 in

5′ UTR	$\Delta G_{mRNA:CsrA}$ (RT) (most likely conformation)	Ensemble Average $\Delta G_{mRNA:CsrA}$ (RT)
glgC	−24.77	−21.38
maeB	−23.06	−19.35
aidB	−19.51	−16.66

Table 3. Model-calculated free energies of CsrA-mRNA binding, where the ensemble includes the 15 most likely CsrA-bound mRNA conformations.

	Total	Large Structural Change ($\Delta G \geq 3$ RT)	Strong Inter-site Cooperativity ($\Delta G \leq -4$ RT)
All Conformations	990	258	518
Conformations with 2 High Affinity Sites	75	49[a]	25[b]
Conformations with 1 High Affinity Site	539	151[c]	286[d]

Table 4. Frequency of types of CsrA-mRNA binding predicted across ensembles. [a]Hypergeometric test relative to condition in all conformations. P-value = 8 E-14, significantly enriched (P-value < 0.01); [b]Hypergeometric test P-value = 3 E-4, significantly depleted (P-value < 0.01); [c]Hypergeometric test P-value = 2 E-2, not significant (P-value ≥ 0.01); [d]Hypergeometric test P-value = 4 E-2, not significant (P-value ≥ 0.01).

Fig. 4A). Similarly, we suspect the presence of weaker, non-consensus sequence binding sites within pocket 4 of *sdiA* (Fig. 4B) and pocket 1 of *hfq* (Fig. 4E), given that these pockets contain 8–9 overlapping non-consensus binding sites. Earlier results from published *in vitro* experiments support this possibility. First, G residues in pocket 4 of *sdiA* may be weakly protected, based on footprint analysis[43]; second, CsrA may bind in pocket 1 of *hfq*, based on toeprint analysis[23]. In any of these cases, consideration of individual binding sites would not have substantially indicated a preference for CsrA to bind in the region. These results indicate that collective analysis of overlapping binding sites as pockets can suggest mRNA regions that bind CsrA. Biophysically, this represents a different concept in that CsrA-mRNA binding may not be strictly defined by rigid consensus sequences, as mostly used in the literature[40,41]. Instead, by capturing binding sites of differential affinities, our model suggests that CsrA binding may be influenced by extended sequences with weak binding affinity (as calculated by the FEM) that *collectively* contribute to CsrA-mRNA binding. In this way, numerous weak, non-consensus sites in close proximity may enhance the strength of the CsrA-RNA interaction at that region by allowing CsrA many opportunities to bind and slide along the stretch of RNA.

Calculated ΔG of CsrA-mRNA Binding is Indicative of Relative Affinities Observed In Vivo. Physiological relevance of the model's calculated free energy values was assessed by a series of *in vivo* fluorescence assays. As a proxy for measuring CsrA-mRNA binding affinity, we measured extent of CsrA-mRNA regulation using a variation of the fluorescent translational reporter assay. Here, fluorescent expression of a 5′ UTR-GFP reporter is measured in the presence of varying CsrA expression levels, induced by a series of plasmids with varying RBS strength (Supplementary Methods) (Fig. 6A,B). Testing three representative targets, *glgC*, *aidB*, and *maeB*, reveals that CsrA shows increased regulation of *glgC*, relative to *maeB* and *aidB* (Fig. 6C). This is in agreement with the model's predicted relative binding affinities: *glgC* has the lowest $\Delta G_{CsrA:mRNA}$ as compared to the other targets (Table 3). This variation of the fluorescent translational reporter assay was also used to test a versions of the *glgC* 5′ UTR with single nucleotide mutations to a footprinted binding site (Supplementary Methods) (Fig. 6D,E). Results are in general agreement with the FEM used in ΔG_{site} calculations as well as conclusions drawn in previous works[18]. Most importantly, these results collectively support the relevance of predicted free energies, both $\Delta G_{mRNA:CsrA}$ and ΔG_{site}, to observable differences in *in vivo* CsrA-mRNA binding and regulation at consensus or near-consensus sites.

Predicted CsrA-Bound mRNA Conformations Highlight Molecular mRNA Features Expected to Contribute to Regulation. To obtain insights into predicted molecular features of CsrA-mRNA interactions at a large scale, we analyzed modeling results obtained for the pool of 66 targets well captured by the canonical biophysical model. Importantly, we make three major observations that provide insights as to how different molecular features of mRNAs impact CsrA binding and regulation. First, we observe the frequent identification of CsrA binding sites within the Shine-Dalgarno region of the mRNAs: 61% of the 66 mRNA targets contain at least one predicted CsrA binding site overlapping the Shine-Dalgarno region (5–15 nucleotides upstream of the start codon) in their most-likely CsrA-bound conformation (Supplementary Table S3). Notably, 24% of the 66 mRNAs contain a predicted binding site that overlaps its start codon by at least one nucleotide; 14% of the 66 mRNAs fit both criteria (i.e., paired sites that overlap both the Shine-Dalgarno and the start codon are predicted in the most-likely CsrA-bound conformation). The remaining 29% of the 66 mRNAs contain predicted binding sites in other portions of the 5′ UTR in their most-likely bound conformations. The high frequency of binding to the Sine-Dalgarno is expected, given the similarity of the CsrA consensus sequence to the Shine-Dalgarno consensus sequence. While start codons have been identified as a major CsrA binding location in *Salmonella*[3], they have only recently been implicated in a CsrA-mRNA interaction in *E. coli*[39].

Second, we observed that the model identifies two high affinity binding sites (ANGGN sequences) for binding of a CsrA homodimer in only 13 (20%) of the 66 mRNAs' most-likely CsrA-bound conformations. Instead, 38 members (58%) present a single high affinity binding site and 15 mRNAs (23%) present no high affinity binding sites (Supplementary Table S3). Given the importance placed on sequence of CsrA binding sites in the literature, the apparent preference towards a single high affinity binding site was unexpected. Counting the number of unique high affinity binding sites identified throughout mRNAs' ensembles of CsrA-bound conformations reveals that this pattern is not due to a lack of ANGGN sites. 24 (~63%) of the 38 mRNA with a single high affinity site in their most-likely conformation have at least two unique high affinity sites in the modeled sequence that could be identified as a pair (Supplementary Table S3). This suggests that CsrA dimers may preferentially recognize (but not always) two binding sites of varying affinity (i.e. one high, one low), even in the presence of multiple possible high affinity binding sites along the target.

A third observation we made is that 46 mRNAs (70% of the 66 well-captured mRNA) are predicted to have strong cooperativity ($\Delta G_{Cooperativity} \leq -4.0$ RT, <25 nt apart) between the identified binding sites in their most-likely CsrA-bound conformation, while only 16 (24%) are predicted to have large structural changes ($\Delta\Delta G_{Structure} \geq 3.0$ RT) (Supplementary Table S3). Considered alongside our second observation, these results suggest that many mRNAs display a single high affinity site in its most-likely CsrA-bound conformation to (i) minimize free energy changes from structural unfolding and (ii) maximize free energy changes from inter-site cooperativity. Table 4 quantifies this pattern for the 15 lowest-energy CsrA-bound ensemble conformations of each of the 66 well-captured mRNA. CsrA-bound mRNA conformations that present two high affinity ANGGN sequences are significantly depleted for strong inter-site cooperativity and enriched in large structural changes (P-value < 0.01 by hypergeometric test). It should be noted that seven mRNAs present most-likely CsrA-bound conformations that are exceptions to this pattern: *hfq*, *pgaA*, *relA*, *dgcZ*, *deoD*, *tnaA*, and *yfgM*. These targets exhibit two high affinity ANGGN sequences that are a short length apart (strong cooperativity) and require minimal unfolding of mRNA structure to be bound.

Patterns in Predicted Binding Site Pockets Suggest Flexibility in CsrA-mRNA Recognition. We next analyzed in detail the fifteen most-likely CsrA-bound conformations predicted for each of the 66 mRNAs captured by the canonical model. Specifically, we looked to determine mRNAs in which CsrA binding could be influenced by extended sequences of weak binding affinity that may collectively contribute to CsrA-mRNA binding as calculated for *glgC*, *sdiA*, and *hfq* above. This was executed by identifying and analyzing pockets of predicted binding sites in each mRNA (Supplementary Table S3) and mapping predicted free energy terms across their most-likely fifteen CsrA-bound conformations (Supplementary Fig. S2). We counted pockets with 7–10 predicted non-consensus binding sites (i.e., not ANGGN sequences) as potential extended regions of low affinity CsrA interactions given that this number of overlapping weak binding sites was predicted in *glgC*, *sdiA* and *hfq*. Importantly, 17 of the 66 mRNA well-captured by the canonical biophysical model (26%) contain a pocket of 7–10 predicted non-consensus binding sites; we propose that these extended regions of low affinity may collectively contribute to CsrA binding and regulation. Some mRNAs in this category, like *rseA*, *sucC* and *sdhA*, contain predicted extended sequences of weak CsrA binding affinity in addition to predicted high affinity binding sites (Fig. 7A–C). Overall, the analysis suggests that a model where CsrA does not always immediately recognize the highest affinity binding site(s) within an mRNA, but sometimes recognizes extended sequences of lower affinity that collectively contribute to binding and regulation, may be relevant to a portion of modeled targets beyond the examples of *glgC*, *sdiA*, and *hfq* (Fig. 4). We term this conceptual model as "flexible" CsrA-mRNA binding and propose that mRNAs exhibiting one or more low frequency pockets of binding sites may be regulated in part by such a mechanism.

This interpretation is especially interesting for the *clpB* mRNA, in which only extended sequences of weak CsrA binding affinity are predicted (i.e., no ANGGNs are identified) (Fig. 7D). Perhaps an ability of CsrA to recognize and bind multiple weak sites in the absence of a strong one may contribute to the observed scope and diversity among CsrA targets. These results suggest that predicted pockets of overlapping low affinity binding sites may be useful for identifying non-traditional CsrA-mRNA binding sites within mRNAs of known CsrA regulation that contribute to flexible binding and regulation.

Analysis of Remaining Modeled mRNA Targets. To elucidate potential molecular diversity within the CsrA targets modeled, we extended our binding site pocket analysis to the 41 mRNAs (out of 107 known to be regulated by CsrA) for which the canonical biophysical model did not correctly predict the mode of regulation. It was expected that predicted binding sites in these targets could still provide insight into CsrA-mRNA interactions because differences between model-predicted and experimental regulation could arise (i) within the translation initiation rate prediction step (as for *flhD*) or (ii) via CsrA regulating translation by destabilizing the transcript[20], a mechanism of CsrA regulation beyond the scope of the canonical biophysical model. Among their most-likely CsrA-bound conformations, the most notable difference between these 41 mRNAs and the 66 for which the model correctly predicted regulation is the relative depletion of pairs of high affinity ANGGN binding sites in the top conformations of the incorrectly predicted mRNAs. Additionally, 44% (18 of 41) of these mRNAs yield a pocket of 7–10 predicted, non-consensus CsrA binding sites, suggestive of an extended non-traditional binding site that may contribute to flexible binding and regulation (Supplementary Table S3 and Supplementary Fig. S2).

Lastly, we analyzed model results for 129 mRNAs that are suspected to be regulated by CsrA (Supplementary Table S3 and Supplementary Fig. S2). The majority of this pool (78 mRNAs) did not show regulation in our fluorescent translational reporter assays (presumed due to improper expression of GFP) while the others fluoresced, but an impact of CsrA on translation could not be determined (46 mRNAs) or were not tested (5 mRNAs).

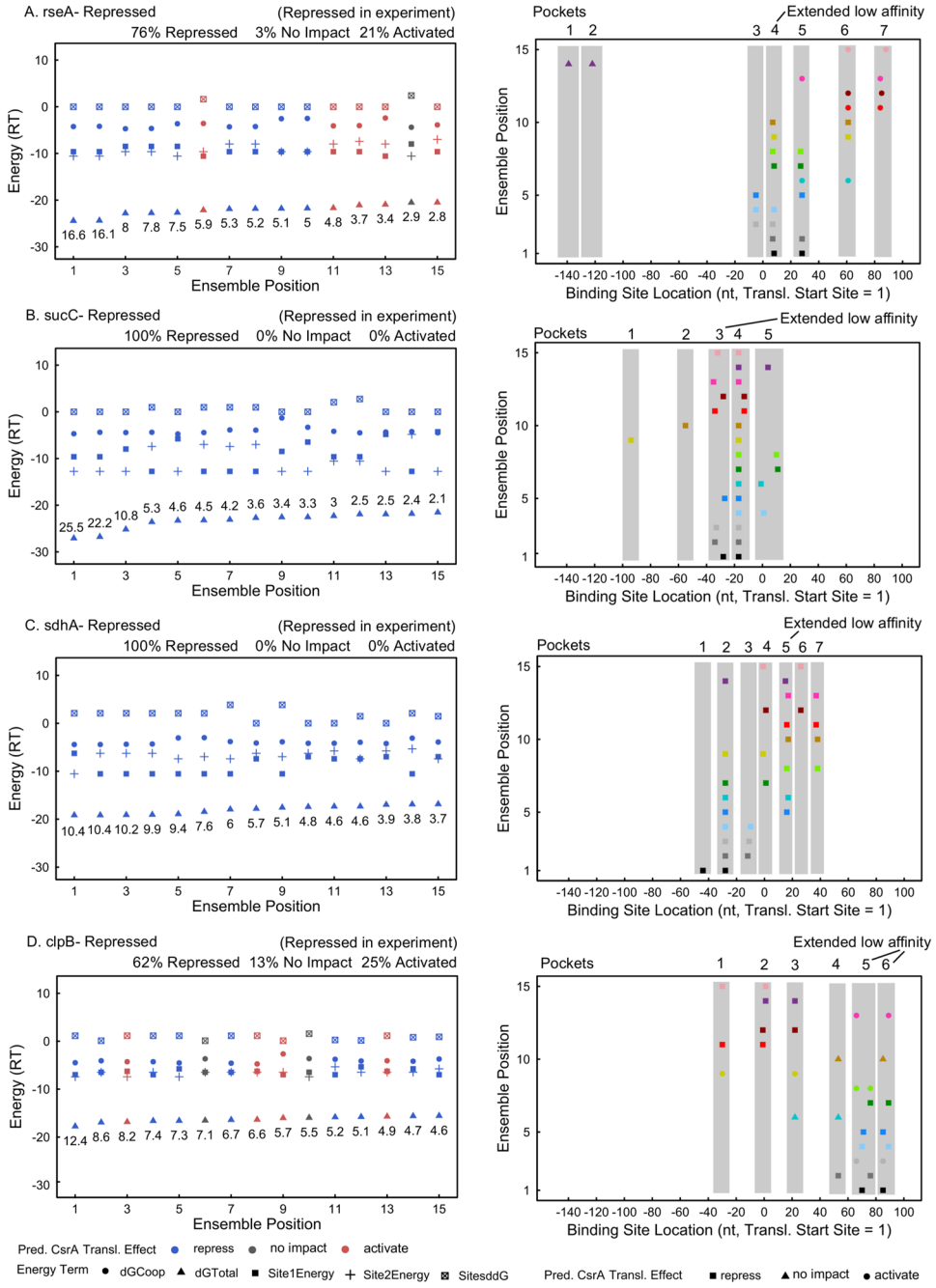

Figure 7. Patterns of binding site frequency within pockets suggest probable strong and weak *in vivo* CsrA-mRNA binding sites. (A, left panel) Free energy terms calculated for the most-likely fifteen members of the ensemble of CsrA-bound *rseA* conformations. Data labels mark the Boltzmann probability of each conformation, scaled such that the total probability of the fifteen most-likely conformations is one hundred percent. Total scaled Boltzmann probabilities of the repressed, not impacted, or activated conformations are noted, as well as the most-likely regulatory outcome as "rseA- Repressed". Regulation observed in fluorescent translational reporter assay is noted in parentheses as "Repressed in experiment". (A, right panel) Distribution of binding sites (by location) predicted in *rseA*. Gray panels highlight pockets of binding sites. Seven pockets are identified, one of which, pocket 4, contains 7–10 non-consensus low affinity predicted binding sites. (**B**) 5 pockets, one of extended low CsrA affinity are predicted for *sucC*. (**C**) 7 pockets, one of extended low CsrA affinity, are predicted for *sdhA*. (**D**) 6 pockets, two of extended low CsrA affinity are predicted for *clpB* in the absence of any high affinity binding sites.

Analysis of predicted binding site pockets in these 5′ UTRs (Supplementary Table S3) presents excellent starting points for designing mutations in likely binding sites for testing CsrA regulation (and elimination thereof) in alternative assays.

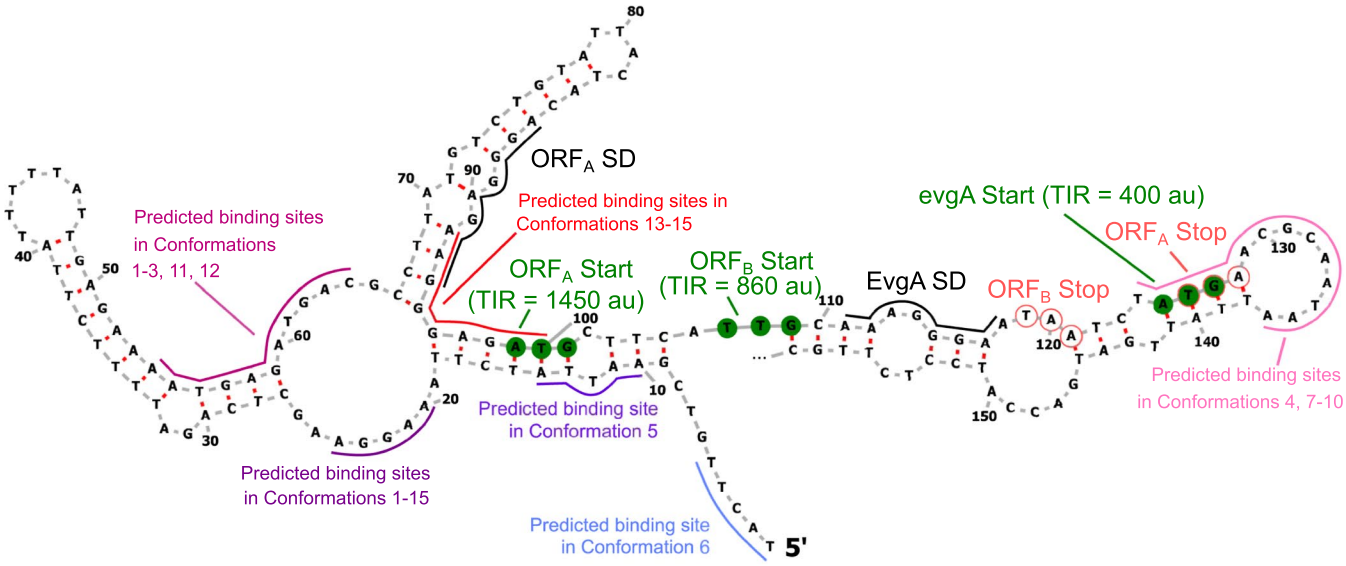

Figure 8. The *evgA* mRNA demonstrates potential co-translational repression mechanism. Two upstream open reading frames are present in the 5′ UTR of *evgA*, both of which are predicted to be translated more quickly than the *evgA* start codon by the RBS calculator (default parameters). The stop codon of ORF$_A$ overlaps with the start codon of *evgA*, a signature of co-translation. Notably, CsrA is predicted to bind the Shine-Dalgarno region of ORF$_A$, rather than the shine-Dalgarno of *evgA*. While these conformations are predicted to not impact or activate translation, CsrA binding here likely represses *evgA* co-translationally with ORF$_A$, a mechanism not incorporated into design of the biophysical model.

Discussion

In this work, we developed a biophysical model of canonical CsrA binding and regulation to investigate how the molecular features of a target mRNA can influence its regulation (Figs 1 and 2). We considered a set of 107 mRNAs known to be regulated by CsrA (Tables 1 and S1 and Fig. 3) and 129 mRNAs suspected to interact with CsrA (Supplementary Table S1). After establishing that the model correctly predicted the regulatory modes of 66 of 107 mRNA targets with known CsrA regulation (Table 2), we demonstrated the ability of the model to (i) correctly predict CsrA binding site locations across a range of binding affinities (Figs 4 and 5). Importantly, when examining ensembles of CsrA-mRNA complexes collectively, the model can differentiate between rigidly defined binding pockets with a single high affinity site and loosely defined low affinity binding pockets containing several overlapping sites (Fig. 7 and Supplementary Fig. S2). In Supplementary Table S3, we include all model calculations, including the locations of potential weak, non-consensus sequence CsrA-mRNA interactions across all 236 mRNAs considered in this study.

For a fraction of the 107 mRNA known to be regulated by CsrA, the model either does not clearly predict a single regulatory outcome (22 mRNAs) or clearly predicts a wrong regulatory outcome (19 mRNAs) across the most-likely fifteen members of their ensemble of CsrA-bound conformations (Table 1). Given that the biophysical model was built from our current understanding of CsrA-mRNA regulatory mechanisms, these CsrA targets offer opportunities to explore unusual or as-of-yet unknown CsrA regulatory mechanisms. Specifically, we were inspired by a recent study of the CsrA-*iraD* mRNA interaction[45] to look for mRNAs that could be repressed indirectly via a translational coupling mechanism. Here, CsrA indirectly represses *iraD* translation by binding and repressing translation of the *idlP* open reading frame, which is located in the same operon and within the 5′ UTR of *iraD* (as defined by *iraD's* annotated promoters). The canonical biophysical model does predict a CsrA binding site in the Shine-Dalgarno region of *idlP* (which also overlaps an *in vitro* CsrA footprint[45]), but it does not predict repression because translational coupling and thus potential co-regulation between coding sequences is not currently accounted for (Supplementary Table S3). The *evgA* mRNA, a well-characterized CsrA target not captured by the model, presents a similar case of potentially complex CsrA interaction and regulation that falls outside the canonical model. Employing the RBS Calculator indicates that the translation rates of two upstream start codons are substantial compared to the translation rate of the annotated *evgA* start codon (Fig. 8). Furthermore, the first upstream start codon reveals an open reading frame (ORF$_A$ in Fig. 8) that contains a stop codon overlapping with the annotated *evgA* start codon (5′-ATGA-3′). The biophysical model identifies a predicted binding site within the SD of ORF$_A$, but predicts this conformation to inconsistently impact *evgA* translation. Importantly, features exhibited by the *evgA* mRNA are commonly associated with the ribosome re-initiation mode of translational coupling[49], a mechanistic feature not currently accounted for in the biophysical model. Biophysical models of translational coupling have previously been developed[49], and could be leveraged in the future to further extend and improve the model predictions here. We anticipate that further investigation of all 41 mRNAs whose regulation is not well-predicted by the canonical biophysical model will uncover similar insights into poorly understood CsrA-mRNA regulatory mechanisms.

More broadly, our results suggest that CsrA participates in more flexible binding interactions than previously discussed in the literature. The presence of many low affinity CsrA binding sites nearby a single high affinity site suggests that CsrA may be initially recruited to mRNAs at one of the low affinity sites, followed by sliding into a high affinity site, which often appear within Shine-Dalgarno sequences with a clear regulatory outcome. Analogously, transcription factors have been proposed to hop or slide across low-specificity DNA sites to accelerate their search for high affinity sites[50,51]. Similarly, RNA-binding proteins like CsrA may utilize similar mechanisms of facilitated diffusion to accelerate their search for consensus, high affinity sites, which become all the more important as mRNAs are more well-mixed inside the cytoplasm, compared to genomic DNA.

This perspective adds an interesting biophysical possibility for mRNA in which CsrA-bound conformations in the ensemble lead to different types of regulation, where multiple modes of binding (using different binding site combinations) could lead to opposing regulatory outcomes. In this sense, it is possible that the specific CsrA-mRNA binding modes that take place *in vivo* are biased by additional factors, such as other regulators, small molecules, or environmental stresses, shifting the preference of one regulatory outcome or the other. For instance, CsrA-bound conformations 1–5 and 7–10 identified in the *rseA* mRNA are predicted to repress translation, while CsrA-bound conformations 6, 11–13, and 15 are predicted to activate translation (Fig. 7A). Although the conformations that lead to activation represent an energetic minority (21%) considering only CsrA-mRNA interactions, it is possible that the binding of other RNA-binding proteins or small molecules may increase the chance that CsrA visits these conformations, activating translation of *rseA*. Dual regulatory modes may also exist for the mRNAs *iscR*, *nnr*, and *dkgA* (Supplementary Fig. S2). It is therefore conceivable that translation of such mRNAs could be controlled by several factors hierarchically; in the absence of the external factor, CsrA may repress the mRNA's translation, but in its presence, the mRNA's translation may be activated.

Future developments in understanding CsrA binding across its broad target pool will continue to provide other valuable insights about CsrA regulation. One important factor to consider about the model described in this work is that it was derived from thermodynamic first principles and parameterization using only canonical CsrA binding sites[17,18,52]. The fact that the model has accuracy with only a minimal amount of input suggests that similar models of regulators could be generated and analyzed in a comparable manner. We envision that the model principles described here may be valuable for other systems of protein-based mRNA translation control, such as the trp RNA-binding attenuation protein (TRAP) of *B. subtilis*[53,54]. More broadly, the concept of thermodynamic models elucidating affinity-based target hierarchies is readily extendable to other types of regulators, including those involved in transcriptional control where target hierarchies have already been experimentally implicated[55–57]. Overall, we anticipate that this work will provide a generalizable strategy as a starting point to model post-transcriptional regulation in a variety of contexts.

Materials and Methods

Development of Model. *Identifying Potential CsrA Binding Sites.* Given an arbitrary mRNA sequence, the model first calculates CsrA's binding free energy to all possible binding sites. To do this, we assign a thermodynamic binding free energy to each nucleotide within a 5-nucleotide binding site, determined by converting the EMSA-determined binding affinities for mutated CsrA binding sites into free energy changes[18]. The resulting position-specific mono-nucleotide free energy model (FEM) is:

$$\begin{array}{r} \\ N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{A}\ (\Delta G_i\ in\ RT) & \boldsymbol{C}\ (\Delta G_i\ in\ RT) & \boldsymbol{G}\ (\Delta G_i\ in\ RT) & \boldsymbol{U}\ (\Delta G_i\ in\ RT) \end{array} \\ \left[\begin{array}{cccc} -2.63 & 0 & 0 & 0 \\ -2.20 & 0 & 0 & 0 \\ 0 & 0 & -3.14 & 0 \\ 0 & 0 & -3.14 & 0 \\ -1.65 & 0 & 0 & 0 \end{array}\right] \end{array} \quad \textit{Nucleotide Position}.$$

All energy units are RT, which is the gas constant multiplied by system temperature. The binding free energy of CsrA to a 5-nucleotide site is then calculated by summing the contributions:

$$\Delta G_{Site} = \sum_{i=N_1}^{N_5} \Delta G_i \tag{2}$$

Position score matrices have been used previously for the identification of CsrA targets[23,43,58,59]. Here, we quantify changes in binding occupancy in terms of free energies to develop a more comprehensive multi-term free energy model that incorporates CsrA binding, inter-site cooperativity, mRNA structural changes, and regulation of translation rate.

Calculating Cooperative Binding Effects. The model next quantifies the cooperative effects of a CsrA homodimer binding to two sites in an mRNA. Previous studies have indicated that the dimeric CsrA protein will bind to two adjacent sites in an mRNA, separated by 10 to 46 nucleotides[17,19], with the potential for cooperative binding[19–22]. We quantify the energy of the cooperative effect of a CsrA homodimer binding two adjacent sites, denoted by $\Delta G_{cooperativity}$, in terms of the inter-site distance d, according to the following formulas:

$$\text{For } 10 \le d \le 55,\ \ \Delta G_{cooperativity} = 0.001(d-5)^2 + 0.05(d-5) - 5.0 \text{ RT units} \tag{3}$$

$$\text{For } d < 10,\ \ \text{Steric penalty of } \Delta G_{cooperativity} = 20 \text{ RT units} \tag{4}$$

Here, if the inter-site distance is too small, then two adjacent sites will sterically clash, resulting in a large, repulsive free energy contribution. If the inter-site distance is 10 nucleotides, the cooperative interaction is strongest ($\Delta G_{cooperativity} = -4.725$ RT). The strength of the cooperative interaction is then proposed to decrease as the inter-site distance increases, reaching zero at 55 nucleotides. These formula fit previous observations of cooperative CsrA binding across relatively long inter-site distances[17].

Calculating CsrA-induced mRNA Structural Changes. When CsrA binds to an mRNA, it has the potential to induce changes in the mRNA's structure. When these changes take place nearby the start codon of a protein coding sequence, it can alter the ribosome's ability to bind the mRNA and initiate translation, changing the mRNA's translation rate. To determine where and when this takes place, we determine the free energy needed to remodel the mRNA's structure by calculating the difference in mRNA folding free energy with and without accessible CsrA binding sites ($\Delta G_{\S}$ versus $\Delta G_{unbound\ mRNA}$). We perform these calculations for each unique pair of the identified CsrA binding sites. For a pair of sites, we first calculate the most stable mRNA structure in the absence of CsrA, using ViennaRNA's dynamic programming algorithm[48] and the Turner 1999 RNA free energy parameters[60]. The default temperature of 37 °C is used. Dangling nucleotide energies are not used. We then repeat the same calculation with the addition of a structural constraint that prevents the two 5-nucleotide CsrA binding sites from participating in any mRNA structure. The CsrA binding sites in the resulting mRNA structure must be single-stranded, but they can be located anywhere within the mRNA, including within a hairpin loop. We then take the difference between these two folding free energies to determine the amount of energy needed to induce this structural change, which must be positive:

$$\Delta\Delta G_{unfold} = \Delta G_{\S} - \Delta G_{unbound\ mRNA} \tag{5}$$

An Overall Free Energy Model for CsrA-mRNA Interactions. Given an arbitrary mRNA sequence, we enumerate all possible CsrA binding sites, evaluate each free energy term, and sum them together to calculate the total free energy change when a CsrA homodimer binds the mRNA at two sites, according to the following formula:

$$\Delta G_{mRNA:CsrA} = \Delta G_{site1} + \Delta G_{site2} + \Delta G_{cooperativity} + \Delta\Delta G_{unfold}$$

where ΔG_{site1} is the binding free energy of a CsrA to an upstream site (negative), ΔG_{site2} is the binding free energy of a CsrA to a downstream site (negative), $\Delta G_{cooperativity}$ is the cooperative free energy whenever CsrA is bound to two sites at most 55 nucleotides apart (positive or negative), and $\Delta\Delta G_{unfold}$ is the free energy needed to unfold one or both CsrA binding sites (positive). Together, the total binding free energy ($\Delta G_{mRNA:CsrA}$) for a particular pair of sites may be positive or negative. For the mRNAs in this study, each has between 1800 to 14,000 possible pairs of CsrA binding sites, comprising an ensemble of possible CsrA-bound mRNA states with varying binding free energies. Formally, the probability of CsrA binding to one of these CsrA binding site conformations can be expressed in terms in a Boltzmann distribution:

$$p(\alpha) = \frac{e^{-\beta\Delta G_{mRNA:CsrA_\alpha}}}{\sum_i^M e^{-\beta\Delta G_{mRNA:CsrA_i}}}, \tag{6}$$

where α is a particular CsrA-bound mRNA conformation and $\Delta G_{CsrA:mRNA,\alpha}$ is the total free energy change in that conformation. M is the total number of possible CsrA-bound mRNA conformations for a particular mRNA sequence.

Predicting CsrA-induced Translation Regulation. When CsrA binds an mRNA and induces structural changes, the mRNA's translation initiation rate can increase *or* decrease, depending on where the CsrA binding site is located and the structures that form after CsrA is bound. To predict these effects, we calculate the ribosome's binding free energy to mRNAs that are either unbound or bound twice by a CsrA protein, employing a previously developed free energy model, called the RBS Calculator[34,35]. For a free mRNA without CsrA bound, the RBS Calculator's free energy model has the following five terms:

$$\Delta G_{total,free} = (\Delta G_{mRNA:rRNA} + \Delta G_{start} + \Delta G_{spacing} + \Delta G_{standby}) - \Delta G_{unboundmRNA} \tag{7}$$

where the first four terms quantify the free energy of the ribosome-mRNA complex in its final state, and the last term quantifies the folding free energy of the mRNA by itself in the initial state. The $\Delta G_{total,free}$ calculation is performed by identifying the most stable initial and final states with their lowest respective free energies, and then computing their difference in free energy. The mRNA's translation initiation rate r is then determined according to:

$$r_{free} \propto \exp(-\beta\Delta G_{total,free}) \tag{8}$$

where β is an empirically measured Boltzmann constant for *E. coli* ($\beta = 0.45$ mol/kcal)[37,61].

When a CsrA protein binds to an mRNA, the $\Delta G_{mRNA:rRNA}$ and $\Delta G_{standby}$ terms will change whenever there is a CsrA-induced mRNA structural change that overlaps with an upstream standby site or the ribosome's footprint, which includes the Shine-Dalgarno sequence, the start codon, and the first 13 nucleotides of the protein coding sequence[62]. For any change in mRNA structure, we recalculate these free energy model terms to determine the change in the ribosome's binding free energy, $\Delta G_{total,CsrA}$, and the corresponding change in the mRNA's translation rate, r_{CsrA}, using the free energy model:

$$\Delta G_{total,CsrA} = (\Delta G_{mRNA:CsrA:rRNA} + \Delta G_{start} + \Delta G_{spacing} + \Delta G_{standby:CsrA}) - \Delta G_{mRNA:CsrA} \quad (9)$$

Calculating the free energy terms $\Delta G_{mRNA:CsrA:rRNA}$ and $\Delta G_{standby:CsrA}$ follows the same procedure as for a free mRNA, except with two structural constraints: first, the mRNA region overlapping with the ribosomal footprint may neither have a bound CsrA nor a folded inhibitory mRNA structure, and second, a CsrA-bound site outside the ribosomal footprint must be single-stranded. In this way, if CsrA is predicted to bind to a site overlapping the ribosomal footprint, the energy required to remove bound CsrA is included in $\Delta G_{mRNA:CsrA:rRNA}$. Calculating $\Delta G_{mRNA:CsrA}$ is as described in the previous section. Importantly, an mRNA will often contain thousands of potential pairs of sites where CsrA can bind; each CsrA-bound mRNA conformation can have different $\Delta G_{mRNA:CsrA}$ free energies. In practice, we consider the most probable 15 double-site CsrA-bound mRNA conformations, i.e., those with the most negative $\Delta G_{CsrA:mRNA}$ free energies, and apply the RBS Calculator's free energy model to calculate their respective translation initiation rates. For each of these i^{th} conformations, we calculate their translation initiation rates using:

$$r_{CsrA,i} \propto \exp(-\beta \Delta G_{total,CsrA,i}) \quad (10)$$

Finally, for each i^{th} CsrA-bound mRNA conformation, we determine whether the mRNA's translation rate was activated or repressed by calculating the ratio:

$$R_i = r_{free}/r_{CsrA,i} \quad (11)$$

Metric for evaluating predictions of CsrA targets. We label each CsrA-bound mRNA conformation's translation rate as "repressed" if $R_i > 1.2$, as "activated" if $R_i < 0.8$, and as "not impacted" otherwise. We then determine the most likely mode of translation regulation by comparing the 15 conformations' calculated $\Delta G_{CsrA:mRNA}$ free energies and ensuring that the most frequent states, energetically weighted, all had the same regulatory label. If predictions yielded CsrA-bound conformations with different regulatory labels with approximately equal frequency, then the mRNA was labeled as having a "heterogeneous" mode of translation regulation. Specifically, we summed the probabilities ($p(\alpha)$, defined above) of all of the repressed, activated or not impacted CsrA-bound conformations of an mRNA; the most likely mode of regulation was taken to represent the ensemble only if it was at least 1.8-fold more likely than the other two possible types of regulation (threshold determined from inspection of classical CsrA target model results). This qualitative analysis is valid across a wide range of CsrA concentrations inside the cell. Using an ensemble-based collective regulation prediction allows us to account more thoroughly for mRNAs with multiple pairs of equally likely CsrA binding sites. However, a higher-energy conformation within an mRNA's most likely 15 CsrA-bound conformations may only be occupied in *in vivo* under conditions of excess CsrA and after the more likely bind site pairs are bound. By weighting regulation predictions with Boltzmann distribution-derived probabilities we aim to account for this discrepancy while still incorporating the diversity of the binding sites predicted by the model within a given mRNA.

Fluorescent Translational Reporter Assay. Fluorescent translational reporter assays were performed as described in the Supplementary Methods.

Titration of CsrA targets. To test whether $\Delta G_{CsrA:mRNA}$ was indicative of the thermodynamics of CsrA binding to transcripts *in vivo*, we constructed variants of pHL 600 that produce a range of CsrA expression (Supplementary Tables S5 and S6). Specifically, the RBS calculator[35] was used to design RBSs with an anticipated 50-fold range of translation rates, including variants that were predicted to yield CsrA expression with putative levels of translation, as calculated by the RBS Calculator, of 2000, 1000, 800, 750, 350, and 200 AU (original plasmid, used in the fluorescent translational reporter assay, was 9000 AU). In order to estimate the level of CsrA produced from each plasmid in the uninduced condition, i.e. from promoter "leakiness", the putative translation level of the RBS was multiplied by 0.000645, a value of promoter expression in the undinduced condition determined in a prior work[63]. Additionally, the putative CsrA translation levels are expected to have error between 50% and 200%[37].

We then paired this set of seven RBS-CsrA variants with three 5′ UTRs-GFP reporters (*glgC*, *aidB*, and *maeB*), each with 5′ UTRs predicted to (1) be repressed by CsrA, (2) have a range of $\Delta G_{CsrA:mRNA}$ values (*glgC*, −24.77 RT units; *aidB*, −19.53 RT units; and *maeB*, −23.06 RT units), and (3) contain one identified binding site of the most likely CsrA-bound conformation in the Shine-Dalgarno region. These 5′ UTRs were cloned into pHL 1756 plasmids previously (Supplementary Table S5). Eight mutations to the *glgC* 5′ UTR-GFP reporter plasmid were also constructed via site directed mutagenesis (Quickchange II protocol for Agilent Technologies) to vary the middle three nucleotides of the footprinted non-Shine Dalgarno CsrA binding site (wild type sequence: ACGGA) (Supplementary Tables S5 and S6). In this way, the mutated sequences created a range of ΔG_{Site} values. Pairing a single 5′ UTR-GFP reporter with the set of RBS-CsrA variants at a time, we took fluorescence measurements in biological triplicates as described in the fluorescent translational reporter assay section above. However, analysis of titration experiment data differs; in these tests, rather than comparing corresponding uninduced and induced conditions, a repression ratio due to CsrA was calculated for each RBS in each condition, with respect to the maximum average mean fluorescence of all of the samples:

$$Repression\ Ratio = \frac{Maximum\ of\ all\ average\ mean\ sample\ fluorescences}{Average\ mean\ sample\ fluorescence}. \quad (12)$$

For example, in the case of *glgC*, the maximum of all average mean sample fluorescences corresponded to the average mean fluorescence of the weakest RBS-CsrA variant (200x) in the uninduced condition. In this way, 14 repression ratios, each corresponding to a different level of CsrA expression (uninduced 200x, induced 200x,

uninduced 350x, induced 350x, etc.), are determined with the minimum repression ratio always equal to 1. To compare results of the titrated 5′ UTR-GFP reporters, the repression ratios measured for each were scaled from 0 to 1, where 0 represents the minimum repression ratio and 1 the maximum repression ratio:

$$Normalized\ Repression\ Ratio = \frac{(Repression\ Ratio - minimum\ Repression\ Ratio)}{(maximum\ Repression\ Ratio - minimum\ Repression\ Ratio)}. \quad (13)$$

Lastly, dissociation constants (K_d) for the *glgC*, *maeB*, and *aidB* titration datasets were approximated as a means to generate fits for their titration curves[64]. Assuming a near constant pool of each 5′ UTR-GFP reporter transcript, K_d can be approximated with,

$$\theta = \frac{[CsrA]}{K_d + [CsrA]}, \quad (14)$$

where θ indicates the fractional extent of CsrA-5′ UTR binding (i.e. normalized repression ratio) and [CsrA] indicates CsrA expression levels (i.e. putative RBS translation levels). The R Package for Statistical Computing non-linear least squares equation fitting function, nls, was employed to fit each titration dataset with the above equation and approximate a K_d for the CsrA-5′ UTR interaction.

Data availability. All data generated or analyzed during the current study are included in this published article (and its Supplementary Information files). Fluorescent translational reporter assay results analyzed in this work but published previously are available as Supplementary Information files at https://doi.org/10.1093/nar/gkx048. Original scripts are available from the corresponding author upon request.

References

1. Sittka, A. *et al*. Deep Sequencing Analysis of Small Noncoding RNA and mRNA Targets of the Global Post-Transcriptional Regulator, Hfq. *Plos Genetics* **4**, e1000163, https://doi.org/10.1371/journal.pgen.1000163 (2008).
2. Edwards, A. N. *et al*. Circuitry linking the Csr and stringent response global regulatory systems. *Mol. Microbiol.* **80**, 1561–1580, https://doi.org/10.1111/j.1365-2958.2011.07663.x (2011).
3. Holmqvist, E. *et al*. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*. *The EMBO Journal*, 991–1011, https://doi.org/10.15252/embj.201593360 (2016).
4. Esquerré, T. *et al*. The Csr system regulates genome-wide mRNA stability and transcription and thus gene expression in Escherichia coli. *Sci. Rep.* **6**, 25057, https://doi.org/10.1038/srep25057 (2016).
5. Sowa, S. W. *et al*. Integrative FourD omics approach profiles the target network of the carbon storage regulatory system. *Nucleic Acids Res.* **45**, 1673–1686, https://doi.org/10.1093/nar/gkx048 (2017).
6. Potts, A. H. *et al*. Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics. *Nature Communications* **8**, 1596, https://doi.org/10.1038/s41467-017-01613-1 (2017).
7. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628, https://doi.org/10.1016/j.cell.2009.01.043 (2009).
8. Soper, T., Mandin, P., Majdalani, N., Gottesman, S. & Woodson, S. A. Positive regulation by small RNAs and the role of Hfq. *Proceedings of the National Academy of Sciences* **107**, 9602–9607, https://doi.org/10.1073/pnas.1004435107 (2010).
9. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C. & Storz, G. The Sm-like Hfq Protein Increases OxyS RNA Interaction with Target mRNAs. *Molecular Cell* **9**, 11–22, https://doi.org/10.1016/S1097-2765(01)00437-3 (2002).
10. Otaka, H., Ishikawa, H., Morita, T. & Aiba, H. PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proceedings of the National Academy of Sciences* **108**, 13059–13064, https://doi.org/10.1073/pnas.1107050108 (2011).
11. Ishikawa, H., Otaka, H., Maki, K., Morita, T. & Aiba, H. The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3′ poly(U) tail. *RNA (New York, N.Y.)* **18**, 1062–1074, https://doi.org/10.1261/rna.031575.111 (2012).
12. Mikulecky, P. J. *et al*. Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nature Structural &Amp; Molecular Biology* **11**, 1206, https://doi.org/10.1038/nsmb858 (2004).
13. Link, T. M., Valentin-Hansen, P. & Brennan, R. G. Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proceedings of the National Academy of Sciences* **106**, 19292–19297, https://doi.org/10.1073/pnas.0908744106 (2009).
14. Robinson, K. E., Orans, J., Kovach, A. R., Link, T. M. & Brennan, R. G. Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Research* **42**, 2736–2749, https://doi.org/10.1093/nar/gkt1171 (2014).
15. Schu, D. J., Zhang, A., Gottesman, S. & Storz, G. Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *The EMBO Journal* **34**, 2557–2573, https://doi.org/10.15252/embj.201591569 (2015).
16. Małecka, E. M., Stróżecka, J., Sobańska, D. & Olejniczak, M. Structure of Bacterial Regulatory RNAs Determines Their Performance in Competition for the Chaperone Protein Hfq. *Biochemistry* **54**, 1157–1170, https://doi.org/10.1021/bi500741d (2015).
17. Mercante, J., Edwards, A. N., Dubey, A. K., Babitzke, P. & Romeo, T. Molecular Geometry of CsrA (RsmA) Binding to RNA and Its Implications for Regulated Expression. *Journal of Molecular Biology* **392**, 511–528, https://doi.org/10.1016/j.jmb.2009.07.034 (2009).
18. Dubey, A. K., Baker, C. S., Romeo, T. & Babitzke, P. RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA (New York, N.Y.)* **11**, 1579–1587, https://doi.org/10.1261/rna.2990205 (2005).
19. Schubert, M. *et al*. Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat Struct Mol Biol* **14**, 807–813, https://doi.org/10.1038/nsmb1285 (2007).
20. Wang, X. *et al*. CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of Escherichia coli. *Mol. Microbiol.* **56**, 1648–1663, https://doi.org/10.1111/j.1365-2958.2005.04648.x (2005).
21. Dubey, A. K. *et al*. CsrA regulates translation of the Escherichia coli carbon starvation gene, cstA, by blocking ribosome access to the cstA transcript. *J. Bacteriol.* **185**, 4450–4460 (2003).
22. Weilbacher, T. *et al*. A novel sRNA component of the carbon storage regulatory system of Escherichia coli. *Mol. Microbiol.* **48**, 657–670 (2003).
23. Baker, C. S. *et al*. CsrA inhibits translation initiation of Escherichia coli hfq by binding to a single site overlapping the Shine-Dalgarno sequence. *J. Bacteriol.* **189**, 5472–5481, https://doi.org/10.1128/jb.00529-07 (2007).
24. Jonas, K. *et al*. The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins. *Mol. Microbiol.* **70**, 236–257, https://doi.org/10.1111/j.1365-2958.2008.06411.x (2008).
25. Baker, C. S., Morozov, I., Suzuki, K., Romeo, T. & Babitzke, P. CsrA regulates glycogen biosynthesis by preventing translation of glgC in Escherichia coli. *Mol. Microbiol.* **44**, 1599–1610 (2002).
26. Soper, T. J. & Woodson, S. A. The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA (New York, N.Y.)* **14**, 1907–1917, https://doi.org/10.1261/rna.1110608 (2008).

27. Gelderman, G., Sivakumar, A., Lipp, S. & Contreras, L. Adaptation of Tri-molecular fluorescence complementation allows assaying of regulatory Csr RNA-protein interactions in bacteria. *Biotechnol. Bioeng.* **112**, 365–375, https://doi.org/10.1002/bit.25351 (2015).
28. Smirnov, A. *et al.* Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proceedings of the National Academy of Sciences* **113**, 11591–11596, https://doi.org/10.1073/pnas.1609981113 (2016).
29. Wright, P. R. *et al.* CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Research* **42**, W119–W123, https://doi.org/10.1093/nar/gku359 (2014).
30. Brewster, R. C., Jones, D. L. & Phillips, R. Tuning Promoter Strength through RNA Polymerase Binding Site Design in *Escherichia coli*. *PLOS Computational Biology* **8**, e1002811, https://doi.org/10.1371/journal.pcbi.1002811 (2012).
31. Amman, F., Flamm, C. & Hofacker, I. Modelling Translation Initiation under the Influence of sRNA. *International journal of molecular sciences* **13**, 16223–16240, https://doi.org/10.3390/ijms131216223 (2012).
32. Rydenfelt, M., Cox, R. S., Garcia, H. & Phillips, R. Statistical mechanical model of coupled transcription from multiple promoters due to transcription factor titration. *Physical Review E* **89**, 012702 (2014).
33. Vazquez, J. *et al.* Optimization of a novel biophysical model using large scale *in vivo* antisense RNA hybridization data in bacteria displays improved prediction capabilities. *Nucleic Acids Res.* **45**, 5523–5538, https://doi.org/10.1093/nar/gkx115 (2017).
34. Salis, H. M. The ribosome binding site calculator. *Methods in enzymology* **498**, 19–42, https://doi.org/10.1016/b978-0-12-385120-8.00002-4 (2011).
35. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res* **42**, 2646–2659, https://doi.org/10.1093/nar/gkt1139 (2014).
36. Espah Borujeni, A. & Salis, H. M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *Journal of the American Chemical Society* **138**, 7016–7023, https://doi.org/10.1021/jacs.6b01453 (2016).
37. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated Design of Synthetic Ribosome Binding Sites to Precisely Control Protein Expression. *Nature biotechnology* **27**, 946–950, https://doi.org/10.1038/nbt.1568 (2009).
38. Auweter, S. D., Oberstrass, F. C. & Allain, F. H. T. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Research* **34**, 4943–4959, https://doi.org/10.1093/nar/gkl620 (2006).
39. Yakhnin, H., Aichele, R., Ades, S. E., Romeo, T. & Babitzke, P. Circuitry linking the global Csr and σE-dependent cell envelope stress response systems. *Journal of Bacteriology*. https://doi.org/10.1128/jb.00484-17 (2017).
40. McKee, A. *et al.* Manipulation of the carbon storage regulator system for metabolite remodeling and biofuel production in Escherichia coli. *Microb. Cell Fact.* **11**, 79 (2012).
41. Kulkarni, P. R. *et al.* A sequence-based approach for prediction of CsrA/RsmA targets in bacteria with experimental validation in Pseudomonas aeruginosa. *Nucleic Acids Res* **42**, 6811–6825, https://doi.org/10.1093/nar/gku309 (2014).
42. Pannuri, A. *et al.* Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA. *J. Bacteriol.* **194**, 79–89, https://doi.org/10.1128/jb.06209-11 (2012).
43. Yakhnin, H. *et al.* CsrA represses translation of sdiA, which encodes the N-acylhomoserine-L-lactone receptor of Escherichia coli, by binding exclusively within the coding region of sdiA mRNA. *J. Bacteriol.* **193**, 6162–6170, https://doi.org/10.1128/jb.05975-11 (2011).
44. Yakhnin, A. V. *et al.* CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. *Mol Microbiol* **87**, 851–866, https://doi.org/10.1111/mmi.12136 (2013).
45. Park, H. *et al.* Translational Repression of the RpoS Antiadapter IraD by CsrA Is Mediated via Translational Coupling to a Short Upstream Open Reading Frame. *mBio* **8** https://doi.org/10.1128/mBio.01355-17 (2017).
46. Wei, B. L. *et al.* Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of Escherichia coli. *Mol. Microbiol.* **40**, 245–256 (2001).
47. Yakhnin, H. *et al.* Complex regulation of the global regulatory gene csrA: CsrA-mediated translational repression, transcription from five promoters by Eσ(70) and Eσ(S), and indirect transcriptional activation by CsrA. *Molecular microbiology* **81**, 689–704, https://doi.org/10.1111/j.1365-2958.2011.07723.x (2011).
48. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for molecular biology: AMB* **6**, 26, https://doi.org/10.1186/1748-7188-6-26 (2011).
49. Tian, T. & Salis, H. M. A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. *Nucleic Acids Research* **43**, 7137–7151, https://doi.org/10.1093/nar/gkv635 (2015).
50. Wang, F. *et al.* The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. *Nature Structural &Amp; Molecular Biology* **20**, 174, https://doi.org/10.1038/nsmb.2472 (2012).
51. Redding, S. & Greene, E. C. How do proteins locate specific targets in DNA? *Chemical Physics Letters* **570**, 1–11, https://doi.org/10.1016/j.cplett.2013.03.035 (2013).
52. Duss, O., Michel, E., Diarra dit Konte, N., Schubert, M. & Allain, F. H. Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucleic Acids Res.* **42**, 5332–5346, https://doi.org/10.1093/nar/gku141 (2014).
53. Yakhnin, H., Zhang, H., Yakhnin, A. V. & Babitzke, P. The trp RNA-Binding Attenuation Protein of Bacillus subtilis Regulates Translation of the Tryptophan Transport Gene trpP (yhaG) by Blocking Ribosome Binding. *Journal of Bacteriology* **186**, 278–286, https://doi.org/10.1128/JB.186.2.278-286.2004 (2004).
54. Babitzke, P., Stults, J. T., Shire, S. J. & Yanofsky, C. 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. *Journal of Biological Chemistry* **269**, 16597–16604 (1994).
55. Durfee, T., Hansen, A.-M., Zhi, H., Blattner, F. R. & Jin, D. J. Transcription Profiling of the Stringent Response in Escherichia coli. *Journal of Bacteriology* **190**, 1084–1096, https://doi.org/10.1128/jb.01092-07 (2008).
56. Brinsmade, S. R. *et al.* Hierarchical expression of genes controlled by the Bacillus subtilis global regulatory protein CodY. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 8227–8232, https://doi.org/10.1073/pnas.1321308111 (2014).
57. Pi, H. & Helmann, J. D. Sequential induction of Fur-regulated genes in response to iron limitation in Bacillus subtilis. *Proceedings of the National Academy of Sciences* **114**, 12785–12790, https://doi.org/10.1073/pnas.1713008114 (2017).
58. Willias, S. P., Chauhan, S., Lo, C.-C., Chain, P. S. G. & Motin, V. L. CRP-Mediated Carbon Catabolite Regulation of *Yersinia pestis* Biofilm Formation Is Enhanced by the Carbon Storage Regulator Protein, CsrA. *Plos One* **10**, e0135481, https://doi.org/10.1371/journal.pone.0135481 (2015).
59. Sridhar, J., Sekar, K. & Rafi, Z. A. CsrA interacting small RNAs in Haemophilus spp genomes: a theoretical analysis. *Archives of Microbiology* **191**, 451–459, https://doi.org/10.1007/s00203-009-0474-y (2009).
60. Turner, D. H. & Mathews, D. H. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res* **38**, D280–282, https://doi.org/10.1093/nar/gkp892 (2010).
61. Hao, Y. *et al.* Quantifying the sequence–function relation in gene silencing by bacterial small RNAs. *Proceedings of the National Academy of Sciences* **108**, 12473–12478, https://doi.org/10.1073/pnas.1100432108 (2011).
62. Espah Borujeni, A. *et al.* Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Research* **45**, 5437–5448, https://doi.org/10.1093/nar/gkx061 (2017).
63. Adamson, D. N. & Lim, H. N. Rapid and robust signaling in the CsrA cascade via RNA–protein interactions and feedback regulation. *Proc. Natl. Acad. Sci. USA* **110**, 13120–13125, https://doi.org/10.1073/pnas.1308476110 (2013).
64. Beckett, D. Measurement and analysis of equilibrium binding titrations: A beginner's guide. *Methods in enzymology* **488**, 1–16, https://doi.org/10.1016/b978-0-12-381268-1.00001-x (2011).

Acknowledgements

We wish to thank Jordan K. Villa and Mia K. Mihailovic for helpful readings of the manuscript and Nicholas Curtis for help with RNA structure predictions. We also acknowledge the Welch Foundation (Grant F-1756 to L.M.C.), the Air Force Office of Scientific Research (Grant FA9550-16-1-0174 to L.M.C.) and the National Science Foundation (Grant MCB 1716777 to L.M.C. and DGE-1610403 to A.N.L.) for funding associated with the work at the University of Texas at Austin. For the work at Penn State University, we acknowledge and thank the following funding sources: Air Force Office of Scientific Research (Grant FA9550-14-1-0089 to H.M.S.), the Office of Naval Research (Grant N00014-13-1-0074 to H.M.S.), and an NSF Career Award (Grant CBET- 1253641 to H.M.S.). Additionally, S.S. thanks a Named Graduate Fellowship from the University of Texas at Austin and A.N.L. thanks the University of Texas at Austin Graduate School Provost and McKetta Department of Chemical Engineering Thrust fellowships for funding.

Author Contributions

G.G., L.M.C. and H.S. designed the study. H.S. developed the biophysical model. A.N.L., G.G., and S.S. analyzed results. G.G., A.N.L., A.M. and S.S. preformed fluorescence assays. A.N.L., S.S. and L.M.C. wrote the manuscript. H.S. and G.G. reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-27474-2.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018