CHAPTER TWO

The Ribosome Binding Site Calculator

Howard M. Salis[*,†]

Contents

Abstract

The Ribosome Binding Site (RBS) Calculator is a design method for predicting and controlling translation initiation and protein expression in bacteria. The method can predict the rate of translation initiation for every start codon

* Department of Chemical Engineering, Pennsylvania State University, University Park, Pennsylvania, USA
† Department of Agricultural and Biological Engineering, Pennsylvania State University, University Park, Pennsylvania, USA

Methods in Enzymology, Volume 498
ISSN 0076-6879, DOI: 10.1016/B978-0-12-385120-8.00002-4

in an mRNA transcript. The method may also optimize a synthetic RBS sequence to achieve a targeted translation initiation rate. Using the RBS Calculator, a protein coding sequence's translation rate may be rationally controlled across a 100,000+ fold range. We begin by providing an overview of the potential biotechnology applications of the RBS Calculator, including the optimization of synthetic metabolic pathways and genetic circuits. We then detail the definitions, methodologies, and algorithms behind the RBS Calculator's thermodynamic model and optimization method. Finally, we outline a protocol for precisely measuring steady-state fluorescent protein expression levels. These methods and protocols provide a clear explanation of the RBS Calculator and its uses.

1. Introduction

The Ribosome Binding Site (RBS) Calculator is a predictive design method for controlling translation initiation and protein expression in bacteria (Salis *et al.*, 2009). In its reverse engineering mode, the method predicts the rate of translation initiation for every start codon in an mRNA transcript. In its forward engineering mode, the method optimizes a synthetic RBS sequence to achieve a targeted translation initiation rate. Using the RBS Calculator, a protein coding sequence's (CDS) translation rate may be rationally controlled across a 100,000+ fold range.

The RBS Calculator employs a thermodynamic model of bacterial translation initiation to calculate the Gibbs free energy of ribosome binding. Using a statistical thermodynamic approach, we relate this Gibbs free energy change to a protein CDS's translation initiation rate. For forward engineering, the thermodynamic model is combined with a stochastic optimization method to design synthetic RBSs according to the desired specifications.

A Web interface to the RBS Calculator is located at http://salis.psu.edu/software. The source code (v1.0) is freely available at http://github.com/hsalis/Ribosome-Binding-Site-Calculator.

1.1. Inputs, outputs, and usage

To use the reverse engineering mode, an mRNA sequence is inputted. The translation initiation rate for each start codon is then predicted on a proportional scale from 0.001 to 100,000+ au. A protein CDS translated at 1000 au will produce 10 times more protein than one translated at 100 au, assuming that all other conditions are equal (e.g., transcription rates and mRNA stabilities). These comparisons in translation initiation rate are unaffected by the proportionality of the scale.

The method will warn you if the prediction may not satisfy a model assumption. The following warnings are generated:

a. *Kinetic trap (K) or not at equilibrium (NEQ)*: the folding of the mRNA transcript *may* experience a kinetic trap, preventing it from reaching an equilibrium and invalidating a key assumption of the thermodynamic model.
b. *Overlapping start codons (OS or OLS)*: the presence of two or more closely spaced start codons may alter the accuracy of the model's prediction.
c. *Short protein CDS*: the model requires at least 35 nucleotides of the protein CDS to make a valid prediction. A longer CDS should be inputted.

A key benefit of the forward engineering mode is the ability to design synthetic sequences that always satisfy the model's assumptions, which leads to higher predictive accuracies.

To use the forward engineering mode, a target translation initiation rate and the first 35 nucleotides of a protein CDS are inputted. The translation initiation rate is selected from a proportional scale that ranges from 0.001 to 100,000+ au. An optional presequence may also be inputted, which is any sequence that must appear upstream (5′) of the RBS. The method then generates a synthetic RBS sequence (30–35 nucleotides) that will initiate translation of the inputted protein CDS at the target rate. The forward engineering mode uses a stochastic optimization method to efficiently search through 4^{35} possible RBS sequences until it identifies one with the desired specifications.

A variant of the forward engineering mode allows you to input an initial RBS sequence and to specify which nucleotides are allowed to be altered according to the UIPAC degenerate nucleotide code. For example, when an *Xba*I restriction site must be located nearby the start codon, the sequence NNNNNTCTAGANNNNNN could be inputted. We refer to these sequences as RBS constraints.

1.2. Considerations

a. For *Escherichia coli* K12, the maximum possible translation initiation rate is 5,687,190 au, which is only obtainable for extremely AT-rich protein CDSs that do not form any secondary structure. The maximum predicted translation initiation rate in the *E. coli* K12 genome is 250,000 au.
b. Reusing the same RBS sequence with different protein CDSs can decrease the translation initiation rate by 500-fold (Salis *et al.*, 2009). A synthetic RBS sequence should be designed for each protein CDS.
c. Typically, there are multiple RBS sequences that equally satisfy an inputted specification, which is referred to as *degeneracy* in an optimization problem. The RBS Calculator's forward engineering mode will return the first synthetic RBS sequence that satisfies the design specifications.
d. A synthetic RBS sequence may not exist given the desired inputs; for example, if the 5′ end of the protein CDS is GC-rich and a high

translation initiation rate is selected (100,000+ au). An RBS constraint with few alterable nucleotides may also not have a solution. In these cases, the optimization method will self-terminate, and an RBS sequence is not returned.

e. The RBS Calculator does not predict the physiological effect of translating a protein CDS at a selected rate. Sufficiently high expression of a protein will result in cell growth inhibition, due to competition for metabolic resources. A protein may also exhibit an activity that causes cell death prior to this threshold.

2. Applications of the RBS Calculator

Proteins are the workhorse of cells. They are responsible for carrying out metabolism, replication, signal transduction, and differentiation. They sense the environment, regulate gene expression, catalyze chemical reactions, and act as therapeutics for treating human disease. Consequently, the ability to rationally control the protein expression level has broad applications in biotechnology.

2.1. Manipulating the protein expression level

Recombinant protein production is a multibillion dollar industry. The RBS Calculator allows you to produce your protein of choice at a selected expression level in bacteria (e.g., *E. coli* BL21 or C43) by balancing the translation rate with the transcription and folding rates (Grunberg and Serrano, 2010). Using the reverse engineering mode, you may also predict a protein CDS's translation rate before cloning it into a vector, allowing you to know in advance whether protein expression will be low. Then, using the forward engineering mode, you can systematically increase the translation rate of a protein CDS in controlled steps (e.g., 1000 to 10,000 to 100,000 au) to identify the optimal expression conditions.

2.2. Optimizing synthetic metabolic pathways

Bacteria employ synthetic metabolic pathways to manufacture high-value chemical products from sugars or plant biomass. The RBS Calculator allows you to efficiently identify the optimal enzyme expression levels of a metabolic pathway, eliminating any bottlenecks in the pathway, and maximizing its productivity (Bujara and Panke, 2010; Holtz and Keasling, 2010; Na *et al.*, 2010).

Both combinatorial and convergent optimization strategies are possible. In combinatorial optimization, the RBS Calculator's forward engineering

mode is used to systematically sample combinations of the enzyme expression levels across their entire range. The optimal enzyme expression levels are then identified using a high-throughput screen or selection. In convergent optimization, the forward engineering mode is used to measure the local gradient of productivity with respect to enzyme expression level, followed by a large leap in expression level space toward the optimal levels. Convergent optimization requires fewer experiments than combinatorial optimization, but is not guaranteed to yield the global optimum. The development and demonstration of these optimization strategies are active areas of research.

2.3. Designing and connecting genetic circuits

Genetic circuits—networks of regulated genes—employ transcription factors, regulatory RNAs, signaling proteins, and cell–cell communication to control gene expression according to a logical program or a dynamical behavior. Genetic circuits have been engineered to exhibit bistability, oscillations, pattern formation, traveling waves, and Boolean logic (Danino *et al.*, 2010; Ellis *et al.*, 2009; Khalil and Collins, 2010; Lu *et al.*, 2009; Purnick and Weiss, 2009; Stricker *et al.*, 2008; Tabor *et al.*, 2009).

The RBS Calculator allows you to rationally tune the expression levels of the proteins in a genetic circuit, either during the initial design phase or when connecting two genetic circuits together. A quantitative model of the genetic circuit can be used to determine which experimental perturbations are needed to successfully obtain a desired behavior; however, the model can only report these perturbations as numbers (e.g., increase a transcription factor's expression level by fivefold). Using the reverse and forward engineering modes, you may convert these numbers into a synthetic RBS sequence, increasing or decreasing a protein's expression level by a selected fold change. The circuit's quantitative model may be physics based or empirical, dynamical or at steady state, stochastic or deterministic; regardless, the RBS Calculator connects the model's numerical solution to a specific DNA sequence.

2.4. Evolutionary robustness of RBSs

During evolutionary mutation and selection, protein expression level changes can result in physiological changes that increase an organism's fitness. The RBS Calculator's reverse engineering mode allows you to perform a sensitivity analysis on a RBS sequence, identifying which nucleotide mutations will most affect its translation initiation rate and protein expression level. Thus, probable paths during evolution can be identified *a priori*. Synthetic RBSs can also be designed for robustness to evolutionary pressure, by ensuring that any nucleotide mutation does not result in a large change in protein expression level.

2.5. Predicting translation initiation rates across a genome

The RBS Calculator's reverse engineering mode can predict the translation initiation rate of every start codon in a bacterial genome. These predictions enable you to find the correct start codons of open reading frames, estimate the expression levels of proteins within an operon, and identify internal start codons that exhibit significant translation and produce variant proteins. The RBS Calculator software uses MPI for parallel programming and its calculations may be distributed across many processors on a supercomputer.

For example, the *E. coli* K12 genome contains 629,738 start codons. Over 600 internal start codons have significant amounts of translation, compared to the annotated open reading frame. In particular, the RBS Calculator correctly predicts the significant translation (6700 au) of a five-amino acid peptide encoded within its 23S rRNA (Tenson *et al.*, 1996).

3. RBSs and Bacterial Translation

Synthetic biology cannot advance as an engineering discipline without clear definitions and predictive rules. In this section, we clearly state the rules that control bacterial translation initiation, including the boundaries of the RBS genetic part and its molecular interactions with the ribosome.

3.1. Translation initiation as a rate-limiting step

Translation is the process by which ribosomes bind to an mRNA sequence and produce a corresponding protein according to the genetic code. The process has three phases: initiation, elongation, and termination (Kozak, 1999; Laursen *et al.*, 2005). Translation initiation is the key rate-limiting step; it controls how many ribosomes begin the elongation process and how many can potentially finish it. To obtain the highest protein production rates, both the translation initiation and elongation rates must be maximized. Codon optimization of protein CDSs is a commonly used method for increasing the translation elongation rate (Welch *et al.*, 2009). However, any reduction in translation initiation will always lead to a reduction in protein production. Thus, by codon optimizing the protein CDS and manipulating its translation initiation rate, we can exert complete control over a protein's production rate across its entire range.

The rate of translation initiation is determined by two factors: global changes in the cell's metabolism and the specific mRNA sequence surrounding a start codon. When bacteria are grown in nutrient-rich media, their number of available ribosomes is increased to support faster protein production rates and self-replication. Conversely, in nutrient-poor media or when protein

production exceeds a critical threshold, bacteria will reduce the overall protein production rate to conserve resources (e.g., the stringent response). These changes are global; they affect the translation of all mRNAs inside the cell.

Bacteria *differentially* regulate the translation of individual mRNAs according to the sequence of the RBS, a ubiquitous genetic part that is located upstream of a protein CDS. By modifying the RBS sequence, one alters a CDS's translation initiation *relative* to all other translated CDSs inside the cell. In other words, the RBS sequence controls the distribution of protein synthesis resources to transcribed protein CDSs (a proportion)—not their absolute translation initiation rate (proteins/mRNA/second).

3.2. The RBS genetic part

We propose the following functional definition of the RBS genetic part and its boundaries:

Definition

(i) The RBS part begins at least 35 nucleotides before the start codon of a protein CDS, up to the start (+1) of the mRNA transcript.
(ii) The translation initiation rate of a protein CDS depends on at least the 35 nucleotides before and after its start codon on an mRNA transcript.

Our RBS part length is longer than previous definitions to include the presence of long-range molecular interactions between the RBS sequence, the CDS, and the ribosome. If the RBS part length is shortened, these interactions would not be incorporated into the prediction and would instead appear as "context effects"—unknown position-dependent molecular interactions that somehow alter the translation initiation rate. Our definition substantially reduces these context effects.

3.3. The rate-limiting molecular interactions of translation initiation

Bacterial translation initiation requires the coordinated assembly of the 30S ribosomal complex onto the RBS at a protein CDS's start codon. This preinitiation complex includes translation initiation factors IF1, IF2, and IF3, and the initiator $tRNA^{fMet}$ (Ramakrishnan, 2002). Successful initiation also requires GTP as an energy source. Once bound, the ribosome protects a large mRNA region from hydroxyl radical attack, consisting of about 35 nucleotides before the start codon and extending to 19–22 nucleotides after the start codon (Hüttenhofer and Noller, 1994).

The RBS region is subdivided into a standby site, a 16S rRNA binding site, a spacer region, and a start codon (Fig. 2.1). Initially, the 30S subunit binds to the standby site, which is a single-stranded region farther upstream of the

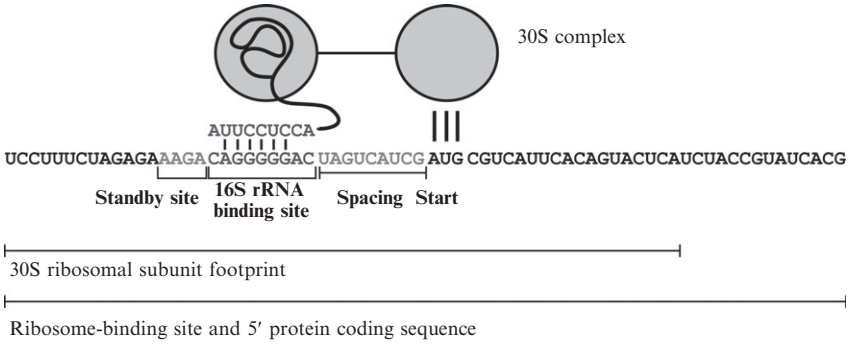

Figure 2.1 The ribosome binding site and 5′ protein coding sequence control an mRNA's translation initiation rate. The RBS region is subdivided into a standby site, a 16S rRNA-binding site, a start codon, and a spacer sequence. The footprint of a bound 30S subunit extends from the beginning of the standby site to about +22 bp, centered on the start codon.

start codon. Once bound, the 30S subunit slides downstream into its position over the RBS, assembles the 30S preinitiation complex, and initiates translation.

The existence of the standby site was hypothesized to solve a paradox (de Smit and van Duin, 2003): how would a freely diffusing, cytoplasmic 30S subunit have sufficient time to bind to a RBS that has been sequestered by an mRNA secondary structure? As the mRNA's structure dynamically unfolds and refolds, it will only be single stranded for a span of microseconds. During this brief time, the 30S must diffuse toward the mRNA and bind to it, which would require it to have an association binding constant of 10^{10} $[\text{Ms}]^{-1}$ or more; however, its actual association binding constant is around 10^{7} $[\text{Ms}]^{-1}$.

The presence of the standby site eliminates this paradox and has been experimentally demonstrated; the 30S binds to a standby site, followed by a nondiffusive downstream slide along the mRNA and assembly at the RBS. mRNAs that contain single-stranded standby sites and sequestered 16S rRNA-binding sites can be translated efficiently; however, if the standby site itself is sequestered by secondary structures (Studer and Joseph, 2006) or bound by sRNAs (Darfeuille *et al.*, 2007), then the mRNA's translation initiation decreases.

As the 30S complex slides across the mRNA into its preinitiation position, many noncovalent bonds are created and broken. The stability of the preinitiation complex and the translation initiation rate is determined by the energetics of these bonds. The complex's assembly rate is decreased by the unfolding of mRNA structures that sequester the 16S rRNA binding site, spacer region, start codon, or ribosome footprint region (Studer and Joseph, 2006). These mRNA structures are composed of intramolecular nucleotide base pairings (hydrogen bonds) that form helices, knots, loops, and bulges. The absence of these mRNA structures will increase the

translation initiation rate. Importantly, both the RBS and protein CDSs can participate in these mRNA structures.

The preinitiation complex may disassemble before translation initiation takes place. Stabilizing interactions will decrease this disassembly rate. These interactions include (a) hybridization between the mRNA and the last nine nucleotides of the 30S complex's 16S rRNA, which is called the anti-Shine–Dalgarno sequence (ASD; Shine and Dalgarno, 1974); (b) hybridization between the $tRNA^{fMet}$ anticodon and the start codon; and (c) attractive interactions between the mRNA and ribosomal proteins (e.g., S1; Aliprandi *et al.*, 2008; Boni, 1991). Conversely, the preinitiation complex is destabilized when it is forced to stretch or compress itself to occupy the RBS in the required position, analogous to a rigid spring. These distortions occur when the aligned distance between the 16S rRNA-binding site and the start codon deviates from an optimal of five nucleotides (Chen *et al.*, 1994).

4. A Biophysical Model and Optimization Method for RBSs

The RBS Calculator uses a statistical thermodynamic model to predict the translation initiation rate of a protein CDS. Given a RBS and a protein CDS, the model calculates the free energy change during the assembly of the 30S complex onto the mRNA (ΔG_{tot}). We then use a statistical ensemble approach to relate the protein CDS's translation initiation rate r to the ΔG_{tot}. The biophysical model bridges the gap between an RBS sequence and its translation initiation rate, creating a quantitative relationship between a string of letters (As, Gs, Cs, and Us) and a number.

The RBS Calculator combines the biophysical model with stochastic optimization to identify a synthetic (nonnatural) RBS sequence that will yield a user-selected translation initiation rate. Importantly, this relationship also depends on the first 35 nucleotides of the protein CDS and the synthetic RBS sequence must be designed with this sequence included.

4.1. Thermodynamics of RNA interactions

Using thermodynamics, the strengths of the intermolecular and intramolecular interactions that govern ribosome binding to mRNA may be calculated. At constant temperature and pressure, the available energy of a chemical species—the maximum amount of internal energy that is convertible to nonmechanical work—is called the Gibbs free energy, G. The change in Gibbs free energy (ΔG) quantifies the strengths of the interactions that cause the system to transition from a well-defined initial molecular state to a well-defined final molecular state. Importantly, the ΔG of a transition is *path*

independent; the ΔG calculation does not depend on the transition rate or the number of intermediate states in between the initial and final state.

The path-independence of thermodynamics allows us to define an arbitrary *reference state* and calculate the ΔG of a transition from state 1 to state 2 using two independent terms: $\Delta G_{12} = \Delta G_2 - \Delta G_1$, where $\Delta G_1 = G_1 - G_{\text{ref}}$ and $\Delta G_2 = G_2 - G_{\text{ref}}$. We are allowed to define the reference state and its energy, but the same reference state must be used in all calculations. For RNA-related ΔG calculations, the reference state is defined as the fully unfolded and unstructured RNA molecule with a free energy of zero ($G_{\text{ref}} = 0$ kcal/mol).

RNA is single stranded and will fold back onto itself to form secondary structures (helices, loops, bulges, and knots; Bevilacqua and Blose, 2008). Multiple RNA molecules will cofold or hybridize together, forming a combination of intermolecular and intramolecular nucleotide base pairings. The Gibbs free energy of an RNA secondary structure is calculated by a semiempirical model that predicts the individual free energies of a structure's helices, loops, bulges, and mismatches (Badhwar *et al.*, 2007; Blose *et al.*, 2007; Christiansen and Znosko, 2008; Mathews *et al.*, 1999; Miller *et al.*, 2008; Vecenie *et al.*, 2006; Xia *et al.*, 1998). For short RNA sequences, the model has a 5–10% accuracy, compared to experimentally measured free energies.

Under equilibrium conditions, a solution of RNA molecules forms a mixture (an ensemble) of secondary structures. The probability of an RNA molecule forming a particular structure with free energy ΔG is proportional to its Boltzmann weight, $\exp(\Delta G/RT)$, where T is the system's temperature and R is the gas constant. The most probable RNA secondary structure has the lowest ΔG (the minimum free energy, MFE). However, there are many suboptimal RNA structures that will coexist with the MFE RNA structure; the most common structures will have energies within $2RT$ (1.24 kcal/mol at 37 °C) of the MFE ΔG.

Algorithms have been developed to calculate the MFE structure and energy, suboptimal structures and free energies, and base pairing probabilities of folded or cofolded RNA molecules. These algorithms use dynamic programming to implicitly enumerate all possible RNA secondary structures, without explicit calculation of every structure's free energy, in order to find ones with minimal or near-minimal energies (Dirks *et al.*, 2007; Gruber *et al.*, 2008; Markham and Zuker, 2008; Mathews and Turner, 2006). More recent algorithms can generate RNA structures from the ensemble in proportion to their Boltzmann weight (Mathews, 2006).

4.2. A free energy model for ribosome assembly

The thermodynamic model calculates the difference in Gibbs free energy before and after the 30S complex assembles onto an mRNA transcript, denoted by ΔG_{tot}. The model considers an mRNA subsequence consisting

of 35 nucleotides before and after a start codon. Given an mRNA subsequence, five free energy terms are calculated and summed together:

$$\begin{aligned}\Delta G_{\text{tot}} &= \Delta G_{\text{final}} - \Delta G_{\text{initial}} \\ &= \left(\Delta G_{\text{mRNA:rRNA}} + \Delta G_{\text{start}} + \Delta G_{\text{spacing}} - \Delta G_{\text{standby}}\right) - \Delta G_{\text{mRNA}}\end{aligned} \tag{2.1}$$

The initial state is the unbound 30S complex and mRNA subsequence. The mRNA subsequence is folded to its MFE secondary structure with a corresponding free energy ΔG_{mRNA}. We do not include the free energy of the unbound 30S complex in the initial state.

The final state is the 30S complex bound to the mRNA subsequence. The strengths of the participating interactions are quantified by four free energy terms (Fig. 2.2). The $\Delta G_{\text{mRNA:rRNA}}$ is the energy released when the last nine nucleotides of the 16S rRNA cofolds and hybridizes with the mRNA subsequence at the 16S rRNA-binding site. The $\Delta G_{\text{mRNA:rRNA}}$ calculation includes both intermolecular nucleotide base pairings between the 16S rRNA and mRNA and intramolecular nucleotide base pairings within the mRNA itself ($\Delta G_{\text{mRNA:rRNA}} < 0$). These intramolecular nucleotide base pairings are mutually exclusive with the ribosome footprint.

The ΔG_{start} is the energy released when the $\text{tRNA}^{\text{fMet}}$'s anticodon hybridizes to the start codon ($\Delta G_{\text{start}} < 0$). The $\Delta G_{\text{standby}}$ is the energy released when the standby site is folded ($\Delta G_{\text{standby}} < 0$); accordingly, $-\Delta G_{\text{standby}}$ is the amount of energy needed to unfold the standby site. We define the standby site as the four nucleotides upstream of the 16S rRNA-binding site.

The $\Delta G_{\text{spacing}}$ is an energetic penalty for a nonoptimal distance between the 16S rRNA-binding site and the start codon ($\Delta G_{\text{spacing}} > 0$). The quantitative relationship between the distance s and $\Delta G_{\text{spacing}}$ was experimentally determined and fit to a simple model (Salis *et al.*, 2009). RBS sequences were constructed to have a high affinity 16S rRNA-binding site, minimal secondary structure, and an aligned distance s that was varied from $s = 0$ to $s = 15$ nucleotides (TCTAGA A_7 TAAGGAGGT A_s ATG ...). These sequences are predicted to have the same $\Delta G_{\text{mRNA:rRNA}}$, ΔG_{mRNA}, ΔG_{start}, and $\Delta G_{\text{standby}}$ free energies. Accordingly, the differences in their translation initiation rates may be directly related to their $\Delta G_{\text{spacing}}$ energies as a function of their aligned spacing s:

$$\Delta G_{\text{spacing}}(s) = \frac{1}{\beta} \log\left(\frac{r_{s=5}}{r_s}\right) \tag{2.2}$$

Here, we define the optimal spacing ($s_{\text{opt}} = 5$) to have a zero $\Delta G_{\text{spacing}}$. The translation initiation rates of these sequences were measured and their $\Delta G_{\text{spacing}}$ energies were calculated and fit to an appropriate equation.

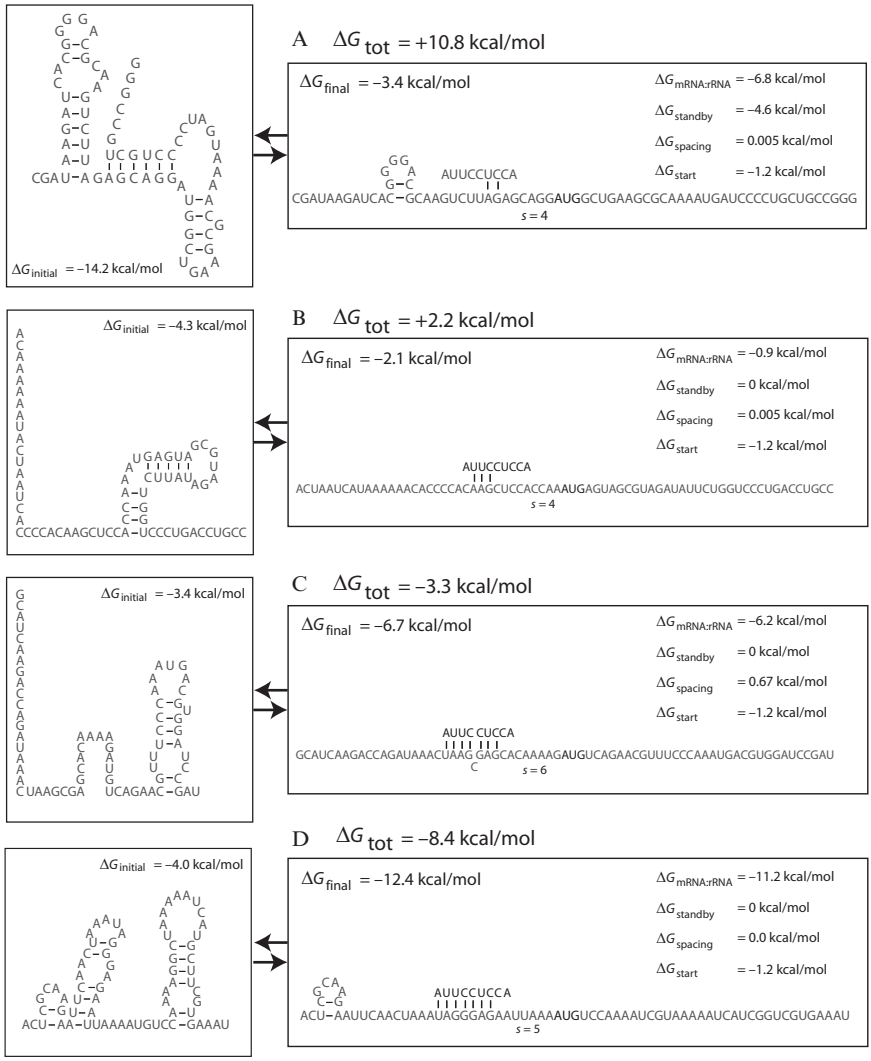

Figure 2.2 The RBS Calculator's forward engineering mode generates synthetic ribosome binding sites for the (A) araC, (B) sucB, (C) aceE, and (D) eno protein coding sequences with targeted ΔG_{tot} energies (translation initiation rates). The minimum free energy structures for the initial (left) and final (right) states are shown with corresponding free energies. The aligned spacing s is also shown.

When the ribosome is stretched ($s > 5$), the $\Delta G_{spacing}$ has a quadratic behavior, well fit by the equation:

$$\Delta G_{spacing} = c_1 \left(s - s_{opt}\right)^2 + c_2 \left(s - s_{opt}\right) \tag{2.3}$$

where $c_1 = 0.048$ kcal/mol/nt^2, and $c_2 = 0.24$ kcal/mol/nt. When the ribosome is compressed ($s < 5$), the $\Delta G_{\text{spacing}}$ has a sigmoidal behavior, fit by the equation:

$$\Delta G_{\text{spacing}} = \frac{c_1}{\left[1 + \exp\left(c_2\left(s - s_{\text{opt}} + 2\right)\right)\right]^3} \tag{2.4}$$

where $c_1 = 12.2$ kcal/mol and $c_2 = 2.5$ nt^{-1}. We next describe the steps to calculate the ΔG_{tot} of an mRNA subsequence.

4.2.1. The free energy calculation of the initial state

Using a dynamic programming algorithm, the mRNA subsequence is folded to its MFE secondary structure with a corresponding free energy ΔG_{mRNA}. This straightforward calculation may be carried out by the MFE program of NuPACK (Dirks *et al.*, 2007), the RNAfold program of ViennaRNA (Gruber *et al.*, 2008), or the hybrid-ss-min program of UNAFold (Markham and Zuker, 2008).

4.2.2. The free energy calculation of the final state

The free energy calculation of the final state begins with cofolding the last nine nucleotides of the 16S rRNA with the mRNA subsequence. MFE and suboptimal mRNA–rRNA structures are efficiently enumerated with the constraint that nucleotides located in the standby site or ribosome footprint are not allowed to base pair. For each structure, the free energies $\Delta G_{\text{mRNA:rRNA}}$, $\Delta G_{\text{spacing}}$, and $\Delta G_{\text{standby}}$ are calculated and summed together. The final state is the mRNA–rRNA structure that minimizes the summation of $\Delta G_{\text{mRNA:rRNA}}$, $\Delta G_{\text{spacing}}$, and $\Delta G_{\text{standby}}$.

Using a lookup table, the ΔG_{start} term is then added to this summation. The ΔG_{start} is -1.194 kcal/mol for AUG, -0.0748 kcal/mol for GUG, -0.0435 kcal/mol for UUG, and -0.03406 kcal/mol for CUG start codons, respectively.

The algorithm finds the physiological 16S rRNA-binding site by identifying the rRNA–mRNA base pairings that will minimize the entire system's Gibbs free energy, including the $\Delta G_{\text{spacing}}$ and $\Delta G_{\text{standby}}$. It is possible that the physiological 16S rRNA-binding site is not the highest affinity one within the mRNA. For example, if a very strong rRNA-binding site ($\Delta G_{\text{mRNA:rRNA}} = -12$ kcal/mol) is located far upstream of the start codon, then its penalty for nonoptimal spacing would be large ($\Delta G_{\text{spacing}} = +12$ kcal/mol), which yields a nonminimal ΔG_{final}. To find the physiological 16S rRNA-binding site with a brute force approach, one would enumerate all potential 16S rRNA-binding sites with a large range in energies, generating millions of possible configurations.

Instead, our algorithm uses an efficient spacing-indexed approach. First, classes of suboptimal mRNA–rRNA structures that have the same aligned spacing s and $\Delta G_{spacing}$ penalty, ranging from 0 to +15 kcal/mol, are enumerated. Each class contains MFE and suboptimal mRNA–rRNA structures where the $\Delta G_{spacing}$ and ΔG_{start} are constant, but the $\Delta G_{mRNA:rRNA}$ and $\Delta G_{standby}$ vary. The number of structures in each class is relatively small in contrast to the brute-force approach. The enumeration procedure may use the subopt program of NuPACK (Dirks *et al.*, 2007) or the RNAsubopt program of ViennaRNA (Gruber *et al.*, 2008) to generate the MFE and suboptimal mRNA–rRNA structures.

Then, for each structure in each class, the $\Delta G_{mRNA:rRNA}$ and $\Delta G_{standby}$ are calculated in a multistep procedure. The positions of the 16S rRNA–mRNA base pairs that are closest and farthest from the start codon are, respectively, labeled as (x_1, y_1) and (x_2, y_2). The borders of the 16S rRNA-binding site are between $[y_2 - x_2 + 1]$ and $[y_1 - x_1 + 9]$ and the aligned spacing is $s = \text{start} - [y_1 - x_1 + 9]$.

Each mRNA–rRNA structure is allowed to form intramolecular base pairs that do not sequester its standby site, its 16S rRNA-binding site, or its ribosome footprint. The mRNA subsequence that begins at position 1 and ends at $([y_2 - x_2 + 1] - N_{standby})$ is folded to its MFE structure, preventing the standby site from participating in base pairings. The number of nucleotides in the standby site is defined as $N_{standby} = 4$. Then, the mRNA subsequence that begins at start + footprint and ends at start + 35 is folded to its MFE structure. The *footprint* is the number of ribosome-bound nucleotides after the start codon. In current practice, we assume that the footprint is large enough to prevent significant base pairing in the protein CDS; this assumption only applies to the final state and may change in the future.

Finally, we add these intramolecular folding energies to $\Delta G_{mRNA:rRNA}$ so that it quantifies the total intermolecular and intramolecular interactions between the mRNA and RNA, including our folding constraints. By introducing our constraints during the folding process, the $\Delta G_{mRNA:rRNA}$ will include the $\Delta G_{standby}$ penalty. We may also calculate the individual value of the $\Delta G_{standby}$ penalty by comparing the free energies with and without the standby site folding constraints. For clarity, Eq. (2.1) explicitly includes the individual $\Delta G_{mRNA:rRNA}$ and $\Delta G_{standby}$ values.

The total Gibbs free energy for each structure in each spacing-indexed class is summed according to $\Delta G_{mRNA:rRNA} + \Delta G_{start} + \Delta G_{spacing} - \Delta G_{standby}$. The mRNA–rRNA structure that minimizes the total Gibbs free energy is identified and labeled as the final state.

4.2.3. Considerations

a. The thermodynamic model assumes a two-state system where both initial and final states have reached equilibrium. Based on kinetic studies

of RNA folding (Chen, 2008), some highly structured mRNAs may exhibit long-lived intermediate states that may invalidate this assumption.

b. The thermodynamic model does not include nonspecific protein–mRNA interactions. For example, the ribosome likely has a strong attractive and nonspecific interaction with all mRNAs. This sequence-independent interaction would appear as a large, negative constant in our free energy model; instead, it is currently incorporated into the proportionality constant that lumps together all sequence-independent effects.

c. The minimum possible value for ΔG_{tot} is -17.2 kcal/mol, yielding the maximum possible translation initiation rate. The translation initiation rate becomes negligibly small when ΔG_{tot} is greater than $+25$ kcal/mol.

d. The mRNA subsequence begins at the mRNA transcript's $+1$ nucleotide if the start codon's position is less than 35 nucleotides. In this case, the $+1$ nucleotide is considered "dangling" and an energetic bonus for dangling nucleotides is included in the free energy calculations of $\Delta G_{mRNA:rRNA}$ and ΔG_{mRNA}.

e. The model for calculating $\Delta G_{spacing}$ has the highest accuracy within the range $2 < s < 12$ nucleotides, where s is the aligned distance between 16S rRNA-binding site and start codon. Outside this range, the translation initiation rate was sufficiently decreased to make precise experimental measurements difficult, leading to uncertainty during model fitting.

f. Multiple configurations of the initial or final states may all have identical minimum free energies (e.g., degenerate structures). In these cases, the calculation of ΔG_{tot} will produce the same result and predict the same translation initiation rate.

g. According to statistical thermodynamics, a system at equilibrium will occupy a mixture of configurations; the majority will have energies between MFE and MFE + $2RT$. The thermodynamic model may be improved by sampling the configurations and computing the statistics of their ΔG_{tot} free energies. The improvement will be limited to about 1–2 kcal/mol.

4.3. A statistical thermodynamic model of translation initiation

We apply statistical thermodynamics to predict the translation initiation rates of all mRNAs inside the cell according to their Gibbs free energies of ribosome binding, ΔG_{tot}. Our system-of-interest is the cellular pool of ribosomes and mRNA transcripts. Ribosomes and mRNA transcripts are dynamically produced and degraded, but during the exponential phase of growth, their numbers will fluctuate around an average. The system achieves a nonequilibrium steady-state condition called *detailed balance*, where statistical thermodynamics may be validly applied.

For each mRNA transcript inside the cell, we describe the 30S ribosomal subunit's binding and assembly according to the following association and dissociation reactions:

$$\mathrm{mRNA}_i + 30\mathrm{S} \leftrightarrow \mathrm{mRNA}_i{::}\ 30\mathrm{S} \tag{2.5}$$

where the index $i = 1 \ldots N$ enumerates overall transcribed protein CDSs with their corresponding RBS sequences. At this point, we make our first assumption.

Assumption #1:

The pool of ribosomes and mRNAs remain at chemical equilibrium.

During chemical equilibrium, ribosomes dynamically bind to mRNAs, initiate translation, or dissociate from the mRNA; however, the *average* number of free ribosomes, free mRNAs, or mRNA–ribosome complexes remains constant. The number of ribosomes R, mRNAs m_i, and 30S–mRNA complexes C_i are then related according to

$$C_i = m_i\, R \exp(-\beta \Delta G_i) \tag{2.6}$$

where ΔG_i is the change in Gibbs free energy before and after the 30S complex of the ribosome assembles onto the ith protein CDS of an mRNA transcript. The total amount of 30S complex R_{tot} is the sum of the free and bound forms, which is

$$R_{\mathrm{tot}} = R + \sum_j C_j = R\left(1 + \sum_j m_j \exp\left(-\beta \Delta G_j\right)\right) \tag{2.7}$$

Equation (2.7) may be rearranged to give the free amount of 30S complex:

$$R = \frac{R_{\mathrm{tot}}}{\left(1 + \sum_j m_j \exp\left(-\beta \Delta G_j\right)\right)} \tag{2.8}$$

Substituting Eq. (2.8) into Eq. (2.6) and rearranging, we obtain the following relationship between the number of 30S–mRNA complexes C_i and the assembly reaction's Gibbs free energy change:

$$C_i = \frac{m_i R_{\mathrm{tot}} \exp(-\beta \Delta G_i)}{1 + \sum_j m_j \exp\left(-\beta \Delta G_j\right)} \tag{2.9}$$

Here, we make our second assumption.

Assumption #2:

The translation initiation rate of a protein CDS is proportional to the number of assembled 30S–mRNA complexes.

This assumption allows us to rewrite Eq. (2.9) in terms of a relative translation initiation rate r:

$$r_i \propto \frac{m_i R_{\text{tot}} \exp(-\beta \Delta G_i)}{1 + \sum_{j=1}^{N} m_j \exp(-\beta \Delta G_j)} \tag{2.10}$$

where the j summation is performed over all transcribed protein CDSs.

Equation (2.10) describes the relative translation initiation rates of the "ribosome ensemble"—the pool of ribosomes interacting with the pool of mRNAs inside the cell—in terms of the Gibbs free energy changes of their individual assembly reactions. It indicates that the translation initiation rate of the ith protein CDS is (a) proportional to the number of mRNA transcripts and ribosomes, (b) higher when the assembly reaction possesses a more negative Gibbs free energy change, and (c) lower when a competing assembly reaction ($j \neq i$) has a more negative Gibbs free energy change. Its denominator quantifies the portion of the total protein synthesis rate that is defined by the cell's genome and transcriptome.

Equation (2.10) also allows us to quantify the effect of expressing new proteins on the cell's total protein synthesis rate. Without any cellular modifications, the total protein synthesis rate is a generally large number. When a new protein is modestly expressed inside the cell, the total synthesis rate and the translation rate of other mRNAs are generally unaffected. However, when one or more new proteins are highly expressed, Eq. (2.10)'s denominator will greatly increase. While the cell contains feedback loops to increase protein synthesis capacity under stress, a sufficiently large demand will result in lower translation rates for all mRNAs inside the cell. The resulting slowdown in global protein production will have many physiological effects, including a longer cell doubling time.

With these potential physiological changes in mind, we may simplify Eq. (2.10) by approximating its denominator as a constant:

$$r_i = K \exp(-\beta \Delta G_i) \tag{2.11}$$

where the proportionality constant K now includes the denominator. Equation (2.11) has two important conclusions: (i) a natural log r versus ΔG_{tot} plot will be linear and (ii) the linear plot's slope is $-\beta$. Experimentally, we confirmed the validity of Eq. (2.11) and measured that $\beta = 0.45 \pm 0.05$ mol/kcal (Salis *et al.*, 2009; Fig. 2.3). We choose $K = 2500$ so that

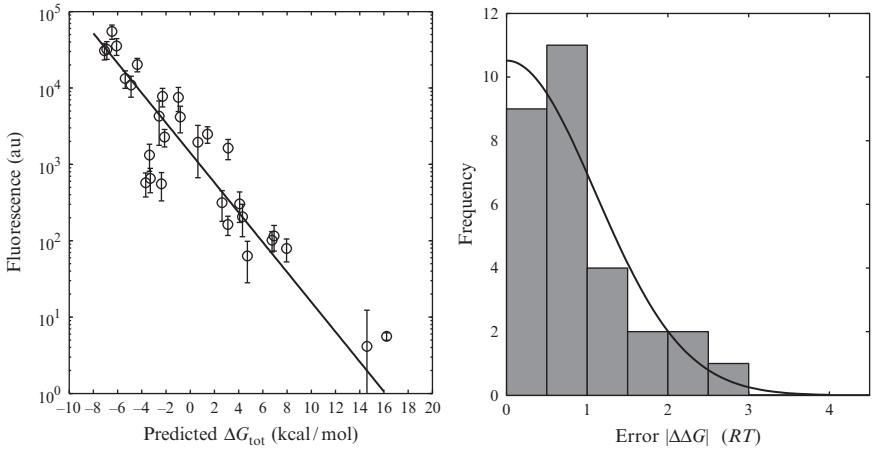

Figure 2.3 The translation initiation rates of 29 synthetic ribosome binding sites are experimentally compared to the RBS Calculator's predictions. According to the theory, a linear relationship between the calculated ΔG_{tot} and the natural logarithm of the measured protein expression level is expected. This theory is validated with an $R^2 = 0.84$. The average error is 1.82 kcal/mol ($0.82RT$).

physiologically possible translation initiation rates r will vary between 0.1 and 100,000.

According to statistical thermodynamics, a system at thermal equilibrium with its surroundings will have a $\beta = (RT)^{-1}$. According to the experimental data, the system's apparent temperature is about 1100 K, which is much higher than the actual system temperature of 310 K. Understanding this discrepancy will require further experimentation; however, our calculation of ΔG_{tot} assumes a two-state model at equilibrium. It is not uncommon for complicated thermodynamic systems to exhibit greater effective temperatures when modeled as a simpler system.

4.4. Optimization of synthetic RBSs

Synthetic RBSs are nonnatural sequences that are optimized to yield a targeted translation initiation rate. The target translation initiation rate and a selected protein CDS are inputted into a stochastic optimization method, called simulated annealing, which then performs iterative rounds of mutation, calculation, and selection until it identifies an RBS sequence that satisfies all constraints. The target translation initiation rate is converted into a target ΔG_{tot} (ΔG_{target}) using Eq. (2.11) and an experimentally measured $\beta = 0.45$ mol/kcal.

The mRNA subsequence is initialized by concatenating a randomly generated 35 nucleotide RBS sequence to the inputted protein CDS.

Its ΔG_{tot} is calculated according to Eq. (2.1) and its objective function is evaluated as $O_{old} = |\Delta G_{tot} - \Delta G_{target}|$. New mRNA subsequences are generated by inserting, deleting, or replacing nucleotides in the RBS, followed by calculation of their ΔG_{tot}, and evaluation of their objective function $O_{new} = |\Delta G_{tot} - \Delta G_{target}|$. The mutation is then accepted or rejected according to the Metropolis criteria (Metropolis *et al.*, 1953), using an annealing temperature T_{SA}, and three additional sequence constraints.

The Metropolis criteria compares the objective functions of the current and new mRNA subsequences and calculates a probability of accepting the mutation according to

$$P = \max\left[1, \exp\left(\frac{O_{old} - O_{new}}{T_{SA}}\right)\right] \tag{2.12}$$

If the new mRNA subsequence is closer to satisfying the target and constraints, then it is immediately accepted with probability one. Otherwise, it is conditionally accepted with probability $P < 1$. The annealing temperature T_{SA} is dynamically adjusted to maintain a conditional acceptance ratio between 5% and 20%.

Sequence constraints are added to prevent the optimization algorithm from generating a synthetic RBS sequence that may invalidate the thermodynamic model's assumptions. If a mutated sequence fails to satisfy these rules, then the mutation is discarded. First, the mutated RBS sequence may not contain start codons. Second, the energy required to unfold the 16S rRNA-binding site must be less than 6 kcal/mol. Third, the MFE structure of the mRNA subsequence may not contain base pairings that are farther than 35 nucleotides apart.

The first rule prevents translation from initiating at additional start codons. The second rule prevents the formation of potentially long-lived intermediate states that may result when the 16S rRNA-binding site is strongly sequestered by secondary structure. The third rule encourages the formation of local secondary structures that reach equilibrium quickly. Conversely, nonlocal secondary structures may encounter kinetic traps that prevent them from reaching equilibrium.

4.5. Accuracy and limitations

The thermodynamic model's predictions have been experimentally tested on 119 natural and synthetic mRNA sequences (Salis *et al.*, 2009). On average, a synthetic RBS sequence may be designed to achieve a targeted translation initiation rate to within a factor of 2.3 over a range of 100,000-fold (Fig. 2.3). The average error in the ΔG_{tot} calculation is 1.82 kcal/mol ($0.82RT$).

The model's ΔG_{tot} error is well fit by a one-sided Gaussian distribution with a variance of 2.44 kcal/mol (1.1RT). However, there are extrema sequences where the model's error exceeds 6 kcal/mol (2.7RT); in these cases, certain molecular interactions may be absent from the model that play an important role in altering the translation initiation rate.

The thermodynamic model has the following limitations: (a) the model does not include the interaction between the mRNA and ribosomal S1 protein; (b) the model does not consider the presence of antisense RNA- or RNAse-binding sites; (c) the model assumes that multiple start codons are independently translated, ignoring the potential for coupling between closely spaced start codons; and (d) the model ignores the potential for translational coupling between protein CDSs in an operon, which likely occurs when an RBS and upstream protein CDS overlap.

5. Precise Measurements of Fluorescent Protein Expression Levels

Fluorescent proteins are versatile reporters that enable the measurement of *in vivo* protein expression levels. The following protocol describes the usage of spectrophotometers and/or flow cytometry to record fluorescent protein expression levels from *E. coli* cultures in 96-well microplate format. Importantly, our protocol uses long culture times of 24–36 h to achieve steady-state conditions so that the average number of proteins per cell is constant over time. This protocol improves the measurement's precision, day-to-day and lab-to-lab reproducibility, and enables a more accurate comparison between model predictions and experimental data.

5.1. Protocol

Equipment: a spectrophotometer with monochromators and incubation capability (e.g., a TECAN M1000); a flow cytometer with appropriate lasers and detectors for measuring selected fluorescent proteins in a high-throughput microplate format (e.g., a BD LSRII or Fortessa).

a. A standard (or deep) 96-well microplate containing 200 μL (or 1 mL) Luria–Bertani (LB) media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) and selective antibiotic is inoculated from single colonies. A nonfluorescent (white) cell culture is also inoculated. Cultures are grown overnight at 37 °C with 250 rpm orbital shaking to an optical cell density (OD_{600}) of 2.0.

b. A fresh 96-well transparent, flat bottom microplate is filled with 198 μL M9 minimal media (1× M9 salts: 6.8 g/L Na_2PO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl; 2 m*M* $MgSO_4$, 100 μ*M* $CaCl_2$; selective

antibiotic; and 0.4% glucose, adjusted to a pH of 7.4). Microplate wells are inoculated by overnight cultures using a 1:100 dilution. The microplate is placed inside the spectrophotometer and incubated at 37 °C with 250 rpm orbital shaking.

c. The spectrophotometer records OD_{600} and fluorescence (FLU) measurements every 10 min. These measurements are used to calculate cell growth rates and steady-state bulk FLU per OD_{600}.

d. Once a culture reaches an OD_{600} between 0.15 and 0.20, 10–20 μL samples of each culture are transferred to fresh transparent, flat bottom microplates containing 180–190 μL prewarmed M9 minimal media and selective antibiotic (a 1:5–1:10 dilution). The new microplate is placed inside the spectrophotometer and incubated at 37 °C with 250 rpm orbital shaking.

e. Additional 10 μL samples of each culture are transferred to a round-bottom microplate containing 190 μL PBS and 2 mg/mL kanamycin for flow cytometry measurements. The excess kanamycin stops bacterial protein production.

f. Repeat steps c, d, and e. Steady-state conditions are not reached until about 4–10 h of culture time. This media replacement strategy can be continued until sufficient protein expression level data have been gathered. Typically, at least three to four serial dilutions are performed for a total culture time of 24–36 h.

5.2. Considerations

a. The M9 minimal media is supplemented with 0.05 g/L leucine when using *E. coli* DH10B cells, which have a leucine auxotrophic phenotype.

b. The M9 minimal media may be replaced with another defined media, such as MOPS minimal media or yeast synthetic defined (SD) media. The key requirement is that the culture's growth rate remains constant between media batches. The formulation of medias that use tryptone, yeast extract, or peptide mixtures will greatly vary between batches.

c. Only the inner 60 wells of a 96-well microplate should be used for cell culture. During an 8- 12-h culture, evaporation will cause the liquid levels in the outer wells to significantly drop. The outer wells should be filled with blank media to buffer this effect.

d. Both the spectrophotometer and flow cytometry use a gain parameter to convert photon emission counts into digitized data. It is essential that this gain parameter remain a constant throughout all experiments. Some machines offer dynamic control of the gain to maximize FLU sensitivity; this dynamic control should be turned off. During preliminary experiments, the gain parameter may be optimized such that (i) the background FLU of media or white cells is twofold higher than detector noise, and (ii)

the data from cell cultures expressing the highest protein expression level are twofold less than the machine's digital overflow value. The optimal gain parameter depends on fluorescent protein brightness and detector path length.

e. The spectrophotometer records cell optical density (OD) and FLU data of cell culture samples, white cells, and blank media over time. For each time point, the average bulk fluorescence per cell (FLPC) is calculated according to

$$\mathrm{FLPC}_{\mathrm{sample}} = \frac{\mathrm{FLU}_{\mathrm{sample}} - \mathrm{FLU}_{\mathrm{media}}}{\mathrm{OD}_{\mathrm{sample}} - \mathrm{OD}_{\mathrm{media}}} - \frac{\mathrm{FLU}_{\mathrm{white}} - \mathrm{FLU}_{\mathrm{media}}}{\mathrm{OD}_{\mathrm{white}} - \mathrm{OD}_{\mathrm{media}}} \qquad (2.13)$$

which correctly subtracts the background OD and accounts for background FLU from both blank media and white cell sources. This measurement is useful when the single-cell FLU distribution is unimodal (single-peaked) and when stochastic contributions to gene expression are not relevant.

f. The flow cytometer records forward scatter (FSC), side scatter (SSC), and FLU data for single cells taken from cell culture samples or white cells. At least 100,000 single-cell measurements should be recorded. Cell debris and clumped cells are eliminated from the data by discarding events ("gating") according to the following criteria: (i) the FLU is greater than zero; (ii) the FSC to SSC ratio is greater than 0.50 and less than 1.50; and (iii) the FSC and SSC values are inside a circle, drawn with its center located at the average FSC and SSC data and with a radius of 10,000. The circle's radius may be adjusted according to the flow cytometer's gain parameter. Automated gating may be performed using MATLAB and the fca_readfcs script (Laszlo Balkay). The arithmetic mean (not geometric) of the FLU distribution should be used to calculate the average FLU per single cell. The variance, kurtosis, and skewness of the distribution, respectively, quantify its variation, peakedness (heavy tailedness), and asymmetry.

g. A cell culture has reached steady state when (i) the rate of change of $\mathrm{FLPC}_{\mathrm{sample}}$ over time is approximately zero for a sufficiently long time, or (ii) the distribution of single-cell FLU does not significantly change between two time points. Once steady-state conditions have been reached, the recorded $\mathrm{FLPC}_{\mathrm{sample}}$ data is time averaged to reduce measurement noise and the FLU distribution's statistics are calculated.

ACKNOWLEDGMENTS

We would like to thank Joseph Wade (Wadsworth Center) for an insightful discussion on genome-scale translation rates. This work was supported by a DARPA Young Faculty Award.

REFERENCES

Aliprandi, P., Sizun, C., *et al.* (2008). S1 ribosomal protein functions in translation initiation and ribonuclease RegB activation are mediated by similar RNA-protein interactions. *J. Biol. Chem.* **283**(19), 13289.

Badhwar, J., Karri, S., *et al.* (2007). Thermodynamic characterization of RNA duplexes containing naturally occurring 1 × 2 nucleotide internal loops. *Biochemistry* **46**(50), 14715–14724.

Bevilacqua, P., and Blose, J. (2008). Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Phys. Chem.* **59**(1), 79.

Blose, J., Manni, M., *et al.* (2007). Non-nearest-neighbor dependence of the stability for RNA bulge loops based on the complete set of group I single-nucleotide bulge loops†. *Biochemistry* **46**(51), 15123–15135.

Boni, I. (1991). Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res.* **19**(1), 155.

Bujara, M., and Panke, S. (2010). Engineering in complex systems. *Curr. Opin. Biotechnol.* **21,** 586–591.

Chen, S. (2008). RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* **37,** 197.

Chen, H., Bjerknes, M., *et al.* (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res.* **22**(23), 4953.

Christiansen, M., and Znosko, B. (2008). Thermodynamic characterization of the complete set of sequence symmetric tandem mismatches in RNA and an improved model for predicting the free energy contribution of sequence asymmetric tandem mismatches†. *Biochemistry* **47**(14), 4329–4336.

Danino, T., Mondragón-Palomino, O., *et al.* (2010). A synchronized quorum of genetic clocks. *Nature* **463**(7279), 326–330.

Darfeuille, F., Unoson, C., *et al.* (2007). An antisense RNA inhibits translation by competing with standby ribosomes. *Mol. Cell* **26**(3), 381–392.

de Smit, M., and van Duin, J. (2003). Translational standby sites: How ribosomes may deal with the rapid folding kinetics of mRNA. *J. Mol. Biol.* **331**(4), 737–743.

Dirks, R., Bois, J., *et al.* (2007). Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.* **49**(1), 65.

Ellis, T., Wang, X., *et al.* (2009). Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**(5), 465–471.

Gruber, A., Lorenz, R., *et al.* (2008). The Vienna RNA websuite. *Nucleic Acids Res.* **36** (Web Server issue), W70.

Grunberg, R., and Serrano, L. (2010). Strategies for protein synthetic biology. *Nucleic Acids Res.* **38,** 2663–2675.

Holtz, W., and Keasling, J. (2010). Engineering static and dynamic control of synthetic pathways. *Cell* **140**(1), 19–23.

Hüttenhofer, A., and Noller, H. (1994). Footprinting mRNA-ribosome complexes with chemical probes. *EMBO J.* **13**(16), 3892.

Khalil, A., and Collins, J. (2010). Synthetic biology: Applications come of age. *Nat. Rev. Genet.* **11**(5), 367–379.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**(2), 187–208.

Laursen, B., Sorensen, H., *et al.* (2005). Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **69**(1), 101.

Lu, T., Khalil, A., *et al.* (2009). Next-generation synthetic gene networks. *Nat. Biotechnol.* **27**(12), 1139–1150.

Markham, N., and Zuker, M. (2008). Software for nucleic acid folding and hybridization. *Methods Mol. Biol.* **453,** 3–31.
Mathews, D. (2006). Revolutions in RNA secondary structure prediction. *J. Mol. Biol.* **359**(3), 526–532.
Mathews, D., and Turner, D. (2006). Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* **16**(3), 270–278.
Mathews, D., Sabina, J., *et al.* (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure1. *J. Mol. Biol.* **288**(5), 911–940.
Metropolis, N., Rosenbluth, A., *et al.* (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**(6), 1087.
Miller, S., Jones, L., *et al.* (2008). Thermodynamic analysis of 5 and 3 single- and 3 double-nucleotide overhangs neighboring wobble terminal base pairs. *Nucleic Acids Res.* **36**(17), 5652.
Na, D., Kim, T., *et al.* (2010). Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.* **13**(3), 363–370.
Purnick, P., and Weiss, R. (2009). The second wave of synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**(6), 410–422.
Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* **108**(4), 557–572.
Salis, H. M., Mirsky, E. A., *et al.* (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**(10), 946–950.
Shine, J., and Dalgarno, L. (1974). The 3-terminal sequence of Escherichia coli 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**(4), 1342.
Stricker, J., Cookson, S., *et al.* (2008). A fast, robust and tunable synthetic gene oscillator. *Nature* **456**(7221), 516–519.
Studer, S., and Joseph, S. (2006). Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Mol. Cell* **22**(1), 105–115.
Tabor, J. J., Salis, H. M., *et al.* (2009). A synthetic genetic edge detection program. *Cell* **137** (7), 1272–1281.
Tenson, T., DeBlasio, A., *et al.* (1996). A functional peptide encoded in the Escherichia coli 23S rRNA. *Proc. Natl. Acad. Sci. USA* **93**(11), 5641.
Vecenie, C., Morrow, C., *et al.* (2006). Sequence dependence of the stability of RNA hairpin molecules with six nucleotide loops†. *Biochemistry* **45**(5), 1400–1407.
Welch, M., Villalobos, A., *et al.* (2009). You're one in a googol: Optimizing genes for protein expression. *J. R. Soc. Interface* **6**(Suppl. 4), S467.
Xia, T., SantaLucia, J., Jr., *et al.* (1998). Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs†. *Biochemistry* **37**(42), 14719–14735.