Ribosome tail ends as "signal detectors" for protein production in prokaryotes

L. Ponnala[1], T.M. Barnes[2], D.L. Bitzer[1], MA. Vouk[1]
[1]North Carolina State University, Raleigh, NC, USA
[2]University of North Carolina at Charlotte, NC, USA

Abstract

The 16s ribosomal tail has been conjectured to play an important role in the regulation of protein production. In previous work, we generated sequences of 13 base pairs as hypothetical ribosome tail ends for E. coli K-12. We found the actual E. coli ribosome tail to be significantly different from a randomly generated one in the magnitude of the lock and synchronization signals, and the signal to noise ratio. Using a Genetic Algorithm to search for optimal tails, we found that the actual E. coli ribosome tail end was among the best by these measures. In our current work, we use this method to detect optimal ribosome tails in thirteen other prokaryotic organisms, to test our hypothesis that signal characteristics are good indicators of protein production in prokaryotes. This work also provides theoretical ribosome tails that could be useful in transgenic protein production.[*]

1 Introduction

Transgenic protein production is an important biotechnological advance, offering a method for producing large quantities of necessary proteins at low cost. Its effectiveness and efficiency, which strongly affect its cost, are determined by adjusting foreign messenger RNA (mRNA) to be acceptable to both the host environment and the host ribosome. Without these adjustments, proteins may not be produced in sufficient quantities. However, the process of determining necessary adjustments is complex and often involves much trial and error [1].

The tail end of the 16s ribosomal subunit appears to play an important role in the translation process in prokaryotic organisms [2]. An improved understanding of this role and the interactions of the 16s tail with mRNA may therefore lead to significant advances in genetic engineering.

An important feature of the ribosome is the strong affinity of its exposed 3' tail end to an identifier called the Shine-Dalgarno sequence that is located roughly 13 bases upstream of the start codon. This interaction can be modeled by calculating the free energy released due to the binding of the ribosome's tail to the messenger RNA. This free-energy release is interpreted as a signal that appears to be a good indicator of the regulation that takes place during protein production, and of the translation efficiency.

Previous work [2] suggests that this regulation has two parts: a "lock" and a synchronization signal. The lock is located at or just before the start codon. It appears to reflect the need to bind or pause the ribosome long enough, close to start codon, for it to lock into the reading frame and to start protein production. Once the lock is achieved, and protein production starts, the synchronization signal must be strong enough, and in the right phase, to maintain the reading frame. In this model, the tail end of the ribosome should be able to detect these features in the genome for optimal translation.

In our previous work [3], we examined this model for *E. coli K-12* by generating random sequences of 13 bases as hypothetical ribosome tail ends, and assessing each based on signal criteria. We also designed and ran a Genetic Algorithm (GA) to search for hypothetical ribosome tails, optimized for these characteristics. Our findings suggest that the actual *E. coli* ribosome tail differs from a randomly generated one in the magnitude of the lock and synchronization signals, and in the signal to noise ratio (SNR). We also designed and ran a Genetic Algorithm (GA) to search for hypothetical ribosome tails, optimized for these characteristics. We found that the actual *E. coli* ribosome tail was among the best candidates in the GA search as well.

These results support the conjecture that the actual ribosome tail may have been selected by nature, using the lock, synchronization, and SNR characteristics, to be effective in translating the genes in *E. coli K-12*. The GA search resulted in a list of candidate ribosome tails which can be useful in two significant ways: 1) to compare with the actual ribosome tail to determine if other characteristics may have been important in natural selection of the actual ribosome, and 2) as candidate ribosome tail ends for efficient transgenic protein production. Because of their similarity to the actual ribosome tail, the hypothetical tails found by the GA also offer additional evidence that the selection of the ribosome tail is not random.

These positive results led us to extend our experiment to include 13 other prokaryotes, as listed in Table 1. As found with *E. coli K-12*, actual ribosome

[*] This work was supported in part by the DOE SciDAC grant DE-FC02-01ER25484, and the IBM Corp. Shared University Research Program.

tails for each species were found to be significantly different from randomly-generated hypothetical ribosome tails in our three signal characteristics, and among the best candidates found by GA search for each species.

Species	Accession Number	3' ribosome tail
Aquifex aeolicus	NC_000918	ATTTCCTCCACTAG
Bordetella bronchiseptica	NC_002927	TTTCCTCCACTAG
Bacillus subtilis	NC_000964	TCTTTCCTCCACTAG
Chlamydophila pneumoniae	NC_000922	TTTTTCCTCCACTAG
Escherichia coli	NC_000913	ATTCCTCCACTAG
Lactobacillus plantarum	NC_004567	TCTTTCCTCCACTAG
Pseudomonas aeruginosa	NC_002516	ATTCCTCCACTAG
Rickettsia conorii	NC_003103	ATTCCTCCATTAG
Salmonella typhimurium	NC_003197	ATTCCTCCACTAG
Salmonella enterica	NC_003198	ATTCCTCCACTAG
Streptococcus pyogenes	NC_006086	TCTTTCCTCCACTAG
Staphylococcus aureus	NC_003923	TCTTTCCTCCACTAG
Streptococcus mutans	NC_004350	TCTTTCCTCCACTAG
Thermoplasma volcanium	NC_002689	CCTCCACTAG

Table 1. List of selected species

In the following sections we introduce the concepts of free energy calculations and Genetic Algorithms. In Section 2, we discuss our methodology for evaluating

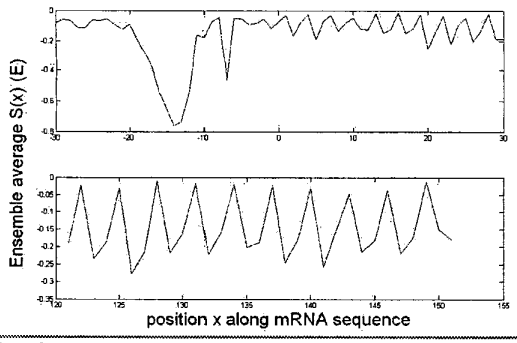

Fig. 1. Ensemble average S(x) for actual E. coli ribosome tail

Sync. signal magnitude	0.1378 E
Lock magnitude	–0.7688 E
Signal to noise ratio	–9.816 dB

Table 2. Signal characteristics for E. coli ribosome tail

hypothetical ribosome tails, estimating the distributions for the set of possible hypothetical ribosome tails for each species, and the fitness function used for the GA. Section 3 presents our findings, and Section 4 concludes our paper and suggests paths for future work.

1.1 Ensemble Average Signal

In this section, we discuss a method to analyze the free energy released during genetic translation using signal processing techniques, as used in [2,3,4,5]. A particular alignment of the 3' exposed tail end against the messenger RNA is referred to as a conformation. The binding energy released in this conformation, also referred to as free energy, is estimated using the method of base-doublets [2]. This calculation penalizes mismatches and rewards consecutive base pairing in the conformation. A shift along the mRNA by one base position results in a new conformation, and the calculation of the free energy estimate is repeated. The binding energies for matching doublets are determined by experiment [6].

The set of free energy estimates for all possible conformations along the mRNA sequence constitutes a discrete signal that can be analyzed using methods of discrete-time signal processing [7]. The signal is calculated for each individual coding sequence along the forward strand in *E. coli*, and the ensemble average of 531 such signals is plotted (See Appendix A for equations). For the remainder of this paper, we refer to this as the ensemble average signal, which, for each conformation, is expressed in units of kcal/mol, referred to as E.

Fig.1 demonstrates the ensemble average signal calculated by averaging the signals obtained by matching the tail end of the *E. coli* 16s ribosomal subunit to the 531 certain coding sequences for *E. coli K-12* available in GenBank [9] (See Appendix A). The dip in this signal, interpreted as a "lock", occurs roughly 13 bases upstream from the start codon, indicating strong affinity of the tail end to the Shine-Dalgarno consensus sequence that resides here [2]. About 90 bases downstream, we observe that the signal becomes strongly 3-base periodic. We will refer to this downstream signal as the "synchronization" signal, since it appears to reflect how the ribosomal subunit moves along the mRNA sequence till the formation of the polypeptide chain is complete [2]. Table 2 summarizes the signal characteristics for the actual tail end in *E. coli.*

1.2 Genetic Algorithms

Genetic Algorithms (GAs) are numerical optimization techniques based on a generalized theory of evolution and natural selection, and have been used to solve a variety of problems such as the selection of optimal convolutional codes [10] and table-based codes for genetic translation initiation [4]. There are 4^{13}

different sequences that may be considered as hypothetical tail ends for the 3' end of the 16s ribosomal subunit, a number which would be prohibitive for performing an exhaustive search. As discussed above, consecutive base-pairings between the ribosome tail end and the mRNA result in higher free energy release, suggesting that the complements of frequent mRNA base sequences will be important patterns in candidate tail ends. Since GAs emphasize patterns such as these in searching for optima, we chose to use a genetic algorithm to search for optimal ribosome tail ends.

2 Methodology

Our current work extends our previous work, as described in Section 1, to include thirteen other prokaryotes. A list of species used, their accession numbers, and their actual 3' ribosome tails is given in Table 1. We used the Gutell Laboratory Database [8] to examine the structure of the 16s subunit of the ribosome. This enabled us to determine where the hairpin starts on the 3' end. Our two-part hypothesis is that: 1) actual ribosome tail ends for each species will differ significantly from a random sample of its hypothetical tails, and 2) our Genetic Algorithm search for candidate tails will find that the actual ribosomes are among the best in signal characteristics in each species.

For each species, we performed the same two-part experiment. First, given the tail length for a species, we generated a random set of 2000 base sequences that represent hypothetical ribosome tails. We then modeled the interaction of each of these hypothetical tails with 500 of that species' verified gene sequences available on GenBank [9], and measure the characteristics of the resulting signal. For the purpose of comparison with the actual ribosome tail and for optimizing our GA, a Gamma distribution was used to approximate the theoretical distribution of the signal characteristics. The Gamma distribution is described by the equation below:

$$f(x) = x^{(a-1)}e^{(-x/b)} / b^a\Gamma(a),$$

where a and b are fit parameters for each species. The values for a and b for each (lock, signal, and SNR) distribution for each species are given in Appendix B.

The probability distribution functions that resulted in this fit for E. coli K-12 and 2000 randomly generated hypothetical ribosome tails are given in Fig. 2. In each probability density function plot, the value for the actual E. coli ribosome is marked by a point. Similar plots are yielded when the actual ribosome tails for each of the fourteen species are compared with their PDFs.

In the second part of our experiment, we ran a Genetic Algorithm for each species to find hypothetical ribosome tails that are optimal with respect to the lock and synchronization signal magnitudes, and the signal to noise ratio. In the following sub-sections, we describe the signal calculations and the parameters used for our Genetic Algorithm.

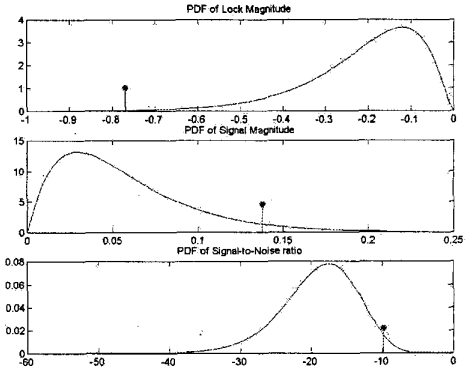

Fig. 2. E. coli Signal Characterisitic PDFs

2.1 Ensemble Average Signal Characteristics

For each candidate ribosome tail end for a given species, we calculate the ensemble average signal (see Appendix A) over a set of 500 certain coding sequences obtained from GenBank [9]. From this signal, we determine three parameters: 1) the magnitude of the synchronization signal, 2) the magnitude of the lock signal (which will be negative, since it represents a free energy "release"), and 3) the signal-to-noise ratio. Table 2 lists these parameters for the actual tail end in *E. coli.*

The magnitude of the synchronization signal (sync) is estimated using a method that takes advantage of our prior knowledge of its periodicity. We calculate running averages of every third position along the signal, and interpolate a sine wave through the three resulting points (see Appendix A). This method of calculating the magnitude works well in the presence of immense noise, which is characteristic of the *E. coli* genome [2].

An estimate of the "pure" signal is obtained using the calculated magnitude and phase. This estimated signal is subtracted from the noisy ensemble-average signal to get the noise signal. The ratio of the variance of the estimated "pure" signal to the variance of the noise yields an estimate of the signal-to-noise ratio (SNR).

The affinity of the 16s tail end to the Shine-Dalgarno sequence is measured by the minimum magnitude of the signal, roughly between positions 16 and 12 bases upstream from the start codon. This is referred to as the "lock" magnitude.

2.2 Genetic Algorithm Parameters

Since we hypothesized that the most important features of a ribosome tail end for optimal translation are the lock and synchronization signal magnitudes, and the SNR of the ensemble average signal, we designed an objective function to simultaneously optimize for these features.

The objective function, or fitness, for a given ribosome tail end is computed in three steps. First, we compute the lock, synchronization, and SNRs as discussed above. We then assume that each of these features is Gamma-distributed and calculate p-values for each feature. Third, the total fitness of a candidate tail end is the sum of these three p-values, and the GA optimizes for the minimum of this sum. Our GA, with population size 100, mutation rate 0.1, and no crossover, was then run using this function to search for optimal ribosome tail ends for each species.

3 Results & Discussion

Results for the Genetic Algorithm searches for each species are given in Tables C1-C14 in Appendix C. In each table, the actual ribosome tail for the species is given in bold italic. Five hypothetical ribosome tails are listed whose fitness value was lower (and therefore "more optimal"), and three are given whose fitness was less optimal. For each signal characteristic, the lock and signal magnitudes, and the SNR for the ensemble average signal (as computed in Appendix A) over 500 confirmed genes in the species are given.

For simplicity, we will discuss results found in Table C1, for the species *Aquifex aeolicus*. The first five rows of Table C1 list five optimal hypothetical ribosome tail ends whose fitness functions were better than that of the actual *Aquifex aeolicus* ribosome tail end. For each of these, the synchronization signal magnitudes and SNRs were better than those of the actual tail. The lock, corresponding to strong complementarity to the Shine-Dalgarno sequence, for the actual ribosome tail in this species, however, is much larger than those that the GA found with higher fitness. This suggests that the lock might be particularly important for translation of the genes in this species. Those hypothetical tails with higher (worse) fitness are still quite close and have similar signal and SNR values, and their lock magnitudes are similar to other top candidates.

As we found in our previous research, the fitness function was able to optimize its search to find hypothetical tails with better performance than the actual ribosome tail in most cases. Since we believe that good performance in all three parameters is important, these results show that the fitness function was appropriate in optimizing for the lock and synchronization signal magnitudes and the signal to noise ratio.

We remark here that most of these species obtained a lock magnitude of -0.57 E or stronger (i.e. more negative), indicating strong complementarity to the Shine-Dalgarno sequence. This finding did not hold for Rickettsia conorii, whose lock magnitude is -0.234 E, and whose signal magnitude is 0.069 E. These values may reflect major biological differences between Rickettsia conorii and other species in this group. Signal magnitudes for all other species, other than *Rickettsia* conorii and Thermoplasma volcanium (-0.054 E), lie between -0.1 and -0.22 E. It is interesting to note that, despite the good performance of the GA in finding hypothetical ribosome tails in other species with stronger lock and signal magnitudes, the GA was unable to find tails that performed significantly better than the actual ribosome for Rickettsia conorii and Thermoplasma volcanium on the 500 genes selected for study. This indicates that the lock may not be so important for Rickettsia, and the signal magnitude may not be so important for either Rickettsia or Thermoplasma. This implies that some other process may be driving the ribosome along the mRNA in these cases. SNR values for the actual ribosomes ranged from -18 dB to -6 dB. It would be interesting to investigate the biological reasons for the difference in the strength of the SNR across several species.

In all the species, we find that the phase of the actual ribosome differs significantly from the phase of the best ribosomes found by the GA. This suggests that phase may play a significant role in the selection of the optimal ribosome tail end.

4 Conclusion

The results of this experiment support our hypothesis that the lock and synchronization signals are important indicators of protein production in most of the fourteen different prokaryotes we selected. We have confirmed that, in these characteristics, each actual ribosome tail seems to differ significantly from a randomly generated hypothetical ribosome tail in these characteristics. Our work has also generated, for each of fourteen species, a set of base sequences that might be used in transgenic protein production. These "best performing" hypothetical tails can also be compared to the actual ribosome tails to determine characteristics that are important to protein translation in each of these species.

Our Genetic Algorithm fitness function was able to distinguish candidate tail ends, and optimize for only

those with favorable values for all three signal criteria. Future work can extend this fitness function to encompass other criteria that may be important in genetic translation. In particular, we plan to investigate the role of the phase of the synchronization signal in these organisms.

In conclusion, our findings suggest that the "ideal" ribosome tail, for most species investigated, needs all of these characteristics: a strong lock to initiate protein production, and a strong synchronization signal, that is well-differentiated from noise, to drive it along.

Acknowledgements

We would like to thank David DeGerolamo, Joshua Starmer and Dr. Anne-Marie Stomp of North Carolina State University for their help in choosing prokaryotic organisms for this experiment.

References

[1] B. E. Schoner et al. "Role of mRNA translational efficiency in bovine growth hormone expression in Escherichia coli." Proc. Natl. Acad. Sci. USA, 81:5403-5407, 1984.

[2] D. I. Rosnick. "Free Energy Periodicity and Memory Model for Genetic Coding," PhD Dissertation, North Carolina State University, 2001.

[3] L. Ponnala, T. M. Barnes, D. L. Bitzer, M. A. Vouk. "The search for the optimal ribosome 3' tail end in *E. coli*", Proceedings of the 26th Annual International Conference of the IEEE EMBS, September 2004.

[4] E. E. May, "Analysis of Coding Theory Based Models for Initiating Protein Translation in Prokaryotic Organisms," PhD Dissertation, North Carolina State University, 2002.

[5] M. Mishra, S. Vu, D. Bitzer, M. Vouk, A. Stomp. "Coding sequence detection and free energy periodicity in prokaryotes", Workshop on Genomic Signal Processing and Statistics (GENSIPS), May 2004.

[6] S. M. Freier et al. "Improved free-energy parameters for predictions of RNA duplex stability," *Proc. Natl. Acad. Sci. USA*, 83: 9373-9377, 1986.

[7] A. V. Oppenheim and R. W. Schafer. *Discrete Time Signal Processing*, Prentice Hall, March 1989.

[8] The Gutell Laboratory at the Institute for Cellular and Molecular Biology, University of Texas at Austin. http://www.rna.icmb.utexas.edu/

[9] Online resource for genomes, GenBank http://www.ncbi.nlm.nih.gov/Genbank/index.html

[10] Tiffany M. Barnes, "Using Genetic Algorithms to Find the Best Generators for Half-Rate Convolutional Coding," North Carolina State University, Raleigh, NC, 1994.

Appendix A: Signal Calculations

Let $e_i(x)$ denote the free energy score estimate for gene i at position x along the mRNA as computed in [2]. Let N denote the number of genes in the sample. The ensemble average signal $S(x)$ at position x is:

$$S(x) = \frac{1}{N}\sum_{i=1}^{N} e_i(x) \tag{1}$$

In the following, we calculate the magnitude M and phase ϕ of the synchronization signal. For a coding sequence of length k codons, we calculate three quantities, A, B and C. We begin these calculations 90 bases, i.e. 30 codons, downstream from the start codon, after which strong periodicity of the signal is observed [4].

$$A = \frac{1}{k-30}\sum_{x=90,93,96,\ldots}^{3k-3} S(x) \tag{2}$$

$$B = \frac{1}{k-30}\sum_{x=91,94,97,\ldots}^{3k-2} S(x) \tag{3}$$

$$C = \frac{1}{k-30}\sum_{x=92,95,98,\ldots}^{3k-1} S(x) \tag{4}$$

These quantities represent the average signal over the entire coding sequence. We subtract the constant DC term from these quantities to remove any bias, resulting in the points a, b, and c, given in (6).

$$DC = (A + B + C)/3 \tag{5}$$

$$a = A - DC, b = B - DC, c = C - DC \tag{6}$$

We interpolate a sine wave (of magnitude M and phase ϕ) through a, b, and c using the formulae given below:

$$a = M\sin(\phi) \tag{7}$$

$$b = M\sin(\phi + 2\pi/3) \tag{8}$$

$$c = M\sin(\phi + 4\pi/3) \tag{9}$$

$$\phi = \arctan(\sqrt{3}a/(a+2b)) \qquad M = a/\sin(\phi) \tag{10}$$

Appendix B: Gamma Distribution Parameters for Each Species

Gamma Distribution Parameters for Each Species						
Species	**Lock Magnitude**		**Signal Magnitude**		**Signal-to-Noise Ratio**	
	a	**b**	**a**	**b**	**a**	**b**
Aquifex aeolicus	2.5109	0.0649	2.6623	0.0236	13.1474	1.2992
Bordetella bronchiseptica	2.1194	0.1379	1.7139	0.0637	14.7797	1.1934
Bacillus subtilis	1.2442	0.2186	2.5182	0.0202	17.5558	1.2187
Chlamydophila pneumoniae	4.0821	0.0459	2.8601	0.0161	15.0623	1.3060
Escherichia coli	2.3505	0.0887	2.1110	0.0267	12.7096	1.4919
Lactobacillus plantarum	1.4671	0.1684	2.3069	0.0273	16.8853	1.3181
Pseudomonas aeruginosa	1.8461	0.1659	1.7235	0.0711	17.6469	1.1049
Rickettsia conorii	7.4221	0.0106	2.8777	0.0136	14.7528	1.2538
Salmonella typhimurium	2.0624	0.1036	2.1754	0.0281	12.6653	1.4377
Salmonella enterica	2.1107	0.1073	1.9115	0.0428	20.6544	1.0545
Streptococcus pyogenes	2.3386	0.0749	2.6105	0.0226	14.0200	1.3669
Staphylococcus aureus	1.2239	0.1892	2.4298	0.0313	12.4596	1.2712
Streptococcus mutans	1.7324	0.1151	2.6984	0.0244	9.5766	1.5520
Thermoplasma volcanium	3.6642	0.0265	2.7745	0.0093	20.6189	1.0921

Table B1. Gamma Distribution Parameters for Each Species

Appendix C: Genetic Algorithm Results for Each Species

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
CTGCTGCTCCTCCT	-0.6744	0.382057	-1.585326	-8.2846	0.009722
GAGCTGCTCCTCCT	-0.7172	0.336574	-1.612809	-8.5777	0.012974
CGGCTGCTCCTCCT	-0.7032	0.351470	-1.589587	-8.6278	0.013756
CAGCTGCTCCTCCT	-0.6932	0.329851	-1.616614	-8.6936	0.014814
CAGCGCCTCCTCCT	-0.9702	0.311578	-1.890936	-8.8379	0.016385
ATTTCCTCCACTAG	**-1.014**	**0.134273**	**2.058083**	**-11.76**	**0.169283**
CCGCGGCCTTCTGA	-0.3026	0.190858	-0.231913	-10.709	0.172611
CGGCGACCTCCTTA	-0.7378	0.156908	-0.549274	-12.373	0.177360
CGGCTCCAACTCTG	-0.6852	0.147070	-1.752580	-12.216	0.177967

Table C1. Acquifex aeolicus

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
CGACCTCCGGTGG	-1.2066	0.524486	0.111048	-9.6922	0.022549
TCCCGTGGCCGCT	-0.9198	0.678074	2.899560	-9.1944	0.022969
TCCCGGGGCCGCT	-1.0950	0.608491	-3.124157	-10.256	0.034719
CCACCTCCGGTGC	-0.8510	0.518414	-0.054488	-9.5182	0.035878
CGACCTGCGGTGG	-0.9902	0.478973	0.428847	-10.076	0.037589
TTTCCTCCACTAG	**-0.6416**	**0.222913**	**0.365266**	**-9.558**	**0.192187**
TCCCGTGACCGTG	-0.5752	0.381477	2.956355	-11.857	0.192316
TCCCGTGGTGGTG	-0.4748	0.418956	2.414503	-9.9878	0.194206
CCCGTGGTGCGGG	-0.8142	0.353023	0.036781	-13.062	0.195571

Table C2. Bordetella bronchiseptica RB50

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
TCGATTGCCTCCACT	-1.3920	0.176753	1.597825	-12.157	0.019019
ACTGGGGCCTCCACG	-1.3074	0.192421	1.602838	-12.308	0.020657
GAGCCGCCTCCGCCT	-1.2854	0.236023	-0.244671	-12.436	0.021344
AGTGGGTCCTCCACT	-1.6900	1.611278	1.611278	-12.286	0.021526
TCGTTTGCCTCCACT	-1.4088	0.172792	1.530784	-12.503	0.024457
TCTTTCCTCCACTAG	**-1.9860**	**0.106768**	**-2.576397**	**-14.402**	**0.126055**
GGCCTCTGGGTCCTC	-0.5998	0.162917	-2.398300	-13.036	0.130705
ACAACGGCCGCCCCT	-0.6882	0.174898	1.201511	-14.395	0.134168
ACAACGCCCGCCCCT	-0.9778	0.149101	0.859330	-15.407	0.134826

Table C3. Bacillus subtilis subsp. subtilis str. 168

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
GTCCTCCTCGAGGTC	-0.5906	0.198543	2.983161	-10.364	0.011638
GCTACTTCCTCTACC	-0.4664	0.167585	0.858716	-9.1699	0.011864
CTAACTTCCTCTACC	-0.4718	0.150007	0.826422	-9.618	0.016190
CCTCCTCCAGGACCA	-0.7520	0.184039	-1.124272	-10.771	0.016225
GCAACTTCCTCTACC	-0.4634	0.150243	0.857856	-9.5434	0.016945
TTTTTCCTCCACTAG	**-0.575**	**0.119317**	**3.129946**	**-11.01**	**0.039270**
GTCCTCCTTGAGGTC	-0.5470	0.151129	3.053628	-12.076	0.049171
CCTCCTCTAGGAACA	-0.7524	0.132249	-0.907145	-12.069	0.052296
CCTCCTCGAGGAACA	-0.6724	0.146960	-1.004224	-12.572	0.062673

Table C4. Chlamydophila pneumoniae CWL029

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
CTCCTCCTCCTCC	-0.9604	0.266911	-1.801841	-8.2120	0.002223
CACCTCCTCCTCC	-0.9036	0.279738	-1.830295	-8.3557	0.003199
GTCCTCCTCCTCC	-0.966	0.252857	-1.833466	-8.4280	0.003875
CGCCTCCTCCTCC	-0.9002	0.292774	-1.905205	-8.5066	0.004066
CTCCTCCTCCTTC	-0.9358	0.226248	-1.690487	-8.2015	0.004094
ATTCCTCCACTAG	**-0.7688**	**0.137845**	**-0.184043**	**-9.8162**	**0.061365**
GTCCTACTCCTTC	-0.4946	0.172841	-1.550298	-9.079	0.061837
GTTCTCCTCCTTT	-0.8730	0.135957	-1.619182	-9.829	0.062063
GTCCTCCTTGTTC	-0.8758	0.138574	-1.646571	-10.041	0.062133

Table C5. Escherichia coli K-12

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
CCCTCCCACTTCCAC	-1.1456	0.250064	2.845609	-11.938	0.011244
CCCTCCCACTTTCAC	-1.1864	0.237842	2.763478	-12.173	0.013781
CACTCCCACTTTCCC	-0.9836	0.218590	2.762998	-12.635	0.027054
CTGCATCTCCGACTC	-0.7710	0.258042	-1.510436	-11.243	0.028061
CTTCATCTCCGACTC	-0.8068	0.225151	-1.541969	-11.580	0.028449
TCTTTCCTCCACTAG	**-1.706**	**0.114735**	**-3.122419**	**-15.82**	**0.211261**
TTTCATCTCCAACTC	-0.7744	0.114041	-1.616498	-15.279	0.217134
TGCAGCTGCCACTTA	-0.3878	0.201618	3.003500	-12.667	0.218464
CGCACTACGCCTTCC	-0.3802	0.225870	0.727093	-12.511	0.219540

Table C6. Lactobacillus plantarum WCFS1

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
CGCCGCCGCCGGT	-1.5948	0.988286	-3.131063	-10.89	0.010810
AGCCGCCGCCGGT	-1.5452	0.9212	3.129184	-10.879	0.010847
CGCCGCTGCTGGT	-1.174	0.717702	3.027511	-10.579	0.012180
CGCCGCCGCTGGT	-1.2914	0.860090	3.080557	-10.881	0.012969
CGCCGCCGGAGGT	-1.3972	0.802813	3.135438	-11.069	0.014325
ATTCCTCCACTAG	**-0.8958**	**0.214904**	**0.304156**	**-12.172**	**0.202380**
AGGCGCCGGTTGT	-0.5518	0.454562	-2.687980	-13.182	0.204978
ACCGCAGCAGCGT	-0.6808	0.345049	0.742828	-14.156	0.210182
CGCCGCCAGTAGG	-0.7600	0.389718	-2.719456	-14.802	0.210740

Table C7. Pseudomonas aeruginosa

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
CTTCTCCTTCTAT	-0.1872	0.172527	-1.691118	-9.3593	0.009519
CTCCTTCTTCTAT	-0.1848	0.171855	-1.688626	-9.4896	0.011273
CTCCTTCTACTAT	-0.1702	0.162764	-1.634890	-9.4639	0.014616
CTTCTTCCTCTAT	-0.1648	0.159159	-1.363288	-9.2443	0.014675
CTCCTACTTCTAT	-0.1638	0.165802	-1.640774	-9.3403	0.015862
ATTCCTCCATTAG	**-0.2346**	**0.069519**	**-0.805266**	**-13.080**	**0.219663**
GTTCTACTTCTTT	-0.0994	0.162670	-1.511706	-9.2938	0.225087
CTCCTTCTAATCT	-0.1674	0.077859	-2.487901	-13.708	0.225683
GTGCCCCTTCTTT	-0.1212	0.083605	-1.544292	-12.663	0.227761

Table C8. Rickettsia conorii

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
ACCTCCTCGACGA	-1.0594	0.350371	2.547725	-7.5023	0.000522
ACCTCCTCCACGA	1.0132	0.321562	2.490761	-7.3398	0.000866
ACCTCCTCGACCA	-0.9784	0.342654	2.536780	-7.4481	0.001014
TCCTCCTCGACGA	-1.079	0.328372	2.539949	-7.6786	0.001046
ACCTCCTCGACTA	-1.0936	0.277153	2.473578	-7.3978	0.001121
ATTCCTCCACTAG	**-0.8776**	**0.146858**	**0.163748**	**-8.357**	**0.049797**
CACCTCGACTGTA	-0.8466	0.172819	-1.805557	-10.078	0.052888
ACCTTCTCGACGA	-0.4954	0.300702	2.624290	-8.1761	0.056588
CACCTCCTGTGTA	-0.8866	0.150701	-1.621553	-9.5096	0.059184

Table C9. Salmonella typhimurium

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
CGCCTCCGCCGCG	-0.7622	0.477802	-2.765245	-13.357	0.025184
TGCCTCCGCCGCG	-0.7460	0.459295	-2.752658	-13.336	0.026007
CGCCTCCGCCGTG	-0.7690	0.408871	-2.986875	-13.506	0.027785
TGCCTCCGCCGTG	-0.7460	0.388768	-2.973449	-13.581	0.031016
AGCCTCCGCCGTG	-0.7356	0.374271	-3.074664	-13.584	0.032212
ATTCCTCCACTAG	**-1.025**	**0.110043**	**0.023202**	**-14.403**	**0.292525**
TGCCTCAGCCTCG	-0.6754	0.155947	-2.562298	-17.477	0.306179
TGTCTCAGCCGTG	-0.3626	0.199222	2.950219	-16.055	0.315351
CGCCTCCGACGTG	-0.7840	0.181455	-3.048442	-18.383	0.318314

Table C10. Salmonella enterica subsp. enterica serovar Typhi str. CT18

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
ACGACCTCGACCTCG	-0.5244	0.296854	0.776419	-8.4715	0.012483
ATGTCCTCGACCTCG	-0.6402	0.223632	0.805434	-9.7355	0.014342
AAGTCCTCGACCTCG	-0.6334	0.217786	0.862119	-10.002	0.018304
TCCATTTCCACCACA	-0.5422	0.202948	0.902848	-9.3158	0.018906
TCAAGCACCTCCACC	-0.7086	0.162665	1.033755	-9.6442	0.025408
TCTTTCCTCCACTAG	**-1.1580**	**0.123829**	**3.031330**	**-11.948**	**0.115257**
CTGCTACCTCTGTCT	-0.5864	0.125092	-0.470355	-11.835	0.115492
TCAGCCTCGACTATC	-0.4254	0.137465	0.868454	-11.523	0.117114
TCAGCCTCGACTAGA	-0.4488	0.138739	0.713807	-11.889	0.118181

Table C11. Streptococcus pyogenes MGAS10394

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
TAATTTTCCACCACG	-1.0542	0.242550	0.600227	-7.2446	0.020937
TTATTTTCCACCACG	-1.0482	0.235333	0.605716	-7.2671	0.022967
TAATCATCCACCACG	-0.8402	0.260988	0.570533	-6.6962	0.026162
TAATAATCCACCACG	-0.8540	0.227407	0.598389	-6.9556	0.033407
TAATTATCCACCACG	-0.8550	0.222408	0.594320	-7.0451	0.035644
TCTTTCCTCCACTAG	**-2.192**	**0.161711**	**2.636484**	**-8.962**	**0.099778**
TAATTAGCCACCACG	-0.5838	0.196353	0.622602	-7.632	0.105360
GTCGAACTCCTCGTT	-1.0762	0.195430	3.037620	-10.037	0.110186
TAATTATCCACGACG	-0.5296	0.244909	0.587440	-7.8147	0.110617

Table C12. Staphylococcus aureus subsp. aureus MW2

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
GACAACCTCCACCAA	-0.8022	0.238462	2.968023	-5.3868	0.010120
GACCTCCACGACTCA	-0.7231	0.252450	2.922096	-5.5102	0.014147
AACCTCCACGACTCA	-0.7478	0.243938	2.931898	-5.6987	0.014478
GTTCCGACTTCCACT	-0.6910	0.292156	0.567677	-5.6655	0.016796
GTTACGACTTCCACT	-0.6358	0.335308	0.704897	-5.0313	0.018740
TCTTTCCTCCACTAG	**-1.2515**	**0.161973**	**2.919068**	**-6.7717**	**0.046366**
AACCTCCACGAGTCA	-0.6910	0.191698	2.941499	-7.2341	0.050069
AACCTCGACGACTCA	-0.4978	0.251193	2.839010	-5.7733	0.056633
AACCTCCAGGACTCA	-0.6597	0.166285	3.028264	-6.8150	0.057808

Table C13. Streptococcus mutans UA159

Ribosome	Lock (E)	Sync Mag (E)	Sync Phase (Rad)	SNR (dB)	Fitness
GACTCCAACT	-0.3508	0.073692	2.944629	-13.9121	0.030111
GACTCCCACT	-0.5528	0.076942	-3.009937	-14.2012	0.032043
GACTCCTGCT	-0.2386	0.089599	2.937780	-13.9432	0.036434
GACTCCTACT	-0.2726	0.082815	3.033298	-14.4529	0.039372
GGCTCCAACT	-0.3310	0.071120	2.919964	-14.3925	0.042197
CCTCCACTAG	**-0.5812**	**0.053719**	**-0.229560**	**-18.1783**	**0.245097**
TGCTCCGGCG	-0.1776	0.065039	2.731275	-17.589	0.248074
GACTACTGCT	-0.1256	0.084407	-3.106916	-12.146	0.251380
TTCTCCGGCC	-0.2256	0.055668	2.757214	-18.255	0.262799

Table C14. Thermoplasma volcanium