Intrinsic Terminator Prediction and Its Application in Synechococcus sp. WH8102

2 authors:

Xiu-Feng Wan
Mississippi State University
181 PUBLICATIONS **2,824** CITATIONS

SEE PROFILE

Dong Xu
University of Missouri
623 PUBLICATIONS **14,184** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project — Feral Swine Adapted Influenza A Virus Risk To Human and Domestic Swine Health View project

Project — National Surveillance for Avian Influenza in Wild Birds View project

Intrinsic Terminator Prediction and Its Application in *Synechococcus sp.* WH8102

Xiu-Feng Wan and Dong Xu[¶]

Digital Biology Laboratory, Department of Computer Science, University of Missouri, Columbia, Columbia, MO 65211
Emails: wanx@missouri.edu; xudong@missouri.edu.

Running title: Intrinsic terminator prediction

[¶]**Corresponding author.**

Abstract

We have developed a new method for intrinsic terminator prediction based on Rnall, an RNA local secondary structure prediction algorithm that we developed recently, and two U-tail score schemas. By optimizing three parameters (thermodynamic energy of RNA hairpin structure, U-tail T weight, and U-tail hybridization energy), our method can recognize 92.25% of known terminators while rejecting 98.48% of predicted RNA local secondary structures in coding regions (negative control) as false intrinsic terminators in *E. coli*. We applied this method to scan the genome of *Synechococcus sp.* WH8102, and we predicted 266 intrinsic terminators, which included 232 protein-coding genes, 12 tRNA genes, and 3 rRNA genes. About 17% of these terminators are located at the end of operons. We also identified 8 pairs of bio-directional terminators. Our method for intrinsic terminator prediction has been incorporated into Rnall, which is available at http://digbio.missouri.edu/~wanx/Rnall/.

1 Introduction

Transcription termination is one of the three key procedures of gene transcription, together with transcription initiation and elongation. Two categories of transcription termination modes have been discovered in eubacteria. The first category is factor-dependent termination. The efficiency of termination in this category depends on the transcriptional termination factor, such as rho protein, which may bind nascent RNA, catch up the RNA polymerase (RNAP) when RNAP is paused by RNA local secondary structure, and further melt the RNA-DNA duplex in the replication bubble to cause termination [1]. The second category of transcription termination uses intrinsic terminator, which is also called rho-independent terminator. It contains an RNA structural motif to terminate transcription without involvement of other factors (e.g. rho protein) [2]. To understand gene transcription in eubacteria, it is essential to study intrinsic terminator, which is ubiquitous in eubacteria. Although the detailed mechanism about rho-independent termination is still unclear, much progress has been made in characterization of this termination mode. Current understanding of the intrinsic terminator is: (1) An intrinsic terminator is composed of a G-C rich RNA hairpin loop and a U-rich tail (Fig. 1) [3]. The hairpin-loop structure may stalk the RNAP to proceed while the loose binding between RNA and DNA due to the rich "U" may result in the detachment of the RNA polymerase from the template, which leads to the termination of transcription process [4]. (2) The RNAP spanning regions in transcription elongation complex (TEC) can be defined as three continuous sites: the single RNA binding site (RBS) (5-8 nts), the RNA-DNA hybrid-binding site (HBS) (about 8 nts), and the double-stranded DNA-binding site (DBS) (about 9 nts) [5] (Fig. 1A). (3) Melting RNA-DNA hybrid is required for rho-independent termination [5]. Mutation of HBS may impair the termination process. (4) Substitution of the nucleotides in the stems of intrinsic terminators may not affect the transcription termination as much as in the HBS [5,6].

The unique features of the intrinsic terminators have been provoking computational studies for more than two decades. As early as 1984, Brendel and Trifonov [7] utilized the dinucleotide distribution matrix to extract the terminator signal. No energy measurement was taken into account in their method. In 1990, Carafa [3] developed a statistical method for rho-independent terminator prediction. An efficient T weight measurement for the HBS based on positional weight matrix was proposed in their research. TransTerm [8] employed this T weight measurement along with energy stability evaluation for the RNA hairpin structure to predict terminator. Furthermore, RNAmotif [9] is the first algorithm to utilize the thermodynamic parameters to measure the stability of hairpin-loop structure and its downstream sequence. The combined stability was assumed to be the determinant factor for the formation of an efficient intrinsic terminator. Although most algorithms deem the U-tail is a necessary component for an efficient transcriptional termination, GeSTer [10] assigned all the DNA palindrome sequences (which form RNA hairpin structures) at the intergenic regions as intrinsic terminators regardless of whether U-tails are present or not. All the above algorithms were able to retrieve known intrinsic terminators effectively. Nevertheless, a common problem for these methods is that their predictions contained many false positive terminators [3, 7-10].

Our terminator study will focus on *Synechococcus sp.* WH8102 (later referred to as *Synechococcus*), an important member of cyanobacteria, which produce approximately

20-40% of chlorophyll biomass and carbon fixation in the ocean [11]. *Synechococcus* is widely present in many marine planktonic ecosystems [12, 13]. The mobility of *Synechococcus* distinguishes itself from most of other cyanobacteria [11]. With unique physiological features with carbon fixation, *Synechococcus* potentially provides a key for global carbon modeling. However, the knowledge about this unique bacterium is still very limited. The genomic sequence of *Synechococcus* [11] has given us an opportunity to analyze this bacterium using computational techniques. For example, Chen *et al.* [14] predicted the operon structures for this bacterium. The computational analyses of the terminators may enhance the construction of the regulatory network of *Synechococcus* and provide further understanding of the biological mechanism of this bacterium.

In this paper, we describe a new intrinsic terminator prediction approach based on our recently developed RNA local secondary structural prediction algorithm [15] and two U-tail score schemas proposed by Carafa *et al.* [3] and Lesnik *et al.* [9]. By optimizing three parameters (hairpin energy, U-tail T weight, and U-tail hybridization energy), our approach can retrieve 92.25% known terminators while rejecting 98.48% of the predicted RNA local secondary structures in coding regions (negative control) as false intrinsic terminators in *E. coli*. We applied this method in *Synechococcus sp.* WH8102 with detailed analyses.

2   Material and Methods
2.1   Datasets
The *Synechococcus sp.* WH8102 and *E. coli* K12 complete genomes and their annotation were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov) updated in February 2004. The 147 known *E. coli* rho-independent terminators were obtained from Carafa *et al.* [3]. By using BLAST [16], 129 terminators were retrieved from the updated *E. coli* K12 genome sequence. We randomly selected 129 gene-coding regions from all of the protein genes with size over 300 bps (100 codons), as the negative control. The first and last 100 bps were removed to reduce the possibility for these sequences to have intrinsic terminators within coding regions [8]. To predict the intrinsic terminators in *E. coli* and *Synechococcus*, we extracted the 50 bps upstream and 280 bps downstream from each stop codon, based on previous statistics of intrinsic terminator distribution along the sequences [8]. If the intergenic region between the gene stop codon and its downstream gene was less than 280 bps, the intergenic sequence (together wither the 50 bps upstream sequence) was extracted instead.

2.2   Algorithm
Recent research demonstrated that the intrinsic terminator functions based on two structural motifs: the hairpin-loop structure followed by a U-rich tail [4]. The stability of the hairpin-loop structure and instability of U-tail block are the two key factors for transcription termination. Some researchers used the additive effects of the two components as the basis for intrinsic terminator prediction algorithms, for example, RNAmotif [9]. However, recent study shows that mutation of hairpin structure may not impair the function of terminator [5, 6]. This demonstrated that the hairpin loop and U-tail may have different roles for rho-independent termination. Here we developed a new intrinsic terminator prediction method based on this argument. To predict intrinsic terminator, we first predicted the hairpin-loop structures using Rnall. Then we filtered the

hairpin-loop structure using two U-tail parameters, i.e., T weight and hybridization energy, as illustrated in Fig. 2.

2.2.1 Hairpin-loop structure prediction

Recently, we have developed a novel algorithm, Rnall (http://digbio.missouri.edu/~wanx/Rnall/), for RNA local secondary structure prediction using symmetric mapping and dynamic programming [15]. This method has been demonstrated to be efficient and effective in hairpin-loop structure retrieval for rho-independent terminators in *E. coli*, and RNA motifs in the HIV genome and large RNA molecules, such as tRNAs and rRNAs. The detail of the method was described in [15], and here we briefly summarize this method.

For a given RNA secondary structure, we can consider it as a mapping between the left half of the structure and the right half of the structure, based on the canonical bond formation rules [5]. The canonical pairs (AU, GC, and GU) of the sequences will form the stem of secondary structure. The "mismatch", "insertion", or "deletion" will form the hairpin loop, bulge loop, and internal loop. To scan the local secondary structures along RNA sequence, a sliding window with size *W* is utilized. Only the structures with the maximum symmetric mapping score for each window will be retrieved. The thermodynamic energy of the predicted structures is computed using *efn2* based on the most recently updated thermodynamic energy parameters [17].

To preserve the biological significance [17, 18], Rnall utilizes a subsequence located in the middle of the window *W* as an initial hairpin loop in the predicted structure. The default size of this subsequence is 4 since more than 99% of experimentally verified rho-independent terminators (146 out of 147) have the hairpin loop size of 4.

Previous analyses of the experimentally verified structures demonstrated three more important features for intrinsic terminators [3]: (1) the stem size verified from 3 to 10 pairs; (2) high GC content; and (3) the hairpin-loop structure may have 1 bulge loop or 1 internal loop with the maximum loop size of 3. Thus, a window size of 30 is selected for the hairpin-loop structure prediction. To keep the high GC content, the 3' terminal U and more than 3 consecutive U's on the 3' stem are prohibited when retrieving the hairpin-loop structures. The predicted hairpin-loop structures are also filtered if the total number of bulge loop and internal loop is more than 1.

2.2.2 U-tail definition

Different feature vector definitions have been proposed for the U-tail. Carafa *et al.* [3] defined a weighted positional weight matrix:

$$T = -\sum_{n=1}^{15} x_n \tag{1}$$

for all U's in the 15 nucleotide segment where $x_0 = 1$ and

$$x_n \begin{cases} x_{n-1} \times 0.9 & \text{if the } n\text{th nucleotide is a U} \\ x_{n-1} \times 0.6 & \text{if the } n\text{th nucleotide is other than U.} \end{cases} \tag{2}$$

Carafa *et al.* [3] and Ermolaeva *et al.* [8] applied this method in intrinsic terminator prediction.

After calculation of the RNA/DNA hybridization parameters using nearest-neighbor model [19], the energy of U-tails was computed using thermodynamic parameters, which was employed by RNAmotif for intrinsic terminator prediction [9]. RNAmotif divided the U-tails into four regions: spacer (the first two nucleotides), proximal part (the 5 nucleotides following the spacer), distal part (the 4 nucleotides following the proximal part), and extract part (3 nucleotides following the distal part) (Fig. 1B). The score of the U-tail was calculated for hybridization energy by the following formulas:

$$T = \boldsymbol{DG}^0_{37(spacer)} + \boldsymbol{DG}^0_{37(proximal\,part)} + 0.5 \times \boldsymbol{DG}^0_{37(distal\,part)} + 0.01 \times \boldsymbol{DG}^0_{37(extract\,part)} \quad (3)$$

As illustrated in Fig. 1B, the information content of "U" decreases along with U-tail from 5' to 3'. Apparently neither of the above two definitions of the U-tail are perfect: (1) The T-weight approach did not differentiate A, G, and C thus it may not be able to distinguish an RNA-DNA hybrid with high energy and the one with low energy, which will affect significantly the hybridization efficiency. For example, UUUAAAAA may have the same T weight as UUUGGGCC, but their hybridization efficiencies are quite different. (2) In contrast, the energy calculation in equation 3 cannot reflect the sequence attribute (e.g. U distribution). To overcome their disadvantages, we used both definitions of U-tail in our algorithm.

2.2.3   Intrinsic terminator retrieval

Given one sequence, many local structures may be retrieved after Rnall prediction and some of them may overlap. To select non-overlapping local secondary structures, we first discard the structures that do not fit the 3 conditions of intrinsic terminators defined in section 2.2.1. The hairpin-loop structures with energy lower than threshold will not be considered in next retrieval procedure. In the end, we will have a pool of hairpin-loop structures for further extraction. Within this pool, we first choose the structure with the lowest energy. All other local secondary structures overlapped with the selected structure, together with the selected structure, are removed from the local secondary structure pool. Then we select the next local secondary structure with the lowest energy in the remaining structure pool. By recurrence, all the non-overlapping RNA local secondary structures are retrieved. Then we filter these hairpin structures by U-tail filter function (equation 1 or 2).

2.2.4   Optimizing computational parameters

The best performance of this algorithm depends the optimization of the three thresholds for hairpin energy (G), T weight (T), and hybridization energy (H). We proposed that all of three parameters (G, T, and H) are necessary to be above their thresholds to obtain an efficient transcription termination. We employed the orthogonal array to optimize these three parameters [20]. Simply, three initial thresholds were taken for G, T, or H. Six different experiment sets were run to determine the best combination of GTH by the optimal sum of sensitivity and specificity. The "guess" of next threshold values depends on previous experiments. After a number of iterations, the approximately optimal thresholds for G, T and H were generated.

2.2.5   Software

The above computational approach for intrinsic terminator prediction was implemented in the C/C++ programming language and incorporated into the RNA local

secondary structure prediction package Rnall. Rnall is available at http://digbio.missouri.edu/~wanx/Rnall/.

3   Results
3.1   Distributions of G, T, and H for known intrinsic terminators in *E. coli*

To optimize the thresholds of G, T, and H, we analyzed the 129 known terminators and 129 gene-coding sequences (see 2.1). The G, T and H of the known terminators ranged from –5.4 to –25 Kcal/mol, 2.15 to 6.16, -3.2 to –13.5 Kcal/mol, respectively. Without any constraints, our algorithm predicted 2,758 RNA local secondary structures for the gene-coding sequences.   To analyze their statistical distributions, a peak or decay regression analysis (p<0.0001) was performed using Sigma Plot (SYSTAT Software Inc., Chicago, IN) for G, T, and H distributions (Fig. 3). From these results, G, T, and H have different distributions in the known terminators from the negative control. The interceptions for G, T and H are –7.8 Kcal/mol, 3.63, and –8.15 Kcal/mol, respectively, which represent the best parameters to differentiate these two datasets using the corresponding single parameter, G, T, or H. For example, -7.8 Kcal/mol would be the best hairpin energy threshold if only G is used to predict the terminator. By using the interceptions for G, T and H as thresholds, we retrieved 93.80%, 96.90%, and 95.25% known terminators and reject 80.09%, 86.33%, and 92.06% structures in the coding regions (negative control) as false intrinsic terminators, respectively.

3.2   Optimization of G, T, and H thresholds

To further improve the prediction accuracy, we used orthogonal array to tune the three thresholds. The final G, T and H thresholds were –7.8 Kcal/mol, 3.00, -11.7 Kcal/mol, respectively. By using these parameters, we retrieved 92.25% in the known terminator dataset and reject 98.48% structures in the coding regions.

3.3   Scanning terminators in *E. coli*

Rnall retrieved 1,193 intrinsic terminators for 1,112 genes among 4,211 protein-coding genes, 22 rRNA genes, 86 tRNA genes, and 48 other RNA genes annotated in the *E. coli* genome. This number is much smaller than 1,883 terminators predicted by GeSter [10], and 2,974 by RNAmotif [9], although it is larger than 567 terminators identified by TransTerm with confidence higher than 98%. The energy of the hairpin is mainly (90.78%) less than –18 Kcal/mol (Fig. 4A). About 73.34% of the terminators are located within 70 nts downstream the stop codons (Fig. 4B). No terminators were found beyond 260 nts downstream of the stop codons. The number of predicted terminators decreases as the T-weight threshold or the hybridization-energy threshold increases (Fig. 4C&4D). Rnall also predicated 1,412 terminators within intragenic regions, which are less than the 2,586 terminators identified in intragenic regions by RNAmotif [9].

3.4   Scanning terminators in *Synechococcus sp.* WH8102

Within the 2,569 genes (2,517 protein-coding genes, 44 tRNA genes, and 8 rRNA genes) annotated in the *Synechococcus* genome [11], we identified 266 intrinsic terminators for 247 (9.61%) genes (232 protein-coding genes, 12 tRNA genes, and 3 rRNA genes). Among these genes, some of their gene homgies were reported with know

intrinsic terminators in *E. coli* [3]. These genes include *gln*A, *glt*S, *gua*B, and *inf*A. The stem of the terminators identified in *Synechococcus* varied from 5-13 nts, and the loop size varied from 3 to 15 nts. About 42% and 27% of the terminators have the loop size of 4 and 5, respectively. Different from the abundant terminator loop patterns (UUCG and GAAA) in *E. coli*, the loop of the intrinsic terminators in *Synechococcus* are much more diverse. Although the GC% in *Synechococcus* (59%) is higher than that in *E. coli* (51%), the hairpin-loop energy distribution in *Synechococcus* is similar to that in *E. coli*, and more than 90% hairpin loops in the predicted terminators have energy less than -18 Kcal/mol (Fig. 5A). The GC% distribution of the predicted terminators in *Synechococcus* and *E. coli* are also similar (data not shown). The T-weight and distance distributions of the predicted terminators in *Synechococcus* are similar to those in *E. coli* as well (Fig. 5B&5C). The average of T weight is 4.22 and 4.31 in *Synechococcus* and *E. coli*, respectively. Similar to *E. coli*, over 61% of the terminators is located within 70 nts downstream the stop codon. However, the average hybridization energy of the terminators in *Synechococcus* (-8.5 Kcal/mol) is lower than the one in *E. coli* (-7.5 Kcal/mol) (Fig. 5D). This may be due to the higher GC% of the genomic sequence in *Synechococcus*.

The bio-directional terminators can terminate the transcription process for two neighboring genes, which are located in different strands of the genomes but shares the 3' intergenic region [2]. They are believed to be one of the most efficient terminators. Such terminators have been reported in *E. coli* [9]. Here we identified 8 putative bi-directional terminators to terminate 8 pairs of genes (Table 1). These genes include psaJ/gmk, masA/rRNA, pgm/hypothetical protein, rpl12/hypothetical protein, heat shock protein/hypothetical protein, cynS/hypothetical protein, ABC transporter/dnaK2, and hemolysin-like protein/apcC. The hairpin-loop structures of these bi-directional terminators have A-rich upstream sequences and U-rich downstream sequences.

Chen *et al.* [14] recently identified 537 operons, which included 1,454 genes in *Synechococcus* by combining multiple types of genomic data, such as intergenic distance, COG gene function, and phylogenetic profile. Among the 266 intrinsic terminators we identified in *Synechococcus*, 71 terminators are associated with 70 operons, including 46 located at the end of the operons whereas 25 within the operons (Table 2). The intrinsic terminators found within operons may suggest transcriptional attenuation mechanisms [2] in *Synechococcus*. Further identification of other features for attenuators and anti-termination signals would shed some light on the roles of these structures, which will not be discussed in this paper.

The 52 RNA genes in *Synechococcus* are composed of 42 tRNA genes and 8 rRNA genes. Within the 52 RNA genes, Rnall identified intrinsic terminators after 12 tRNA genes (4 tRNA-Ala, 2 tRNA-Pro, 2 tRNA-Ser, 1 tRNA-Arg, 1 tRNA-Met, 1 tRNA-Phe, and 1 tRNA-Asn) and 3 rRNA genes (rrnA, 5s rRNA and rrnB) (Table 3). The intrinsic terminator of tRNA-Ala (start position: 1874314) is located within the operon tRNA-Ala–tRNA-Ile. It is interesting that this terminator is located downstream of the operon tRNA-Ile–tRNA-Ala (start position: 2081254). The 5s rRNA and rrnA shares the same intrinsic terminator at the end since these two genes shares the same stop codon. This terminator may regulate the divergent expression of these genes. The other terminators are listed in Table 4.

Rnall also predicted 792 terminators in the intragenic regions, and 192 (24%) of these terminators are located within the 100 bp downstream the start codon (84, 11%) and upstream the stop codon (98, 13%), both of which might be true intrinsic terminators. It is interesting to note that Rnall identified 15,726 local secondary structures in *Synechococcus* whereas 13,819 in *E. coli* although the former genome size is only about half of the latter [15]. However, there are about 26% genes in *E. coli* have been predicted as intrinsic terminators whereas less than 10% in *Synechococcus*.

4   Discussion

In this paper, we developed a new computational approach to detect intrinsic terminator based on Rnall, an RNA local secondary structural prediction algorithm, and two U-tail score measurements. By using the known terminators in *E. coli*, we optimized the threshold parameters for G, T, and H. Then we applied this method to explore the intrinsic terminators in *Synechococcus*. We identified 266 intrinsic terminators for 247 genes among 2,569 genes annotated in *Synechococcus*.

In the past two decades, several computational methods have been developed for intrinsic terminator predictions. The main differences among these methods are on the ways in which they calculated the score for hairpin-loop structure or U-tail sequence. None of these methods were able to avoid false positive results. By combining the two score schemas, T weight [3] and hybridization energy [8], which complement each other, we were able to reduce greatly the false positives (the number of terminators within intragenic regions). For example, the predicted terminators in the intragenic regions are less than 55% of those generated by RNAmotif [8]. We also explored the filtering processes by using combinations of two threshold, i.e., hairpin + T weight or hairpin + hybridization energy. With hairpin energy (-7.8 Kcal/mol) and U-tail T weight (3.0) for thresholds, Rnall predicted 2,521 terminator. With hairpin energy (-7.8 Kcal/mol) and U-tail hybridization energy (-11.7 Kcal/mol) for thresholds, Rnall predicted 1,317 terminators. Meanwhile, the terminators predicted within the intragenic regions by these two methods are 57% and 296% more than the combination of G, T, and H. Thus, combination of three parameters (G, T, and H) increases the specificity of the predictions significantly.

Our computational approach can retrieve 92.25% terminators in the known terminator dataset while rejecting 98.48% structures in the intragenic regions (negative control). However, there are still some limitations of our approach, which include: (1) The thresholds of G, T, and H are optimized approximately, which can be improved. (2) The prediction is in binary mode (yes or no for terminator). There is no statistical assessment for predicted terminators. Future work will include discrimination analysis, such as support vector machine (SVM) to optimize the thresholds for G, T and H, and a statistical measurement to assess the probability for a predicted structure to be a true intrinsic terminator.

Wang et al. [21] found that the GeSTer [10] and GCG [7] are highly comparable for intrinsic terminator prediction and suggest the GeSTer may be one of the best algorithm for intrinsic terminator prediction. However, not all the hairpin loops (even with very stable/ low thermadynamic energy) function as intrinsic terminators. We identified many hairpin loop structures in the intragenic regions in WH8102. These structures would be less likely to an intrinsic terminator. Based on our results, the

combination of G, T, and H in a more stringent parameter set may result in a potentially lower false positive rate.

The assumptions for Rnall included the two structural patterns of the intrinsic terminators: the hairpin loop structure and U-trail. However, recent research has found that U trail may not be necessary pattern of the intrinsic terminators [4, 22–25]. Since no rho gene homology was present in *Synechococcus*, it is possible that *Synechococcus* may use very different transcription termination machineries. Thus, Rnall may result in a high false negative rate regard to intrinsic terminator prediction. Another disadvantage for Rnall is that utilization of the parameters generated from *E. coli* cross species might not be appropriate if different patterns exist due to evolution. Future analyses of the sequence around the stop codon for those genes with experimentally verified intrinsic terminators may shed some light for terminator patterns within different species.

The comparison of the attributes of the terminators predicted in *E. coli* and *Synechococcus* indicates that the abundance of intrinsic terminators in *Synechococcus* is much less than that in *E. coli*. We found that, compared with *E. coli*, there is no sharp energy drop around the stop position of the open reading frames in *Synechococcus* (data not shown), which is similar to *Synechocystis* PCC6803 [26]. These results suggest less intrinsic terminators with hairpin loop structures might be present in *Synechococcus*. The GC% composition shows little effects on the G or T of the predicted terminators. Instead, the average H of *Synechococcus* is higher than that in *E. coli* and this reflects the GC% of the genomes has an impact on U-tail hybridization stability. Further experimental verification of these terminators will give a more informative picture for the gene regulation in *Synechococcus*.

Previous research demonstrated *Synechococcus* is abundant (about 57% genes) in operon structures [14]. The predicted 46 terminators at the end of the operons control 148 genes. This reflects the economic termination mechanism of this bacterium. It is interesting to characterize the 25 terminators predicted within the operon structures, which may reflect the transcription attenuation mechanisms in *Synechococcus* [27]. Systematic analyses of these terminators may give a better understanding of transcription attention and anti-termination in *Synechococcus*, e.g., whether the attention mechanism in Trp operon of *E. coli* is present in *Synechococcus* [27].

Legends:

Fig. 1. Illustration of intrinsic terminators. (A) Intrinsic terminators are composed of two functional units: a hairpin-loop structure and a U-rich tail. (B) U-tail can be separated into three components: U section from 1 to 7 nt, distant section from 8 to 11 nt, and extra section from 12 to 14 nt.

Fig. 2. Workflow for Rnall to predict intrinsic terminators.

Fig. 3. Distributions of hairpin energy (A), T weight (B), and hybridization energy (C) for the known terminators and intragenic regions (negative control).

Fig. 4. The attributes of putative terminators predicted by Rnall in *E. coli* K12. (A) Histogram of hairpin energy. (B) Histogram of distance from gene stop codon. (C) Histogram of T weight. (D) Histogram of hybridization energy.

Fig. 5. The attributes of putative terminators predicted by Rnall in *Synechococcus sp.* WH8102. (A) Histogram of hairpin energy. (B) Histogram of distance from gene stop codon. (C) Histogram of T weight. (D) Histogram of hybridization energy.

Fig. 1

A

U tail

0-3

Stem: 5-14
Loop: 3-10

B

U-tail
(1-7)

Distant
(8-11)

Extra
(12-14)
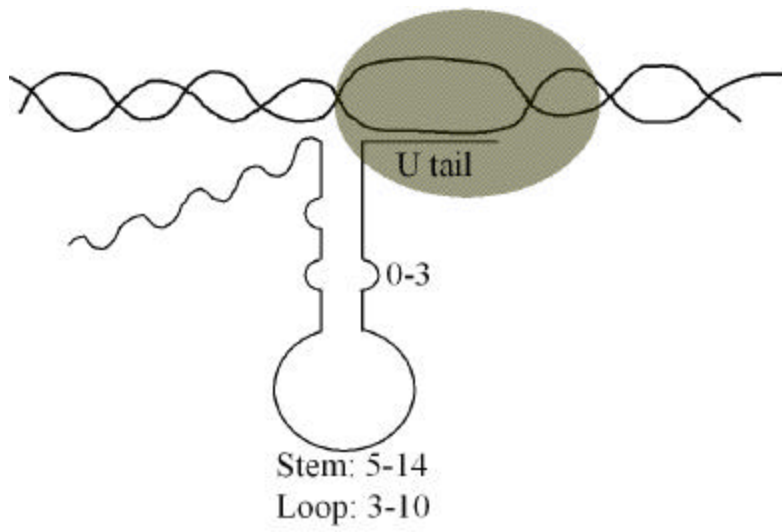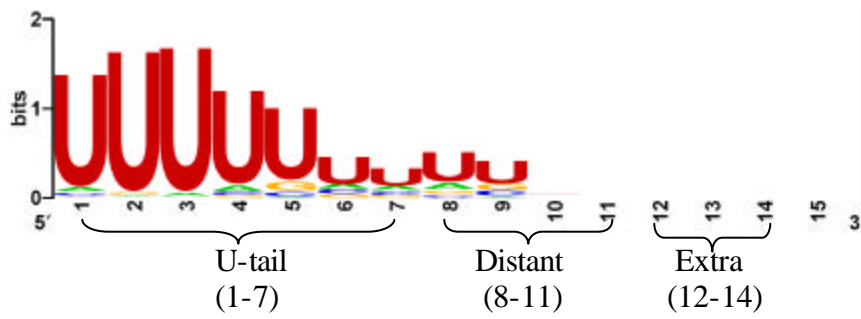
Fig. 2
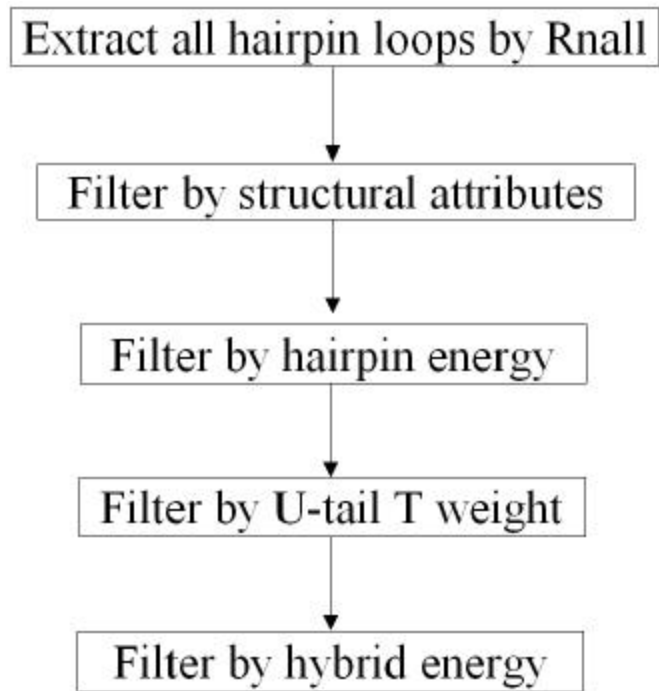
Fig.3

A
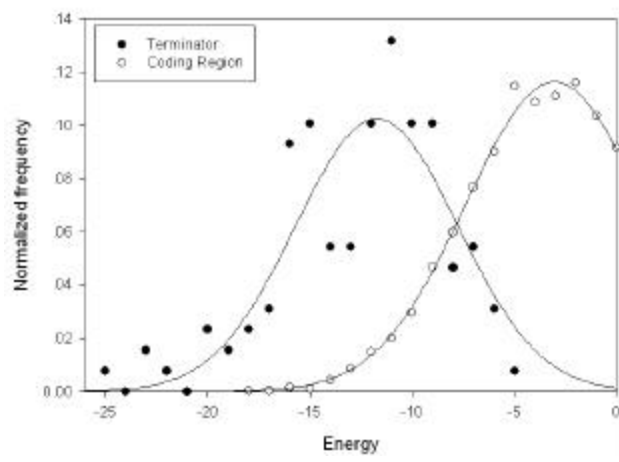
B
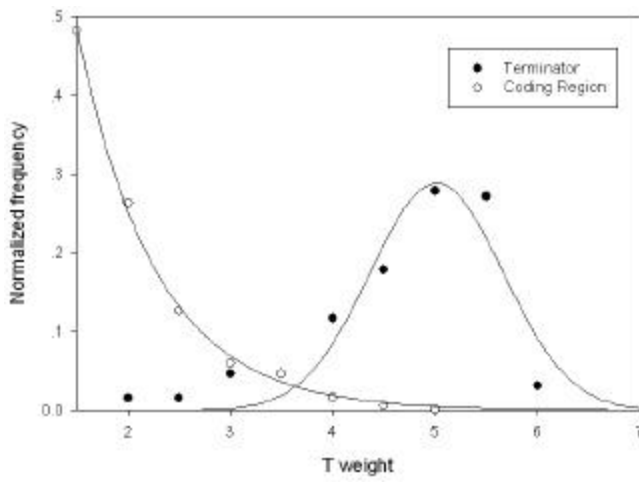
C
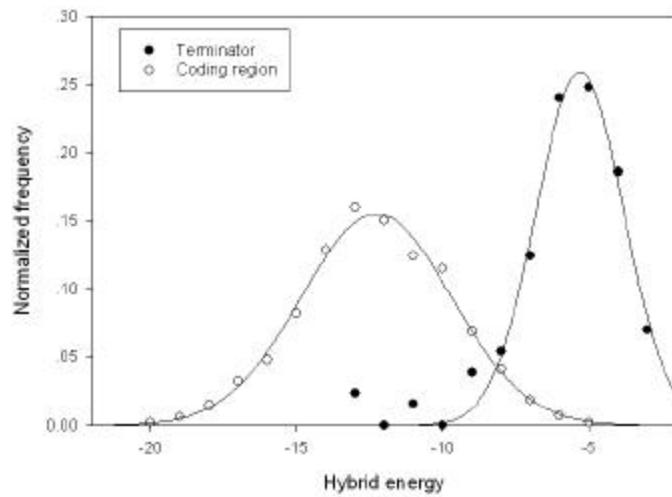
Fig. 4

A
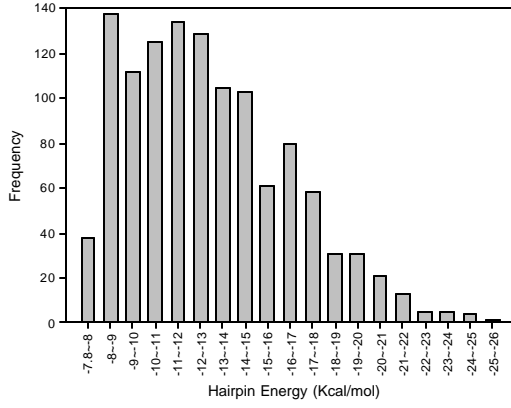
B
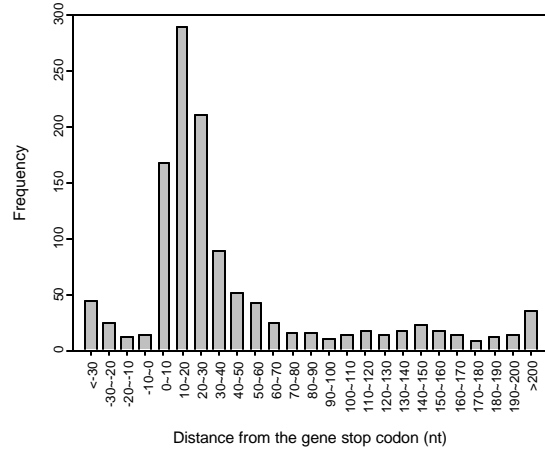
C
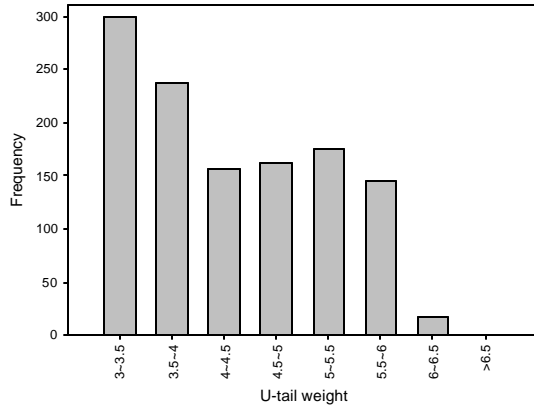
D
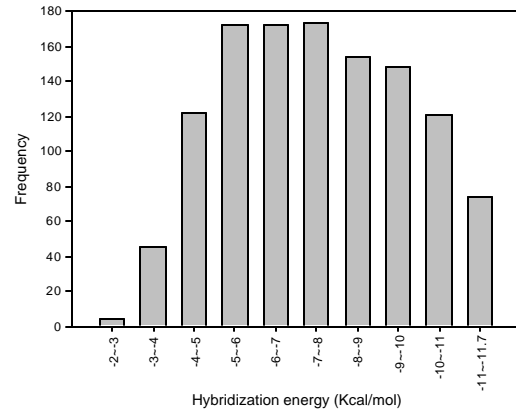
Fig. 5

A
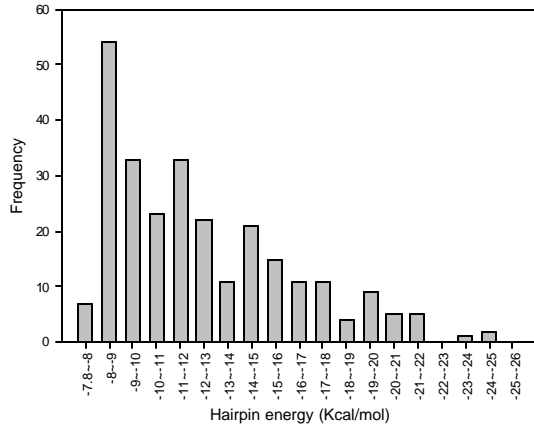
B
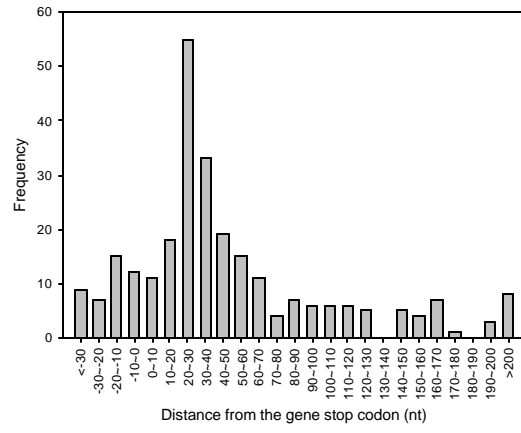
C
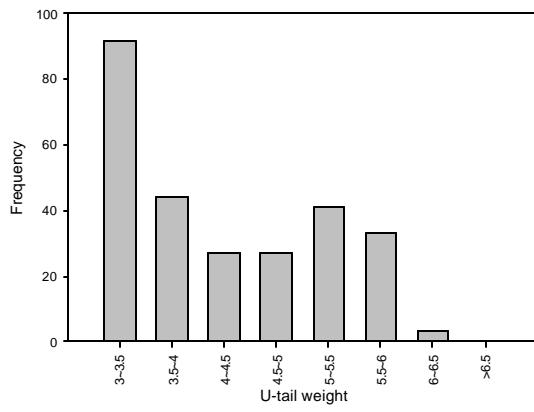
D
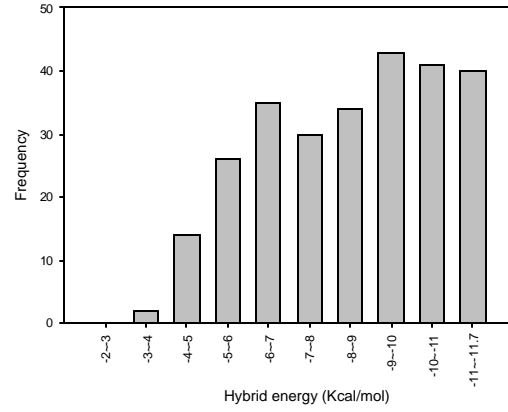
Table 1. Putative bi-directional intrinsic terminators identified in *Synechococcus sp.* WH8102.

gene name$	start	stop	strand	dis@	5' A-tail	5'stem	loop	3' stem	3' U-tail	G#	T#	H#
photosystem I reaction center subunit IX (*psaJ*)	1754127	1754243	+	29	CAAUAUCAAAAAAA	GAGGGGG	AUUUA	UCCCCUC	UUUUUUUAUGCCAUC	-12.2	5.63	-5.5
Guanylate kinase (*gmk*)	1754303	1754851	-	15	AUGGCAUAAAAAAA	GAGGGG	AUAAAUC	CCCCUC	UUUUUUUGAUAUUGG	-9.9	5.63	-4.7
putative enolase-phosphatase E-1 (*masA*)	1869880	1870617	+	56	CAAGACACAAAAAA	CCACCUC	AUAGU	GAGGUGG	UUUCUUCGUGGGGGU	-11.2	4.25	-10.5
RRNA	1870762	1870876	-	73	CCCCCACGAAGAAA	CCACCUC	ACUAU	GAGGUGG	UUUUUUGUGUCUUGC	-11.2	5.47	-6.3
Phosphoglucomutase (*pgm*)	2045204	2046862	+	40	CCUGUCAUGCAAAA	GCCACCC	CUCG	GGGUGGC	UUUUUUGAUGGGUUC	-15.1	5.18	-8.1
CHP	2046937	2047233	-	20	AACCCAUCAAAAAA	GCCACCC	CGAG	GGGUGGC	UUUUGCAUGACAGGA	-14.1	4.18	-9.7
CHP	2241872	2242618	+	157	CCACUUAAAAAAAA	CCCCUGC	CUUCG	GCGGGGG	UUUUUUAUUGAGCAA	-12.1	5.45	-5.0
50S ribosomal protein L7/L12 (*rpl12, rplL, rpl7*)	2242893	2243288	-	102	UGCUCAAUAAAAAA	CCCCCGC	CGAAG	GCAGGGG	UUUUUUUUAAGUGGA	-8.1	5.81	-3.7
putative small heat shock protein	2272670	2273056	+	45	CGACGAACGAAAAA	CCCCCGC	CGUG	GCGGGGG	UUUUUCGUUGCCCUG	-14.4	4.87	-8.4
putative integral membrane protein	2273142	2274509	-	26	AGGGCAACGAAAAA	CCCCCGC	CACG	GCGGGGG	UUUUUCGUUCGUCGU	-15.3	4.91	-7.4
CHP	2392602	2393327	+	3	GUUUAGAACCUAAA	GCCCAAU	AAGA	AUUGGGC	UUUAGAUAAGUCCAA	-9.0	3.67	-8.3
cyanate lyase (cynS)	2393373	2393816	-	28	UGGACUUAUCUAAA	GCCCAAU	UCUU	AUUGGGC	UUUAGGUUCUAAACU	-9.4	3.81	-9.0
ABC transporter, substrate binding protein, phosphate	2412101	2413111	+	34	CACAACACAAAAAA	GGGGCCC	GUAU	GGGCCCC	UUUUUUGAUGGCUAU	-15.8	5.17	-8.0
Molecular chaperone DnaK2, heat shock protein hsp70-2 (dnaK2)	2413189	2415102	-	29	UAGCCAUCAAAAAA	GGGGCCC	AUAC	GGGCCCC	UUUUUUGUGUUGUGG	-15.8	5.52	-6.0
hemolysin -like protein	483260	484075	+	17	GGACAACAAAAAAA	GCCCGACC	GCAGC	GGCCGGGC	UUUUCAGCUGAUCGU	-11.0	4.16	-9.9
linker polypeptide, allophycocyanin-associated (apcC)	484135	484335	-	25	CGAUCAGCUGAAAA	GCCCGGCC	GCUGC	GGUCGGGC	UUUUUUUGUUGUCCA	-15.5	5.89	-4.8

$: HP denotes hypothetical proteins and CHP denotes conserved hypothetical proteins.

@: dis represents the distance from the gene stop codon.

#: G, T, and H represent the thresholds for hairpin loop (Kcal/mol), U-tail T weight, and hybridization energy (Kcal/mol), respectively.

Table 2. Putative intrinsic terminators identified for operons in *Synechococcus sp.* WH8102.

gene name[$]	start	stop	strand	dis[@]	5' A-tail	5'stem	loop	3' stem	3' U-tail	G[#]	T[#]	H[#]
Putative glycosyltransferase	81057	82268	+	102	UAAAAAUUACUGAC	UGAUGUUGCC	CUUGA	GGCCAUAUCA	UUAAAUAUAGUAUUU	-8.3	3.17	-7.3
HP	110275	110730	+	-4	GCUUUGGCAGCAGG	CACUGAGGC	ACCUGGCCCGCA	GCGUCAGUG	AUUUCUGGUCUGUCC	-10.5	3.02	-10.8
putative sugar-binding protein	122501	123790	+	61	GAAGCGCUUUCGCC	CCAAUCCGUUC	GGGUU	GAACGGUAUGG	UUUCAACUCGUUAAG	-10.7	3.64	-9.1
CHP	178048	179151	+	-15	CGAGAUGCGAUCUC	UUCUCAGG	GGGA	CUUGAGAG	UUUAUUCCUCUUUUU	-8.7	4.4	-7.5
putative photosystem II reaction center J protein (*psbJ*)	206387	206587	-	28	GUUUGAUUGUUCAA	CCGGCC	UUCG	GGCCGG	UUUUUUUGUGGCCUC	-12.0	5.63	-7.0
putative cysteine desulfurase or selenocysteine lyase	318704	319984	+	-2	CUGAGGGAGCACGG	UUAACACGUUCAC	UGAUG	GUGACGUGUUGG	UUCUGACGUUCAGAC	-11.6	3.36	-11.2
HP	358046	358405	+	122	CAGUUGUUGAUUUC	GGCUUGGUCC	AUAG	GGGGUGAGUC	UUUGAAAAGAUCCCC	-8.4	3.54	-9.1
HP	358627	360171	-	45	AAGACGUACAACAG	UCCCCGG	AGGUGAGAGC	CCGGGGA	UCUUUUCAAAGACUC	-12.5	3.64	-8.0
Putative acylneuraminate cytidylyltransferase	448496	449197	+[%]	-35	CAACUGGUUGAAUU	GCUCAUGC	GUCAACAACA	GCAGGAGC	UUAAUCGAUGAACGU	-8.0	3.18	-9.6
allophycocyanin beta chain (*apcB*)	484341	484829	-[%]	-15	CGACUACAUCUGCU	CCGGCCUGGGC	AACUGAG	GCCCAUGCGG	UUGUUCAAAGUCACC	-10.6	3.64	-9.3
ATP synthase subunit gamma (*atpC*)	494640	495590	+	24	UCAAACCUGAUCGA	GCCGGUCC	CCGG	GGGCCGGC	UUUUUUAAUGGAGUG	-15.1	5.16	-6.0
GroEL chaperonin (*groEL*)	510706	512340	+	31	CUGGAGAGCAUUGA	CCCCCGC	UAAG	GCGGGGG	UUUUUUUAUUGGUGA	-15.4	5.84	-4.6
CHP	521288	521671	+	-11	GACACCAGGAUUGA	UCAGCGU	GACUGAUCAAUGG	ACGCUGA	UUGAUUCCUUCCACC	-8.8	3.45	-10.2
putative Dethiobiotin synthase (*bioD*)	613206	613865	-	31	UGAUCAGCUUGGUG	CUGGUGAUCCUC	UGCG	GAGGAAUUCUGG	UGUUGUUCACCGAUG	-9.4	3.39	-10.7
possible cobalt transport protein	643820	644698	+[%]	-36	UUGAGCCUGUUGG	CCGUGCUGGGUC	UGAG	GACUCGGUACGG	UGCUUUGUGAUGGAG	-18.5	3.04	-11.1
photosystem II chlorophyll-binding protein CP43 (*psbC*)	661320	662708	-	33	ACGAUCAUCAUUCA	CCCUCGUC	UACG	GGCGGGG	UUUUUAUUUGGCACG	-13.8	5.27	-6.2
phosphoribulokinase (*prk*)	765446	766348	+[%]	33	UCUGAUCCCUAAAG	UGAGUUCAUCUGA	GCGA	UCAGGUGGGCCCA	UUUCCUCUUCAAAGC	-13.0	3.92	-9.2
Fructose-bisphosphate/sedoheptulose-1,7-bisphosph ate aldolase (*cbbA,cfxA,fbaA,fda*)	769896	770969	+	24	ACCCAGUGAUCAGA	GGGGGC	UUCG	GCCCCC	UUUUUUUGUGUCAAG	-12.9	5.72	-5.5
CHP	783901	784518	-[%]	-45	GACAACCACCAGCG	UUGACGGACCC	GAACCA	GGGUCCAAAA	UUAUUAAUGAGACAA	-9.5	3.72	-7.4
CHP	797976	798242	+	36	GCAGAUACCAGUAA	CUACCGCUUC	AACGAUC	GAAGCGGUAG	UUACUGGUUAGUUUG	-15.6	3.32	-11.3
HP	841710	841961	-	80	CACAGCGUCCCCGC	UUCAGGCUCA	AUCCC	UGAACCUGGA	UUCCGAAUGGUUUUU	-7.9	3.02	-11.6
putative glycerol dehydrogenase	900736	901836	+[%]	-12	CGCCGAACCCAGUC	CCGUCGCGC	CUUGAAAAGCCU	GCUCGAUGG	UUUGAUCCCCGAGCU	-8.4	3.73	-11.5
CHP	1045313	1045630	+	69	GCUGCUGCUGUCUC	CAGCGAUGC	UCCUGUCC	GUGUUGCUG	UUGUUCACCUUCGCC	-9.8	3.68	-10.0

CHP	1053683	1054687	-	158	CAAGGAGCGUGGGA	GUUGGGUC	UCUU	GGCCCUGC	UGUUUCUAAGCCUGC	-8.5	3.53	-10.0
CHP	1114904	1115269	+%	-35	CCGCUGGAGGCACU	GGGACGGC	GGCU	GCGGUCCC	UUUUGCAUCACGACU	-11.9	4.18	-9.8
CHP	1194715	1195080	-%	31	UCAAGGAAUCCAGC	CGCCCUGGUU	UUGC	GGCCGUGGUG	UUCCUUCUCCCGGCC	-8.2	3.44	-11.1
CHP	1228577	1228849	-	24	UUGAACGGCCCCAG	GCUCUGUUGGGG	UCAC	CCCUGACAGAGC	UUCUCCAUGGUGUCG	-19.4	3.39	-11.6
1-deoxy-D-xylulose 5-phosphate synthase (*Dxs*)	1289301	1291232	-%	-36	GGCGGACGCGACU	UCGCCGGUGCU	GCAUC	GGUACCGGCGA	UUCAAUCCUGAACGC	-19.1	3.1	-10.6
HP	1358010	1358246	-	-14	GGUUGUUACACCUU	CAAGCGCCAGA	AGUAGUUCG	UCCGCGCUUG	UUCACUGUUGAUCUC	-8.4	3.21	-10.3
cytochrome P450 family protein	1380815	1382059	+%	-29	CCAACGCCUCGGGA	UGGGCUGCUGGU	AAGG	GCCACAGCCCG	UUGAUCAACAGAUGU	-16.4	3.15	-10.5
CHP	1390783	1391250	-%	16	CGAUGGAUCUCCAG	CGCCAUCC	CUUGCC	GGAGGCG	AUUGUUUCUAUCGCA	-8.8	3.08	-7.7
DnaJ domain-containing protein	1398600	1399307	-	81	CAUUGGUGCCGGCA	UCGCCGUCGCCA	UCUCCC	UGGCGGUUGCGG	UGUUAUUCGAACGUU	-11.9	3.4	-9.2
predicted membrane protein	1412125	1412625	-%	31	CGACUUGAUCAGUU	CGGGUCCCGUC	CAUGC	GGCGGGAUCG	UUGCUGUUAGCACAU	-12.9	3.36	-11.3
possible dihydroneopterin aldolase (*folB*)	1436174	1436545	-%	-10	AGCGCGCUCGCCAU	GGGGCU	CCAUGACAA	GGCCCC	UUGUGCUUGUCCAUG	-9.2	3.57	-11.5
				27	GUGCUUGUCCAUGG	CCUGUGGG	AUAC	CCCACGGG	UGUUUACGCUGAAUG	-13.6	3.39	-10.8
Type I copper blue protein: plastocyanin (*petE*)	1451284	1451643	+	46	UAACAGUUCAAAAA	CCCCGGACUGUG	UCAAA	CAUGGUCCGGGG	UUCUUUCAUUUGUCG	-20.9	4.19	-6.9
possible cytochrome C6 (soluble cytochrome F) (cytochrome c553)(*petJ*)	1451769	1452125	+%	2	AACUCGGCUGGUAA	CGCGAUGAGC	CGAGAAGC	GCUCAGCG	AUUUUCUACGAGCGG	-8.1	3.22	-8.0
possible acetyltransferase	1483524	1484027	-	-38	GGUUGGGAGCUGGA	GCCGCAGC	AGCAUC	GCUGCGC	GUUUUGGUACGCCAG	-10.4	3.15	-10.9
HP	1520810	1521193	-	80	UGGUUUGAAUUGAG	UGACCAACUGCA	UUGGU	UGCAGUAGGUCA	UUCAGUAUCACGAGG	-15.7	3.12	-10.2
putative circadian phase modifier CpmA homolog	1547754	1548404	-	-1	UGCUGCGGAAUCGU	UGAGGGUCAG	CCUGC	CUGGCCGUCG	UUGAUCUCUCCCAAU	-11.2	3.32	-11.2
30S Ribosomal protein S16 (*rps16*)	1554833	1555255	-%	15	AUCUCCCCGGUGCC	UGAGGACGGC	GCUCGCA	GCCGCUUCA	UUUUUGAUCUUCCCG	-11.7	4.88	-6.7
putative nickel-containing superoxide dismutase precursor (NISOD)	1564941	1565414	+%	29	GUUCUCUCCCCUAG	CAGCAGCAGGUC	UUCAG	GAUCUGCUGCUG	UUUUUUACCGCCCUC	-20.0	5.01	-8.0
putative carboxysome peptide B	1646561	1646812	-	-5	CACUGGAAUCCUGA	CGGAUAGGGAG	CCUUCC	CUCCUUCCGCCG	UUUCCGUUCCGCUUC	-8.1	3.77	-11.0
Ferrochelatase (*hemH*)	1678942	1680117	-%	36	GCCCUUAACCUUGA	GGGUCAU	UCCUGACCAGUUGCC	GUGACCC	UUACUUCCGCCUCAA	-7.9	3.35	-11.3
CHP	1688385	1688768	-%	-12	GCGCGACAUCAAAA	UGGGUGUGA	AUUGA	UCACCCCCG	UUGUCGUCUCUGCAG	-8.2	3.5	-11.2

CHP	1719743	1720168	-	12	UUGAUGCUGCGGUA	UGCGGAUUUCGU	GGCCU	GCGGAAACGCA	UUGUUCACUCCAGCG	-8.8	3.68	-10.4
CHP	1852184	1853080	+	169	AAGCCCGCGACAUA	GGCCUUCUUUU	CACCA	GGAAGAGCGGCC	UUCCUCUUUAGGAAG	-10.6	3.46	-10.1
photosystem II reaction center T protein (*psbT*)	1890726	1890821	+	18	GAUCGAUUCAAAAA	CCCCGC	CACG	GCGGGG	UUUUUUGUGACCUUU	-12.0	5.29	-7.3
phycoerythrobilin:ferredoxin oxidoreductase (*pebB*)	1920322	1921110	+	2	AGGACGCCGCUUAA	CUCCCUUG	UCGUAC	CAAGGGAG	UUAAGCAUUAUGAAC	-11.8	3.03	-9.8
Anthranilate synthase component I and chorismate binding enzyme (*trpE*)	1940531	1942051	+%	-5	CUCAACCCGGAGCG	GCCAUGAGCG	AAGGGCUGC	UGCUCAAGGGC	UUUGAAGUGGAGCUG	-7.9	3.6	-11.1
CHP	1960238	1961869	-	96	ACCGGAUCAUUGGG	CCGUCGCC	AAGCGGUAA	GGCAGCGG	GUUUUGGUCUCGCCA	-8.7	3.2	-10.2
50S ribosomal protein L15 (*rpl15*)	1972019	1972477	+	41	CCCUAAGGUCUGAG	CCGUCCGCUGGAU	GUCA	AUCCACGGGCGG	CUUUUCUGCAUUCAG	-18.6	3.28	-7.9
HP	2005030	2005374	-	-33	GACGCUCAAGGUGU	UCUGGGCU	GAUCCUGGAACA	AGCCACAGG	UUCUAAAUCGUUCCA	-8.3	3.4	-8.4
photosystem I P700 chlorophyll a apoprotein subunit Ib (*psaB*)	2006259	2008472	-	50	UUCCUCGUCCUUCA	CCCCGUCCGG	UUUA	CCGGGCGGGG	UUUUUUGAUGCAUCC	-20.8	5.17	-7.0
photosystem I P700 chlorophyll a apoprotein subunit Ia (*psaA*)	2008494	2010797	-	-8	GCCCACAUUCUUGU	GGUCGGCUG	ACCU	CACUGACC	UUUCCCUCUAAUGGC	-7.8	3.73	-10.3
Ferredoxin-dependent glutamate synthase, Fd-GOGAT (*glsF*)	2018953	2023554	-%	-7	ACCAGCAGGCGGUG	GCGGCCU	GAAC	AGGCCGC	UUUCCAUGCCCUGGU	-12.6	3.66	-11.4
inositol monophosphate family protein	2083683	2084486	-	79	GCGCAAGACCGAAG	CCUGAGGG	GCGAAAG	CCUUGAGG	UGUCUUGGUUCACUU	-8.3	3.19	-11.7
photosystem II extrinsic protein (psbU)(psbU)	2089405	2089812	+%	31	CCCAUCACCCGAAA	GCCGUCGG	ACUAU	CCGGCGGC	UUUUUUUGACUGCCU	-14.6	5.48	-5.8
6-pyruvoyl-tetrahydropterin synthase homolog	2099753	2100673	-%	35	GCGCCCCACGGUUC	GGUUGGUG	CUGGC	CAUCAGCC	UUGAUGGUAAAAAAU	-8.5	3.21	-11.4
Dihydroorotate dehydrogenase (*pyrD*)	2239926	2241116	-	164	UUUUUGAUGAAAAC	CAGGUGA	AUGAG	UCGCCUG	UAUUGCCUAUAAAAU	-8.4	3.13	-10.6
CHP	2272027	2272503	+	114	GACCGCGCAGAUUG	UCGAUGCGGGC	GAUCAACCCA	GUCCGCAGA	UGUUCACUCUUCGGA	-9.6	3.16	-10.4
putative urea ABC transporter, urea binding protein (*urtAl*)	2347549	2348868	-%	32	UGAGUGCUCUUCAG	GGAGGAGCCCG	GUGA	CGGGCUCCUCC	UUUUGUCGUUUCACU	-24.9	4.62	-8.0
putative asparagine synthetase protein (*asnB*)	2356512	2358536	-%	45	UGACGACAGGGCGU	UUGGGUCGCUG	CGGCGA	UGGUGAUUCGA	UCUUUUGAUCUGAAG	-8.4	3.78	-8.4
formyltetrahydrofolate deformylase (*purU*)	2409579	2410418	-	24	CGGCGCAUCGGGGG	CCAGGC	GUUUCG	GCUUGG	GUUGUUCUUUAAGGC	-8.5	3.05	-7.7
CHP	2424284	2425780	-	162	CCCUAAAUACAAAC	CAACGGAGUC	GCUC	GAUUCCGCUG	UUGUUUGAAGGGAUC	-9.3	3.91	-10.5

$: HP denotes hypothetical proteins and CHP denotes conserved hypothetical proteins.

@: dis represents the distance away from the gene stop codon.

#: G, T, and H represent the thresholds for hairpin loop (Kcal/mol), U-tail T weight, and hybridization energy (Kcal/mol), respectively.

%: intrinsic terminators located within the operon.

Table 3. Putative intrinsic terminators identified at the end of RNA genes in *Synechococcus sp.* WH8102.

gene name[$]	start	stop	strand	dis[@]	5' A-tail	5'stem	loop	3' stem	3' U-tail	G[#]	T[#]	H[#]
tRNA-Arg	258586	258659	+	10	GGCGCGCUUGAAAA	CUGCCC	AAUC	GGGCAG	UUUUUUUUUGGACAG	-10.1	6.07	-4.4
tRNA-Ala	298547	298619	-	18	CCACUCUUGCGGUU	CCGAUGGGUC	GGUGCUG	GAUCCUCGG	UGUUGUCGUGAUGCU	-11.4	3.30	-11.3
tRNA-Met	806346	806422	+	19	CUGAUUGAAUUUGA	GAUCCCU	CCAUC	GGGGAUC	UUUUUUUUGUCCGUG	-9.0	5.96	-4.6
tRNA-Pro	858432	858505	-	39	ACCGCUUGCGCGAA	CCCGGUGC	GAAA	GCAUCGGG	UUUUUUAUUGGCCUU	-14.1	5.48	-6.2
tRNA-Pro	1017787	1017860	-	48	GCGACAGGUCACAA	GCCGGCAC	UUCC	GUGCCGGC	UUUUUAAUGCCCUGA	-15.9	4.74	-7.6
tRNA-Ser	1141506	1141592	-	29	CCACCGCUUCAAAG	CCCCGU	CGUUG	ACGGGG	UUUUUUAUUGCGCUG	-9.0	5.47	-5.7
				140	AACCUUGUGCAACG	GUGCAAACGGGG	UAGCU	UUUCGGGCGCGC	UUCGUUGUCUGAAGU	-8.8	3.50	-10.1
tRNA-Ser	1273817	1273906	-	15	ACCUAACCUCCAAA	CCCCUUG	CUAAG	CAAGGGG	UUUUUUAUUGAAAUU	-10.4	5.48	-4.6
tRNA-Ala	1524848	1524920	-	37	GCACUGGCAUUUGA	GCGGUC	UACGGU	GGCCGC	UUUUUUAGUGCAAAA	-9.8	5.15	-6.6
				251	UUAGCGGCACUGAG	CAGCCAGU	CCACCGUAAA	ACUGGCUG	UUUUCUGUUUACUCU	-11.0	4.83	-7.3
tRNA-Ala	1874314	1874386	-	166	AUGCUGGGCUCGAU	UUCAGUGAUCAA	GCAA	UUGGUCAUUGGA	UUUGAAGUUCUGGCA	-12.4	3.69	-9.4
tRNA-Phe	1898273	1898345	+	36	CGUGUCAUUCAUCU	CACCUGAU	GCAGGCACAG	AUCGGGUG	UUCGAUAACGCCCGU	-9.6	3.05	-11.5
tRNA-Ala	2081254	2081326	-	166	AUGCUGGGCUCGAU	UUCAGUGAUCAA	GCAA	UUGGUCAUUGGA	UUUGAAGUUCUGGCA	-12.4	3.69	-9.4
tRNA-Asn	2133291	2133362	+	9	CGGGGAGCUUGAAA	GCCACAGC	GUUU	GCUGUGGC	UUUUUUAUCGACAAA	-14.5	5.24	-5.8
rrnA/5s rRNA	1870762	1870876	-	73	CCCCCACGAAGAAA	CCACCUC	ACUAU	GAGGUGG	UUUUUUGUGUCUUGC	-11.2	5.47	-6.3
rrnB	2077703	2077817	-	73	CCCCCACGAAGAAA	CCACCUC	ACUAU	GAGGUGG	UUUUUUUGUGCUUUG	-11.2	5.76	-6.0
				121	CUUUGCUCUUCGUU	CACCAGUUUUU	UAGU	GAGUCGCUGGUG	UUAACGUUCUCAUAA	-8.5	3.12	-9.3

Note: same as Table 1.

Table 4. Other putative intrinsic terminators identified in *Synechococcus sp.* WH8102

gene name$	start	stop	strand	dis@	5' A-tail	5'stem	loop	3' stem	3' U-tail	G#	T#	H#
glyceraldehyde 3-phosphate dehydrogenase (NADP+) (*gap2*)	30250	31275	+	23	UCGUUGAUCAUUCA	UCCCUUCC	UUCG	GGAAGGGA	UUUUUUUUGGCCAAU	-13.5	5.76	-6.1
CHP	35685	36590	-	-8	CAGGAAUCUCAGUC	CUGCUCCUGAUC	GCCGAU	GAUCAAGCGG	UUGAAAUCAAUCGGC	-9.3	3.02	-9.7
CHP	51683	52234	+	6	ACUCUGGUGAAGAC	UGGGUGAUCG	UUGAAU	CGAUCAGACCCG	UUUUUCAAGCUUUUU	-10.9	4.67	-7.0
CHP	72760	73518	+	29	ACGGAUCCACAUAG	GCUGAGCCAC	CGCCCG	GUCGGCUCAGC	UCUUUCAUGCCUGCC	-14.9	3.47	-9.8
SwmA-cell surface protein required for swimming motility(swmA)	82533	85040	-	29	AAGUCAAAUAUUCA	GCCCCUCC	UUUA	GGAGGGGC	UUUUUUAUUGGCAAC	-16.5	5.45	-6.2
Type II alternative RNA polymerase sigma factor, sigma-70 family	102314	103312	+	107	AUCAUCAAGGAACU	GAUGCGCUGCAU	GAAU	GUGCAGAACGUC	UUUUACUUCUUGGAA	-11.1	4.56	-6.6
thiamin biosynthesis protein (*thiC*)	135007	136458	+	29	AAUUGGCUCGCAUG	CCCGGCUUCU	GAUUC	AGGAGCCGGG	UUUUUUGUUGAACAU	-17.3	5.45	-5.8
transketolase(tktA,)	136580	138589	-	28	CAAUCGAAAUUUCA	GACCCCUUGC	UACG	GCAGGGGGUC	UUUUUUGUUCAACAA	-19.9	5.45	-5.5
photosystem I iron-sulfur center subunit VII (*psaC*)	140259	140534	+	29	ACGCCAUAGGCUCG	GCCCGGC	UUGU	GCUGGGC	UUUUUUUAUGUGUGG	-13.4	5.77	-4.5
CHP	161067	163178	+	25	CCGGAUCAUGCUCA	GCCCCCUC	AUUC	GAGGGGGC	UUUUUGCCUUGCUC	-17	5.18	-7.8
Ammonium transporter family (*amt1*)	253620	255092	-	25	GGCUUCAGUUCACU	GCCCGGCC	CUCAG	GGCCGGGC	UUUUUGAUGGCCUGC	-17.7	4.74	-9.3
CHP	298147	298347	-	68	UCCCCAUCCCCCGG	CCCGGG	CAACU	CCCGGG	UUUUUCUUGAAUUCC	-10.4	5.14	-5.8
photosystem II manganese-stabilizing polypeptide (*psbO*)	300581	301411	+	33	UUCAACGUCAUCAA	GGGGGC	UUCG	GCCCCC	UUUUUUGAUGCCUGG	-12.9	5.17	-7.6
CHP	311031	312137	-	46	UGCGUUGGUUUGUG	GCCAUCACCCCU	CUUGC	AGGGGCGAUGGC	UUUUCCGCUGAUCGU	-19.1	4.16	-11.0
phycobilisome rod-core linker polypeptide cpcG (L-RC 28.5) (*cpcG1*)	312299	313057	-	26	CAAGAGUCCGAAAA	GCGGGGC	CUCA	GCCCCGC	UUUUUUUGUGUCGCA	-14.5	5.72	-5.5
				88	CUUCUUCCCAGCUC	UCCAGGGUCAG	GAUCCA	CUGACCCUGGG	UUCGUUAACGGUUCG	-21	3.36	-10.2
HP	345482	346312	+	67	GCCAAGUGAAAUAA	UUCAGGCGG	UGCGCAAGUUU	CCCCUGAA	UUUCACGCUUUUAUA	-8.1	3.68	-10.1
CHP	346837	347112	-	53	AUGCGAGCAAUUUA	GGCCGUCUUUCC	ACCG	GGAAGGGCGGCC	UUCCUCUCUUGCUGA	-21.5	3.39	-11.3
HP	347901	348713	+	120	UUCUGAAGUGAUGG	UCUGGUUGAU	UUCA	GUCACACCAGG	AUUUUUGGAAAACGA	-8.3	3.34	-8.4
weak similarity to phage integrase family	361958	363160	-	67	CACUCGAAAGUUGC	CCCGCCGAU	CGUAU	AUCCGGCGGG	UUCCCUUUCUGAGAA	-13.4	3.33	-9.9
possible integrase/recombinase	363372	364283	+	216	GACAACACAAGAAA	GGGGGCGGU	CGAAACA	ACCAGCCCCC	UUUCUUUGUGGAAGU	-14.9	4.56	-8.3
putative urea binding protein (*urtA2*)	367336	368616	-	31	AUAUUUGCGUUCAG	GCCUAGCGGCA	UUAAA	UGCCGCUAGGC	AUUUUAAUGCCAAAA	-21.2	3.15	-7.7
HP	381629	382033	-	26	GUAAGGUCAGCGGA	CGGGUGGGUGG	CAUGG	CCGCCAGCCCG	UCUUUUUCAUGGUGU	-16.2	4.01	-7.9
CHP	384325	384627	+	13	UAAAUCAAGUGUCC	CUAUCACCGG	CCCAGUU	CUGGGUGGUUGG	UUUAGCUCAGCUGAC	-11.5	3.66	-11.1
HP	402762	403415	-	24	CAGAUUCAGUUCAA	GACCCCUC	ACUC	GAGGGGUC	UUUUUUUAUGGUUUC	-14.9	5.76	-5.1

Annotation	Start	End	Strand	Pos						ΔG1		ΔG2
anchor polypeptide LCM (*apcE*)	485764	488655	-	64	UUCUCGAUCAAUUU	GCCUGGGUGAAG	UGCUG	CUUCAUCCGGGC	UCUUUUGCUGUCAUC	-19.3	3.79	-9.9
CHP	507521	508057	+	27	UUAGUUUUCGAUGA	UCCCGCC	GCAA	GGCGGGG	CUUUUUUUAUGCCAU	-13.4	3.96	-5.0
CHP	524944	525819	-	41	UCAAAUCCCGAAAU	CAACCGAC	AGUGCAC	GUUGGUUG	UCUUCGUUUCGGGCG	-8.9	3.33	-10.3
Ferredoxin(petF4)	528955	529254	+	34	GGUUUUCACUUCAA	CCACCCC	UGCG	GGGGUGG	UUUUUUGUUGCUCAG	-14.7	5.51	-6.4
Cytochrome c-550(psbV)	529796	530341	-	28	AUCGGAUUCACUCA	CCCCAGUGGCU	UUAG	GGCCGCUGGGG	UUUUCUGUCUCAGCA	-21.4	4.61	-8.2
50S ribosomal protein L27 (*rpl27*)	536682	536948	-	31	CCUCAAAUCCAUGA	GCCCCAGAACC	UGCA	GGUAUCUGGGGC	UUUUGAAUGGCGCCU	-17.9	4.18	-9.9
CHP	562165	563136	-	172	CUUCCACCAGUUGC	GCGCGGC	CAAC	GCUGUGC	UUCCAUUUUUGCGCG	-9.7	3.45	-8.4
DNA binding protein HU	569614	569889	+	28	UCUCUGAUCGAAUG	GGCGGUC	CAGG	GGCCGCC	UUUUUCAUGCCCGUU	-11.8	4.73	-8.4
putative ribose 5-phosphate isomerase A (*rpiA*)	588890	589609	+	15	AGCCGCUUGCGGAC	GGACUCCGC	AUGGCU	GUGCAGUCC	UUCGCUUUCCGCCAG	-8.8	3.28	-11.4
possible ABC transporter involved in polysaccharide efflux	622992	624236	+	81	AAGGCCUAUUGGGA	GUUUGGAUGAG	AAUGA	UUCGUUCCGAGC	UUAUAGAUUAAAUUG	-8.0	3.44	-7.7
O-acetylserine (thiol)-lyase A (*cysK1*)	657491	658477	-	29	AGCAAAUCUCUACG	GCCCCGUG	GCAA	CACGGGGC	UUUUUGCUGCUUUUA	-16.9	4.87	-8.7
N2,N2-dimethylguanosine tRNA methyltransferase-like protein	666609	667784	+	107	UUGUUGAAGCUGCA	GGGGGA	CUGA	UCCCCC	UUUGUCACCCGAGCC	-10.1	3.86	-11.3
translation initiation factor IF-1 (*infA*)	670786	671055	+	1	CCGCGUCGUCGUUA	GGCCGC	CAGCGAUUCAAUC	GCGGCC	UUGAAUUCGCUGCGU	-10.1	3.19	-10.2
possible multidrug efflux transporter, MFS family	691135	692436	-	53	CAACCAGAACAUUG	CCAGCUUGG	AGGU	CUAAGCGGG	UCUUGAUGUCGUAAG	-8.2	3.21	-10.7
putative IMP dehydrogenase (*guaB*)	706903	708066	-	59	CCUCACACCCCCCG	GCCCGG	CGCGC	UCGGGC	UUUCACUCUUCCUGC	-8.4	3.78	-9.7
ferredoxin --NADP reductase (FNR) (*petH*)	732773	733942	-	34	AUCAAGCUCUGCGA	GGACUGCC	CAUC	GGCAGUCC	UUUUUUUUCGCUUGG	-14.3	5.86	-4.8
CHP	734081	734623	+	43	GUUCACUCUGAUUU	UGAUCACCG	GCGAU	CGGUGAAUCA	UUAAUACCCCCAAGG	-9.9	3.14	-11.2
CHP	736728	737138	+	39	UGAAGGUUUCUCAG	GGGCCAGUCG	CAUU	CGGCUGGUCC	UUUUUUUUGAAAAGG	-16.1	5.75	-3.9
Sodium:alanine symporter family:Permease for amino acids and	798285	799637	-	-17	AAACUAACCAGUAA	CUACCGCUUCG	AUCGU	UGAAGCGGUAG	UUACUGGUAUCUGCA	-16.3	3.27	-11.4
CHP	837567	838649	+	117	GUCCACGAUCCCAG	CGCCGC	GAUGCCA	GCGGCCG	UUUUCAGUCCAGCGG	-8.7	4.18	-9.1
Sodium/glutamate symporter (*gltS*)	847605	848834	+	108	AUCUGGCAAGCAAU	GCCUCUGU	AAAU	ACAGGGGC	UUCACACUAUCUAAA	-11.4	3.01	-10.6
HP	852392	852652	+	142	GAAUUGUGCAAAAG	GUCGGAUGCG	UACC	UGCGUCCAGC	UUUGUUCGCAUCGGU	-9.2	4.18	-9.6
CHP	853766	855001	+	27	CUGCAAACCCAUAG	CCCCGUUA	GCAA	UGGCGGGG	UUUUUGAUGGGUACC	-11.9	4.75	-8.9
tRNA	858432	858505	-	39	ACCGCUUGCGCGAA	CCCGGUGC	GAAA	GCAUCGGG	UUUUUUAUUGGCCUU	-14.1	5.48	-6.2
possible photosystem II PsbY protein (*psbY*)	861099	861230	+	-18	GCUGCAGCUGCUGA	UCAAGCGC	AGCC	GCGCCUGA	UUUCAACGCACUCCC	-7.9	3.54	-11.0
				37	GCAGGCAUCCGUAG	GCUGAACGC	UGCUU	GCUUCAGC	UUCUGUUGUCUGUUG	-9.0	3.84	-9.7
putativeRNA-binding protein RbpD (*rbpD*)	883208	883498	+	48	AGAUCGAUCUCCGG	CAGGGGUCAGCG	AUGUG	UGCUGAUCCCUG	UUGUUUUGUUUGUUG	-19.7	4.63	-6.7
HP	912128	912412	+	16	UAUUGUAAGUUGUG	CUCUGGCCG	GGGAUAGGAA	UGUGCCAGGG	UUUUACAGUUGUUUG	-9.9	4.26	-8.0

Annotation	Start	End	Strand	Pos	Seq1	Seq2	Seq3	Seq4	Seq5	ΔG1	ΔG2	ΔG3
photosystem II D1 protein form II (*psbA2*)	971755	972834	-	63	AACUGAAUCGGAAA	GCCCUCACC	AAUC	GGUGGGGGC	UUUUUGUUGUGCCAA	-17.1	5.19	-7.2
CHP	991930	993123	+	21	UCGUCCGGCUCGAU	GGCCAUCUGUUC	AUCAGC	GAACAAGGUC	UCUUCUGAACAUCUC	-8.0	3.23	-10.4
ABC transporter, substrate binding protein, phosphate	1015895	1016869	+	14	GAUGACUAGCUGAG	CCGUCGG	UUGAU	CUGACGG	UUUUGAGCAGUUCCA	-9.7	4.11	-10.0
				51	GAGCAGUUCCAAGG	GGGGCC	UACG	GGCCCC	UUUUUUUGUUCUUUU	-13.8	6.04	-4.7
tRNA	1017787	1017860	-	48	GCGACAGGUCACAA	GCCGGCAC	UUCC	GUGCCGGC	UUUUUAAUGCCCUGA	-15.9	4.74	-7.6
HP	1051574	1051960	-	93	ACAUACACUCAUCA	UCCCCGUCA	GCAA	UGGCGGGA	UUUUCGUUGCCUCCU	-10.5	4.44	-8.8
Glutamine synthetase, glutamate-ammonia ligase (*glnA*)	1060651	1062072	-	25	UGGGUGUUGAUCGG	CCCGCCUC	CACG	GAGGCGGG	UUUUUUUGUGGAUCA	-16.5	5.64	-6.3
CHP	1075191	1075346	+	27	UUCUGACGUUCAGC	GGGGGACC	UCAA	GGUCCCCC	UUUUUGUUGCUUUUA	-16.5	5.28	-6.4
geranylgeranyl hydrogenase (*chlP*)	1086597	1087949	-	35	GAUCACUGACGCUC	UGCGGCCG	CCUUGC	UGGCGGCA	UCUUUGAUCCGCUCG	-9.3	3.46	-10.6
ribulose-5-phosphate 3-epimerase (*rpe*)	1107531	1108307	-	29	CUGACCCCCUAGGA	GCCCCGGC	CUCA	GCCGGGGC	UUUUUUUAUGGAGAC	-17.8	5.61	-5.2
HP	1127972	1128364	+	24	CCGUAAGGACAUGG	CCCCUGC	UACG	GCGGGGG	UUUCUUUUUGACAUC	-14.0	4.87	-6.0
CHP, phage associated	1156514	1157080	+	86	AAGGCUGUUUAGAG	UGGUCUCUAGGGU	AUUU	AUCCUGUUGACCA	CUUUUAACGGCUUCU	-12.0	3.07	-9.3
HP	1167583	1167747	-	-28	CCAAUGGCGGCGGU	GGUGGCGGU	CUGAUGCA	ACCGGUCGCC	UGAUUUUUUUGACUU	-10.4	3.52	-6.3
HP	1184977	1185144	+	22	UCAGUUAUCUCCAA	UGGACUGCUG	UUCGG	CGGCAGCCA	UUUGUUGAACGAGGU	-11	4.15	-9.3
thioredoxin peroxidase (*tpx*)	1204065	1204667	-	38	CUGCUGAUUCAGGA	CCCUGUCGC	CACUG	GCGGCAGGG	UUUUUUGCAUCAGUC	-15.9	5.1	-7.4
allophycocyanin alpha-B chain (*apcD*)	1221449	1221943	+	-22	UCGACUACCUGAUU	CAGGGGAUG	CAGA	CAUCCACCUG	AUUUGUUUGCCGGUC	-11.4	3.25	-8.5
				30	UCAAUCAAGACUCG	CCCCUGGUCGC	GCGAU	GCGUCCAGGGG	UUUUUUAGUGAGAUU	-19.6	5.16	-6.3
possible GTP cyclohydrolase II / 3,4-dihydroxy-2-butanone 4-phosphate synthase (*ribA/ribB*)	1260179	1261882	+	-13	CCCUCCUGCACUGG	UCAGCCAGGG	ACUGAUCAG	UCCUGAAUGA	UUACCUUAUUGAUCU	-8.5	3.31	-8.7
heat shock protein HtpG (*htpG*)	1276760	1278664	+	-13	UGGGAACGCUGAUG	CAGAGAGGG	AUGUGA	CCUUCGCUG	UUUGAUAAGCUAGUA	-8.4	3.75	-8.8
HP	1323028	1323285	-	59	AAUCGCGGCUCCAG	CGAUCAGG	CCCGCAA	UCUGAUCG	UUCAUAUUGCUUUUU	-8.1	3.47	-7.9
Possible phosphoribulokinase/uridine kinase family protein	1334517	1335128	-	-8	UCUGUUGCUUCAGA	CGACGUCUGAGU	UGGG	ACCCAGACGUUG	UCUUGAUUGGCACGG	-12.6	3.24	-10.7
CHP	1335418	1337838	+	55	CCCCUUCUUUUCCA	CCCCUACUUCU	UAGU	AGUAGUAGGGG	UUUUGAAGACUGGGG	-13.8	4.1	-9.3
phage integrase family	1338474	1339034	-	54	GGCCCAAAAGAAAA	CCCCUAG	UGGGGUUU	CUAGGGG	UCUUCUCCCUGAAUG	-11.5	3.25	-11.6
				237	UCAAUAUCAGGCAU	CCGAGUGGC	UAAC	GCUGCCCG	UUUGCUCUGCGUCGU	-8.9	3.86	-11.6
HP	1344826	1345071	+	23	GUCAACAGCAACGA	GCGGGGG	ACACU	CCUCCGC	UUUUUUUGUGCUCAG	-11.8	5.67	-6.0
HP	1347608	1347925	+	-2	CGAUAUGCAGAGAG	CUGAUCGCGUU	CAGAGUC	AGCGUGUCGG	UUCAAGUUCCAAAGC	-7.8	3.07	-9.5
RNA-binding protein	1352434	1352691	-	32	CGCUUGCCGAUGAC	CUCUUCCCAGCU	GAUC	AGUUGGGAAGAG	UUUUUUGUGCUGGUUG	-18.5	4.95	-9.2
HP	1352831	1353166	+	198	UUCUUCAGUCCCUU	UGCUUGAGG	UUCCAACUC	CCUGAAGCA	UUACUGAUCAGGUCG	-8.7	3.22	-11.2
similar to zeta-carotene desaturase	1369594	1371222	-	-20	GGUCUGCGCUGGCC	CUGGUUCGCG	GGUUCA	CGUUGAGCCGG	UAGUUGUUUGUUGUG	-12.3	3.13	-8.2
CHP	1379402	1379710	+	153	GUUACAAAGAUGGA	CAGGGCAAUG	ACAU	CAAUGCCCUG	UUAUCUGUAAAAGCA	-12.4	3.59	-8.7

Gene	Start	Stop	Strand	Pos	Seq1	Seq2	Seq3	Seq4	Seq5	ΔG1	V2	ΔG3
HP	1383192	1383440	-	97	ACACAAGAACAGGA	CGCUUGCUGU	UGGUG	GCAGCAGCG	UUGUUCUGGACGUAC	-11.2	3.8	-11.1
HP	1401121	1401327	+	99	UGCGUCGAUUGACG	GGAGGG	CCUGGGAAA	CCCUCC	UUUUUUGUGCCUGAA	-9.1	5.28	-7.4
HP	1404069	1404434	-	110	GUUGCCUGAUGGCA	UCACUCUCCC	GUUGAU	GGCAGGGUGA	UCUUCUUGAAUGUCA	-10.1	3.43	-9.2
30S ribosomal protein S21 (*rps21*)	1408581	1408757	+	27	CGCAAGAUUGUUCA	GCGGGCC	AAAG	GGCCCGC	UUUUUUUUUGCGGUC	-14.9	6.09	-4.6
possible high light inducible protein (*hli1*)	1411230	1411433	-	142	CGCCGGUGUGCCCA	GACCCUGC	CUUCG	GCGGGGUC	UUUUUUUAGCGCUGC	-13.4	5.42	-6.3
photosystem II D1 protein form I (*psbA1*)	1420584	1421660	-	63	AACUGAAUCGGAAU	GCCUCCACC	GCAA	GGUGGGGGC	UUUUUGCUGGUGCAA	-17.9	4.77	-9.6
CHP	1421802	1423130	-	97	AUAGUCUUGAUCCA	CCGAUCUCCU	AUU	GGGAGAGCGG	UUUAUCCGUCUUCUU	-11.5	3.99	-9.1
CHP	1463159	1463530	+	59	GGCAAAGUCCUAUU	GGCCCUGG	UUGUU	CCAGGGCC	UUUUUGAUGUGGGAU	-16.7	4.81	-8.3
ABC transporter, ATP binding component, possibly iron transporter (*futC*,*sfuC*)	1487175	1488335	-	-12	AGCUCUUCGGCCGA	UGCCGCCCGCA	UGAACAGAU	UGCUGCGGCG	UUCCUCUCCGAUGCU	-11.5	3.27	-11.7
CHP	1489335	1489568	+	129	AAAAAGCCGCUGCA	GGGCAACAUCA	GUUCUGA	UGAUGUUGUUC	AUUUUUCCGCCUCCG	-12.6	3.36	-9.1
possible pfkB family carbohydrate kinase	1492086	1492919	-	20	CUUGGAUGUCUGGC	CCCGUCA	GCAA	UGGCGGG	UUUUUUAUUGCUGCU	-10.7	5.52	-5.3
HP	1492903	1493142	+	226	AUGCCAGGCAGAUC	CCCUGC	AGGAACUAAGGA	GCAGGG	CUUUUAUUCAUAAUG	-9.0	3.43	-5.2
CHP	1502689	1503033	+	5	AGAUGUCCUGACAA	CGGGCUGGGGGC	AAUUAC	GCUCCCAACCCG	UUCACUGACUUUCCA	-19.6	3.12	-10.2
CHP	1503533	1505344	+	111	UUGUCUUUUCACCA	GCCCCGUCA	GCAA	UGGCGGGGC	UUUUAUGCAAGGGCA	-17.4	4.37	-9.8
CHP	1506298	1509762	+	28	ACUGAACACAUUCU	CCCGUCA	GCCA	UGGCGGG	GUUUCUUAUUGCAUC	-10.7	3.27	-7.4
HP	1512732	1513076	+	195	ACUCCUGAAGUGCU	GGUGGUUGAUC	UCAGU	GAUUGACCACC	UGUUCACAUUGCAGG	-15.3	3.09	-11.1
possible phage integrase	1513971	1515110	-	161	UAAAAACCGCAAGA	GAUCAGCAGA	GAACAAAA	UCGCUGAUC	UUCUUGUGCCUGAAC	-10.2	3.82	-10.8
CHP	1553030	1553761	-	54	AUUCUCUUCCUGCA	GCCCCGG	AUGAU	CCGGGGC	UUUUUUGUUGGUUAG	-14.8	5.56	-6.6
Pyruvate dehydrogenase E1 alpha subunit (*pdhA*)	1558919	1560004	+	-15	GCUCACCCGCUACA	UCUGGGCCG	AAGAC	UGACCCAGA	UCUGUUUGGCUGCUC	-8.4	3.28	-11.2
HP	1594600	1595028	-	114	UUCCCUCGUGAUUG	CGCUCGACGCG	AAAU	CGUGUUGCCCG	UUUUUCGUCGUGAGA	-8.5	4.77	-8.1
CHP	1596030	1596164	-	73	ACAUGAGAACGGCG	GACCCAGGAG	GUCGAACA	CUCCUGGGUC	UUUUCUUGGAAAUCG	-20.2	4.65	-7.9
HP	1602043	1603305	-	98	UGUGCGAGUUUGAU	CUUGCAAGC	ACCUGCGC	GUUUGCAAG	UUCCCAAUAUUUUCA	-10.6	3.05	-9.7
putative tRNA (guanosine-2'-O-)-methyltransferase	1617506	1618195	+	250	CGCUGAAUCGUGAG	CCGCUGA	AUCA	UCAGUGG	UUUGAUCGCGACGC{	-8.5	3.73	-11.3
NifU-like protein	1623675	1623920	+	42	CACCUGUUAGUCGG	GGCCCUUGCCG	AGGC	UGGAGGGGCC	UUGAAUCAGGUCGAC	-14.2	3.07	-11.1
UDP-N-acetylmuramyl-tripeptide synthetase (*murE*)	1626144	1627646	-	55	CAAUAUCGUGUAAU	GGCUGUGGUC	AGUGCCUC	GGCUUCAGCC	UCUGUAGUAUUCCUG	-12.3	3.02	-10.2
HP	1634015	1634347	-	164	AUCCCCAUCACCCG	GCCGCC	CGCAGA	GGUGGC	UUUUUUGUUGGCCUG	-9.5	5.47	-7.4
Ribulose bisphosphate carboxylase, small chain (*rbcS*)	1651326	1651667	-	-17	GGUGCAUGCUUCGU	GGUUUUCGAAGGA	CGCUGA	UCCUUCGAACC	UUUGGUUCCGAGCCC	-11.3	3.9	-11.2
				18	CUUCGAACCUUUGG	UUCCGAGCCCUGA	CCUGA	UCAGGGCUCCAG	UUUUUUCCGGAGGGG	-12.7	5.01	-8.5
transaldolase(tal)	1688849	1690021	-	3	GACCACGGCUUGAU	CCGUUGGGU	GACGA	AUCGAGCGG	UCUUUACUUCCUUUU	-8.9	3.62	-8.8
				101	CGCCCUUGACACGA	UGCGCGCUCUUG	CGAG	CAAGAGCGAAGUG	UUGUGUUAAGCACUCA	-10.9	3.67	-9.5
Putative principal RNA	1712270	1713685	+	237	CGGAAGAGAUCGAA	CCCUGG	CUGAAA	CCGGGG	UUUUUUUCUAGAUCA	-9.1	5.64	-5.0

Description	Start	End	Strand	Pos	Seq1	Seq2	Seq3	Seq4	Seq5			
polymerase sigma factor (*sigA*)												
putative inorganic pyrophosphatase	1717053	1717640	-	37	CCUCAGGGGAUCAA	GGCGGGG	CAAGG	CCCCGCC	UUUUUCAUGGCAAGC	-14.4	4.72	-8.2
CHP	1722529	1722690	+	26	AAAGCCCACUCCGA	UCCCCUGC	CAACG	GCGGGGGA	UUUUUUUUGCACGUU	-14.0	5.78	-5.1
CHP	1738577	1738816	+	152	UGUCCACAGUCACC	CUGGGCUUAA	AAAUG	UUGCGCCCGG	UUUUCAGGUAGAGAA	-9.0	4.14	-10.1
ABC transporter, substrate binding protein, phosphate	1739194	1740171	-	15	AUUUUUAUUGGGUU	CUGAUUAAUC	UGAUUAC	GAUUAAUCAG	UUUGUCGAUUUUUUG	-9.7	4.12	-8.4
possible 50S ribosomal protein L19 (*rpl19*)	1745275	1745739	+	44	AGCAUCCUCGAUGU	UGGCAGGGGU	UCAUC	GCCUUGCGCCG	UUAGUUCAGUUGGUA	-8.6	3.41	-8.7
CHP	1745913	1746101	+	26	GAUUCCUCCGCAAA	CCCCCGAUGA	CUCCG	UCAUCGGGGG	UUUUGUUGUGACCAA	-18.0	4.77	-8.2
possible high light inducible protein (*hli8*)	1752259	1752438	-	13	UGAUCAACUCCUUC	CCCGCAGGGC	CUGGGCCA	GCUCAUUGGG	UUUGAGACUGACUUC	-8.2	3.58	-11.5
30S ribosomal protein S14 (*rps14*)	1808090	1808392	+	41	AGUCGUUAUCCAAG	GGGAGC	UUCG	GCUCCC	UUUUUUCAUGCGCGG	-10.8	5.15	-6.9
HP	1814062	1814301	-	-43	UUGCCUUGAGCUCC	UUUCGACCAAG	ACAAAGUA	CUUGGUUGGAA	UGUUCUGAACGUUGA	-12.1	3.23	-11.2
photosystem II D1 protein form II (*psbA4*)	1824940	1826019	-	63	AACUGAAUCGGAAA	GCCCUCACC	AAUC	GGUGGGGGC	UUUUUGUUGUGCCAA	-17.1	5.19	-7.2
CHP	1826708	1826977	+	-21	UGUGGACUACCUCG	UCUCUGCCG	ACGAGAA	CGGCUAGAGG	UUCAGCUUCAAAGGA	-10.1	3.07	-10.5
HP	1846017	1846385	+	-13	CGCAUGAAUACAUA	GCCUCGC	AUUUCUA	GCGGGGC	UUUUUUUGAUACUUG	-11.6	5.58	-5.3
				24	UUUUUGAUACUUGA	UAGCCGGU	UAAC	ACCGGCUG	UUUUUUAGUGGUCGA	-11.7	5.19	-6.8
HP	1854064	1854777	+	115	AAGAAACAAGCCUG	GCUUCGCG	GCUACUAAG	CGCGUAGC	UUAAUUCGGUAGCAG	-8.3	3.37	-9.4
				198	GGAGCGCAGUCAAG	CCAGGCC	AUGGAACGG	GGACCUGG	AUUUCUUCGAGGAAC	-8.7	3.09	-9.3
CHP	1855261	1855485	+	30	CUAGAUGCAUUAAA	GCCCAUGC	AGAU	GCAGGGGC	UCUUUUGUUUAUUAA	-10.2	4.17	-6.7
cytochrome b6-f complex subunit 4 (17 kd polypeptide) (*petD*)	1878339	1878821	-	26	GAUGUUUUGUUGGA	CCCCGGCUC	CUUUG	GAGUCGGGG	UUUUUGUUGCUCAG	-16.1	5.51	-6.4
Possible phycobilisome linker polypeptide	1896172	1897074	+	36	CCAGGCUGACCAAA	CGUAGGACUUGC	GACAC	GCAAGUCCUACG	UUUUCGAAGUCAAAU	-20.2	4.11	-8.8
CHP	1899145	1899813	+	39	CCGUGCAUGUCGAA	UCACCCGCC	AACCA	GGCUGCGGUGA	UUGUUUUAGGGAACA	-9.7	4.17	-8.6
Phycobilisome linker polypeptide	1903591	1905237	+	26	AGUCCUGCCAUAGG	GGGCCGC	CGCCUGAA	GCGGCCC	UUAUAGAUUUCAAAC	-14.6	3.47	-7.7
C-phycoerythrin class II alpha chain (*mpeA*)	1910879	1911376	+	3	CUCCCUGGGCUGAU	UAUCAGUCC	UUAAC	GGAUUGGUA	UUUACGACUAGUCGU	-9.2	3.59	-10.4
				45	GACUAGUCGUCUCA	GGGGGUGAGC	AAAUA	GCUCACCCCC	UUUUUUUAGCAGUAU	-21.2	5.43	-5.4
C-phycoerythrin class II gamma chain, linker polypeptide (*mpeC*)	1911741	1912622	+	56	GUGUGUGAUGACCA	GCUCCCUU	UCAA	AAGGGAGC	UUUUUUUUGUCCAAGA	-12.1	5.61	-5.6
C-phycoerythrin class I alpha chain(cpeA)	1917288	1917782	-	23	ACGGCUGUGAGAGA	GGGGGGGGC	GAUAA	GCCCCCCCC	UUUUUUGAUGACUGC	-23.1	5.17	-6.9
HP	1918383	1918874	+	145	CCCCUCCAGUCAUA	UCCACCA	AUGACGA	UGGUGGA	UUUUUUCGAAGUCAG	-9.2	5.03	-6.8
R-phycocyanin II alpha chain (*rpcA*)	1921830	1922318	+	25	CCACAGAUGUGACA	GCCCUCC	UCAU	GGAGGGC	UUUUUUUGUUCAAAC	-14.3	5.79	-4.7
possible Pex protein (*pex*)	1933158	1933634	+	-35	AUGUCCCGUGGCA	CAGCUGGCUGG	AAUCCCA	CCGGCUGGUG	UUGAACUAACGCGAU	-9.7	3	-11.5
CHP	1949349	1950497	+	-28	ACGCCGAACAGGGC	GUGGACCUC	AAACGG	GGGGUUGAC	UAUUAAUUCAAUGGG	-9.3	3.26	-6.6

Annotation	Start	End	Strand	Pos	Seq1	Seq2	Seq3	Seq4	Seq5	ΔG1	ΔG2	ΔG3
CHP	1957744	1958100	+	26	CGCCUCAUCCGCAG	UGGCCUUCGGC	GCCGCG	GCCGCCGCCG	UUUUUACUUAGGACU	-11.0	4.87	-7.2
HNH endonuclease family protein	1980276	1980776	-	12	CUGAGCUGAGUUCA	GUCCUCGUUC	UGUU	GGCUGAGGGC	UUCCAGUUUCUGCCG	-9.3	3.16	-10.0
CHP	1987261	1987626	+	29	GAAGACAGUCUUGC	CCCUAUGGUUC	CCGUAC	GGUCCAUGGGG	AUUUUUAUGACCCGA	-11.9	3.49	-6.8
CHP	1989574	1989822	-	123	AAUGCGUCGAACUG	GGGCGGG	GUGGGA	CCUGCCC	UUUUUUGUGCUCAGC	-13.9	5.31	-7.1
putative photosystem I reaction center subunit XI] (*psaL*)	2000166	2000657	+	64	CCCAACGCUGACCC	CCGGCUGUU	CCAUCAAGA	DACAGCCGGGGG	UUUCUGACUGUCAGG	-8.3	3.95	-10.4
elongation factor EF-Tu (*tufA*)	2028879	2030078	+	27	UGAUCACUGAUGGA	UGGGGGGA	GUCAUCCCCCCA	UCCCCUACA	UUUGUUUUGAACCUC	-8.2	4.66	-6.7
				83	UCGACAACUGAAUU	CAGGACUCC	CUUG	GGAUCCUG	UUCGCUUCUUUCUUA	-8.3	3.37	-10.1
30S ribosomal protein S10 (*rps10*)	2030195	2030515	+	-3	CAUUGAAGUGAAGC	UCUGAACGC	UUCA	GCGUCAGA	UUCUGCUUCCUAGGA	-8.5	3.54	-10.6
photosystem II D1 protein form II (*psbA3*)	2043226	2044305	+	63	AACUGAAUCGGAAA	GCCCUCACC	AAUC	GGUGGGGGC	UUUUUGUUGUGAAAA	-17.1	5.19	-6.5
possible high light inducible protein (*hli6*)	2074038	2074178	+	22	GCGAUCUAAAAAAG	CCCCUUCGG	CCUGG	CCGGAGGGG	UUUUUGCUGUCUGCU	-15.9	4.85	-8.6
				144	CCGUCGUGAUGGCU	CCAGGUUGCU	GAGC	AGUGCCUGG	UUAUUUUCACUAUCG	-8.6	4.24	-6.1
possible porin (*som*)	2119133	2120638	-	53	CCUCACACAGAUCA	GCCCCCC	UUUCA	GGGGGGC	UUUUUGUGUUUAACA	-15.2	5.2	-6.2
possible porin (*som*)	2123030	2124475	-	41	UCUUUAUUCCUAAA	CCCACCG	CUCUG	CGGUGGG	UUUUUUUUCGUCUUU	-12.0	5.97	-4.3
CHP	2124656	2126284	+	-23	GAUCGACCACUCCC	UGCCGAC	CCAC	GUCGGCG	UUCCUGAACAAAACG	-10.1	3.14	-11.1
photosystem II D2 protein (*psbD2*)	2131112	2132167	-	23	GACCAUAGUUACCA	GCCUCCC	GAAA	GGGAGGC	UUUUUUAUGUAAUGA	-14.5	5.4	-4.7
HP	2143657	2144055	+	9	CGCAUGAAUACAUA	GCCUCGC	AUUUCUA	GCGGGGC	UUUUUUUGAUACUUG	-11.6	5.58	-5.3
				46	UUUUUGAUACUUGA	UAGCCGGU	UAAC	ACCGGCUG	UUUUUUUAGUGGUCG	-11.7	5.56	-5.5
putative RNA-binding protein (RRM domain)	2153621	2154055	+	37	UUAAGCUCAAUUAA	CCCCGGAUGU	CCAG	GCCUUCGGGG	UUUUUCGAUGGGCCG	-11.0	4.66	-9.5
CHP	2197435	2203479	-	27	CGCAAGGCUUUCAA	CAACCCUCAUGGC	UCAG	GCCAUGAGGGUUG	UUUUAUGGACUUCCA	-24.6	4.43	-8.9
50S ribosomal protein L7/L12 (*rpl12*)	2242893	2243288	-	8	UCAAGUGAUCCUUU	UUGGGCUGG	UCUCA	CCGGCCCGA	UUCACCUCUUCCACA	-14.4	3.08	-11.6
				54	UCCACAAAACCAAG	CCGGUCC	CCAG	GGGCCGG	UUUUUUGUUGGACAC	-12.9	5.45	-6.8
glutamate N-acetyltransferase / ornithine N-acetyltransferase (*argJ*)	2256316	2257545	-	223	CGCUUUUGUGGAAU	UCAUGACCCU	AAGUUGG	AGGGAAGUGA	UUCCCUACUUUCAUG	-7.8	3.17	-10.9
putative alkaline phosphatase/5' nucleotidase	2297203	2299455	-	41	UCAACCAAUCAUUG	CCCCUCGC	GUCA	GCGAGGGG	UUUUUUAUGUGCUUA	-16.9	5.42	-6/0
putative alkaline phosphatase	2299701	2301431	-	-21	UGACACCUACGCAA	UCAGUCAGGC	UGAUCUU	GCUUGAUCGA	UUUAGGUUGCACUU	-8.7	4.28	-9.7
				27	UUGCACUUCCUUGA	CCCCUCGC	UUUG	GCGAGGGG	UUUUUGUUUAGUAG	-15.6	5.68	-5.1
CHP	2341603	2342880	+	32	GUAAACGAAUCAAA	UUGGCCUCAAUC	GUGA	GAUUGGGGCCAA	UUUCUAUUCACUUCA	-19.3	4.24	-7.2
nitrate transporter, MFS family	2368191	2369063	+	38	GGCGCUGCAUCACG	GCCCGGAGC	UUCG	GCUUCGGGC	UUUUUCUCUUUCGUAA	-17.0	4.82	-7.2
cyanate lyase (*cynS*)	2393373	2393816	-	28	UGGACUUAUCUAAA	GCCCAAU	UCUU	AUUGGGC	UUUAGGUUCUAAACU	-9.4	3.81	-9.0
Shikimate / quinate 5-dehydrogenase (*aroE*)	2415218	2416087	+	43	UGGAACCCCCCCUC	UGGCAGCG	ACUGGU	CGCGCCA	UUGAUGUAUCUCCUG	-8.8	3.36	-9.7

| threonine synthase (*thrC*) | 2433061 | 2434167 | + | 26 | GUCUGAUCCCUCAC | GACCCC | GAAA | GGGGUC | UUUUUUUUGCGAACC | -11.2 | 5.75 | -5.1 |

Note: same as Table 1.

References

[1] Richardson J P. Loading Rho to terminate transcription. *Cell*, 2003, 114: 157–159.

[2] Henkin T M. Control of transcription termination in prokaryotes. *Annu. Rev. Genet.,* 1996, 30: 35–57.

[3] d'Aubenton Carafa,Y, Brody, E, and Thermes, C. Prediction of rho‑independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem‑loop structures. *J. Mol. Biol.,* 1990, 216: 835–858.

[4] Yarnell W S, Roberts J W. Mechanism of intrinsic transcription termination and antitermination. *Science*, 1999, 284: 611–615.

[5] Gusarov I, Nudler E. The mechanism of intrinsic transcription termination. *Mol. Cell*, 1999, 3: 495–504.

[6] Grundy F J *et al.* Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: disparity between conservation and functional requirements. *Nucleic Acids Res.,* 2002, 30: 1646–1655.

[7] Brendel V, Trifonov E N. A computer algorithm for testing potential prokaryotic terminators. *Nucleic Acids Res.,* 1984, 12: 4411–4427.

[8] Ermolaeva M D *et al.* Prediction of transcription terminators in bacterial genomes. *J. Mol. Biol.*, 2000, 301: 27–33.

[9] Lesnik E A *et al.* Prediction of rho‑independent transcriptional terminators in Escherichia coli. *Nucleic Acids Res.,* 2001, 29: 3583–3594.

[10] Unniraman S, Prakash R, Nagaraja V. Conserved economics of transcription termination in eubacteria. *Nucleic Acids Res.,* 2002, 30, 675–684.

[11] Palenik B *et al.* The genome of a motile marine *Synechococcus*. *Nature*, 2003, 424: 1037–1042.

[12] Hess W R. Genome analysis of marine photosynthetic microbes and their global role. *Curr Opin Biotechnol.,* 2004, 15: 191–198.

[13] Scanlan D J. Physiological diversity and niche adaptation in marine Synechococcus. *Adv Microb Physiol.,* 2003, 47: 1–64.

[14] Chen X, Su Z, Dam P, Palenik B, Xu Y, Jiang T. Operon prediction by comparative genomics: an application to the *Synechococcus sp.* WH8102 genome. *Nucleic Acids Res.,* 2004, 32: 2147–57.

[15] Wan X-F, and Xu, D. Rnall: a novel algorithm for RNA local secondary structure prediction. *Nucleic Acid Research*, submitted.

[16] Altschul S F *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.,* 1997, 25: 3389–3402.

[17] Mathews D H *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.,* 2004, 101: 7287–7292.

[18] Mathews D H *et al.* Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.,* 1999, 288: 911–940.

[19] Sugimoto N *et al.* Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. Biochemistry, 1995, 34: 11211–11216.

[20] Sun Z, Xia X, Guo Q, Xu D. Protein structure prediction in a 210-type lattice model: parameter optimization in the genetic algorithm using orthogonal array. *J Protein Chem.*, 1999, 18: 39–46.

[21] Wang L, Trawick J D, Yamamoto R, Zamudio C. Genome-wide operon prediction in Staphylococcus aureus. Nucleic Acids Res., 2004, 32: 3689–3702.

[22] Ingham C J, Hunter I S, Smith M C. Rho-independent terminators without 3' poly-U tails from the early region of actinophage oC31. Nucleic Acids Research, 1995, 23: 370–376.

[23] Reynolds R, Chamberlin M J. Parameters affecting transcription termination by Escherichia coli RNA. II. Construction and analysis of hybrid terminators. J Mol Biol., 1992, 224: 53–63.

[24] Abe H, Aiba H. Differential contributions of two elements of rho-independent terminator to transcription termination and mRNA stabilization. Biochimie, 1996, 78: 1035–1042.