FEBS Letters 580 (2006) 6909–6914

On the interplay of gene positioning and the role of rho-independent terminators in *Escherichia coli*

Nozomu Yachie[a,b], Kazuharu Arakawa[a,*], Masaru Tomita[a,b,c]

[a] *Institute for Advanced Biosciences, Keio University, 5322 Endo, Fujisawa 252-8520, Kanagawa, Japan*
[b] *Bioinformatics Program, Graduate School of Media and Governance, Keio University, Fujisawa 252-8520, Japan*
[c] *Department of Environmental Information, Keio University, Fujisawa 252-8520, Japan*

Received 14 September 2006; revised 7 November 2006; accepted 21 November 2006

Available online 30 November 2006

Edited by Takashi Gojobori

Abstract **The majority of intrinsic rho-independent terminator signals, reported to consist of stable hairpin structures followed by T-rich regions, possess the potential to operate bi-directionally and to induce transcription terminations on both strands of the DNA duplex in *Escherichia coli*. By using RNAMotif software, we investigated the distributions of termination motifs around the 3′-ends of overlapping and non-overlapping genes at the genomic level. We suggest that the positions of compactly encoded *E. coli* genes and rho-independent terminators are optimized to terminate the adjoining genes on their antisense strands efficiently, and not to mis-terminate overlapping transcripts, due to their bi-directional properties.**

Keywords: Rho-independent terminator; Overlapping genes; Operon; RNAMotif; Bioinformatics

1. Introduction

In bacterial species, rho-independent terminators are widely accepted as canonical intrinsic termination signals in the DNA that function in minimal in vitro transcription systems [1]. Although bacterial transcription termination is controlled by many factors, genomic analyses that have taken advantage of the availability of sequence data have suggested that most of the transcription terminations of annotated protein-encoding and non-coding RNA transcription units are regulated by intrinsic rho-independent terminators [1]. Intrinsic terminators are characterized by a GC-rich palindromic structure followed by a tail of A-rich region on the template DNA strand [2,3]. The palindromic region forming a stem–loop structure in the nascent RNA is reported to pause RNA polymerases [4–6] and weaken the interaction between elongated oligo RNA and template DNA [7,8]. Owing to the weak hybridization energy between rU and dA [9], final release may also be facilitated by the following U-rich region [10].

In view of these subgenomic features, many algorithms have been developed to identify the intrinsic terminators [11–16]. RNAMotif software is used to search RNA motifs defined by the nucleotide sequences and structural constraints given by model file arguments named 'descriptor' [17]. Descriptor files of the intrinsic terminator have been reported by Lesnik et al. [15], based on a model consisting of stable hairpin structures followed by T-rich regions as proposed by d'Aubenton Carafa et al. [13]. As demonstrated by Livny et al. [18], RNAMotif misses out a fraction of documented intrinsic terminator motifs detectable by other resources such as TransTerm [19]. Nevertheless, RNAMotif has been widely utilized in many bioinformatics approaches; for example, to define transcription terminations and to predict novel transcription units [18,20–23]. Following these findings and applications, a bi-directional rho-independent terminator in the intergenic region (IGR) between the *tonB* and *P*14 genes has been identified; this could function bi-directionally by sharing the two complementary hairpins on the double-stranded genomic DNA, terminating the transcription of both genes [15,24]. Lesnik et al. also suggested that many bi-directional rho-independent terminators might exist in the *Escherichia coli* genome [15].

Here, we discuss how transcription units and intrinsic terminators are coordinately organized and distributed within the *E. coli* genome based on the prediction of possible intrinsic terminator motifs at the genomic level using RNAMotif. In order to analyze the positioning characteristics of the predicted motifs in light of the patterns of the 3′-ends of genes overlapping on neighboring genes or adjoining IGRs, either uni-directionally or bi-directionally, we constructed bioinformatics workflows using G-language Genome Analysis Environment, a generic analysis workbench for bioinformatics [25,26]. We first present that the majority of rho-independent terminators are potentially bi-directional owing to the symmetric features of their hairpins, with low free energies in both the genetic complements. Further, by comparing the analysis of the 3′-ends of overlapping and non-overlapping genes with those in the operon transcripts, we demonstrate computationally the circumvention and utilization of the intrinsic terminations associated with adjacent genes and with the bi-directional properties of these terminators. We suggest that the loci of intrinsic rho-independent terminators may have been optimized to be efficiently positioned in the downstream regions of genes, compactly encoded on the *E. coli* genomic DNA.

*Corresponding author. Fax: +81 466 47 5099.
E-mail address: gaou@sfc.keio.ac.jp (K. Arakawa).

Abbreviations: IGR, intergenic region; tRNA, transfer RNA; rRNA, ribosomal RNA; *E. coli*, *Escherichia coli*; nt, nucleotides; bp, base pairs

0014-5793/$32.00
doi:10.1016/j.febslet.2006.11.053

2. Materials and methods

2.1. Genome annotation and sequence data

The complete genome sequence of *E. coli* strain K-12 MG1655 (RefSeq: NC000913) containing 4639675 bp was obtained from the National Center for Biotechnology Information (NCBI) ftp server (ftp://ftp.ncbi.nlm.nih.gov). The annotations and positions of 4441 documented genes, including tRNA, rRNA, and other non-protein-coding RNAs, and 997 identified transcription units of *E. coli* were downloaded as flat files from the EcoCyc database, Encyclopedia of *Escherichia coli* K-12 Genes and Metabolism (http://www.ecocyc.org, accessed 23 July 2006) [27].

2.2. Preparation of data on 3′-end positioning and operon categories

The 3′-ends of every gene were grouped into four categories, and the genes included in the operon transcripts were divided into two categories. Among the 3′-ends of the 4441 genes, 2335 were non-overlapping and adjoined IGRs with the adjacent downstream gene on the same DNA strand (defined as 'tail-to-head neighboring'), 1230 adjoined IGRs with the adjacent downstream gene on the complementary strand ('tail-to-tail neighboring'), 690 overlapped with their adjacent genes uni-directionally ('tail-to-head overlapping'), and 186 overlapped bi-directionally ('tail-to-tail overlapping'). Out of 977 transcription units stored in EcoCyc, 355 were prepared as operon transcripts, each encoding more than two genes. The endmost genes of the respective operon transcripts were defined as 'operon-end' genes, and the remaining 884 genes were defined as 'within-operon' genes.

To determine the validity of our data set, which included annotated but experimentally unidentified genes, we repeated the same analysis on another data set excluding genes encoding 'hypothetical proteins', as annotated in GenBank, and filtered by a Gene Prediction Accuracy Classification (GPAC) test [28] (see Supplementary Materials for details).

2.3. Prediction of rho-independent terminators

Using RNAMotif, the sequences and structural motifs of rho-independent terminators containing hairpins followed by T-rich regions were predicted at the genomic level. The original RNAMotif model files were constructed by machine learning of the nucleotide composition around experimentally confirmed rho-independent terminators, which included an A-rich region at the 5′-side of the unidirectional motif in addition to the experimentally suggested hairpin structure with a 3′-side T-rich region, for more efficient prediction of the terminators. However, this model with the 5′-side A-region inherently results in preferential prediction of bi-directional terminators. Therefore, to determine the bi-directional properties of rho-independent terminators by using unbiased criteria, we omitted the 5′ A-region from the model and adopted only the total ΔG^0 scores of hairpins with a 3′ T-region for the prediction of each motif at its optimal cut-off score of $\leqslant -4$, which is reported to significantly cover the majority of experimentally identified intrinsic terminators [15]. In order to compare the bi-directional properties of predicted motifs with different scores, the predicted motifs were further prepared at three ΔG^0 thresholds of -4, -8, and -12. The positions of transcription terminations were defined accordingly as 7 nt upstream of the 3′-ends of the predicted motifs. Within the predicted terminators, those having possible hairpin structures in their complementary DNA strands were further screened for bi-directionality by seeking a hairpin structural motif located directly opposite of the original hairpin position; that is, symmetric bi-directional rho-independent terminators were defined if the positions of the partial base-stacking pairs of two complementary hairpin structures on either of the two strands coincided with each other.

2.4. Characteristic analyses of predicted intrinsic terminators

The physical properties to potentially form stable hairpins with low free energies reside in both DNA strands at a locus. However, pairs of significant hairpin structures of the intrinsic terminators predicted by RNAMotif in the same loci on opposite strands are not always symmetric, when considering the conditions for other sequence attributes in the motif and/or hybridization of rU and rG. Therefore, in order to estimate the bi-directional properties of rho-independent terminators, we performed not only the prediction of symmetric bi-directional terminators, but also the analysis of the lengths between loop medians and termination positions of the predicted motifs and the distances between the termination positions of two motifs encoded on the complementary DNA strands. The two-sided 95% confidence intervals of the lengths between loop medians and termination positions of the predicted motifs were calculated. Then, for each predicted terminator, we counted the number of antisense terminator motifs, whose termination positions were within twice the ranges between the two-sided 95% confidence intervals upstream of the termination position in the sense strand.

2.5. Estimation of occurrence ratio of predicted rho-independent terminators with respect to each position around 3′-end of genes

Distributions of the termination positions of the predicted rho-independent terminators were analyzed bi-directionally around the 3′-ends of the genes within the each 3′-end positioning categories. For every gene in each category, termination positions of motifs were counted using 20-bp sliding windows with 1-bp displacement, and the motif count at each distance (bp) was normalized and defined as 'Ratio (%)' by dividing by the total number of genes in the category, and by the window size, i.e. 20. In order to discuss how terminator motifs were characteristically distributed in the each 3′-end positioning category, the highest 'Ratio (%)' of both DNA strands defined as peak ratios and their relative positions from the 3′-end of gene were paralleled with those of the other categories.

3. Results and discussion

In this study, the optimization of gene positioning associated with the bi-directional properties of intrinsic rho-independent terminators in *E. coli* was discussed. We initially predicted the rho-independent terminators at genomic level by using RNAMotif software, and suggested many of the rho-independent terminators possibly have bi-directional properties that not only mis-terminates an overlapping transcription units on the terminators in the same DNA strand but also those in the complementary strand. By analyzing distributions of predicted rho-independent terminator motifs around the 3′-ends of genes categorized into the four 3′-end positioning types of 'tail-to-head neighboring', 'tail-to-tail neighboring', 'tail-to-head overlapping' and 'tail-to-tail overlapping', our results suggested that the positioning of intrinsic terminators has minimized mis-termination of overlapping genes on the same strand and the complementary strand. Furthermore, bi-directional intrinsic terminators are speculated to distribute in such a way as to maximize co-ordinated termination of adjoining genes on opposite strands. Though 'within-operon' genes are not terminated in the proximity of downstream region of the 3′-ends, we indicated that the minimization of mis-termination were independent from the biases of the 'within-operon' genes. This is confirmed by observing the distribution pattern around the 3′-ends utilizing the information of identified operon transcripts and by comparing the numbers of respective 3′-end positioning type genes in the two operon categories of 'within-operon' and 'operon-end' with those of all genes.

3.1. Comparison of 3′-end positioning types within each operon category

Genes in each of the 3′-end positioning categories were counted and compared with respect to the corresponding operon category (Table 1). Among the total 884 'within-operon' genes, the numbers of 'tail-to-head neighboring' and 'tail-to-head overlapping' 3′-ends were 611 (69.12%) and 263 (29.75%), respectively. As expected, these proportions were both higher than 'operon-end' genes and all genes, because

Table 1
Numbers of genes with each 3′-end positioning type within the categories of 'within-operon', 'operon-end', and all genes

Positioning type of 3′-end	'Within-operon'genes	'Operon-end'genes	All genes
Tail-to-head neighboring	611 (69.12%)	203 (57.18%)	2335 (52.58%)
Tail-to-tail neighboring	9 (1.02%)	133 (37.46%)	1230 (27.70%)
Tail-to-head overlapping	263 (29.75%)	8 (2.25%)	690 (15.54%)
Tail-to-tail overlapping	1 (0.11%)	11 (3.10%)	186 (4.19%)
Total	884	355	4441

the operon encodes multiple closely packed genes transcribed together.

Only one gene of the 'tail-to-tail overlapping' 3′-end type was found in the 'within-operon' category, and its proportion of 0.11% was markedly lower than the 3.10% of these genes in the 'operon-end' category and the 4.19% among all genes (Table 1). The percentage of genes with 'tail-to-tail neighboring' 3′-ends (1.02%) in the 'within-operon' category was also markedly lower than in the other categories (37.46% for 'operon-end' genes and 27.70% for all genes); 133 out of the 355 'operon-end' genes had 'tail-to-tail neighboring' 3′-ends – the highest proportion among those of 'tail-to-tail neighboring' 3′-ends in all categories. The same tendencies were observed upon removal of the hypothetical genes (see Supplementary Materials).

3.2. Bi-directional properties of intrinsic rho-independent terminators

Using RNAMotif, 15665, 5445, and 1585 rho-independent terminators were predicted at the respective ΔG^0 thresholds of −4, −8, and −12. We analyzed the properties of the predicted terminators by analyzing their lengths (Fig. 1a–c), the distances between the termination positions of two motifs encoded on the complementary DNA strands (Fig. 1d–f), and the numbers of rho-independent terminators that were bi-directional (Fig. 1g–i).

The maximum and minimum lengths between the loop medians and termination positions of all predicted motifs shown in Fig. 1a–c were 43 and 10.5 nt, respectively, and the average size was 19.57 nt. The regions between the two-sided 95% confidence intervals ranged from 13 to 26 nt, from 13.5 to 26 nt, and from 15 to 26 nt, for ΔG^0 thresholds of −4, −8, and −12, respectively. The possible distances of the predicted termination positions in the antisense DNA strand from those in the sense strand are displayed in Fig. 1d–f. The total frequency of antisense terminators distributed in the regions between −52 and −26 nt away from the termination position of terminator motif in the sense strand at ΔG^0 threshold −4 was 133.04%. The region between positions −52 and −26 nt was defined in accordance with the two-sided 95% confidence interval regions from positions 13 to 26 nt as described above, and by taking twice the length to estimate the equivalent length of the bi-directional motifs, in order to account for the antisense structures. According to the same procedure, the total frequencies of antisense terminators were calculated to be 189.35% and 219.74% at ΔG^0 thresholds −8 and −12, respectively. Within the sets of complementary terminators at ΔG^0 scores −4, −8, and −12 and below, 3375 (21.54%), 1644 (30.19%), and 529 (33.38%), respectively, were bi-directional rho-independent terminators, sharing double-stranded DNA regions to form symmetrical hairpin structures with ΔG^0 scores −4 and below. The distances between the termination positions of symmetrical complementary pairs were distributed around 40 bp. Within the putative terminators predicted by RNAMotif, those of lower ΔG^0 scores were suggested to have higher bi-directional properties. This result provided the rationale that the observations of predicted terminators having low ΔG^0 scores were reasonable to discuss their bi-directionalities in further analyses, conducted as follows. Although a comprehensive listing of 439 candidate bi-directional terminators has already been reported by Lesnik et al. using a model suited for the identification of bi-directional motifs as explained above [15], we showed here additionally using a uni-directional motif model, that the majority of intrinsic terminator motifs with

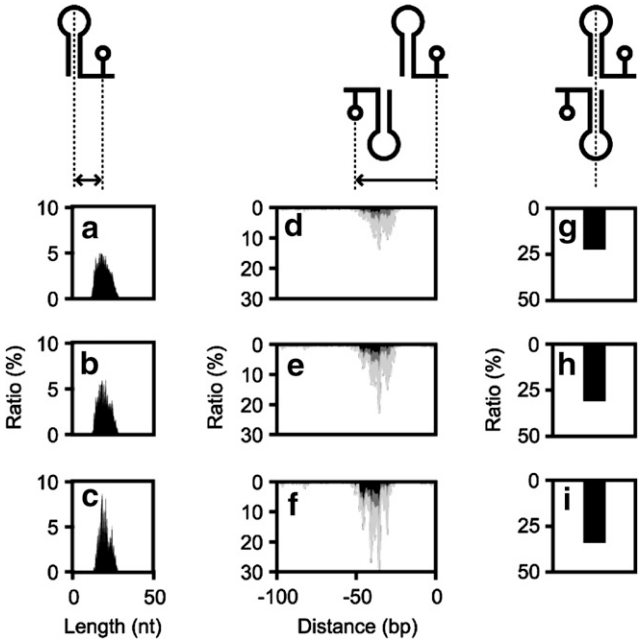

Fig. 1. Characteristic analysis of predicted rho-independent terminators. Lengths between loop medians and termination positions of predicted rho-independent terminator motifs with ΔG^0 scores of ⩽−4, −8, and −12 are displayed respectively in (a), (b), and (c). Count of motifs at each length (nt) was normalized as 'Ratio (%)' by dividing by the total number of motifs predicted at the respective threshold. Distances of predicted transcription termination positions in the antisense DNA strand from those in the sense strand are shown in (d–f). (d), (e), and (f) represent distributions of predicted motifs in the antisense strand from motifs in the sense strands with ΔG^0 scores of ⩽−4, −8, and −12, respectively. In the stacked histogram, the count of antisense motifs at each distance (bp) was normalized as 'Ratio (%)' by dividing by the total number of motifs predicted at the respective threshold. ΔG^0 scores of antisense motifs are indicated by light gray (⩽−4 and >−8), gray (⩽−8 and >−12), and black (⩽−12). Counts of bi-directional rho-independent terminators with sense strand motifs with ΔG^0 scores of ⩽−4 (g), −8 (h), and −12 (i) were also normalized as 'Ratio (%)' by dividing by the total number of motifs predicted by the respective threshold.

sufficiently low ΔG^0 have antisense complements and thus have a high likelihood of forming bi-directional terminators. Therefore, most rho-independent terminators probably have highly bi-directional properties, as a result of the symmetry-based low free energies and hairpin formations on the double-stranded DNA.

3.3. *Optimization of* E. coli *gene positioning associated with intrinsic terminators*

The termination positions marked by the predicted rho-independent terminators were analyzed bi-directionally around the 3′-ends of the genes. The distances of the terminations of the predicted terminators and the bi-directional terminators from the 3′-ends of all genes in both DNA strands are shown in Fig. 2a–d, including certain fractions of those having lower ΔG^0 scores of $\leqslant -8$ and $\leqslant -12$, and their bi-directional properties were also suggested by focusing on the downstream parts of the 3′-ends of the genes. The predicted termination positions were distributed at around +33 bp with the peak ratio of 3.92% in the sense strand, and at about −4 bp with the peak ratio of 2.66% in the antisense strand, as calculated from the normalized count per gene per position (for details, see Section 2.5). The terminations of bi-directional terminators were distributed at around +34 bp in the sense strand, where the peak ratio was 1.53%; in the antisense strand they were distributed at around 0 bp, where the peak ratio was 1.23%.

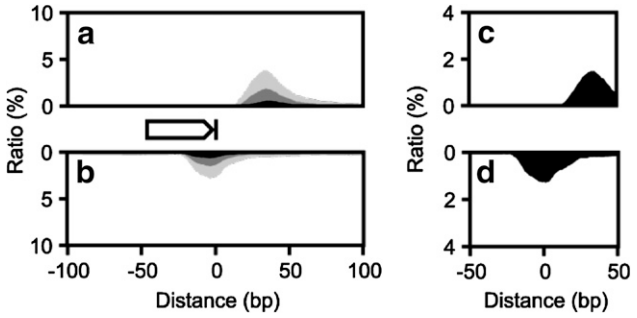

Fig. 2. Distributions of rho-independent terminators around 3′-ends of all genes. Distances of termination positions of rho-independent terminator motifs distributed in sense and antisense DNA strands from the 3′-ends of genes are shown in (a) and (b), respectively. Frequency of the motifs with respect to the position around 3′-end of genes were normalized as 'Ratio (%)' (for details, see the Section 2.5). ΔG^0 scores are indicated by light gray ($\leqslant -4$ and > -8), gray ($\leqslant -8$ and > -12), and black ($\leqslant -12$) bars in the stacked histogram. From the histograms, bi-directional rho-independent terminators (not colored by motif scores) were extracted to (c) sense strand and (d) antisense strand.

Similarly to the result for all genes, the termination positions were distributed around the 'tail-to-head neighboring' and 'tail-to-tail neighboring' 3′-ends (Fig. 3a–d and e–h). The peak ratios of 3.66% (1.09% bi-directional) in the sense strand and 1.36% (0.79% bi-directional) in the antisense strand, calculated using all predicted terminators, were distanced +34 bp (+37 bp bi-directional) and −4 bp (+2 bp bi-directional), respectively, from the 3′-ends of the genes in the 'tail-to-head neighboring' category. Likewise, for the 'tail-to-tail neighboring' category, the peak ratios were 7.07% (3.63% bi-directional) in the sense strand and 6.86% (3.19% bi-directional) in the antisense strand, distanced +32 bp (+32 bp bi-directional) and −4 bp (0 bp bi-directional) from the 3′-ends of the genes. The peak ratios of the 'tail-to-tail neighboring' category, including higher fractions of lower ΔG^0 scores of $\leqslant -8$ and $\leqslant -12$, were significantly higher than those of the 'tail-to-head neighboring' and of all genes (Fig. 2), suggesting the marked adoption of bi-directional properties that can terminate a pair of 'tail-to-tail neighboring' genes with one shared bi-directional terminator.

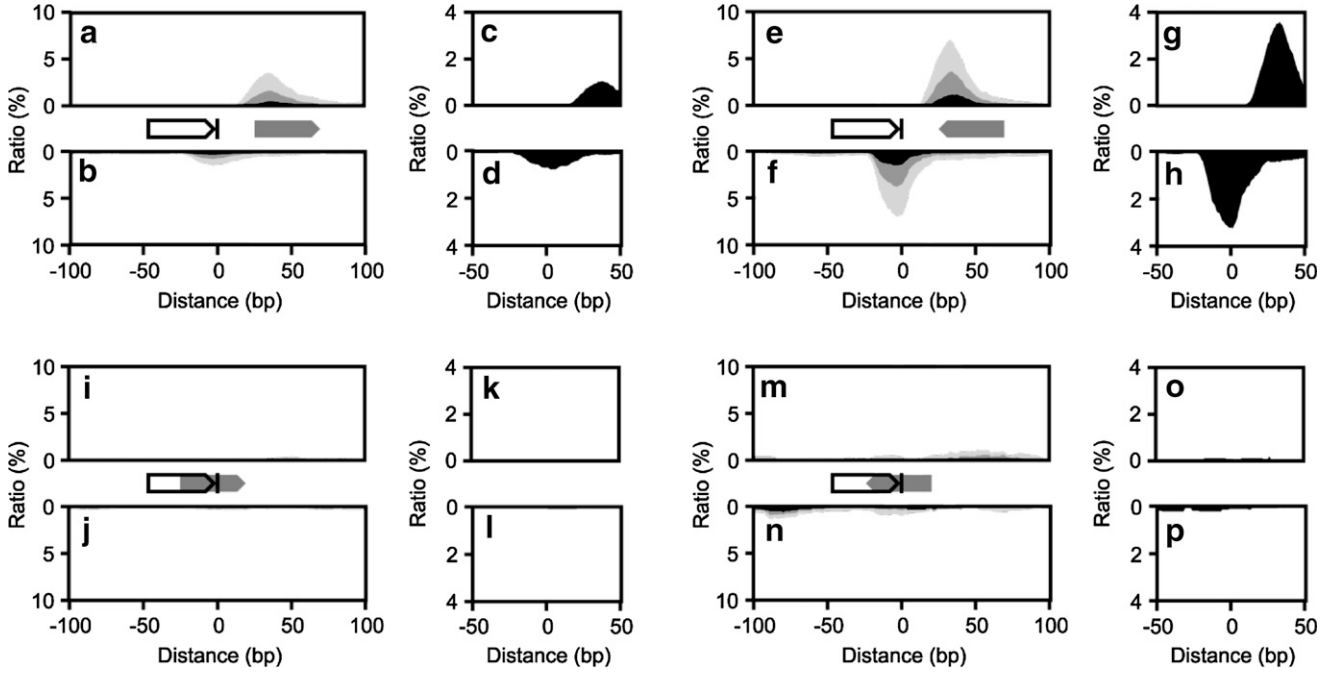

Fig. 3. Distributions of rho-independent terminators around 3′-ends of 'tail-to-head neighboring' (a–d), 'tail-to-tail neighboring' (e–h), 'tail-to-head overlapping' (i–l), and 'tail-to-tail overlapping' (m–p) genes. For details, see caption to Fig. 2.

In the categories of 'tail-to-head overlapping' and 'tail-to-tail overlapping', marked distributions of terminators were not observed, compared with those described above (Fig. 3i–l and m–p). We suggest that the result in the 'tail-to-head overlapping' category, showing a circumvention of the terminator structures, might be designed to prevent mis-termination, pausing the transcription in the middle of another overlapped gene in co-orientation. Genes in this category may be terminated by other intrinsic motifs that are not given in the model used here, or by other transcription termination factors [1]. However, a comparison of the 'within-operon' and 'operon-end' categories (Fig. 4) showed that the 'within-operon' category also lacked marked distributions, and many of the 'tail-to-head overlapping' genes were suggested to be involved in the operon transcripts (see Table 1 for details). We thus speculate that some of the 'tail-to-head overlapping' genes are 'within-operon' genes, and therefore are not terminated at the proximal downstream regions of their 3′-ends.

From the results of the 'tail-to-tail overlapping' category, we further consider that the uni-directional intrinsic motifs terminating the 'tail-to-tail overlapping' genes are avoided to decrease the chance of mis-termination of their adjacent genes in antisense strands by their possible bi-directional properties. Only a negligible number of genes in this category belonged to the 'within-operon' category (Table 1), and therefore some kind of termination other than by the intrinsic signal is expected – for example, by rho-dependent termination [1]. Negative selective pressure works against gene formation in the antisense strands of operon transcripts (Table 1), and this may be partly attributable to the circumvention of mis-termination of the antisense gene caused by the bi-directional terminator. On the other hand, in the analyses of the 'operon-end' category, the high peak ratios of 8.07% (3.66% bi-directional) of the sense strand and 4.19% (2.75% bi-directional) of the antisense strand were distanced at +32 bp (+34 bp bi-directional) and −2 bp (+5 bp bi-directional) from the 3′-ends of the genes (Fig. 4); we considered that this result was caused by the inclusion of many 'tail-to-tail neighboring' genes in the 'operon-end' category, as suggested in Table 1, or was due to the high sensitivity of the experimentally confirmed transcription terminations.

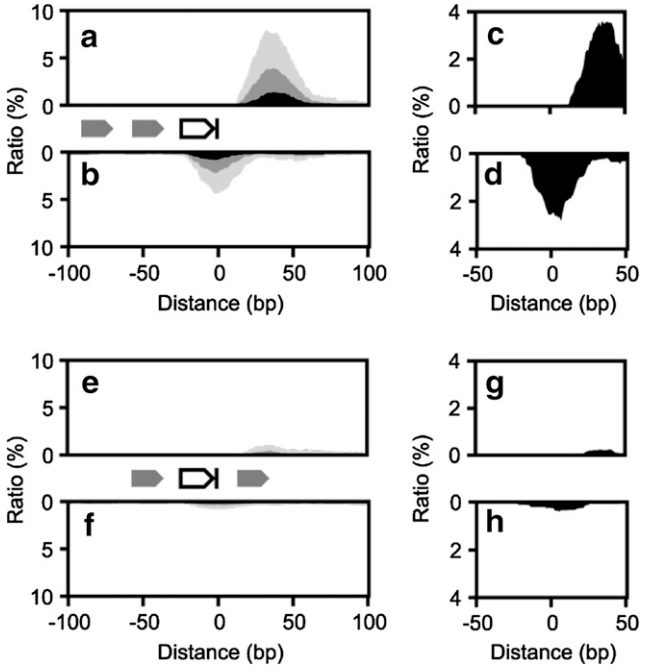

Fig. 4. Distributions of rho-independent terminators around 3′-ends of 'operon-end' (a–d) and 'within-operon' (e–h) genes. For details, see caption to Fig. 2.

The same procedures were repeated for every positioning pattern of the 3′-ends after filtering by the GPAC test and for randomly sampled sets of 100 genes by a bootstrap test, but the results showed no marked change (see Supplementary Materials). We therefore conclude that of our results were not biased by variations in the annotation accuracy of genes or by the numbers of genes in the respective categories.

We conducted a bioinformatics analysis to investigate gene positioning in relation to the intrinsic termination motifs, and we discussed here the optimal positioning of bi-directional terminators in utilizing efficient terminations of head-on and non-overlapping gene pairs in a dual role and in circumventing the mis-termination of overlapping transcripts. A previous study has indicated that such structure-dependent intrinsic terminators appear to be employed in only a few bacterial species, as suggested by the fact that the calculation of free energies around the 3′-ends of their genes resulted in the lowest distributions in regions downstream of 'tail-to-tail neighboring' genes [29]. We also suggest that our data support the concept of dynamic gene positioning by a combination of intrinsic terminators that has possibly been optimized in *E. coli*.

Acknowledgements: The authors are grateful to the members of MGSP at the Institute for Advanced Biosciences, Keio University, for their critical discussions. This research was supported in part by the Japan Society for the Promotion of Science (JSPS).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2006.11.053.

References

[1] Nudler, E. and Gottesman, M.E. (2002) Transcription termination and anti-termination in *E. coli*. Genes Cells 7 (8), 755–768.
[2] Platt, T. (1986) Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. 55, 339–372.
[3] Yarnell, W.S. and Roberts, J.W. (1999) Mechanism of intrinsic transcription termination and antitermination. Science 284 (5414), 611–615.
[4] Farnham, P.J. and Platt, T. (1981) Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. Nucleic Acids Res. 9 (3), 563–577.
[5] Wang, D., Severinov, K. and Landick, R. (1997) Preferential interaction of the his pause RNA hairpin with RNA polymerase beta subunit residues 904–950 correlates with strong transcriptional pausing. Proc. Natl. Acad. Sci. USA 94 (16), 8433–8438.
[6] Artsimovitch, I. and Landick, R. (1998) Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. Genes Dev. 12 (19), 3110–3122.
[7] Arndt, K.M. and Chamberlin, M.J. (1990) RNA chain elongation by *Escherichia coli* RNA polymerase. Factors affecting the stability of elongating ternary complexes. J. Mol. Biol. 213 (1), 79–108.
[8] Wilson, K.S. and von Hippel, P.H. (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proc. Natl. Acad. Sci. USA 92 (19), 8793–8797.

[9] Lynn, S.P., Kasper, L.M. and Gardner, J.F. (1988) Contributions of RNA secondary structure and length of the thymidine tract to transcription termination at the operon attenuator. J. Biol. Chem. 263 (1), 472–479.
[10] Martin, F.H. and Tinoco Jr., I. (1980) DNA–RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. Nucleic Acids Res. 8 (10), 2295–2299.
[11] Brendel, V. and Trifonov, E.N. (1984) A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. 12 (10), 4411–4427.
[12] Brendel, V., Hamm, G.H. and Trifonov, E.N. (1986) Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J. Biomol. Struct. Dyn. 3 (4), 705–723.
[13] d'Aubenton Carafa, Y., Brody, E. and Thermes, C. (1990) Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem–loop structures. J. Mol. Biol. 216 (4), 835–858.
[14] Ermolaeva, M.D., Khalak, H.G., White, O., Smith, H.O. and Salzberg, S.L. (2000) Prediction of transcription terminators in bacterial genomes. J. Mol. Biol. 301 (1), 27–33.
[15] Lesnik, E.A., Sampath, R., Levene, H.B., Henderson, T.J., McNeil, J.A. and Ecker, D.J. (2001) Prediction of rho-independent transcriptional terminators in *Escherichia coli*. Nucleic Acids Res. 29 (17), 3583–3594.
[16] Unniraman, S., Prakash, R. and Nagaraja, V. (2002) Conserved economics of transcription termination in eubacteria. Nucleic Acids Res. 30 (3), 675–684.
[17] Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A. and Sampath, R. (2001) RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res. 29 (22), 4724–4735.
[18] Livny, J., Brencic, A., Lory, S. and Waldor, M.K. (2006) Identification of 17 *Pseudomonas aeruginosa* sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. Nucleic Acids Res 34, 3484–3493.
[19] Jacobs, G.H., Stockwell, P.A., Tate, W.P. and Brown, C.M. (2006) Transterm – extended search facilities and improved integration with other databases. Nucleic Acids Res. 34, D37–D40.
[20] Chen, S., Lesnik, E.A., Hall, T.A., Sampath, R., Griffey, R.H., Ecker, D.J. and Blyn, L.B. (2002) A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. Biosystems 65 (2–3), 157–177.
[21] Pichon, C. and Felden, B. (2003) Intergenic sequence inspector: searching and identifying bacterial RNAs. Bioinformatics 19 (13), 1707–1709.
[22] Livny, J., Fogel, M.A., Davis, B.M. and Waldor, M.K. (2005) sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. Nucleic Acids Res. 33 (13), 4096–4105.
[23] Yachie, N., Numata, K., Saito, R., Kanai, A. and Tomita, M. (2006) Prediction of non-coding and antisense RNA genes in *Escherichia coli* with Gapped Markov Model. Gene 372, 171–181.
[24] Postle, K. and Good, R.F. (1985) A bidirectional rho-independent transcription terminator between the *E. coli tonB* gene and an opposing gene. Cell 41 (2), 577–585.
[25] Arakawa, K., Mori, K., Ikeda, K., Matsuzaki, T., Kobayashi, Y. and Tomita, M. (2003) G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining. Bioinformatics 19 (2), 305–306.
[26] Arakawa, K. and Tomita, M. (2006) G-language System as a platform for large-scale analysis of high-throughput omics data. J. Pestic. Sci. 31 (3), 282–288.
[27] Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M. and Karp, P.D. (2005) EcoCyc: a comprehensive database resource for *Escherichia coli*. Nucleic Acids Res. 33, D334–D337.
[28] Arakawa, K., Nakayama, Y. and Tomita, M. (2006) GPAC: Benchmarking the sensitivity of genome informatics analysis to genome annotation completeness. In Silico Biol. 6 (1–2), 0006.
[29] Washio, T., Sasayama, J. and Tomita, M. (1998) Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. Nucleic Acids Res. 26 (23), 5456–5463.