THEORETICAL EVALUATION OF TRANSCRIPTIONAL PAUSING EFFECT ON THE ATTENUATION IN *trp* LEADER SEQUENCE

HIROMI SUZUKI, TAKASHI KUNISAWA, AND JINYA OTSUKA
Department of Applied Biological Science, Faculty of Science and Technology, Science University of Tokyo, Noda (278), Japan

ABSTRACT The effect of transcriptional pausing on attenuation is investigated theoretically on the basis of the attenuation control mechanism presented by Oxender et al. (Oxender, D. L., G. Zurawski, and C. Yanofsky, 1979, *Proc. Natl. Acad. Sci. USA.* 76:5524–5528). An extended stochastic model including the RNA polymerase pausing in the leader region is developed to calculate the probability of relative position between the RNA polymerase transcribing the *trp* leader sequence and the ribosome translating the transcript. The present study results in a new rationale that the transcriptional pausing site in the leader sequence makes the attenuation control both more sensitive as an on/off switch and less sensitive to variations in the concentration of cellular metabolites not connected with the need for expressing, or not expressing, the particular operon. It is also proposed that the transcriptional pausing diminishes the dependence of attenuation control characteristics on the number of nucleotides in the leader sequence. This result may be useful for understanding the attenuation control efficiencies of other amino acid leader sequences with different lengths of nucleotides.

INTRODUCTION

Transcription of most amino acid biosynthetic operons in enteric bacteria is controlled by attenuation as well as by repression. In each of these operons, the leader sequence to be transcribed is preceded by the promoter and is followed by structural genes. The attenuator is located in the downstream region of the leader sequence. Whether RNA polymerase terminates transcription at or reads through over the attenuator is regulated by the levels of charged and uncharged $tRNA^{Trp}$ (Morse and Morse, 1976; Yanofsky and Soll, 1977). The 3'-half of terminated tryptophan (*trp*) leader transcript exhibits the extensive secondary structure in vitro (Lee and Yanofsky, 1977) and the studies of *trp* leader mutants indicate that the capacity to form this secondary structure is essential for normal regulation of transcription termination at the attenuation site (Bertrand et al., 1977; Stauffer et al., 1978; Zurawski et al., 1978*b*). The tandem Trp codons in the leader peptide coding region are possibly to be base-paired with a more distal region of the transcript, and this and other possible secondary structures that the *trp* leader RNA can form help explain the physiological response of the operon as well as the behavior of regulatory mutants (Oxender et al., 1979). The DNA sequence of the regulatory region for *E. coli trp* operon transcription is represented in Fig. 1, together with the outline of the regulation control model.

The time required for the potential pairing regions of mRNA to form a stem and loop structure, 10^{-4} s (Pörschke, 1974), is much shorter than the time for RNA polymerase to elongate the RNA chain, (1/45) s/nucleotide (Manor et al., 1969; Rose et al., 1970), or the time for the ribosome to translate a codon in mRNA, (1/15) s/codon (Lacroute and Stent, 1968). It is also suggested that RNA polymerase terminates transcription after transcribing the poly A or T sequence neighboring with a palindromic sequence, probably due to the formation of a stem and loop structure in the transcript and the weak U—A or A—T hydrogen bonds between the transcript and the template (Farnham and Platt, 1980; Otsuka and Kunisawa, 1982). Thus, the problem to be further investigated is about the probability that the ribosome stalls at the Trp codon (10th codon), masking only the region 1 of the transcript (under starvation for tryptophan), or proceeds to the stop codon (15th codon) to mask the regions 1 and 2 (under excess tryptophan) when RNA polymerase arrives at the attenuation site (140th base). Although a simple stochastic model has been proposed for the relative position between the ribosome and RNA polymerase (Manabe, 1981), this model does not contain the transcriptional pausing in the region between the tandem Trp codons and the attenuator found by Winkler and Yanofsky (1981).

The main purpose of the present investigation is to clarify the effect of the transcriptional pausing on the attenuation control efficiency in the *trp* leader sequence.

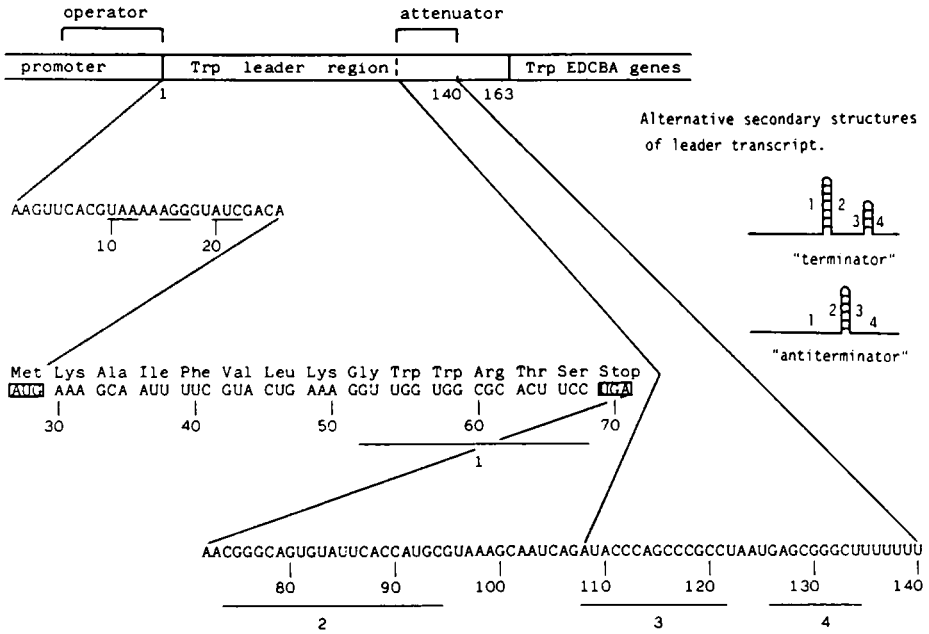

FIGURE 1 Outline of the regulatory region of the tryptophan *trp* operon transcription in *E. coli*. The nucleotide sequence of terminated *trp* leader RNA is shown below where translation start and stop codons are boxed, together with the amino acid sequence of the leader peptide. The RNA bases complementary to 3'-end of 16S rRNA are underlined. Potential pairing regions in the leader transcript are shown by lines with numbers 1, 2, 3, and 4. These four regions can form alternative secondary structures, terminator or antiterminator, depending on the position of the translating ribosome. On the basis of the assumption that the formation of a stem and loop 3·4 is essential for the transcription termination at the attenuation site, Oxender et al. (1979) have proposed the regulation model of *trp* operon transcription that the ribosome stalled at Trp codon(s) under the starvation for tryptophan excludes the formation of a stem and loop 1·2 and facilitates the formation of a stem and loop 2·3, thereby resulting in read-through transcription over the attenuation site.

Although a need for the polymerase pausing site in the attenuator is already discussed (von Heijne, 1982), the discussion still remains within the framework of the formula derived by Manabe (1981). In this paper, a more extended stochastic model is developed to incorporate explicitly the transcriptional pausing into the calculation of the relative position between the ribosome and RNA polymerase. Under some representative conditions of excess and starved amino acids, the probabilities of relative position are calculated on the basis of the stochastic model developed here, and the comparison between the presence and absence of transcriptional pausing is carried out in detail for each condition. The present investigation reveals that the RNA polymerase pausing during the transcription of leader sequence is an ingenious device for the stochastic processes of transcription and translation, enhancing and stabilizing the attenuation control efficiency of *trp* operon transcription.

METHODS

The translation by the ribosome consists of many elementary processes, such as attachment of specific aminoacyl-tRNA molecule corresponding to the codon placed at "A" site, peptide bond formation by peptidyl transferase, ejection of tRNA from "P" site, peptidyl-tRNA translocation from A to P site, and simultaneous movement of mRNA to place the next codon at the A site (Watson, 1977). For the present purpose, however, we will not go into these elementary processes, but only follow the codons of mRNA placed successively at the A site of the ribosome. Correspondingly, we treat the transcription as a stochastic process, only focusing on the base number to be transcribed, although the transcription by RNA polymerase may probably be divided into many elementary processes of polymerase-DNA-nascent RNA complex.

We consider the position probability $W(k, l)$ that RNA polymerase arrives at the kth base in the template while the ribosome stays at the lth codon of the leader RNA. As mentioned in the Introduction, the translation is not faster than the transcription under the usual biological condition, and transcription and translation may be regarded as independent. Then, the position probability $W(k, l)$ may be expressed as the product of these two processes, i.e.,

$$W(k, l) = \int_0^\infty \lambda_{k-1} P_{k-1}(t) Q_l(t) \, dt. \qquad (1)$$

$P_{k-1}(t)$ is the probability that RNA polymerase is on the $(k-1)$th base at time t and λ_{k-1} is the transition rate from the $(k-1)$th to the kth base. Thus, $\lambda_{k-1} P_{k-1}(t) \, dt$ represents the probability that RNA polymerase reaches the kth base at the time interval of t and $t + dt$ after transcribing the $(k-1)$th base on the template. $Q_l(t)$ is the probability that the ribosome is on the lth codon at time t.

As is indicated in Fig. 1, most of the 13 bases, 10th to 22nd bases, in the transcript are complementary to the 3'-end of 16S rRNA, and these bases seem to constitute a recognition signal for the ribosome to bind to the transcript. The initiation codon (AUG) is situated downstream, 27th to 29th bases. It is also known that the ribosome translating a codon masks the further downstream region corresponding to 10 or more nucleotides

(Steitz and Jakes, 1975). Therefore, it seems necessary for the translation initiation by the ribosome that RNA polymerase has transcribed at least the first 39 nucleotides. With this situation in mind, we only consider the transcription process from the 40th base by the following set of differential equations:

$$\begin{cases} \frac{d}{dt}P_{39}(t) = -\lambda P_{39}(t), \\ \frac{d}{dt}P_{40}(t) = \lambda P_{39}(t) - P_{40}(t), \\ \quad \cdot \\ \quad \cdot \\ \frac{d}{dt}P_{91}(t) = \lambda P_{90}(t) - \lambda^* P_{91}(t) \\ \frac{d}{dt}P_{92}(t) = \lambda^* P_{91}(t) - \lambda P_{92}(t), \\ \quad \cdot \\ \frac{d}{dt}P_k(t) = \lambda P_{k-1}(t) - \lambda P_k(t), \\ \quad \cdot \end{cases} \tag{2}$$

with the initial condition $P_{39}(0) = 1$. Since the transcription by RNA polymerase is a dissipative process, the back reaction is neglected. If all the transition rates are taken to be a common value, the solution of Eq. 2 obeys the Poisson distribution as is assumed in Manabe's work (1981). In our case, however, the transition rate from the pausing site is distinguished from the others; i.e., this rate is set to λ^*, and other rates are commonly taken to be λ. Although at least two other pausing sites (90th and 92nd bases) have been experimentally indicated on the *trp* leader sequence (Winkler and Yanofsky, 1981; Fisher and Yanofsky, 1983), the number of pausing sites is not seriously important in the elucidation of the role of transcriptional pausing in the control of transcription termination at the attenuator. Since the pausing sites are localized around the 91st base, these effects are approximately expressed by the single pausing site if the value of transition rate λ^* is suitably adjusted. The elongation rates may be somewhat different by the kind of nucleotides, but the difference is not essential for the transcription process of such a long chain of 100 nucleotides. The calculation procedure of $P_k(t)$ from Eq. 2 is given in the Appendix, together with the analytical solution for $P_k(t)$. For the probability $Q_l(t)$, we consider the following stochastic equations.

$$\begin{cases} \frac{d}{dt}Q_0(t) = -\gamma_0 Q_0(t), \\ \frac{d}{dt}Q_1(t) = \gamma_0 Q_0(t) - \gamma_1 Q_1(t), \\ \quad \cdot \\ \quad \cdot \\ \frac{d}{dt}Q_l(t) = \gamma_{l-1} Q_{l-1}(t) - \gamma_l Q_l(t). \\ \quad \cdot \\ \quad \cdot \end{cases} \tag{3}$$

Actually $Q_0(t)$ represents the probability that the translational initiation site is free of ribosomes at time t. Thus, γ_0 is the rate for the ribosome to form the translation-initiation complex with the transcript and γ_l means the transition rate of the ribosome from the lth to $(l+1)$th codon. $Q_l(t)$ is calculated from Eq. 3 with the initial condition $Q_0(0) = 1$. This calculation procedure and the succeeding calculation of $W(k, l)$ in Eq. 1 are also described in the Appendix. Although our calculation method is applicable for the case that translation rates depend on the kind of amino acids, the present calculation is carried out for a simple case that all transition rates of the ribosome from codon to codon are chosen to be common with a notation γ, except for the initiation complex formation rate γ_0 and the rate γ_L from the special codon(s) corresponding to the starved amino acid(s).

RESULTS

To investigate the effect of transcriptional pausing on the attenuation control, we calculate the position probability $W(k, l)$ for appropriate sets of parameter values of λ, λ^*, γ, γ_0, and γ_L, especially for the cases of $k = 140$ and $l = 10$, 11, and 12. For this purpose, we first estimate the values of λ, λ^*, γ, and γ_0. As for the estimation of transcription rates λ and λ^*, the time course of appearance and disappearance of RNA species (PL-RNA) in the single-round transcription experiment (Fisher and Yanofsky, 1983) is analyzed with the use of Eq. 2 modified to start from the first nucleotide. The rates of λ and λ^* are evaluated to be 45 nucleotide/s and 0.015 nucleotide/s, respectively, from the experimental data at 37°C and in the presence of *nusA* protein (*L*-factor), which is known to enhance the transcriptional pausing. The value of λ thus estimated is consistent with the value estimated previously in vivo (Manor et al., 1969). The translation rate γ has been estimated to be 15 codon/s on the *lac* mRNA (Lacroute and Stent, 1968). With the use of this translation rate value the optimal formation rate of 70S translation-initiation complex is estimated to be 1.0/s, which produces the maximum read-through probability under the starvation for tryptophan without the transcriptional pausing (Manabe, 1981). Thus, we consider some values of γ and γ_0 around the estimated values mentioned above.

The values of position probability $W(140, 10)$ calculated with the use of λ and λ^* values in the presence of *nusA* protein at 37°C are plotted against various values of γ_L (or $1/\gamma_L$) in Fig. 2 for several sets of γ_0 and γ values, where γ_L represents the transition rate of the ribosome from the first Trp codon (10th codon) to the second Trp codon (11th codon). For reference, $W(140, 10)$ in the absence of the pausing site is also shown in this figure. As is clearly seen from Fig. 2, the position probability $W(140, 10)$ in the presence of the pausing site is insensitive to the values of γ_0 and γ, while $W(140, 10)$ is very sensitive to the values of γ_0 and γ in the absence of the pausing site. Furthermore, the effect of the transcriptional pausing is reflected upon the dependence of $W(140, 10)$ on the values of γ_L; in the presence of the pausing site the value of $W(140, 10)$ becomes higher under the starvation for tryptophan (in the

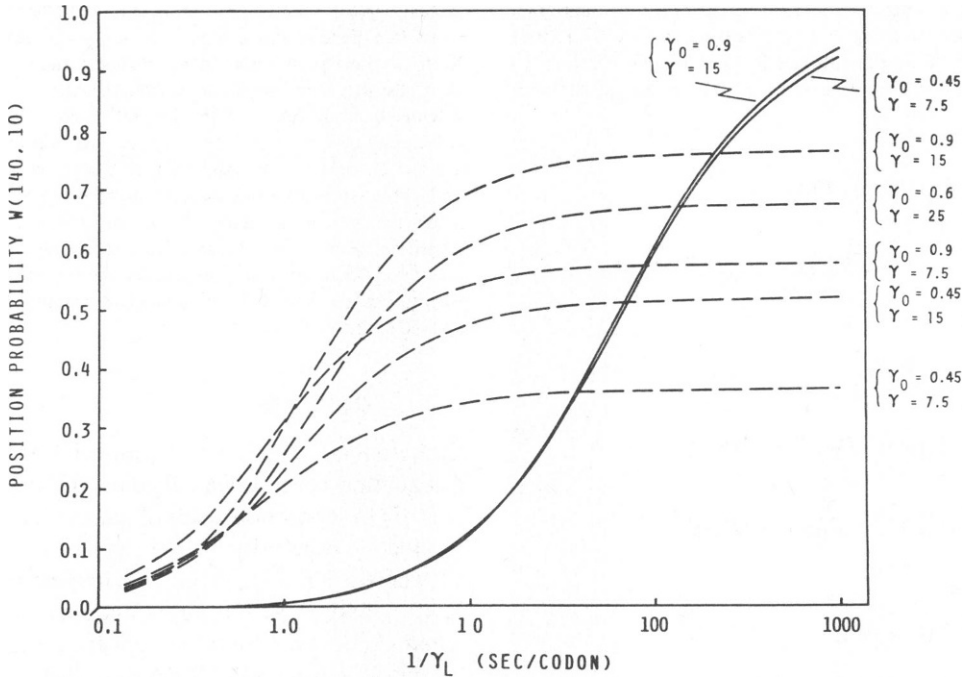

FIGURE 2 The position probabilities $W(140, 10)$ against the transition rate γ_L from the first Trp codon to the next one. The smaller values of γ_L correspond to the physiological condition of the starvation for tryptophan, and larger values of γ_L to the condition of excess tryptophan. Used values of the translation-initiation complex formation rate γ_0 and the other transition rates γ are indicted for each calculated curves of $W(140, 10)$. Solid and broken curves are the position probabilities in the presence and absence of transcriptional pausing. The calculations are carried out with the use of transcription rate λ and λ^*, estimated from the experimental data in the presence of *nusA* protein (*L*-factor) at 37°C (Fisher and Yanofsky, 1983); $\lambda = 45$ nucleotide/s and $\lambda^* = 0.015$ nucleotide/s.

region of larger values of $1/\gamma_L$), and lower under the excess tryptophan (in the region of smaller values of $1/\gamma_L$) in comparison with the case of no pausing site. This result clearly shows that the transcriptional pausing enhances and stabilizes the attenuation control efficiency of *trp* operon transcription. Although λ and λ^* are estimated to be 10 nucleotide/s and 0.007 nucleotide/s from the experimental data in the absence of *nusA* protein at 22°C (Winkler and Yanofsky, 1981), the result of our calculation with the use of these values shows almost the same feature indistinguishable from Fig. 2.

The effect of starvation for various amino acids on the read-through transcription to the structural genes of *trp* operon (Zurawski et al., 1978*b*) has indicated that either tryptophan or arginine starvation especially relieves transcription termination at the *trp* attenuator. In this connection, we also calculate the sum of the position probabilities $W(140, 10)$, $W(140, 11)$, and $W(140, 12)$, which we call the read-through probability hereafter. The calculated read-through probability is plotted against $1/\gamma_L$ in Fig. 3, where the transition rates from the two Trp codons are set to γ_L, and the transition rates from other codons including Arg codon are γ. The read-through probability thus calculated shows essentially the same feature as seen in the position probability $W(140, 10)$, except for the transition to full read-through in the region of slightly smaller values of $1/\gamma_L$. The read-through probability in the absence of the pausing site and with the use of $\gamma_0 = 0.9$/s and $\gamma = 15$ codon/s almost corresponds to the homogeneous case for amino acids calculated already by Manabe (1981), although a slight difference due to γ_0 and γ values is present. It is also clearly seen in this figure that the read-through probability becomes efficient by the presence of the pausing site. The read-through probability under the starvation for arginine is also calculated under the condition that the transition rate from the 12th Arg codon is only set to γ_L. The read-through probability in this case is almost the same as $W(140, 10)$ shown in Fig. 2, and is not shown in this paper.

In the treatment of the transcriptional pausing limited within the framework of Manabe's formula (1981), the role of a pausing site is only considered to increase the effective number of nucleotides for the attenuator and to lower the nonstarved read-through level to ~5% (von Heijne, 1982). The present study reveals the more essential role of the polymerase pausing site in the attenuation control mechanism. The main points clarified in the present study are summarized in the following way. In comparison with the case of no pausing site, (*a*) the attenuation control efficiency is much enhanced by the presence of the pausing site in either the higher read-through probability under the starvation for tryptophan or more complete termination under the excess tryptophan, and (*b*) this efficiency is stable in the sense that it is insensitive to the

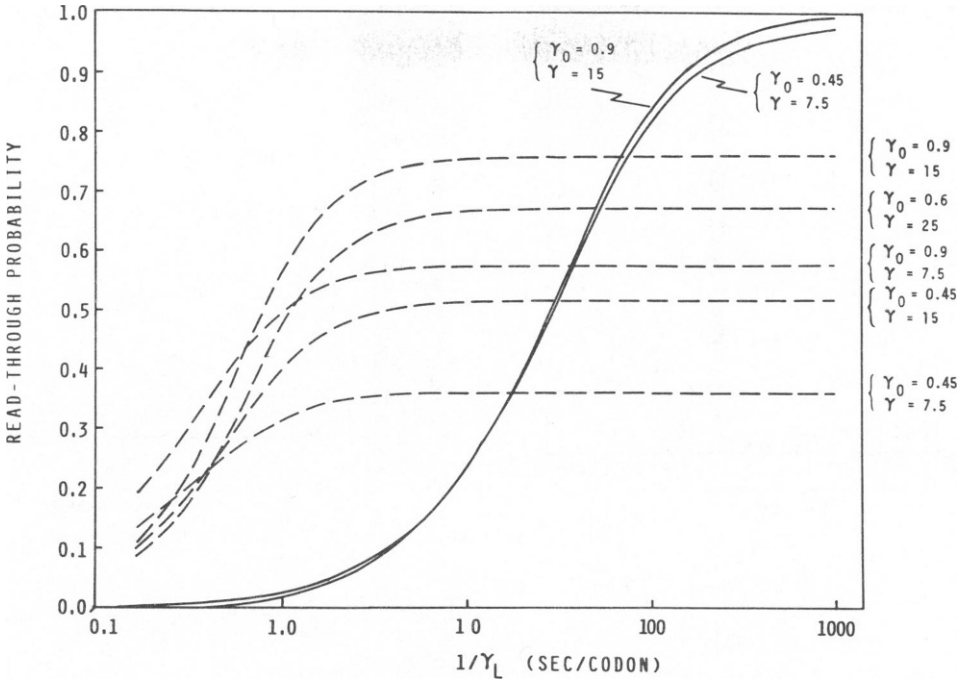

FIGURE 3 The read-through probabilities against the transition rate γ_L from Trp codons. The read-through probability is the sum of the position probabilities $W(140, 10)$, $W(140, 11)$, and $W(140, 12)$. These probabilities are calculated with the same values of λ and λ^* as those in Fig. 2. Solid and broken curves indicate the read-through probabilities in the presence and absence of the pausing site.

formation rate of 70S translation-initiation complex and to the translation rate by the ribosome.

Here, we will explain in detail how the transcriptional pausing affects the characteristics of the attenuation. During the transcription and translation of the leader sequence, the positions of the RNA polymerase on the template and of the ribosome on the transcript are diffused in a stochastic sense. As an example, the position distributions of the ribosome at the times $t = 0.5$ s and $t = 2.0$ s are illustrated in Fig. 4 *a* for the case of $\gamma_0 = 0.9$/s, $\gamma = 15$ codon/s and $\gamma_L = 0.1$ codon/s, where the translation rate γ_L is applied for two Trp codons. At the shorter time $t = 0.5$ s, most of the ribosomes are still unbound to the mRNA, although the remainders are distributed rather widely, showing a peak at the fourth codon. At the later time $t = 2.0$ s, the probability of the ribosome at the first Trp codon becomes highest, since the transition rate γ_L from the Trp codon is much smaller than γ. The position distribution of RNA polymerase for $\lambda = 45$ nucleotide/s and $\lambda^* = 0.015$ nucleotide/s is also shown in Figs. 4 *b* and 4 *c*, corresponding to the cases of the absence and presence of the pausing site, respectively. In the absence of the pausing site (Fig. 4 *b*), the position probability of RNA polymerase exhibits a Poisson-like distribution in the upstream region from the attenuation site. Because of the diffusion such as the Poisson distribution, there is some leak probability that RNA polymerase has already arrived at the attenuation site before the arrival of the ribosome at the Trp codon or RNA polymerase does not arrive at the attenuation site during the stay of the ribosome on Trp codon(s) or on Arg codon, even if an optimal set of the ribosome parameters γ_0 and γ is chosen. This is the reason why the read-through probability under the starvation for tryptophan, e.g., $\gamma_L \lesssim 10^{-3}$ codon/s, is calculated by Manabe (1981) to be only 70% at most. It is also due to the diffused Poisson distribution of RNA polymerase that some degree of read-through probability remains even under the condition of excess tryptophan, (e.g., $\gamma_L \gtrsim 10$ codon/s). In the presence of the pausing site (Fig. 4 *c*), on the contrary, the Poisson-like distribution of RNA polymerase in the upstream region of leader template is condensed at the pausing site, since the transition rate λ^* from the pausing site is $\sim$ 1,000 times as small as λ. In practice, the probability of RNA polymerase at the pausing site reaches the maximum value, nearly one, at a time T. The time T is somewhat later than the time when the peak of RNA polymerase arrives at the pausing site, although the peak arrival time is approximately given by 52 nucleotides/λ. Then, the transition from the pausing site takes place very slowly with a decay factor of the order of λ^*. Thus, the probability that RNA polymerase arrives at the attenuation site continues approximately for a long time of the order of $1/\lambda^*$, with a short time lag due to the transcription of nucleotides in the region between the pausing site and the attenuation site. This situation is schematically illustrated in Fig. 5. The position probability $W(140, 10)$ hardly depends on the values of γ_0 and γ, as long as these ribosome parameter values are chosen in such a range that

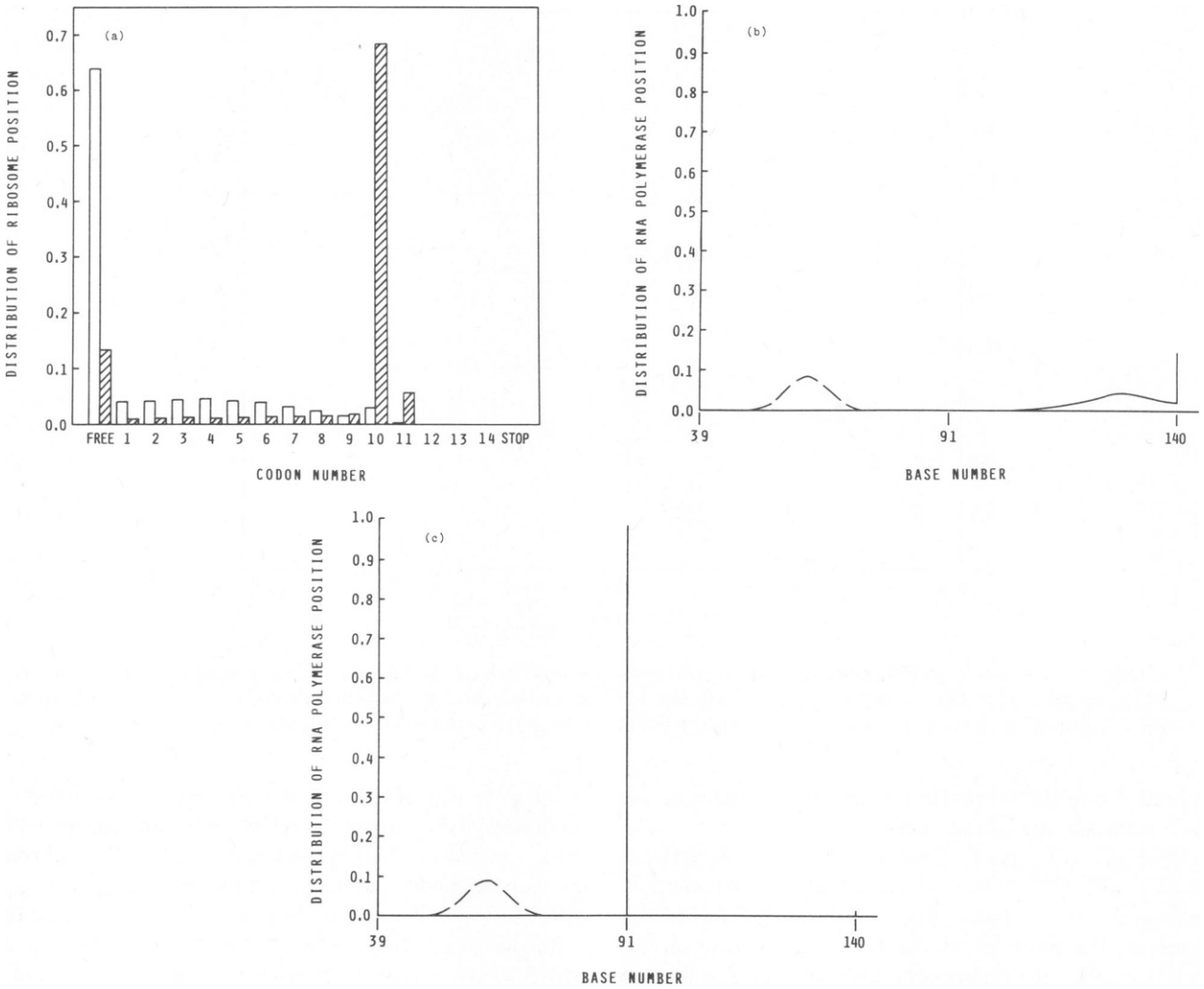

FIGURE 4 Some examples of the position distributions of ribosome and RNA polymerase. (*a*) The ribosome position distributions are represented for the case of $\gamma_0 = 0.9$/s, $\gamma = 15$ codon/s, and $\gamma_L = 0.1$ codon/s at the times $t = 0.5$ s (open rods) and $t = 2.0$ s (striped rods). The numbers in abscissa represent the codon positions in the *trp* leader sequence. (*b*) and (*c*) The position distributions of RNA polymerase in the absence and presence of the pausing site are plotted for the case of $\lambda = 45$ nucleotide/s, $\lambda^* = 0.015$ nucleotide/s, and λ^{**} (the transition rate from the attenuation site) $= 0.165$ nucleotide/s at the times $t = 0.5$ s (broken curve) and $t = 2.0$ s (solid curve). This value of λ^{**} is also estimated from the analysis of single-round transcription experiment (Fisher and Yanofsky, 1983). The numbers in abscissa represent base positions from the transcription initiation site.

the ribosome arrives at the Trp codon during the long stay of RNA polymerase at the pausing site. Within this range of γ_0 and γ values, the position probability $W(140, 10)$, and thus the read-through probability, take higher values than those in the absence of the pausing site, when γ_L is small enough to allow the ribosome to stay at the Trp codon during the arrival of RNA polymerase at the attenuation site (starvation for tryptophan). In other words, the higher value of the read-through probability in the presence of the pausing site is absolutely due to the time integral over a long time of $1/\lambda^*$ for the product of $P_k(t)$ and $Q_l(t)$. As the value of γ_L increases, the read-through probability in the presence of the pausing site decreases more remarkably than in the absence of the pausing site. For a sufficiently large value of γ_L (excess tryptophan), the ribosome passes through the Trp codon during the period that RNA polymerase stays at the pausing site. Approximately, the transition from termination to full read-through occurs in the range of γ_L value satisfying $T < 1/\gamma_L < T + 1/\lambda^*$. The time T is determined by the transcription rate λ^* from the pausing site as well as the transcription of nucleotides between the ribosome binding site and pausing site, but the transcription of nucleotides between the pausing site and attenuation site has less influence on the transition since the number of these nucleotides is much smaller in comparison with the difference between λ and λ^*.

Fig. 4 also shows that the stochastic Eq. 3 with no restriction is approximately allowed for the *trp* leader

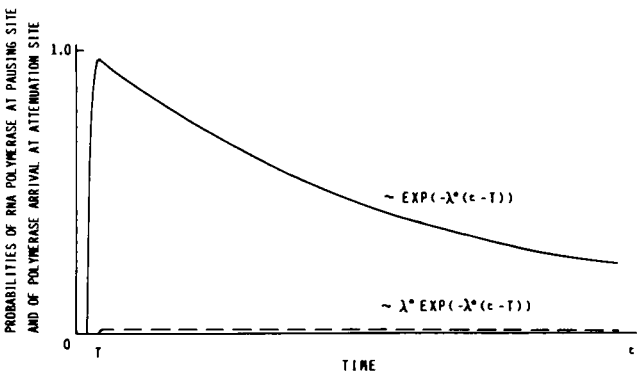

FIGURE 5 Schematic representation of the time course of RNA polymerase at the pausing site (solid curve) and of RNA polymerase arrival at the attenuation site (broken curve). After the probability of RNA polymerase at the pausing site reaches the maximum value ~1 at a time T, the transition from the pausing site takes place very slowly with a single decay factor λ^*, and the arrival of RNA polymerase at the attenuation site continues for a long time of the order of $1/\lambda^*$ with a short time lag after T.

sequence. Although transcription and translation are treated as independent diffusion processes in this paper, it is an extremely low probability that the translation unrealistically gets ahead of transcription. This is easily seen from the comparison of Fig. 4 *a* with Fig. 4 *b* or Fig. 4 *c*. At $t = 0.5$ s, the slight ribosome population on the 9th and 10th codons overlaps with the tail of distribution of RNA polymerase position or of the 3'-end of transcribed mRNA. Such an overlapping, however, is scarcely recognized in either absence or presence of the pausing site at $t = 2.0$ s. Therefore, the result obtained in this paper may hold if a more laborious calculation is carried out under the strict restriction that translation always takes place after transcription.

DISCUSSION AND CONCLUSION

The mechanism revealed in this paper provides a new rationale for the existence of transcriptional pausing sites in the leader regions. Under some changes of circumstances, it is natural to think that the translation rate γ is altered by the intracellular level change of each charged tRNA. And it is also natural to consider that the formation rate of 70S translation-initiation complex γ_0 is altered by the concentration changes of both the ribosome and charged formylmethionyl-tRNA. In these cases, if there is no transcriptional pausing, the altered value of either γ_0 or γ should be sensitively reflected upon the control efficiency of operon transcription. As shown in this paper, transcriptional pausing plays a role in stabilizing the control efficiency for these changes in circumstance. In this case, the read-through probability mainly depends on the ribosome movement from the Trp codon(s). If the nonstarved

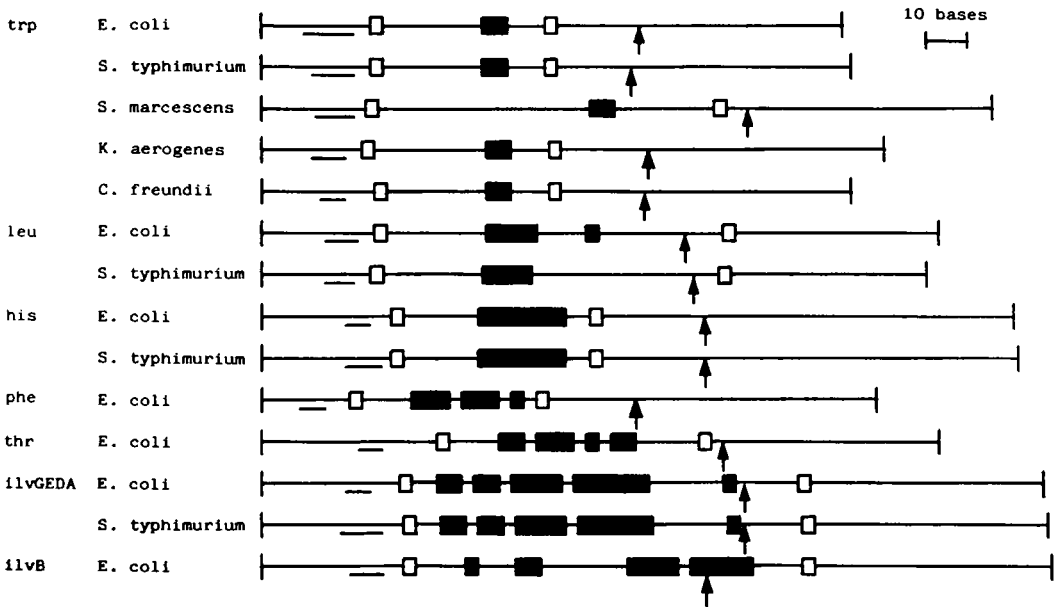

FIGURE 6 Comparison between the leader regions of the biosynthetic operons for amino acids in enteric bacteria. The transcription initiation sites are vertically aligned on the left side, and the right side of each leader sequence corresponds to the attenuation site. The ribosome binding site is underlined, and the pausing site is indicated by an arrow. Initiation and stop codons are indicated by open boxes. Closed boxes represent the own codons. Sequence data are from Lee and Yanofsky (1977) (*E. coli trp*), Lee et al. (1978) (*S. typhimurium trp*), Miozzari and Yanofsky (1978) (*S. marcescens trp*), Blumenberg and Yanofsky (1982*a*, *b*) (*K. aerogenes trp* and *C. freundii trp*), Wessler and Calvo (1981) (*E. coli leu*), Gemmill et al. (1979) (*S. typhimurium leu*), Verde et al. (1981) (*E. coli his*), Barnes (1978) (*S. typhimurium his*), Zurawski et al (1978*a*) (*E. coli phe*), Parsot et al. (1982) (*E. coli thr*), Lawther and Hatfield (1980) (*E. coli ilvGEDA*), Taillon et al. (1981) (*S. typhimurium ilvGEDA*), and Friden et al. (1982) (E. coli *ilvB*).

read-through level is chosen in the vicinity of 5% (von Heijne, 1982) and the quotient of maximal read-through vs. no starvation is taken to be 10 to 20 (Crawford and Stauffer, 1980), the transition to full read-through takes place at ~100-fold reduction in the ribosome movement at the Trp codons relative to the average translation rate (Fig. 3). This would presumably correspond to a ~100-fold reduction in charged $tRNA^{Trp}$ under tryptophan starvation, although physiological data of the relation between the ribosome movement and the tryptophan or charged $tRNA^{Trp}$ concentration are not available at the present time.

Another comment will be given for the numbers of nucleotides and of "own" codons in the leader sequences. The own codon means the codon of the amino acid whose biosynthetic genes are encoded downstream. Some examples of leader sequences are listed in Fig. 6, together with their putative positions of the ribosome binding site and pausing site. The ribosome binding site is assigned as the base sequence complementary to the 3'-end sequence of 16S rRNA. The transcriptional pausing is experimentally detected in *trp* leader sequences of *S. typhimurium* and *K. aerogenes* as well as *E. coli* (Winkler and Yanofsky, 1981), *thr* leader sequence of *E. coli* (Gardner, 1982) and *ilvB* and *ilvGEDA* leader sequences of *E. coli* (Hauser et al., 1985). From the comparison of these experimental data with the base sequence pattern, the pausing sites of other leader sequences are also presumed as the position of ~ 20 bases downstream from the center of G-C rich palindromic sequence, with the consideration of other possible stem and loop structures that may serve as "terminator" and "antiterminator." In these leader sequences, the efficiency of attenuation control severely depends on the numbers of nucleotides between the ribosome binding site, own codons and the attenuation site, if the pausing site is absent. By the presence of the pausing site, this dependency is much diminished. For example, the *trp* leader sequence of *S. marcescens* is longer by ~ 30 bases than the other *trp* leader sequences, but this difference in the number of nucleotides may bring about no essential alteration in the attenuation control efficiency revealed here for *E. coli trp* leader sequence; ~20 nucleotides longer sequence between the ribosome binding site and the pausing site makes the RNA polymerase accumulation time T somewhat later, but the own codons positioned ~20 nucleotides downstream seem to compensate for the time lag, and ~10 nucleotides longer sequence between the pausing site and the attenuation site has hardly any influence on the control efficiency as long as the transcription rate λ^* at the pausing site is ~100-fold smaller than the other average transcription rate λ. In contrast to the two own codons in *trp* leader sequences, the number of own codons is increased in the leader sequences of other amino acid operons. Theoretically, the increased number of own codons shifts the transition from termination to full read-through to the range of smaller values of $1/\gamma_L$ (Figs. 2, 3); in other words, the transition occurs at higher concentration of charged tRNA. Although the ratio of each kind of charged tRNAs is not exactly known in vivo, the codon catalog usage (Grantham et al., 1981) and the correlation between the abundance of tRNAs and the occurrence of respective codons in protein genes (Ikemura, 1981*a*, *b*) suggest that the abundance of *E. coli* charged tRNAs is of the order of $tRNA^{Trp} \lesssim tRNA^{Phe} \lesssim tRNA^{His} < tRNA^{Thr} < tRNA^{Ile} \lesssim tRNA^{Val} \lesssim tRNA^{Leu}$. In practice, our preliminary investigation on the sequence pattern suggests that the ribosome covers the region 1 to facilitate the formation of antiterminator when the ribosome stalls at any of the tandem four leucine (Leu) codons, but the ribosome stalled at the fifth Leu codon in *E. coli* does not seem to be related to the read-through. Therefore, the four Leu codons in the *leu* leader sequence may be a reflection of the physiological requirement that the $tRNA^{Leu}$ is more abundant than $tRNA^{Trp}$, although the biosynthesis of leucine is also under the control of enzymes that are encoded in *ilvB* and *ilvGEDA* operons discussed below. In *his, phe,* and *thr* leader sequences, much more own codons are present, but all of these own codons do not seem to be associated with the attenuation control; possible stem and loop structures suggest that the codons located suitably for the attenuation control are only the last three codons in *his* leader sequence, the second tandem three codons in *phe* leader sequence, and the tandem three codons in *thr* leader sequence. For the leader sequences of *ilvB* and ilvGEDA operons that encode structure genes for the enzymes catalyzing biosynthesis of isoleucine, valine and leucine, the possible secondary structures proposed for the attenuation control are somewhat different from those in *trp, leu, his, phe,* and *thr* leader sequences (Lawther and Hatfield, 1980; Hauser et al., 1985). On the basis of the proposed secondary structures, it seems possible that all own codons except for the third tandem codons and the last own codon in the *ilvGEDA* leader sequence are related to the attenuation control. Thus, we can recognize a correlation that the number of own codons effectively used for the attenuation control is approximately proportional to the intracellular amount of charged tRNAs.

In conclusion, the RNA polymerase pausing in the leader region makes the attenuation control both more sensitive as an on/off switch and less sensitive to variations in the concentration of cellular metabolites not connected with the need for expressing or not expressing the particular operon. The number of own codons including their arrangement in the leader region seems to be closely associated with the intracellular content of charged tRNAs, although this association should be further investigated in relation to the elongation rate from the pausing site in respective leader sequences, possible secondary structures of stem and loop and more detailed examination of intracellular content of charged tRNAs.

APPENDIX

Eq. 2 is reduced to the following form by the change of the base number.

$$\begin{cases} \dfrac{d}{dt}P_0(t) = -\lambda P_0(t), \\ \dfrac{d}{dt}P_1(t) = \lambda P_0(t) - \lambda P_1(t), \\ \quad\cdot \\ \quad\cdot \\ \dfrac{d}{dt}P_i(t) = \lambda P_{i-1}(t) - \lambda^* P_i(t), \\ \dfrac{d}{dt}P_{i+1}(t) = \lambda^* P_i(t) - \lambda P_{i+1}(t), \\ \quad\cdot \\ \dfrac{d}{dt}P_k(t) = \lambda P_{k-1}(t) - \lambda P_k(t), \end{cases} \tag{A1}$$

when all the transition rates are commonly taken to be λ except for the transition rate λ^* from pausing site i. The last equation of A1 can be formally integrated as

$$P_k(t) = \lambda \int_0^t d\tau_{k-1} e^{-\lambda(t-\tau_{k-1})} P_{k-1}(\tau_{k-1}). \tag{A2}$$

The similar expression for $P_{k-1}(\tau_{k-1})$ is also obtained from the integration of the equation of the last but one in Eq. A1. By the substitution of this expression into Eq. A2, we obtain

$$P_k(t) = \lambda^2 \int_0^t d\tau_{k-1} \int_0^{\tau_{k-1}} d\tau_{k-2} e^{-\lambda(t-\tau_{k-2})} P_{k-2}(\tau_{k-2}). \tag{A3}$$

The iteration of such procedures leads to the final expression for $P_k(t)$ as

$$P_k(t) = \frac{\lambda^* \lambda^{k-1}}{\lambda - \lambda^*} \left\{ \frac{e^{-\lambda^* t}}{(\lambda - \lambda^*)^{k-1}} - e^{-\lambda t} \sum_{n_1=0}^{k-1} \frac{1}{(\lambda - \lambda^*)^{n_1}} \frac{tk - 1 - n_1}{(k - 1 - n_1)!} \right\}, \tag{A4}$$

if the initial condition $P_0(0) = 1$ is used. As is easily ascertained, Eq. A4 approaches the Poisson distribution

$$P_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \tag{A5}$$

when $\lambda^* \to \lambda$.

With the same iteration procedure as for Eq. A1, the probability $Q_l(t)$ in Eq. 3 is easily integrated to be

$$Q_l(t) = \int_0^t d\tau_{l-1} \cdots \int_0^{\tau_1} d\tau_0 f_0(\tau_0) f_1(\tau_1 - \tau_0) \cdots f_l(\tau_l - \tau_{l-1}), \tag{A6}$$

where $f_j(t) = \gamma_{j-1} \exp(-\gamma_j t)$ $(j = 1, 2, \cdots, l)$ and $f_0(t) = \exp(-\gamma_0 t)$ with the initial condition $Q_0(0) = 1$. Although the integrals in Eq. A6 can be calculated in a straightforward way as for Eq. A4, it is more transparent to use the Laplace transformation for a further calculation of the position probability $W(k, l)$ defined in Eq. 1. The Laplace transform $\tilde{f}(p)$ of an arbitrary function $f(t)$ is defined as

$$\tilde{f}(p) = \int_0^\infty f(t) e^{-pt} dt. \tag{A7}$$

Since the probability $Q_l(t)$ is represented by the convolution integrals, this is simply expressed as the following complex inversion integral:

$$Q_l(t) = \frac{1}{2\pi i} \int_{C-i\infty}^{C+i\infty} e^{pt} \tilde{Q}_l(p)\, dp, \tag{A8}$$

with

$$\tilde{Q}_l(p) = \tilde{f}_0(p) \tilde{f}_1(p) \tilde{f}_2(p) \cdots \tilde{f}_l(p). \tag{A9}$$

For simplicity, we assume that all translation rates are commonly set to be γ except for the initiation complex formation rate γ_0 and the translation rate γ_L of starved amino acid. For the present investigation of the effect of tryptophan concentration on the read-through probability, we consider the position probabilities $W(140, 10)$, $W(140, 11)$, and $W(140, 12)$ corresponding to the following three cases; only γ_l is chosen to be γ_L in case 1, where l corresponds to 10, both γ_{l-1} and γ_l are to be γ_L in case 2, where l corresponds to 11, and both γ_{l-2} and γ_{l-1} are to be γ_L, but γ_l is to γ in case 3 where l corresponds to 12. In these cases, $\tilde{Q}_l(p)$ in Eq. A9 takes the following forms, respectively:

$$\tilde{Q}_l(p) = \begin{cases} \dfrac{\gamma_0}{p+\gamma_0} \dfrac{\gamma^{l-1}}{(p+\gamma)^{l-1}} \dfrac{1}{p+\gamma_L}, & : \text{case 1} \\ \dfrac{\gamma_0}{p+\gamma_0} \dfrac{\gamma^{l-2}}{(p+\gamma)^{l-2}} \dfrac{\gamma_L}{(p+\gamma_L)^2}, & : \text{case 2} \\ \dfrac{\gamma_0}{p+\gamma_0} \dfrac{\gamma^{l-3}}{(p+\gamma)^{l-2}} \dfrac{\gamma_L^2}{(p+\gamma_L)^2}. & : \text{case 3} \end{cases} \tag{A10}$$

With the use of the expressions Eqs. A4, A8, and A10, the position probability $W(k, l)$ defined by Eq. 1 yields

$$\begin{cases} \lambda^* \lambda^{k-2} \gamma_0 \gamma^{l-1} \left\{ \dfrac{1}{\lambda^* + \gamma_0} \dfrac{1}{(\lambda+\gamma_0)^{k-2}} \cdot \dfrac{1}{(\gamma-\gamma_0)^{l-1}} \dfrac{1}{\gamma_L - \gamma_0} \right. \\ \quad + \dfrac{1}{\lambda^* + \gamma_L} \dfrac{1}{(\lambda+\gamma_L)^{k-2}} \dfrac{1}{(\gamma-\gamma_L)^{l-1}} \dfrac{1}{\gamma_0 - \gamma_L} \\ \quad + \sum_{n_1=1}^{l-1} \dfrac{1}{(\lambda^*+\gamma)^{n_1}} \sum_{n_2=1}^{l-n_1} \dfrac{1}{(\gamma-\gamma_L)^{n_2}} \sum_{n_3=1}^{l+1-n_1-n_2} \\ \quad \cdot \dfrac{1}{(\gamma-\gamma_0)^{n_3}} \dfrac{1}{(\lambda+\gamma)^{k-2+l+1-n_1-n_2-n_3}} \\ \quad \times \dfrac{k-3+l+1-n_1-n_2-n_3}{k-3} \\ \quad \left. \cdots \dfrac{1+l+1-n_1-n_2-n_3}{1} \right\}, & : \text{case 1} \end{cases}$$

$$W(k,l) = \begin{cases} \lambda^* \lambda^{k-2} \gamma_0 \gamma_L \gamma^{l-2} \\ \times \left\{ \dfrac{1}{\lambda^* + \gamma_0} \dfrac{1}{(\lambda+\gamma_0)^{k-2}} \dfrac{1}{(\gamma-\gamma_0)^{l-2}} \dfrac{1}{(\gamma_L-\gamma_0)^2} \right. \\ + \dfrac{1}{(\gamma-\gamma_L)^{l-3}} \sum_{n_1=1}^{2} \dfrac{1}{(\lambda^*+\gamma_L)^{n_1}} \sum_{n_2=1}^{3-n_1} \dfrac{1}{(\gamma_L-\gamma)^{n_2}} \\ \times \dfrac{l-3+n_2-1}{l-3} \cdots \dfrac{1+n_2-1}{1} \\ \times \sum_{n_3=1}^{4-n_1-n_2} \dfrac{1}{(\gamma_L-\gamma_0)^{n_3}} \dfrac{1}{(\lambda+\gamma_L)^{k-2+4-n_1-n_2-n_3}} \\ \times \dfrac{k-3+4-n_1-n_2-n_3}{k-3} \\ \left. \cdots \dfrac{1+4-n_1-n_2-n_3}{1} \right\}, \quad : \text{case 2} \\ - \dfrac{1}{\gamma-\gamma_L} \sum_{n_1=1}^{l-2} \dfrac{1}{(\lambda^*+\gamma)^{n_1}} \sum_{n_2=1}^{l-1-n_1} \dfrac{n_2}{(\gamma-\gamma_L)^{n_2}} \sum_{n_3=1}^{l-n_1-n_2} \\ \times \dfrac{1}{(\gamma-\gamma_0)^{n_3}} \dfrac{1}{(\lambda+\gamma)^{k-2+l-n_1-n_2-n_3}} \\ \times \dfrac{k-3+l-n_1-n_2-n_3}{k-3} \\ \cdots \dfrac{1+l-n_1-n_2-n_3}{1} \end{cases} \tag{A11}$$

$$\begin{cases} \lambda^* \lambda^{k-2} \gamma_0 \gamma_L^2 \gamma^{l-3} \\ \times \left\{ \dfrac{1}{\lambda^* + \gamma_0} \dfrac{1}{(\lambda+\gamma_0)^{k-2}} \dfrac{1}{(\gamma-\gamma_0)^{l-2}} \dfrac{1}{(\gamma_L-\gamma_0)^2} \right. \\ + \dfrac{1}{(\gamma-\gamma_L)^{l-3}} \sum_{n_1=1}^{2} \dfrac{1}{(\lambda^*+\gamma_L)^{n_1}} \sum_{n_2=1}^{3-n_1} \dfrac{1}{(\gamma_L-\gamma)^{n_2}} \\ \times \dfrac{l-3+n_2-1}{l-3} \cdots \dfrac{1+n_2-1}{1} \\ \times \sum_{n_3=1}^{4-n_1-n_2} \dfrac{1}{(\gamma_L-\gamma_0)^{n_3}} \dfrac{1}{(\lambda+\gamma_L)^{k-2+4-n_1-n_2-n_3}} \\ \times \dfrac{k-3+4-n_1-n_2-n_3}{k-3} \\ \cdots \dfrac{1+4-n_1-n_2-n_3}{1} \\ - \dfrac{1}{\gamma-\gamma_L} \sum_{n_1=1}^{l-2} \dfrac{1}{(\lambda^*+\gamma)^{n_1}} \sum_{n_2=1}^{l-1-n_1} \dfrac{n_2}{(\gamma-\gamma_L)^{n_2}} \sum_{n_3=1}^{l-n_1-n_2} \\ \times \dfrac{1}{(\gamma-\gamma_0)^{n_3}} \dfrac{1}{(\lambda+\gamma)^{k-2+l-n_1-n_2-n_3}} \\ \times \dfrac{k-3+l-n_1-n_2-n_3}{k-3} \\ \left. \cdots \dfrac{1+l-n_1-n_2-n_3}{1} \right\}, \quad : \text{case 3} \end{cases}$$

REFERENCES

Barnes, W. M. 1978. DNA sequence from the histidine operon control region: seven histidine codons in a row. *Proc. Natl. Acad. Sci. USA.* 75:4281–4285.

Bertrand, K., L. J. Korn, F. Lee, and C. Yanofsky. 1977. The attenuator of the tryptophan operon of *Escherichia coli*. Heterogeneous 3'-OH termini *in vivo* and deletion mapping of functions. *J. Mol. Biol.* 117:227–247.

Blumenberg, M., and C. Yanofsky. 1982*a*. Regulatory region of the *Klebsiella aerogenes* tryptophan operon. *J. Bacteriol.* 152:49–56.

Blumenberg, M., and C. Yanofsky. 1982*b*. Evolutionary divergence of the *Citrobacter freundii* tryptophan operon regulatory region: comparison with other enteric bacteria. *J. Bacteriol.* 152:57–62.

Crawford, I. P., and G. V. Stauffer. 1980. Regulation of tryptophan biosynthesis. *Annu. Rev. Biochem.* 49:163–195.

Farnham, P. J., and T. Platt. 1980. A model for transcription termination suggested by studies on the *trp* attenuator *in vitro* using base analogs. *Cell.* 20:739–748.

Fisher, R., and C. Yanofsky. 1983. A complementary DNA oligomer releases a transcription pause complex. *J. Biol. Chem.* 258:9208–9212.

Friden, P., T. Newman, and M. Freundlich. 1982. Nucleotide sequence of the *ilvB* promoter-regulatory region: a biosynthetic operon controlled by attenuation and cyclic AMP. *Proc. Natl. Acad. Sci. USA.* 79:6156–6160.

Gardner, J. F. 1982. Initiation, pausing, and termination of transcription in the threonine operon regulatory region of *Escherichia coli*. *J. Biol. Chem.* 257:3896–3904.

Gemmill, R. M., S. R. Wessler, E. B. Keller, and J. M. Calvo. 1979. *leu* operon of *Salmonella typhimurium* is controlled by an attenuation mechanism. *Proc. Natl. Acad. Sci. USA.* 76:4941–4945.

Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier. 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucl. Acids Res.* 9:r43–r74.

Hauser, C. A., J. A. Sharp, L. K. Hatfield, and G. W. Hatfield. 1985. Pausing of RNA polymerase during *in vitro* transcription through the *ilvB* and *ilvGEDA* attenuator regions of *Escherichia coli* K12. *J. Biol. Chem.* 260:1765–1770.

von Heijne, G. 1982. A theoretical study of the attenuation control mechanism. *J. Theor. Biol.* 97:227–238.

Ikemura, T. 1981*a*. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* 146:1–21.

Ikemura, T. 1981*b*. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* 151:389–409.

Lacroute, F., and G. S. Stent. 1968. Peptide chain growth of β-galactosidase in *Escherichia coli*. *J. Mol. Biol.* 35:165–173.

Lawther, R. P., and G. W. Hatfield. 1980. Multivalent translational control of transcription termination at attenuator of *ilvGEDA* operon of *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. USA.* 77:1862–1866.

Lee, F., K. Bertrand. G. Bennett, and C. Yanofsky. 1978. Comparison of the nucleotide sequences of the initial transcribed regions of the tryptophan operons of *Escherichia coli* and *Salmonella typhimurium*. *J. Mol. Biol.* 121:193–217.

Lee, F., and C. Yanofsky. 1977. Transcription termination at the *trp* operon attenuators of *Escherichia coli* and *Salmonella typhimurium*: RNA secondary structure and regulation of termination. *Proc. Natl. Acad. Sci. USA.* 74:4365–4369.

Manabe, T. 1981. Theory of regulation by the attenuation mechanism: stochastic model for the attenuation of the *Escherichia coli* tryptophan operon. *J. Theor. Biol.* 91:527–544.

Manor, H., D. Goodman, and G. S. Stent. 1969. RNA chain growth rates in *Escherichia coli*. *J. Mol. Biol.* 39:1–29.

Miozzari, G. F., and C. Yanofsky. 1978. The regulatory region of the *trp* operon of *Serratia marcescens*. *Nature (Lond.)*. 276:684–689.

Morse, D. E., and A. N. C. Morse. 1976. Dual-control of the tryptophan operon is mediated by both tryptophanyl-tRNA synthetase and the repressor. *J. Mol. Biol.* 103:209-226.

Otsuka, J., and T. Kunisawa. 1982. Characteristic base sequence patterns of promoter and terminator sites in ϕX174 and fd phage DNAs. *J. Theor. Biol.* 97:415–436.

Oxender, D. L., G. Zurawski, and C. Yanofsky. 1979. Attenuation in the *Escherichia coli* tryptophan operon: role of RNA secondary structure involving the tryptophan codon region. *Proc. Natl. Acad. Sci. USA.* 76:5524–5528.

Parsot, C., I. Saint-Girons, and P. Cossart. 1982. DNA sequence change of a deletion mutation abolishing attenuation control of the threonine operon of *E. coli* K12. *Mol. Gen. Genet.* 188:455–458.

Pörschke, D. 1974. Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophys. Chem.* 1:381–386.

Rose, J. K., R. D. Mosteller, and C. Yanofsky. 1970. Tryptophan messenger ribonucleic acid elongation rates and steady-state levels of tryptophan operon enzymes under various growth conditions. *J. Mol. Biol.* 51:541–550.

Stauffer, G. V., G. Zurawski, and C. Yanofsky. 1978. Single base-pair alterations in the *Escherichia coli trp* operon leader region that relieve transcription termination at the *trp* attenuator. *Proc. Natl. Acad. Sci. USA.* 75:4833–4837.

Steitz, J. A., and K. Jakes. 1975. How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* 72:4734–4738.

Taillon, M. P., D. A. Gotto, and R. P. Lawther. 1981. The DNA sequence of the promoter-attenuator of the *ilvGEDA* operon of *Salmonella typhimurium*. *Nucl. Acids Res.* 9:3419–3432.

Verde, P., R. Frunzio, P. P. DiNocera, F. Blasi, and C. B. Bruni. 1981. Identification, nucleotide sequence and expression of the regulatory region of the histidine operon of *Escherichia coli* K-12. *Nucl. Acids Res.* 9:2075–2086.

Watson, J. D. 1977. Molecular Biology of the Gene. Third ed. W. A. Benjamin, Inc., Menlo Park CA. 333 pp.

Wessler, S. R., and J. M. Calvo. 1981. Control of *leu* operon expression in *Escherichia coli* by a transcription attenuation mechanism. *J. Mol. Biol.* 149:579–597.

Winkler, M. E., and C. Yanofsky. 1981. Pausing of RNA polymerase during *in vitro* transcription of the tryptophan operon leader region. *Biochemistry.* 20:3738–3744.

Yanofsky, C., and L. Soll. 1977. Mutations affecting $tRNA^{Trp}$ and its charging and their effect on regulation of transcription termination at the attenuator of the tryptophan operon. *J. Mol. Biol.* 113:663–677.

Zurawski, G., K. Brown, D. Killingly, and C. Yanofsky. 1978*a*. Nucleotide sequence of the leader region of the phenylalanine operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* 75:4271–4275.

Zurawski, G., D. Elseviers, G. V. Stauffer, and C. Yanofsky. 1978*b*. Translational control of transcription termination at the attenuator of the *Escherichia coli* tryptophan operon. *Proc. Natl. Acad. Sci. USA.* 75:5988–5992.