Kow, Y. W., & Wallace, S. S. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8354–8358.
Langfinger, D., & von Sonntag, C. (1985) *Z. Naturforsch. 40C*, 446–448.
Leadon, S. A. (1987) *Br. J. Cancer 55* (Suppl. VIII), 113–117.
Liuzzi, M., Weinfeld, M., & Paterson, M. C. (1989) *J. Biol. Chem. 264*, 6355–6363.
Lown, J. W., & Chen, H. H. (1981) *Can. J. Chem. 59*, 390–395.
Maniatis, T., Fritsch, R. R., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Mouret, J. F., Odin, F., Polverelli, M., & Cadet, J. (1990) *Chem. Res. Toxicol. 3*, 102–111.
Paterson, M. C. (1978) *Adv. Radiat. Res. 7*, 1–53.
Randerath, K., Reddy, M. V., & Gupta, R. C. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6126–6129.
Reuvers, A. P., Greenstock, C. L., Borsa, J., & Chapman, J. D. (1973) *Int. J. Radiat. Biol. 24*, 533–536.
Rosier, J., & Van Peteghem, C. (1988) *J. Chromatogr. 434*, 222–227.
Setlow, P. (1975) Nucleic Acids, in *Handbook of Biochemistry and Molecular Biology*, 3rd ed. (Fasman, G. D., Ed.) Vol. II, p 317, CRC Press, Cleveland, OH.
Sharma, M., Box, H. B., & Paul, C. R. (1990) *Biochem. Biophys. Res. Commun. 167*, 419–424.
Stubbe, J., & Kozarich, J. W. (1987) *Chem. Rev. 87*, 1107–1136.
Tullius, T. D. (1987) *Trends Biochem. Sci. 12*, 297–300.
von Sonntag, C. (1987) *The Chemical Basis of Radiation Biology*, Taylor & Francis, London.
Wallace, S. S. (1988) *Environ. Mol. Mutagen. 12*, 431–477.
Ward, J. F. (1975) *Adv. Radiat. Biol. 5*, 182–239.
Weinfeld, M., Liuzzi, M., & Paterson, M. C. (1989a) *Nucleic Acids Res. 17*, 3735–3745.
Weinfeld, M., Liuzzi, M., & Paterson, M. C. (1989b) *J. Biol. Chem. 264*, 6364–6370.
Weinfeld, M., Liuzzi, M., & Paterson, M. C. (1990) *Biochemistry 29*, 1737–1743.
West, G. J., West, I. W.-L., & Ward, J. F. (1982) *Radiat. Res. 90*, 595–608.

# A Thermodynamic Analysis of RNA Transcript Elongation and Termination in *Escherichia coli*†

Thomas D. Yager and Peter H. von Hippel*
*Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97403*
*Received August 14, 1989; Revised Manuscript Received September 26, 1990*

ABSTRACT: In the first part of this paper we present a thermodynamic analysis of the elongation phase of transcription in *Escherichia coli*. The stability of the elongation complex is described by a "free energy of formation" function ($\Delta G_f°$) that is a sum of terms for forming (i) a locally denatured 17-base-pair DNA "bubble"; (ii) a constant-length hybrid between the 3′-terminal 12-nucleotide residues of the RNA transcript and the corresponding region of the DNA template strand; and (iii) a set of binding interactions between the polymerase and certain DNA and RNA residues within and near the "transcription bubble". The transcriptional elongation complex is very stable at most positions along a natural DNA template and moves in a highly processive fashion. At these positions, the $\Delta G_f°$ function provides a *quantitative* measure of the stability of the elongation complex. Besides allowing for the polymerization of the RNA transcript, the elongation complex also serves to define the context within which transcript termination occurs. In the second part of the paper the thermodynamic analysis is extended to discriminate between template positions at which the elongation complex is stable and positions at which it is rendered relatively unstable by the presence of a string of rU residues at the 3′-terminus of the RNA together with the formation of a specific RNA hairpin just upstream of this point. Most factor-independent (intrinsic) termination events are thermodynamically disallowed at the former positions and are thermodynamically allowed at the latter positions. The extended form of the analysis closely predicts the exact sites of termination at a number of intrinsic terminators (and attenuators) in the *E. coli* genome. It also correctly predicts bidirectional function for a number of bidirectional terminators. In some cases it may identify terminators that are similar to the intrinsic type but that require additional protein factors, unusual polymerase–nucleic acid interactions, or rate-limiting conformational changes in order to function. Finally, it successfully locates intrinsic terminators within a number of *E. coli* operons and discriminates between these terminators and the surrounding DNA sequence.

Transcription of RNA in prokaryotes occurs in six mechanistically distinct phases, which can be described as follows.

(*i*) *Promoter Location.* The RNA polymerase holoenzyme locates a promoter, perhaps in part by facilitated diffusion, and correctly "articulates" with this structure [see, e.g., von Hippel et al. (1984) and McClure (1985)]. This process occurs spontaneously, implying a (favorable) free energy decrease.

(*ii*) *Melting In.* The polymerase holoenzyme melts into (partially open) promoter DNA, which enables the template

†This work was supported in part by USPHS Grants GM-15792 and GM-29158 (to P.H.v.H.) and by USPHS Individual Postdoctoral Fellowship GM-10227 (to T.D.Y.). P.H.v.H. is an American Cancer Society Research Professor of Chemistry.

0006-2960/91/0430-1097$02.50/0 

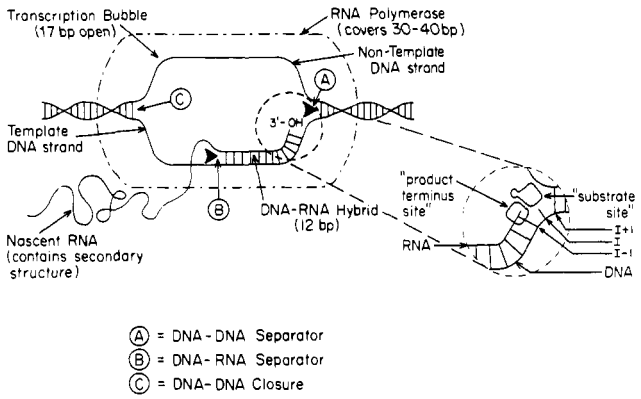



FIGURE 1: Transcription bubble paradigm. Transcription bubble in the elongation mode. The 17 ± 1 base pair DNA "bubble" is from Gamper and Hearst (1982). The 12 ± 1 base pair RNA–DNA hybrid is from Kumar and Krakow (1975) and Hanna and Meares (1983). DNA–RNA "separator" (B) is from Richardson (1975). DNA–DNA "separator" (A) is at present, only a *postulated* structure. We choose not to postulate a structure to close the rear end of the DNA bubble (C), because this seems unnecessary. Placement of the RNA–DNA hybrid within the DNA bubble is from Shi et al. (1988). Extent of RNA polymerase–DNA contact is from Shi et al. (1988) and Telesnitsky and Chamberlin (1989b). The insert (dashed circle) provides a closer view of the "active sites" of the polymerase. The "substrate site" and "product terminus site" are as defined by Krakow and Fronk (1969); see also Yager and von Hippel (1987), Chuknyisky et al. (1990), and Beal et al. (1990).

strand to be correctly "read" (Siebenlist et al., 1980; Kirkegaard et al., 1983). This event is complex and consists of a change in the nature of the polymerase–DNA contacts, a change in the conformation of the polymerase, initiation of the DNA bubble "melting defect", and a subsequent breakage of approximately 13 DNA base-pair stacking interactions [see Gill et al. (1990)]. The process of melting in occurs spontaneously above a critical temperature and is accompanied by a (favorable) free energy decrease.

(*iii*) *Abortive Initiation.* The polymerase begins to join together the first few nucleotide residues of the RNA transcript. This event is also complex. The RNA–DNA duplex is initiated, and is then extended by the formation of new base pairs between the growing RNA chain and the template DNA strand. Some changes in the nature of the contacts between the polymerase and the DNA and RNA strands undoubtedly accompany this process. During abortive initiation the transcription complex is unstable; in this mode the nascent RNA chain may be released prematurely and transcription reinitiated *without* detachment of the holoenzyme from the template (Carpousis & Gralla, 1980; Reznikoff et al., 1982).

The abortive initiation phase ends after the nascent RNA chain has grown to a length of about 8–12 nucleotide residues. (The critical length may vary for different promoters.) At this point the core polymerase separates from the promoter and from the $\sigma$ factor and switches into the elongation mode (Hansen & McClure, 1980; Straney & Crothers, 1985). This switch may be accompanied, and may be thermodynamically driven, by another conformational change in the polymerase [see Gill et al. (1990)].

(*iv*) *Elongation.* During elongation the transcription complex moves processively along the DNA and synthesizes an RNA transcript that is complementary to the template DNA strand. In this phase of transcription the ternary complex of polymerase, DNA, and RNA appears to maintain a fixed geometry. The size of the melted DNA bubble is held constant at 17 ± 1 base pairs, and the size of the RNA–DNA hybrid is held constant at 12 ± 1 base pairs. This geometry is summarized in the "transcription bubble paradigm", which is presented schematically in Figure 1.

The transcription bubble paradigm is based on the following lines of evidence [see Yager and von Hippel (1987)].

(a) Gamper and Hearst (1982) measured the decrease in linking number experienced by closed, double-stranded DNA circles upon transcription by RNA polymerase (in the presence of topoisomerase). They conclude that, in the elongation mode, the DNA is melted to an invariant extent of 17–18 base pairs.[1] This amount of melting appears to be constant, regardless of the length of the RNA transcript and regardless of whether or not the polymerase is "paused" at a particular sequence.

(b) Kumar and Krakow (1975) treated ternary elongation complexes with ribonuclease and determined the lengths of the protected RNA fragments. They found a limiting protected length of 12 ± 2 nucleotides. Also Hanna and Meares (1983) determined the profile of reactivity that the 5′-end of a particular RNA transcript displays toward the template and nontemplate DNA strands in the transcription bubble during the initial phase of a transcription reaction. (The 5′-end of the transcript was made photochemically reactive by the presence of an azido group.) These workers found that the 5′-end of the transcript never passes close enough to the nontemplate DNA strand to react with it. It passes within a "reactive distance" of the template DNA strand only at positions located less than 12 nucleotide residues upstream of the 3′-terminus of the RNA.

In addition, Monforte et al. (1990) have recently described experiments in which they have transcribed an artificial DNA construct that was designed to encode a (self-cleaving) "hammerhead" structure in the RNA. The hammerhead was found to be unable to fold correctly (or to catalyze self-cleavage) until transcription had proceeded at least 15 nucleotide positions past the cleavage site. A control experiment showed that the hammerhead was functional in free RNA that had been extended only 3 nucleotides past the cleavage site. Taken together, all these experiments suggest that the RNA–DNA hybrid is approximately invariant in length at 12 ± 1 base pairs.

(c) Recent work by Shi et al. (1988) has shown that, in the elongation mode, the 3′-OH terminus of the RNA transcript appears to coincide with the first melted position in the DNA bubble. These workers introduced a single psoralen cross-link between the two strands of a DNA template at a particular nucleotide position downstream of a promoter. They then initiated transcription at the promoter and observed that the polymerase "stalls" upon encountering the cross-link. Sequencing of the RNA transcript showed that the 3′-OH terminus of the RNA is located one nucleotide residue upstream of the psoralen cross-link. This result is consistent with the proposal that the RNA–DNA hybrid lies at the extreme downstream end of the transcription bubble, as drawn in Figure 1.

(*v*) *Encounter with a DNA Sequence That Can Potentially Lead to Termination.* Certain sequences within the DNA have the potential to specify transcription termination events. These sequences typically are located at the end of a gene or between genes in an operon; sometimes they are also found just downstream of a promoter. In rare cases they are located within genes, in which case they are called "cryptic" sites to signify that their biological function is not well understood.

A particular sequence in the DNA may specify a termi-

[1] To convert from a "change in linking number" to a value for "total base pairs melted", Gamper and Hearst used a conversion factor derived from studies on the binding of intercalating dyes to DNA. See Gamper and Hearst (1982) for details.

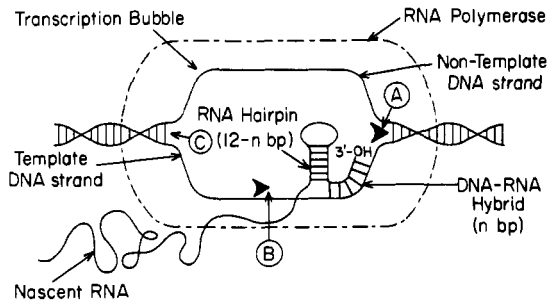



FIGURE 2: Intrinsic termination mode. Postulated form of the transcription bubble in the intrinsic termination mode. Figure 1 has been modified to indicate the displacement of part of the RNA–DNA hybrid by a hairpin in the RNA transcript during intrinsic termination.

nation event in either of two ways.[2] (a) The RNA transcript may be released through the intervention of an "RNA release" factor, such as $\rho$ protein (Roberts, 1969; Pinkham & Platt, 1983; Mott et al., 1985; Brennan et al., 1987). (b) The RNA transcript may be released *spontaneously* without the influence of additional protein factors. We will be concerned here with this latter type of termination. We will use equilibrium thermodynamics to construct a model to predict whether or not intrinsic termination is thermodynamically possible at an arbitrary template position. There is, of course also a kinetic aspect to intrinsic termination [see von Hippel and Yager (1991)]. A possible "transition-state" structure for the intrinsic termination reaction is schematized in Figure 2.

*(vi) Ultimate Dissolution of the Transcriptional Complex.* At those positions for which intrinsic termination is thermodynamically possible, kinetic factors will dictate whether or not it actually occurs. This kinetic competition between elongation, pyrophosphorolysis of the transcript, and intrinsic termination is discussed in detail elsewhere (von Hippel & Yager, 1991). In an intrinsic termination event, the disruption of the RNA–DNA hybrid, closure of the DNA bubble, and dissociation of the polymerase are roughly coincident in time, occurring within a span of a few seconds or less (Andrews & Richardson, 1985; Arndt & Chamberlin, 1988). Under "real-life" conditions, a large entropy of dilution accompanies the dissolution of the ternary complex. Also, when the 3'-end of the RNA transcript is released into the intracellular environment, it becomes vulnerable to "nibbling" by exonucleases. As a consequence of these effects, the probability of spontaneous reassembly of a terminated complex must be virtually nil, and so transcript release must be considered a thermodynamically irreversible process.

In this paper we present thermodynamic analyses of elongation and intrinsic termination, based on the transcription bubble paradigm. We have described preliminary versions of this approach in earlier publications (von Hippel et al., 1987; Yager & von Hippel, 1987; Gill et al., 1990). A method is derived to calculate the relative thermodynamic stability of the elongation complex at an arbitrary DNA template position. This approach provides information about the distribution of stabilizing forces within the elongation complex. We next extend the approach to intrinsic termination. A termination event is considered to be a "virtual" transition in a hypothetical "free energy phase space", with a thermodynamic cost that can be calculated in terms of a "stability function". This stability function experiences a large and abrupt change of value at positions of known intrinsic termination events. Thus the positions of termination are predicted with high accuracy.

The quantitative *efficiencies* of intrinsic termination events are predicted with less accuracy by this analysis. This finding reflects the fact that our thermodynamic approach ignores kinetic effects. In reality, of course, intrinsic termination events must occur within a kinetic context; at each template position, elongation, pyrophosphorolysis of the RNA, and termination must be considered as three mutually exclusive and competitive rate processes. We find that the choice between these competing processes may be analyzed by use of an approach based on transition-state theory (von Hippel & Yager, 1991). This analysis shows that, in contrast to termination *position*, the *efficiency* of termination is *not* a robust parameter. It is instead quite sensitive to small changes in the thermodynamic stability of the transcriptional complex. This would be expected if termination efficiency were subject to regulatory control by small changes in environmental conditions, or by small changes in the concentrations of protein factors. We find that under a fixed set of environmental conditions the efficiencies of many "intrinsic" terminators can be predicted with reasonable accuracy, using a modified form of the thermodynamic approach we describe here (von Hippel & Yager, 1991).

Our thermodynamic analysis of elongation and intrinsic termination correctly predicts the behavior of the RNA polymerase molecule at many positions within natural DNA sequence. However, there remain some intrinsic termination sites for which the analysis fails (see Conclusions). We believe that at such sites exceptional polymerase–nucleic acid interactions may occur, protein termination factors may play a role, or equilibrium between the elongation and termination structures (Figures 1 and 2) may not be achieved. In the latter case kinetic effects would become dominant. It appears that such deviations from the predictions of the thermodynamic analysis may be used to identify unusual termination sites or events.

## METHODS

*(I) Assembly of a Collection of Terminators.* To obtain the sequences of many terminators, we referred to reviews by Adhya and Gottesman (1978), Rosenberg and Court (1979), Brendel and Trifonov (1984), Platt (1986), Lindahl and Zengel (1986), Morgan (1986), Brendel et al. (1986), Friedman et al. (1987), and Landick and Yanofsky (1987). We obtained additional sequences by searching the literature and the EMBL and GenBank nucleotide sequence databases.

We consider only those terminators for which the DNA sequence is known exactly. In our summary tables (Tables V–IX) we try to evaluate the quality of the available information about each terminator on the following scale. *(i) Good to excellent*: In vitro and in vivo efficiency and site(s) of termination have been quantitated, reverse orientation of the terminator has been examined, comparative data are available for mutant or trans-species variants, operon structure and function are well understood, and effects of protein factors on termination have been documented. *(ii) Moderate*: Approximate termination efficiencies or site(s) have been determined in vitro or in vivo. *(iii) Poor*: function has only been inferred from DNA sequence analysis or from genetic studies.

[2] These two classes of termination sites are not always mutually exclusive. For example, under some circumstances *Escherichia coli* $\rho$ protein can enhance the efficiency of release at certain "intrinsic" terminators, such as the "oop" terminator of phage $\lambda$ (Howard et al., 1977), the $t_1$ terminator of phage $\lambda$ (Oppenheim et al., 1982; Schmeissner et al., 1984), and the "central" terminator of phage fd when reversed in orientation (Gentz et al., 1981). In addition, the function of both intrinsic and $\rho$-dependent terminators may be modulated by the presence of other protein factors [reviewed in Yager and von Hippel (1987) and in Friedman et al. (1987); see Grayhack et al. (1985) and Whalen et al. (1988) for specific examples]. Nonetheless, the division of termination sites into intrinsic and $\rho$-dependent classes provides a useful conceptual framework within which to develop specific models.

(*II*) *Definition of the Thermodynamic Standard State.* All thermodynamic data and computations are reported for the same experimentally defined standard state, so that comparisons will be valid. This standard state is defined as 1 M NaCl, pH 7, 37 °C, with 1 M concentrations of all relevant macromolecular species.

(*A*) *Temperature Adjustment.* The thermodynamic data that we use for DNA nearest-neighbor stabilities have been measured at 25 °C rather than 37 °C (Breslauer et al., 1986). To adjust these data to the 37 °C condition, we take the usual approach, which is to assume that $\Delta H_m$ and $\Delta S_m$ (the enthalpy and entropy of melting) are temperature-independent, and then to use $\Delta G_m = \Delta H_m - T\Delta S_m$. The use of DNA and RNA nearest-neighbor stability "basis sets" is, in general, dependent upon this same assumption, because such basis sets derive from the melting of different model oligonucleotides at different $T_m$ values. Further arguments and justification for this approach can be found in Breslauer et al. (1986) and in Freier et al. (1986a). To perform the temperature adjustment, appropriate values of $\Delta H_m$ and $\Delta S_m$ were taken from Breslauer et al. (1986).

(*B*) *Salt Conditions.* Thermodynamic data on nucleic acid nearest-neighbor stabilities have generally been presented for a solution state of 1 M NaCl (Freier et al., 1986a; Breslauer et al., 1986). However, almost all in vitro studies of transcriptional elongation and termination have been performed in an ionic environment of 60–100 mM KCl and 4–10 mM $Mg^{2+}$.[3] Is it appropriate to use the 1 M NaCl "basis set" data to make predictions about transcription in vitro, when the ionic conditions are so different?

We argue that this application is justified. We have carried out thermal melting experiments on model nucleic acid duplexes, such as poly[d(A-T)]·poly[d(A-T)], poly[d(I-C)]·poly[d(I-C)], and poly[r(A-U)]·poly[r(A-U)], as a function of KCl and $Mg^{2+}$ concentrations (unpublished data). We observe (in the absence of divalent cations and at moderate concentrations of KCl) the classic linear increase of $T_m$ with the logarithm of KCl concentration [see Schildkraut and Lifson (1965)]. In the presence of 10 mM $Mg^{2+}$, however, the melting temperature of each nucleic acid duplex is rendered virtually independent of the monovalent salt concentration and becomes equivalent to what is observed in ~1 M KCl or NaCl.

This result is expected, according to the principles elucidated by Record and co-workers [e.g., Record et al. (1977)], since the $Mg^{2+}$ cation has a much higher affinity for duplex nucleic acids than do monovalent cations. In addition, our thermal melting experiments contain high total concentrations of $Mg^{2+}$ relative to phosphodiester binding sites. Therefore, the $Mg^{2+}$ cation effectively saturates the nucleic acid lattice and largely excludes the binding of monovalent cations. Consequently, there should be little apparent effect on $T_m$ when the monovalent salt concentration is varied. By this argument we expect that calculations based on 1 M NaCl data should adequately predict the stabilities of nucleic acid duplexes under in vitro transcriptional conditions (60–100 mM KCl, 4–10 mM $Mg^{2+}$). We note that there *is* a significant competition between divalent and monovalent cations for binding to a nucleic acid lattice at low $Mg^{2+}$ to phosphodiester ratios (Dove & Davidson, 1962). However, transcriptional studies are generally conducted at high ratios of these moieties, where this competitive situation does not apply.

[3] In Tables V–X, we refer to experimental studies of in vitro transcription termination. Most of these employ conditions of 60–150 mM KCl and 4–10 mM $Mg^{2+}$. We note that Gentz et al. (1981), Reynolds (1988), and Arndt and Chamberlin (1990) find changes in the efficiencies of various intrinsic terminators as $K^+$ and $Mg^{2+}$ concentrations are varied out of this range. See Results and Discussion.

Table I: Thermodynamic Basis Sets for Formation of Stacked Nearest-Neighbor Pairs and for Initiation Events (1 M NaCl, pH 7, 37 °C)

| nearest-neighbor pair | $\Delta G_f°$ (kcal/mol) RNA[a] | DNA[b] | RNA-DNA[c] |
|---|---|---|---|
| 5′-A-A→ / 3′←U-U- | −0.9 | −1.6 | −0.4 ($^d_r$), −1.3 ($^r_d$) |
| 5′-A-U→ / 3′←U-A- | −0.9 | −1.2 | −1.3 |
| 5′-U-A→ / 3′←A-U- | −1.1 | −0.7 | −1.3 |
| 5′-C-A→ / 3′←G-U- | −1.8 | −1.7 | −1.7 |
| 5′-C-U→ / 3′←G-A- | −1.7 | −1.4 | −1.7 |
| 5′-G-A→ / 3′←C-U- | −2.3 | −1.4 | −1.7 |
| 5′-G-U→ / 3′←C-A- | −2.1 | −1.1 | −1.7 |
| 5′-C-G→ / 3′←G-C- | −2.0 | −3.3 | −2.9 |
| 5′-G-C→ / 3′←C-G- | −3.4 | −2.8 | −2.9 |
| 5′-G-G→ / 3′←C-C- | −2.9 | −2.8 | −2.9 |
| average | −1.9 | −1.8 | −1.7 |
| initiation of duplex | nr[d] | nr | +3.0[e] |
| creation of "bubble defect" | nr | −3.0[e] | nr |

[a] Freier et al. (1986a). Data were given for 1 M NaCl, pH 7, 37 °C, so no adjustments were necessary. [b] Breslauer et al. (1986). Data were given for 1 M NaCl, pH 7, 25 °C; temperature adjustment was made as described under Methods. [c] This basis set was interpolated from the RNA–RNA and DNA–DNA basis sets of this table, by using the melting data of Chamberlin (1965), Riley et al. (1966), and Martin and Tinoco (1980), as described under Methods. [d] Not relevant to our calculations. [e] Assumed value (see Methods).

(*III*) *Thermodynamic Basis Sets.* (*A*) *The RNA Basis Set.* Freier et al. (1986a) have published a "basis set" of thermodynamic parameters for the formation of stacked nearest-neighbor pairs in double-stranded RNA, as well as for the formation of internal, bulge, and hairpin loops in RNA. Since the thermodynamic standard state that these workers use is identical with ours (1 M NaCl, pH 7, 37 °C), we simply reproduce, in Tables I–III, those data that we require for our calculations.

(*B*) *The DNA Basis Set.* Breslauer et al. (1986) have published a similar "basis set" of thermodynamic parameters for the formation of stacked nearest-neighbor pairs in double-stranded DNA. Their data are reported at 1 M NaCl, pH 7, 25 °C. We have adjusted their data to 37 °C (our standard-state temperature) by the method described above and present the adjusted data in Table I.

(*C*) *An Approximate RNA–DNA Basis Set.* Almost no thermodynamic data exist on the stability of stacked nearest-neighbor pairs in RNA–DNA hybrids. We obtain an approximate basis set in the following way.

(*1*) *Stacked Nearest-Neighbor Pairs Containing Only G and C.* The polynucleotide melting data of Chamberlin (1965) and Riley et al. (1966) suggest the following stability ranking of structures (assuming $\Delta H_m° \simeq$ constant):[4]

(d)-G-G→ / (d)←C-C- < (d)-G-G→ / (r)←C-C- < (r)-G-G→ / (d)←C-C- < (r)-G-G→ / (r)←C-C-

Table II: Thermodynamic Basis Set for RNA–RNA Nearest-Neighbor Pairs Involving GU Base Pairs (1 M NaCl, pH 7, 37 °C)[a]

GU Mismatch in First Position

| first position | second position: 5′-U→ 3′←A | 5′-A→ 3′←U | 5′-G→ 3′←C | 5′-C→ 3′←G |
|---|---|---|---|---|
| 5′-U→ 3′←G | −0.5[b] | −0.7 | −1.5 | −1.3 |
| 5′-G→ 3′←U | −0.7 | −0.5[b] | −1.5 | −1.9 |

GU Mismatch in Second Position

| first position | second position: 5′-U→ 3′←G | 5′-G→ 3′←U |
|---|---|---|
| 5′-U→ 3′←A | −0.5[b] | −0.7 |
| 5′-A→ 3′←U | −0.7 | −0.5[b] |
| 5′-G→ 3′←C | −1.9 | −1.3 |
| 5′-C→ 3′←G | −1.5 | −1.5 |

GU Mismatches in First and Second Positions

| first position | second position: 5′-U→ 3′←G | 5′-G→ 3′←U |
|---|---|---|
| 5′-U→ 3′←G | −0.5[b] | −0.6 |
| 5′-G→ 3′←U | −0.5[b] | −0.5[b] |

[a] All $\Delta G_f°$ values are in kcal/mol and are from Freier et al. (1986a). The average ± standard deviation of all values is $\Delta G_f° = -0.96 \pm 0.52$ kcal/mol. [b] Relatively uncertain value.

However, an analysis using the basis set values of Table I yields the nearly identical estimates $\Delta G_f° = -2.8$ kcal/mol for

(d)-G-G→
(d)←C-C-

(the least stable of the Chamberlin polymers) and $\Delta G_f° = -2.9$ kcal/mol for

(r)-G-G→
(r)←C-C-

(the most stable of these polymers). We therefore assume approximately equal thermodynamic stabilities for all structures containing only G and C. We combine the six values given in Table I to obtain $\Delta G_f° \simeq -2.9 \pm 0.5$ kcal/mol (mean ± standard deviation).

*(2) Stacked Nearest-Neighbor Pairs Containing Only A·T or A·U.* The polynucleotide melting data of Chamberlin (1965) and Riley et al. (1966), and the oligonucleotide melting data of Martin and Tinoco (1980), suggest the following stability ranking for structures containing only A·T or A·U:

(d)-A-A→ / (r)←U-U- << (r)-A-A→ / (r)←U-U- < (r)-A-A→ / (d)←T-T- < (d)-A-A→ / (d)←T-T-

[4] (d) and (r) indicate deoxyribonucleotide and ribonucleotide chains, respectively. Arrows indicate the 5′-to-3′ direction.

Table III: Thermodynamic Basis Set for Internal Loops, Bulge Loops, and Hairpin Loops in RNA (1 M NaCl, pH 7, 37 °C)[a]

| loop size | internal loop $\Delta G_f°$ (kcal/mol) | bulge loop $\Delta G_f°$ (kcal/mol) | hairpin loop $\Delta G_f°$ (kcal/mol) |
|---|---|---|---|
| 1 | nd[b] | +3.3 | nd[b] |
| 2 | +0.8 | +5.2 | nd[b] |
| 3 | +1.3 | +6.0 | +7.4 |
| 4 | +1.7 | +6.7 | +5.9 |
| 5 | +2.1 | +7.4 | +4.4 |
| 6 | +2.5 | +8.2 | +4.3 |
| 7 | +2.6 | +9.1 | +4.1 |
| 8 | +2.8 | +10.0 | +4.1 |
| 9 | +3.1 | +10.5 | +4.2 |
| 10 | +3.6 | +11.0 | +4.3 |
| 12 | +4.4 | +11.8 | +4.9 |

[a] All basis set data are from Freier et al. (1986a). An "internal loop" of length $n$ is defined by the formula

$$\cdots \begin{matrix} 5'\text{WWWW(X)}_{n_1}\text{WWW3}' \\ 3'\text{CCCC(Y)}_{n_2}\text{CCCC5}' \end{matrix} \cdots$$

where W and C denote complementary (Watson and Crick) nucleotides, where X and Y denote noncomplementary nucleotides, and where $n_1 + n_2 = n$. A "bulge loop" of length $n$ is defined by the formula

$$\cdots \begin{matrix} 5'\text{WWWW(X)}_n\text{WWW3}' \\ 3'\text{CCCC-----CCCC5}' \end{matrix} \cdots$$

and a "hairpin loop" of length $n$ is defined by the formula

$$(\text{X})_n \begin{matrix} \text{WWWW5}' \\ \text{CCCC3}' \end{matrix} \cdots$$

where W, C, and X have the same meaning as above. [b] nd = not defined.

This ranking assumes $\Delta H_m° \simeq$ constant, in accordance with the van't Hoff enthalpy data of Martin and Tinoco (1980). An analysis using the basis set values of Table I yields $\Delta G_f° = -0.9$ kcal/mol for

(r)-A-A→
(r)←U-U-

and $\Delta G_f° = -1.6$ kcal/mol for

(d)-A-A→
(d)←T-T-

We will assume the intermediate value $\Delta G_f° = -1.3$ kcal/mol for

(r)-A-A→
(d)←T-T-

and also for the structures

(d)-T-A→ / (r)←A-U- and (d)-A-T→ / (r)←U-A-

To estimate a stability value for the weakest structure

(d)-A-A→
(r)←U-U-

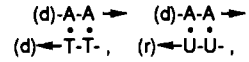
we take the average of two different approximate calculations. In one calculation we assume that $\Delta H_m° = +9.1$ kcal/mol, which is the value for

(d)-A-A→
(d)←T-T-

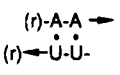
(Breslauer et al., 1986). In the other calculation we assume that $\Delta H_m° = +6.6$ kcal/mol, which is the value for

(r)-A-A→
(r)←U-U-

(Freier et al., 1986a). Using Chamberlin's (1965) $T_m$ data for

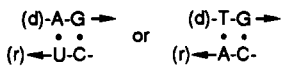


and



and the formulas $\Delta G_m^\circ = \Delta H_m^\circ - T\Delta S_m^\circ$ and $\Delta S_m^\circ = \Delta H_m^\circ / T_m$, we calculate $\Delta G_f^\circ \approx -0.4$ kcal/mol at 37 °C.

(*3*) *Stacked Nearest-Neighbor Pairs Containing A, G, C, and T or A, G, C, and U.* No thermodynamic data exist for structures such as



in which an A·T or A·U base pair resides next to a G·C base pair. To obtain a rough estimate of $\Delta G_f^\circ$, we have averaged the eight $\Delta G_f^\circ$ values in Table I that relate to structures of this type in DNA or in RNA. We obtain $\Delta G_f^\circ = -1.7 \pm 0.4$ kcal/mol (mean ± standard deviation).

(*D*) *The "Initiation" Parameters* $\Delta G^\circ_{\text{RNA-DNA duplex initiation}}$ *and* $\Delta G^\circ_{\text{DNA bubble initiation}}$. To our knowledge, no experimental determinations have been made of the $\Delta G^\circ$ value for RNA–DNA duplex initiation or the $\Delta G^\circ$ value for DNA "bubble" initiation as defined below. Both of these quantities are entropic in origin. The latter can be estimated from melting studies on a homologous series of DNA duplexes containing internal (mismatch) loops of different sizes (Gralla & Crothers, 1973). We assume the approximate equality $\Delta G_f^\circ{}_{\text{RNA-DNA duplex initiation}} \approx -\Delta G_f^\circ{}_{\text{DNA bubble initiation}} \approx +3$ kcal/mol independent of sequence. If, in future experiments, it is shown that these two initiation terms should be assigned different values, then the $\Delta G_f^\circ{}_{\text{pol binding}}$ term of eq 5B may take on a somewhat different value. Since the $\Delta G_f^\circ{}_{\text{pol binding}}$ term is somewhat uncertain anyway (see below), this possibility does not appear too worrisome.

(*E*) *Summary and Caveat.* In Tables I–III we present all approximate basis sets used in this paper. We recognize that these values will eventually be improved in light of future experiments, especially on the thermal melting of RNA–DNA hybrid duplexes (K. J. Breslauer, personal communication). Our final calculations appear robust with respect to significant uncertainties in the basis sets (see, e.g., Figure 8). Thus, we expect our general conclusions to remain approximately valid even after the basis sets are improved.

(*IV*) *Thermodynamic Calculations.* We have used the following algorithms to compute thermodynamic stability ($\Delta G_f^\circ$) functions.

(*A*) *Calculation of* $\Delta G_f^\circ{}_{\text{DNA bubble}}$. Consider a 17-base-pair DNA bubble with the first open base pair located at position *i* on a DNA template. To form this bubble, an initial "melting defect" must be created, 16 "internal" nearest-neighbor stacking interactions must be completely disrupted, and 2 "terminal" nearest-neighbor stacking interactions must be partially disrupted. To calculate the value of $\Delta G_f^\circ{}_{\text{DNA bubble}}$ at template position *i*, we add up the $\Delta G_f^\circ$ values for creating the initial melting defect and for disrupting the 16 internal nearest-neighbor stacking interactions. Next, the $\Delta G_f^\circ$ values for disrupting the two terminal stacking interactions are each weighted by the factor $f = 0.5$ and are added in. After tabulating the result, the DNA bubble is moved one position downstream and the calculations are repeated.

It is hard to estimate the degree of disruption of the two terminal stacking interactions and thus to determine their thermodynamic contributions. They each would appear to resemble a "terminal mismatch pair" on an oligonucleotide duplex (Jaeger et al., 1989). Thus, each may have some thermodynamic stability because of partial hydrogen-bonding and base-stacking properties (Freier et al., 1986b). We assume (rather arbitrarily) that each terminal interaction has half of the thermodynamic stability of its unperturbed analogue in an intact oligonucleotide duplex ($f = 0.5$).

(*B*) *Calculation of* $\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}}$. For this calculation we use the above "sliding window" approach, but add up the $\Delta G_f^\circ$ values for the initial duplex initiation event, and for the creation of $(n-1)$ "internal" stacking interactions. Position "*i*" refers to the first melted position of the DNA bubble and also to the first position of the RNA–DNA hybrid. This relative placement of the DNA bubble and the RNA–DNA hybrid is based on the experimental findings of Shi et al. (1988) and is shown in Figure 1.

(*C*) *Calculation of* $\Delta G_f^\circ{}_{\text{RNA hairpin}}$. The computational method of Tinoco et al. (1973) is used, except that the RNA–RNA basis set values are from our Tables I–III. Perfect hairpins were analyzed automatically, with a computer program of our own design. Hairpins that have internal loops or bulge loops in their stems were analyzed by hand, because considerable judgment is sometimes needed to find the most stable conformation (see below).

(*V*) *Search of Genetic Sequences.* We have tested our analysis by asking it to discriminate between intrinsic terminators and surrounding regions of natural genetic sequence that do not support termination. This has been done by the following sequence of operations: (i) scanning a natural genetic sequence for hairpins; (ii) ranking the hairpins in order of decreasing stability; and (iii) carrying out a full thermodynamic analysis on the most stable of these hairpins, according to the methods described in this paper.

The initial detection of hairpins is sometimes problematic. We have used commercial programs from UWGCG and Amersham ("Staden-Plus") to locate hairpins with loops between 3 and 8 nucleotide residues in length and with stems satisfying either a less restrictive criterion ($S \geq 6$ and $Q \leq 15$) or a more restrictive criterion ($S \geq 8$ and $Q \leq 18$). Here $S$ denotes the number of base pairs in the stem and $Q$ is a score for the stability of the stem:

$$\begin{aligned} Q = {} & (3)(\text{number of G·C base pairs}) + \\ & (2)(\text{numberof A·U base pairs}) + \\ & (1)(\text{number of G·U base pairs}) - \\ & (1)(\text{number of mismatch base pairs}) - \\ & (2)(\text{number of bulge nucleotides}) \end{aligned} \quad (1)$$

The two commercial programs we have used were sometimes unable to pick out irregular hairpins involving bulge loops or overlapping alternative conformations. For example, both programs missed the bulge-containing $t_1$ terminator hairpin of the *rps*U-*dna*G-*rpo*D operon—even though the preferred conformation of this hairpin has a calculated stability of −15.7 kcal/mol! Because the genetic regions we examined (Table X) are well characterized in the literature, we are confident of our conclusions in these cases. However, the fallibility of commercial "hairpin search" programs should be kept in mind when scanning a poorly characterized sequence.

## Theory and Approach

(*I*) *Thermodynamic Analysis of Elongation.* The first goal of this paper is to develop a method for calculating the thermodynamic stability of the elongation complex, position by position, along an arbitrary DNA sequence. Thermodynamic stability is defined in terms of the standard free energy

of formation of the elongation complex from its isolated components. This free energy of formation can be written as the sum of three terms:

$$\Delta G_f^\circ{}_{\text{elongation complex}} = \Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}} + \Delta G_f^\circ{}_{\text{pol binding}} \quad (2)$$

To calculate a value for $\Delta G_f^\circ{}_{\text{elongation complex}}$, we must define a state "I" consisting of three macromolecules (P, D, and R) in solution, each at the standard-state concentration of 1 M. "P" is the polymerase *core* enzyme in its "native" free-solution conformation. "D" is a stretch of double-stranded DNA sufficiently long to accommodate all binding interactions with the core polymerase. "R" is the 3′ portion of an RNA transcript, long enough to accommodate 12 positions of hybridization with the template strand of the DNA, plus any self-complementary base pairing that might *compete* with this DNA–RNA hybrid.

Let a transition occur from state I to another state "E", which also exists at 1 M concentration. State E is depicted in Figure 1. The I → E transition is defined by the following properties. (i) A DNA "bubble" is opened to an extent of 17 ± 1 base pairs, with a free energy of formation equal to $\Delta G_f^\circ{}_{\text{DNA bubble}}$. The quantity $\Delta G_f^\circ{}_{\text{DNA bubble}}$ consists of a "bubble defect initiation" component that is entropic and sequence-independent [see Gralla and Crothers (1973)] and a component for the disruption of "stacked nearest-neighbor pairs" that is sequence-dependent (Table I). (ii) An RNA–DNA hybrid is formed to an extent of 12 base pairs, with a free energy change equal to $\Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}$. The quantity $\Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}$ consists of a "duplex initiation" component that is entropic and sequence-independent and a component for the formation of "stacked nearest-neighbor pairs" that is sequence-dependent (Table I). (iii) The polymerase changes in conformation and also forms a set of contacts with the DNA and RNA in (and near) the transcription bubble. These alterations occur with a total free energy change equal to $\Delta G_f^\circ{}_{\text{pol binding}}$. In principle, the $\Delta G_f^\circ{}_{\text{pol binding}}$ term can be resolved into its two components. However, in practice it may be difficult or impossible to *directly measure* the $\Delta G_f^\circ{}_{\text{pol binding}}$ term or either of its components.

*(II) Value for $\Delta G_f^\circ{}_{\text{pol binding}}$ Is Assumed by Defining a "Reference State".* Because the $\Delta G_f^\circ{}_{\text{pol binding}}$ term is difficult or impossible to measure directly, some assumption must be made about its magnitude. This eliminates the possibility of using an absolute stability scale and forces us to use a *relative* stability scale instead. To fix the "zero point" of the relative scale, we consider the structure depicted in Figure 2 to have 0 kcal/mol relative thermodynamic stability. The rationale for this choice of reference state will become apparent below.

*(III) Thermodynamic Analysis of "Intrinsic" Termination.* During termination of transcription at an intrinsic site, some portion of the RNA within the RNA–DNA hybrid probably unbinds from the template DNA strand and becomes part of an RNA hairpin instead. The evidence for this assertion has been critically evaluated in Yager and von Hippel (1987). It has given rise to the model shown in Figure 2, which depicts the transcription complex at a crucial point in its encounter with an intrinsic terminator. We argue that the model of Figure 2 may represent the structural "transition state" for the intrinsic termination reaction.

*(A) How Much of the RNA–DNA Duplex Is Disrupted?* Our analysis requires the RNA hairpin to displace a *significant* portion of the RNA–DNA duplex during intrinsic termination. We can estimate the extent of displacement that actually occurs, by examining canonical intrinsic terminators for which the *exact* sites of termination are known. Consider the following thought experiment. (i) Imagine an intact transcription bubble in which the 3′-OH terminus of the RNA transcript coincides with an observed termination site. (ii) Allow the 3′-end of the RNA transcript to fold into its final stem–loop structure, disrupting all the RNA–DNA hybrid base pairs that are incompatible with this structure. (iii) Count the number of base pairs ($n$) remaining in the RNA–DNA hybrid. If $n < 12$, then the RNA–DNA hybrid has been disrupted (at least partially) by formation of the RNA hairpin.

We consider 13 exactly mapped structures (*Escherichia coli* unless indicated otherwise; references in Tables V, VI, VIII, and IX): *thr*, *frd-ampC*, *his*, *his* (*Salmonella typhimurium*), *leu*, *trp*, and *trp* (*S. typhimurium*) attenuators; *ton*b, *P14*, *M1*, *supB-E*, "central" (phage fd), and "oop" (phage λ) terminators. For these 13 structures, an average of 2.6 RNA–DNA base pairs remain undisrupted upon complete formation of an RNA hairpin, when the 3′-OH terminus of the RNA coincides with an observed termination site. The number of undisrupted RNA–DNA base pairs ranges from 0 to 6 for this set of structures. Because a *significant* disruption of the RNA–DNA duplex occurs during an intrinsic termination event, we are justified in proceeding to the next stage of the analysis.

*(B) An Idealized Transcription Process.* To calculate the positions where termination is *thermodynamically possible*, we postulate an idealized transcription process that has the following two properties. (i) At each template position, the polymerase can (in principle) exist in either of two states, E or T. State E is described by Figure 1, and state T is described by Figure 2. No intermediate states are allowed, in accordance with the idea that the transition between states E and T is cooperative. (ii) In a large population of idealized transcription complexes, each located at the same template position, there will be a Boltzmann distribution between states E and T:

$$f_T/f_E = e^{-(G_T - G_E)/RT} \quad (3)$$

Here $f_T$ and $f_E$ = the fractional occupancies of states T and E, respectively, and $(G_T - G_E)$ = the free energy difference between the two states.

The free energy difference $(G_T - G_E)$ in eq 3 can be calculated as follows, assuming that the "DNA bubble" and "polymerase binding" terms remain constant:

$$(G_T - G_E) = \Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}} + \Delta G_f^\circ{}_{\text{RNA-RNA [(12-}n\text{)-mer]}} - \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}} \quad (4)$$

In calculating the $\Delta G_f^\circ{}_{\text{RNA-RNA [(12-}n\text{)-mer]}}$ term in eq 4, we consider *only* that fraction of RNA secondary structure that directly involves the $(12 - n)$ nucleotide residues previously bound up in the RNA–DNA hybrid. We note that this term will be sensitive to factors that affect the stability of RNA hairpins, such as the substitution of I for G or base-pair mismatches in the stem of the hairpin.

*(C) The "Stability Function": An Estimator of the Location of an Intrinsic Termination Event.* By referring to our idealized model, we next define a "population-average" standard free energy of formation for the RNA–DNA hybrid, at any template position:

$$\Delta G_f^\circ{}_{\text{RNA-DNA hybrid}} = (f_E)\Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}} + (f_T)\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}} \quad (5A)$$

This in turn allows us to define a population-average "stability function" for the transcriptional complex at any template position:

$$\textit{stability function} = \Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{pol binding}} + \Delta G_f^\circ{}_{\text{RNA-DNA hybrid}} \quad (5B)$$

where $\Delta G_f^\circ{}_{\text{RNA-DNA hybrid}}$ is defined by eq 5A. The stability

function has several properties that help us to predict the location of spontaneous "intrinsic" termination events.

(i) At template positions outside of the terminator region, where the probability of elongation approaches unity, we have $f_E \approx 1$ and $f_T \approx 0$. At such positions the stability function reduces to $\Delta G_f^\circ{}_{\text{elongation complex}}$ as defined in eq 2.

(ii) At positions near, but not coincident with, the exact sites of termination, we expect (in an idealized population of transcription complexes) a Boltzmann distribution between states E and T according to eq 3. Here the stability function will describe the population-average extent to which the unfavorable DNA bubble is compensated by the two different forms of RNA–DNA hybrid (12 base pairs long and $n$ base pairs long, respectively). The $\Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}$ term will be weighted more than the $\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}}$ term at positions where state E is favored over state T and will be weighted less at positions where the converse situation holds. Thus an idealized, population-averaged transcriptional complex would display properties that are intermediate between bona fide elongation and termination complexes.

(iii) At template positions where the probability of termination approaches unity, we have $f_E \approx 0$ and $f_T \approx 1$. At such positions the stability function reduces to a definition for the stability of the termination complex:

$$\Delta G_f^\circ{}_{\text{termination complex}} = \Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{pol binding}} + \Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}} \quad (6)$$

At template positions where the value of eq 6 rises above 0 on our relative free energy scale, we postulate that the ternary complex should be thermodynamically unstable and spontaneous termination should be thermodynamically allowed.

*(IV) The Broader Context in Which Termination Occurs.* Thus far, we have ignored two critical aspects of intrinsic termination that, strictly speaking, lie outside the domain of our model. These aspects however are of paramount importance.

*(A) Kinetic Competition between Elongation and Termination.* In reality, kinetic constraints may operate at each template position where termination is thermodynamically allowed. It is possible to formulate a description of the kinetic competition between elongation and termination by using transition-state theory (von Hippel & Yager, 1991). Such an analysis predicts that the *efficiency* of termination (which is a direct indicator of the competition between elongation and termination) should display significant sensitivity to small changes in the free energy difference $(G_T - G_E)$. This, in turn, suggests a potential mechanism for in vivo regulation of termination efficiency.

*(B) The Ultimate Transcript Release Event Is Irreversible under Real-Life (Non-Standard-State) Conditions.* Consider a termination complex for which $\Delta G_f^\circ$ is calculated to equal 0 according to eq 6. Although such a termination complex is unstable, its components have not yet separated and diffused apart. When this happens, the DNA bubble disappears, the RNA and DNA strands lose contact with each other, and the polymerase is released into solution. Under real-life (non-standard-state) conditions, a substantial entropy of dilution is then registered because the actual molar concentrations of the polymerase, DNA, and RNA are much less than the 1 M values defined in the standard state. [The exact value of the dilution entropy cannot easily be calculated, because we do not know the extent of "macromolecular crowding" (volume exclusion) effects under in vivo conditions; see Minton (1981).] Also, upon release into solution, the 3′-end of the RNA transcript will become vulnerable to cellular exonucleases and may be "nibbled back". Because of these effects there will be a near-zero probability for spontaneous "reattachment" of the components of the transcriptional complex, in correct register, after termination has occurred. Thus the last microscopic step in intrinsic termination (leading to dissolution) should in reality be irreversible. This reemphasizes the fact that our thermodynamic analysis is based on equilibrium considerations in the standard state in vitro and cannot be expected to apply exactly to real-life conditions.

## Results and Discussion

*(I) Thermodynamic Analysis of Elongation on a Natural DNA Sequence.* We are now in a position to analyze the transcription of a natural DNA sequence. We have selected the early portion of the *thr* operon of *E. coli*, because this sequence contains a well-studied canonical transcriptional attenuator,[5] and because it appears to be free of complexities such as antitermination or RNA processing (Gardner, 1979, 1982; Lynn et al., 1982, 1985; 1988; Yang & Gardner, 1989). Figure 3 displays a thermodynamic analysis according to eq 2 over template positions −113 to +219 (where +1 denotes the start of in vitro transcription). The upper curve (open points) denotes $\Delta G_f^\circ{}_{\text{DNA bubble}}$, which is the free energy cost of creating a 17-base-pair "bubble" of melted DNA. The lower curve (solid points) denotes the sum $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$, which is the free energy cost of creating the 17-base-pair DNA bubble as *compensated* by the favorable free energy of forming the 12-base-pair RNA–DNA hybrid. At this point, because the $\Delta G_f^\circ{}_{\text{pol binding}}$ term has not yet been considered, the vertical scale corresponds to *absolute* free energy values.

Several thermodynamic properties of this DNA sequence may be relevant to its transcription.

(i) A transcription complex potentially could exist in the elongation mode over sequence positions +12 to +219 (where position +1 = the start of transcription). By a "sliding window" approach we calculate that, over this region, $\Delta G_f^\circ{}_{\text{DNA bubble}}$ should have the average value of +26.5 kcal/mol, with a standard deviation of ±2.7 kcal/mol. A priori we expect $\Delta G_f^\circ{}_{\text{DNA bubble}}$ to have an average value of (−3.0 kcal/mol) + (+1.8 kcal/mol of stacks)(16 + 2$f$ stacks) = +27.6 kcal/mol, over any region of random-sequence DNA with 50% G+C content (Table I). (The *E. coli* genome has approximately this composition; Phillips et al., 1987). Therefore, this region of the *E. coli thr* operon may be described approximately (at the level of a 17-base-pair sliding window) as a *thermodynamically* random collection of nearest-neighbor pairs.

(ii) In the region of the transcriptional attenuator (positions +120 to +170) the cost of forming the DNA bubble does not differ significantly from the average cost of forming the bubble in other regions of the sequence.

(iii) Positions +12 to +161 and +174 to +219 lie away from the oligo(dA-rU) portion of the transcriptional attenuator. Over these 196 positions, the sum of $\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}$ fluctuates about a mean value of +8.9 kcal/mol, with a standard deviation of ±2.0 kcal/mol. The expected mean value of this sum, as computed over a random sequence of 50% G+C content, is (+1.8 kcal/mol of stacks)(16

[5] In this context a transcriptional attenuator is a specialized termination structure that is found directly downstream of the promoter in certain biosynthetic operons [see Landick and Yanofsky (1987)]. For our purposes the essential characteristic of an attenuator is that, under certain physiological conditions, the RNA hairpin that induces termination is prevented from forming, and consequently the polymerase reads through the site with ~100% probability.

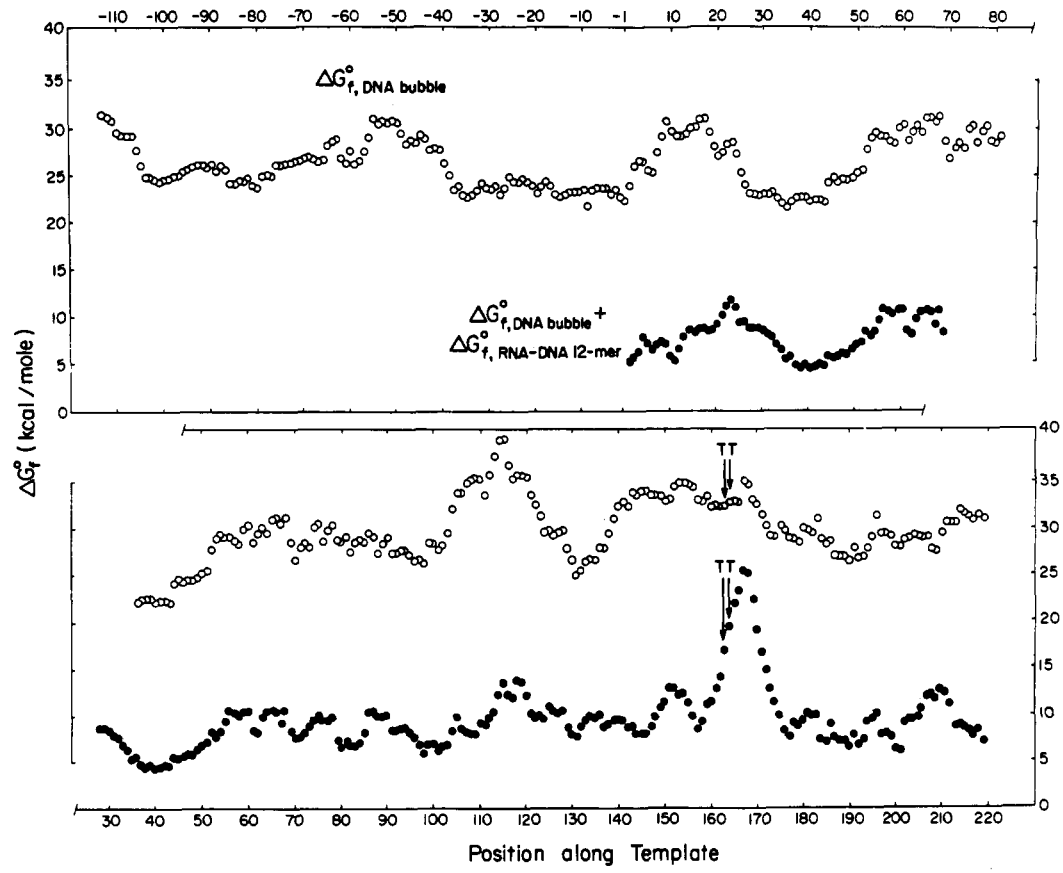



FIGURE 3: Thermodynamic profile of the attenuator region of the *E. coli thr* operon. Plot of $\Delta G_f°_{\text{DNA bubble}}$ and of the sum $\{\Delta G_f°_{\text{DNA bubble}} + \Delta G_f°_{\text{RNA-DNA (12-mer)}}\}$. $\Delta G_f°_{\text{DNA bubble}}$ is calculated with a 17-base-pair "sliding window" and is represented by the upper curve (open points). In the plot of $\Delta G_f°_{\text{DNA bubble}}$, the entropic free energy term for creation of the "bubble defect" has not been added in because of its relatively uncertain value. To put the plot of $\Delta G_f°_{\text{DNA bubble}}$ on an *absolute* scale, one must assume some free energy value for creation of the bubble defect and then shift all the points by a corresponding amount. $\Delta G_f°_{\text{RNA-DNA (12-mer)}}$ is calculated with a 12-base-pair sliding window. The first base pair of the RNA–DNA hybrid is assumed to be coincident with the first melted position of the DNA bubble (Shi et al., 1988). The sum $\{\Delta G_f°_{\text{DNA bubble}} + \Delta G_f°_{\text{RNA-DNA (12-mer)}}\}$ is represented by the lower curve (solid points). In the plot of $\{\Delta G_f°_{\text{DNA bubble}} + \Delta G_f°_{\text{RNA-DNA (12-mer)}}\}$ the two entropic free energy terms for creation of the DNA "bubble defect" and for initiation of the RNA–DNA hybrid duplex have been assumed equal in magnitude and opposite in sign; thus they cancel out. If in the future the sum of these two entropic terms is found experimentally to not equal zero, then all the points in the plot of $\{\Delta G_f°_{\text{DNA bubble}} + \Delta G_f°_{\text{RNA-DNA (12-mer)}}\}$ will be shifted up or down by the appropriate amount. "Position along template" refers both to the first open position of the DNA bubble and to the first base pair of the RNA–DNA hybrid. Position +1 is equated with the start point of the mRNA. The analysis is given over positions −113 to +219. Termination in vitro (vertical arrows and "T" symbols) is observed to occur at positions +163 and +164 (Gardner, 1982; Lynn et al., 1985).

+ 2*f* stacks) + (−1.7 kcal/mol of stacks)(11 stacks) = +14 kcal/mol.

(iv) At positions +162 to +173 the sum $\{\Delta G_f°_{\text{DNA bubble}} + \Delta G_f°_{\text{RNA-DNA (12-mer)}}\}$ is anomalously high (mean ± standard deviation = +17.8 ± 4.6 kcal/mol). This is shown clearly as a "bump" in the solid points of Figure 3. This thermodynamic anomaly is due to the extensive $(dA{\cdot}rU)_n$ stretch that is found in this region [see Table I and also Chamberlin (1965), Riley et al. (1966) and Martin and Tinoco (1980)]. The sum $\{\Delta G_f°_{\text{DNA bubble}} + \Delta G_f°_{\text{RNA-DNA (12-mer)}}\}$ reaches a maximum of +25.7 kcal/mol at position +167, which is 3–4 nucleotides downstream of the two in vitro termination sites. The importance of the $(rU{\cdot}dA)_n$ stretch in helping to determine the position of termination will be discussed further below.

*(II) Boundary Condition for the Stability Function.* As described under Theory and Approach, the stability function provides a criterion for distinguishing between a transcriptional complex that is thermodynamically stable and one that is thermodynamically unstable. However, a problem arises because the $\Delta G_f°_{\text{pol binding}}$ term (within the stability function) cannot be measured or calculated from first principles. This means that one must *assume* some value for the $\Delta G_f°_{\text{pol binding}}$ term. Accordingly, the stability function can provide only a relative measure of stability and not an absolute measure. In this section we place a boundary condition on the stability function that allows one to assign a reasonable value to the $\Delta G_f°_{\text{pol binding}}$ term.

Consider the "encounter" between a transcription complex and a DNA sequence that can potentially lead to intrinsic termination. This is phase v of the transcription process as defined in the introduction. We postulate that, in this phase, the structure depicted in Figure 2 exists transiently on the reaction coordinate that leads to an efficient intrinsic termination event. We adjust the relative free energy scale such that $\Delta G_f°_{\text{termination complex}} = 0$ for this structure. This boundary condition places a restriction on the magnitude of the $\Delta G_f°_{\text{pol binding}}$ term because eq 5B must be satisfied. This boundary condition is relevant only to phase v of the transcription process as defined in the introduction. It is independent of the exact route and end point of dissolution experienced by the ternary complex in phase vi. If a different boundary condition were used, then our estimate of $\Delta G_f°_{\text{pol binding}}$ would change (see below).

*(III) An Estimate of $\Delta G_f°_{\text{pol binding}}$.* Having established a boundary condition for the stability function, we now attempt to estimate the $\Delta G_f°_{\text{pol binding}}$ term. A reasonable value can be assigned to this term by analysis of *E. coli* transcriptional attenuators.

*(A) Estimate of Lower Bound.* The "attenuation paradigm" asserts that the polymerase will read through a transcriptional

attenuator when the RNA hairpin is prevented from forming [see, e.g., Landick and Yanofsky (1987)]. Here we imagine a "virtual" attenuation event in which *only* thermodynamic effects (and not kinetic effects) are important. When the RNA hairpin is prevented from forming, then (in this idealized event) the constraints $f_E \approx 1$ and $f_T \approx 0$ are *forced* to hold, and consequently the stability function $< 0$ (see eqs 5A,B). It follows from eq 2 that

$$|\Delta G_f^\circ{}_{\text{pol binding}}| > \{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\} \quad (7A)$$

From the thermodynamic analysis given in section I under Results and Discussion we know that, in the region of the *thr* attenuator, the sum $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$ rises to a maximum value of +25.7 kcal/mol (Figure 3). This maximum is attained when the RNA hairpin is prevented from forming, and its high value reflects the instability of the extensive $(dA{\cdot}rU)_n$ stretch. By eq 7A, $|\Delta G_f^\circ{}_{\text{pol binding}}|$ must be $\geq$25.7 kcal/mol here.

We can improve our estimate of this lower bound by applying the same argument to a number of other transcriptional attenuators that the polymerase efficiently reads through in the absence of RNA hairpin formation. Figure 4A is a histogram plot of a distribution of high values of $\{G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$ for 13 *E. coli* attenuators and one *S. typhimurium* attenuator (see figure legend). For each attenuator the six highest values of $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$ are plotted, making a total of 84 values in the histogram. We see from this figure that 26–27 kcal/mol is a statistically reasonable estimate of the lower bound of $|\Delta G_f^\circ{}_{\text{pol binding}}|$.

(*B*) *Estimate of Upper Bound.* We next try to estimate an upper bound for $|\Delta G_f^\circ{}_{\text{pol binding}}|$. To do this, we imagine a hypothetical termination event that is extremely efficient. We apply eq 6 to the termination complex and assume that RNA hairpin formation is so favorable that $f_T = 1$ and $f_E = 0$. We rearrange eq 6 and argue that $\Delta G_f^\circ{}_{\text{termination complex}} \geq 0$ during the termination event. It follows that

$$\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}}\} \geq |\Delta G_f^\circ{}_{\text{pol binding}}| \quad (7B)$$

Thus we may use the sum $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}}\}$ as an upper bound for the value of $|\Delta G_f^\circ{}_{\text{pol binding}}|$.

To obtain a histogram plot of our estimate for this upper bound, we note the equality:

$$\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}}\} = \{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\} + \{\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}} - \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\} \quad (7C)$$

This equality holds at *each* template position within an arbitrary terminator. Thus we may obtain the upper bound histogram by adding some (statistical) estimate of the quantity $\{\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}} - \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$ to the histogram of Figure 4A. We present a statistical estimate of the quantity $\{\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}} - \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$ in Figure 4B, again in the form of a histogram. This has been calculated by the method described in the figure legend. The mean value of the distribution in Figure 4B is +8 kcal/mol. Since our best estimate for the lower bound is +26 kcal/mol (from Figure 4A), this implies that our best estimate for the upper bound should be (+26 kcal/mol) + (8 kcal/mol) = +34 kcal/mol. Because of the breadth of the distribution in Figure 4B, our uncertainty in the estimate of the upper bound of $|\Delta G_f^\circ{}_{\text{pol binding}}|$ is considerably greater than our uncertainty in the estimate of the lower bound of this quantity.

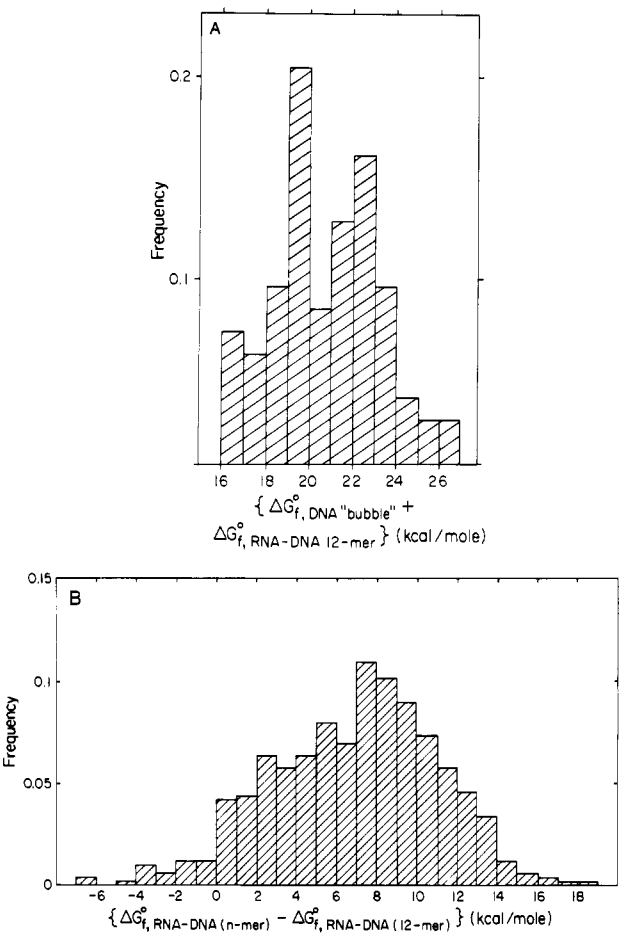



FIGURE 4: Lower and upper bounds for $|\Delta G_f^\circ{}_{\text{pol binding}}|$. (A) Histogram plot of the distribution of high values of the function $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$. The six highest values of this function were plotted for each of 14 efficient attenuators [*E. coli* unless indicated otherwise; references in Tables V, VI, and VIII and in Kolter and Yanofsky (1984)]: *pyr*BI, *thr*, *ilv*GMEDA, *trp*, *trp*L126 mutant, *leu*, *ilv*BN, *phe*A, *his*, *his* (*S. typhimurium*), *rps*O-*pnp*, *orf*E-*pyr*E, *frd-amp*C, and *phe*ST. The upper limit of this distribution is an estimate of a lower bound for $|\Delta G_f^\circ{}_{\text{pol binding}}|$. (B) Histogram plot of the distribution of differences $\{\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}} - \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$. For each attenuator listed in (A), we tabulated the six values of $\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}}$ and the six values of $\Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}$ that had the greatest negative magnitudes. We then computed the difference between each *possible* pair of these values. Thus, for each attenuator, 36 differences were calculated. For each attenuator, this set of differences describes *statistically* the degree to which the $\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}}$ function is different from the $\Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}$ function, *close to* the sites where the termination probability is maximal. To construct the histogram of Figure 4B, the 14 sets of differences were pooled [(14)(36) = 504 values total]. This histogram provides an estimate of the difference between the lower and upper bounds of $|\Delta G_f^\circ{}_{\text{pol binding}}|$.

(*C*) *Combining the Two Results.* It thus appears that $-34$ kcal/mol $\leq \Delta G_f^\circ{}_{\text{pol binding}} \leq -26$ kcal/mol during normal transcript elongation. Taking a simple average of these bounds, we estimate that $\Delta G_f^\circ{}_{\text{pol binding}} \simeq -30$ kcal/mol (at least over the attenuation and termination regions that we have examined).[6]

[6] Ultimately one must consider the dependence of the $\Delta G_f^\circ{}_{\text{pol binding}}$ term on the DNA sequence in and near the transcription bubble (Jin et al., 1988; Telesnitsky & Chamberlin, 1989b), and on particular ions in the milieu (Gentz et al., 1981; Reynolds, 1988; Arndt & Chamberlin, 1990). However, this raises nontrivial questions of how to directly *measure* the $\Delta G_f^\circ{}_{\text{pol binding}}$ term. The general success of our method suggests that the $\Delta G_f^\circ{}_{\text{pol binding}}$ term is approximately constant for many intrinsic terminators (see text).

Table IV: Nine Biological Categories of Terminators

| class | defining characteristics | typical efficiency | example | table in this paper |
|---|---|---|---|---|
| I | attenuator of amino acid biosynthesis operon | 90–100%, modulated down to 0–20% | *trp* attenuator (Stauffer et al., 1978) | V |
| II | attenuator of nucleotide biosynthesis operon | 90–100%, modulated down to unknown level | *pyr*BI attenuator (Turnbough et al., 1983) | VI |
| III | intrinsic mono- or bidirectional operon terminator | 70–100%, unmodulated | *ton*b terminator (Postle & Good, 1985) | VII |
| IV | intergenic attenuator (between two genes in an operon) | 50–90%, modulated up or down in certain cases | *rps*O-*pnp* "attenuator" (Régnier & Portier, 1986) | VIII |
| V | terminator with efficiency enhanced by protein factor(s) | 5–90%, modulated upward by variable amount | λ $t_{R2}$ terminator (Schmidt & Chamberlin, 1987) | |
| VI | terminator with efficiency decreased by antitermination system | 90–100%, modulated down to ≤80% | λ *6S* terminator (Grayhack et al., 1985) | |
| VII | terminator of RNA primer (for DNA replication) | ~100%; stable RNA–DNA hybrid remains after polymerase dissociates | terminator for ColE1 primer RNA (RNA II) (Tomizawa & Masukata, 1987; Masukata & Tomizawa, 1990) | |
| VIII | terminator of plasmid or transposon gene (highly noncanonical sequence and structure) | unknown in general | *tra*S terminator (Jalajakumari et al., 1987) | |
| IX | site of RNA hairpin-induced pausing, where ρ causes termination | 0%, modulated up to 90–100% by the presence of ρ | λ $t_{RI}$ (Morgan et al., 1983a,b, 1984) | |

This analysis raises a fundamental point about the elongation mode of transcription. Let us assume the average values $\Delta G_f^\circ{}_{\text{DNA bubble}} \simeq +28$ kcal/mol, $\Delta G_f^\circ{}_{\text{RNA-DNA hybrid}} \simeq -17$ kcal/mol, and $\Delta G_f^\circ{}_{\text{pol binding}} \simeq -30$ kcal/mol. We then calculate from eq 5B that $\Delta G_f^\circ{}_{\text{elongation complex}} \simeq -19$ kcal/mol. This suggests that, during elongation, the polymerase is trapped in a "potential well" that is ~19 kcal/mol deep. A potential well of this depth, coupled with a significant free energy of activation barrier for dissociation [see von Hippel and Yager (1991)] is certainly sufficient to maintain the processivity of the transcription complex, even in the face of large thermal fluctuations. It also indicates the magnitude of the free energy change required to bring about termination of transcription.

*(IV) Analysis of Intrinsic Termination.* In the preceding two sections, we have defined boundary conditions for the stability function and have estimated the magnitude of the $\Delta G_f^\circ{}_{\text{pol binding}}$ term. We are now almost in a position to proceed with the quantitative analysis of intrinsic termination.

*(A) Terminator Classification Scheme.* The last prerequisite is a scheme for classifying the many terminators of *E. coli* and related bacteria, and of their phages, plasmids, and transposons. This is necessary because different classes of terminators could have different structural and thermodynamic properties, as well as different biological functions. We propose that all such transcription terminators be divided into the nine categories defined in Table IV.

This classification scheme is useful because it simplifies the problem of choosing terminators for analysis. However, it has the shortcoming that its nine categories are not all mutually exclusive. A few terminators can be placed into more than one category. For example, the *nus*A-*inf*B attenuator is an intergenic attenuator (class IV) but is also modulated in function by NusA protein (class V).

The mechanism of termination appears to vary profoundly between certain of the classes, as two examples will illustrate. (i) The *rpo*C terminator is a typical intrinsic terminator (class III), that is not responsive to antitermination signals of the type present in the *rrn*G operon (Albrechtsen et al., 1990). Thus it must differ in some fundamental way from terminators that superficially appear similar but that are sensitive to *rrn*-encoded antitermination signals (class VI; Berg et al., 1989). (ii) The terminator for the ColE1 "primer RNA" (class VIII) apparently can function in two modes. In one mode the RNA transcript is released, while in the other mode the transcript remains part of a stable RNA–DNA hybrid

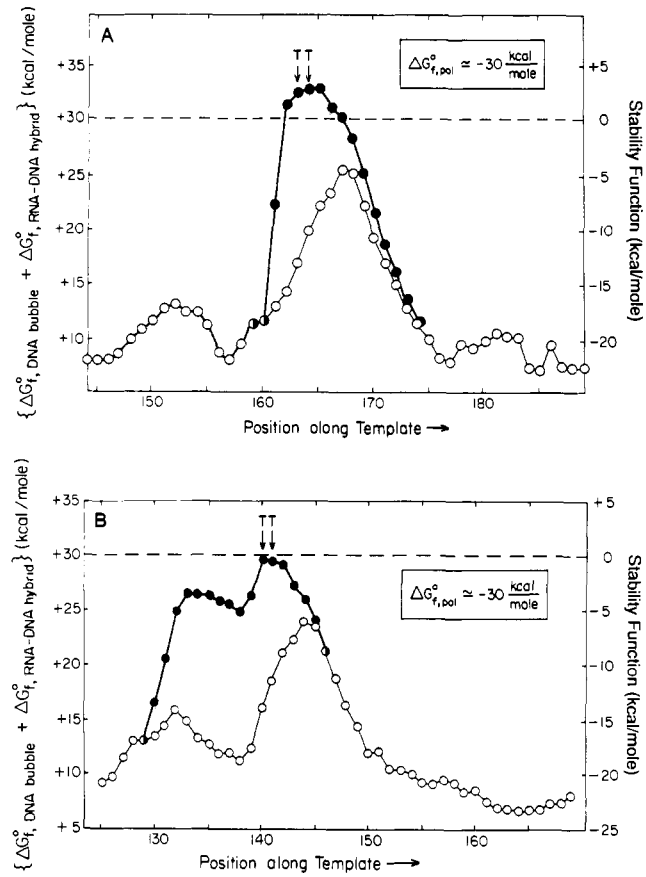



FIGURE 5: Components of the stability function for two canonical terminators of class I. (A) Wild-type *thr* attenuator of *E. coli*. Lower line (open points) represents the sum $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$. Upper line (solid points) represents the stability function defined by eq 6B. The sites of in vitro termination are marked by vertical arrows and "T" symbols. Numbering system on horizontal axis: +1 = start of in vitro transcription. (B) Wild-type *trp* attenuator of *E. coli*. Symbols have the same meaning as in (A).

(Tomizawa & Masukata, 1987; Masukata & Tomizawa, 1990). No other known terminator exhibits this bimodal behavior.

We will not attempt to analyze all nine classes of terminators in this paper but will restrict attention to the intrinsic terminators of classes I–IV. The analysis may be extended in a later publication.

(*B*) *Analysis of Two Intrinsic Terminators.* We now present a complete thermodynamic analysis of the *thr* and *trp* attenuators of *E. coli*, which are the best studied canonical structures of class I. Panels A and B, respectively, of Figure 5 describe key components of the stability function for these two structures. In each figure, the left vertical axis indicates the sum of free energy terms for forming the DNA bubble and the RNA–DNA hybrid. The right vertical axis indicates the complete stability function, which is the sum of these two terms plus the $\Delta G_f^\circ{}_{\text{polymerase binding}}$ term (assumed to equal −30 kcal/mol).

Two functions are plotted position by position along the template sequence in Figure 5. They differ in their treatment of the free energy term for the RNA–DNA hybrid. The lower line (open points) assumes the full RNA–DNA 12-mer always remains intact and, thus, that $\Delta G_f^\circ{}_{\text{RNA-DNA hybrid}} = \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}$. The upper line (solid points) assumes that the RNA–DNA 12-mer is subject to partial displacement by an RNA hairpin and, thus, that $\Delta G_f^\circ{}_{\text{RNA-DNA hybrid}} = \{(f_E)\Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}} + (f_T)\Delta G_f^\circ{}_{\text{RNA-DNA (}n\text{-mer)}}\}$.

Upstream of the locus of termination the full 12-base-pair length of RNA–DNA hybrid always exists, so just the lower line (open points) is found. As the polymerase moves into the region of DNA that specifies the terminator, a progressively greater fraction of the potential hairpin at the terminator becomes encoded as RNA. It folds up and displaces an increasing fraction of the RNA–DNA hybrid. Therefore, in this region of the plot, the contribution of the RNA–DNA hybrid to the stability of the transcription complex grows progressively smaller; the upper line (solid points) separates from the lower line and rises toward a maximum. If the polymerase were to transcribe past the locus of termination, then the RNA hairpin would fold up entirely *outside* of the transcription bubble, the 12-base-pair RNA–DNA hybrid would remain intact, and elongation of the RNA transcript would continue undisturbed. Thus eventually, at some position downstream of the terminator, only the lower line (open points) is again found.

In our model, termination should become thermodynamically allowed at the template position where the stability function first rises above 0 kcal/mol. At this template position the stabilizing free energy contribution made by the polymerase–nucleic acid interactions ($\Delta G_f^\circ{}_{\text{pol binding}}$) should be counterbalanced, and consequently the termination complex should have no thermodynamic stability. In Figure 5 the sites of in vitro termination are marked by arrows. In each case the observed termination sites coincide closely with the template position at which the stability function first rises above zero.

(*C*) *The Intrinsic Terminator as a Binary Switch.* Far upstream of a terminator, the full 12 base pairs of RNA–DNA hybrid is always maintained and the stability function reduces to eq 2. We expect the stability function to have the average value of −19 kcal/mol over this region. Thus the elongation complex should be "trapped" in a potential energy well of 19 kcal/mol depth (see above).

As the polymerase "enters" the termination sequence, the RNA–DNA hybrid is no longer exclusively 12 base pairs long. The stability function must now be described by eqs 5A,B. Precisely at what position will the polymerase no longer be trapped in the potential energy well? An examination of 11 exactly mapped termination structures (see figure legend) reveals that, at positions −1 and −2 nucleotide residues upstream of the first termination site, the stability function has the average value of −13 kcal/mol (Figure 6A). Thus at these positions the elongation complex remains trapped in a potential energy well that is 13 kcal/mol deep. At the exact sites of termination the situation is radically different. Here the stability function has an average value of +1.2 kcal/mol (Figure 6B), and the elongation complex no longer is trapped in the potential energy well.

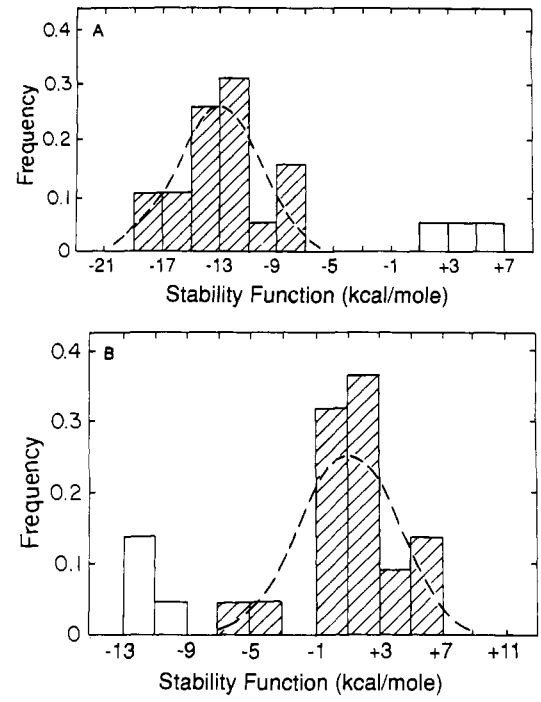



FIGURE 6: Values of the stability function in the proximity of observed intrinsic termination sites. Experimental data on intrinsic termination sites were taken from the following 11 exactly mapped termination and attenuation structures (*E. coli* unless indicated otherwise; references in Tables V and VII): *trp*, *leu*, *thr*, *his*, and *his* (*S. typhimurium*) attenuators; *sup*B-E, M1 (RNase P), P14, *ton*b, central (phage fd), and "oop" (phage λ) terminators. Within this set of 11 structures a total of 26 termination sites have been experimentally mapped. (A) Histogram of the values of the stability function one and two positions upstream of the *first* observed site of termination. This histogram contains 22 entries taken from 11 terminators and attenuators. After discarding the three anomalously high entries (unhatched bars), we calculate the histogram's mean ± standard deviation to be −13.0 ± 3.0 kcal/mol. The dashed line describes a Gaussian-shaped distribution of equal area calculated with these parameters. (B) Histogram of the value of the stability function at the observed sites of termination. This histogram contains 26 entries taken from 11 terminators and attenuators. [This histogram contains more entries than (A) because some terminators exhibit >1 termination sites.] After discarding the four anomalously low entries (unhatched bars), we calculate the histogram's mean ± standard deviation to be +1.2 ± 3.1 kcal/mol. The dashed line describes a Gaussian-shaped distribution of equal area calculated with these parameters.

From the results of Figure 6 it appears that the elongation complex is highly stable during its approach to a terminator but then undergoes an abrupt and complete loss of thermodynamic stability as the *exact* site of termination is traversed. This discrete "off–on" behavior is the consequence of three effects acting in concert: (i) the Boltzmann distribution between states E and T is an *exponential* function of the energy difference between these states; (ii) the RNA–DNA 12-mer is disrupted and replaced by the RNA hairpin to a maximal extent; and (iii) the stability of the remaining portion of RNA–DNA hybrid is very low, because of the preponderance of rU·dA base pairs.

(*D*) *Predicting the Location of an Intrinsic Termination Event.* The "binary switch" property suggests it may be possible to predict, with great accuracy, the positions at which intrinsic termination events are thermodynamically allowed. We have examined this question by reference to the 11 pre-

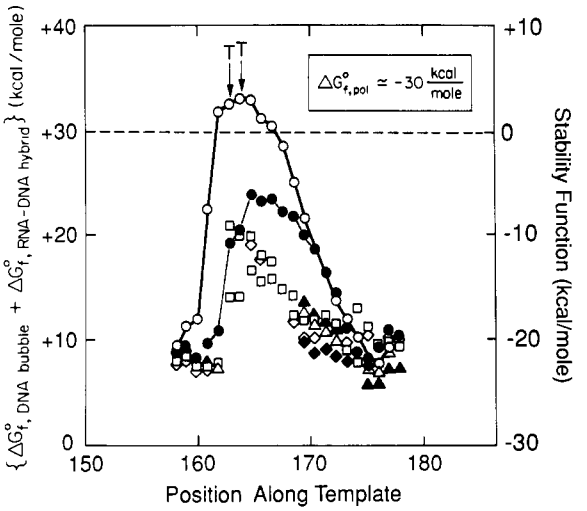



FIGURE 7: Computer simulations of the stability function for the *thr* attenuator of *E. coli*, allowing the oligo(rU·dA) stretch to change to a random sequence. Six simulations (and the wild-type data) are shown in this figure. Each position in the $(rU{\cdot}dA)_n$ stretch has been reassigned as A, U, C, or G at random, with $p = 0.25$ for each choice. Thermodynamic calculations were performed as for Figure 5A. (The "highest" of these sets is outlined with a light line.) Also shown is the stability function of the wild-type *thr* attenuator (from Figure 5A) (heavy line). Axes are defined as in Figure 5A.

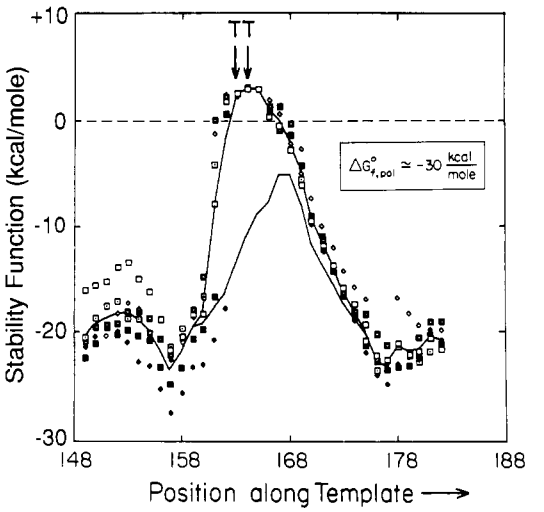



FIGURE 8: Envelope of six stochastic calculations of the $\Delta G_f^\circ{}_{\text{termination complex}}$ function for the wild-type *thr* attenuator of *E. coli*. Each RNA–DNA basis set value was allowed to vary at random about the value given in Table I. The random variation falls in a uniform distribution with range ±50%. No stochastic component was present in the DNA–DNA and RNA–RNA basis sets, because relatively accurate values are available for these parameters (Tables I–III). Six simulations are superimposed in the figure. The lower and upper lines, respectively, are calculations of $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$ and of the stability function and represent the original data of Figure 5A.

cisely mapped terminators and attenuators of Figure 6.

(i) Of the 11 terminators in this data set, 8 display a stability function that rises above 0 kcal/mol. For these 8 structures, 20 termination sites have been exactly mapped in a minimal in vitro transcription system. The sites of termination are found to be Gaussian-distributed at 0.8 ± 1.7 template positions downstream of the site at which the 0 kcal/mol threshold is first crossed (calculations not shown).

(ii) All 11 structures in the data set display a global maximum in the stability function. For these 11 structures, 26 termination sites have been exactly mapped in vitro. The sites of termination are found to be Gaussian-distributed at 0.6 ± 1.6 template positions upstream of the stability function maximum (calculations not shown).

Thus the "threshold predictor" underestimates the position of termination by ~0.8 nucleotides, while the "maximum predictor" overestimates the position of termination by ~0.6 nucleotides. An unbiased predictor of the site(s) of intrinsic termination may be taken as a simple average of these two values.

*(E) The Crucial Importance of the $(rU{\cdot}dA)_n$ Stretch.* There is experimental evidence which suggests that the instability of the $(rU{\cdot}dA)_n$ stretch may be critical for intrinsic termination [see, e.g., Martin and Tinoco (1980) and Lynn et al. (1988)]. We have obtained additional support for this hypothesis through computer simulations. Figure 7 shows a set of five simulations performed on the *thr* attenuator of *E. coli*. To conduct a simulation, each position in the oligo(rU) tail was allowed to change to rA, rC, rG, or rU with equal probability ($p = 0.25$). A full thermodynamic analysis (as in Figure 5A) was then done. The results of this set of simulations are presented in Figure 7. The thermodynamic model predicts the oligo(rU·dA) stretch to be of critical importance for the function of an intrinsic terminator.

The existence of an $(rU{\cdot}dA)_n$ stretch is not however sufficient to cause termination by itself. It does create a "bump" in the thermodynamic profile of $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$ as shown in the lower curve of Figure 3. However, the "binary switch" character of an intrinsic terminator depends on the joint occurrence of *three* effects: (i) the distribution between states E and T is an exponential function of the states' energy difference; (ii) part of the RNA–DNA hybrid is cooperatively displaced by the RNA hairpin; and (iii) the remaining part of the RNA–DNA hybrid is destabilized by its high rU·dA content.

*(F) After a Terminator Switches On, an Irreversible Dissociation Pathway Is Entered.* In our idealized model, a "virtual" termination event occurs completely within the domain of equilibrium thermodynamics. However, in reality the termination mode of Figure 2 should be viewed as a transition state or "bridge" into a subsequent irreversible process. The "termination mode" is characterized by a fully formed RNA hairpin and a short remaining sequence of RNA–DNA hybrid. If the RNA–DNA hybrid were to fall apart, then under real-life conditions the separated components would become diluted into the solution at large, and the 3′-end of the RNA might be "nibbled back" by exonucleases. Thus spontaneous reassociation of all components, in correct register, would be exceedingly improbable. At this point, the domain of equilibrium thermodynamics would be exited, and termination would become irreversible.

*(V) Critique of the Thermodynamic Approach to Intrinsic Termination.* We now describe and evaluate some of the assumptions and limitations that are implicit in our thermodynamic approach.

*(A) Uncertainty in the Basis Set of RNA–DNA Stability Values.* We have assessed the effects of this uncertainty with a computer simulation, using the *thr* attenuator of *E. coli* as a test sequence. Instead of holding the $\Delta G_f^\circ$ values in the RNA–DNA basis set constant, we introduce a random variability. Each basis set value is assumed to fall in a uniform distribution about the value given in Table I, with an allowed range of ±50%. A number of separate computations of the stability function are performed, six of which are shown in Figure 8. Clearly, each value in the RNA–DNA basis set can carry a large random error, without much affecting the final computation of the stability function. This lends confidence to predictions based on the somewhat uncertain

RNA–DNA basis set of Table I.

(*B*) *Uncertainty in the DNA Bubble Size.* Another concern is that the size of the DNA bubble might not actually equal 17 base pairs or that it might fluctuate to values other than 17 base pairs. Although these possibilities seem rather unlikely given the linking-number experiments of Gamper and Hearst (1982), we decided to reexamine the question with a computer simulation.

We have computed a stability function for the *thr* attenuator of *E. coli* assuming that the DNA bubble size is either 18, 15, or 12 bp. In all cases the first melted position in the DNA bubble was aligned with the 3′-OH terminus of the RNA transcript, in accordance with the experimental findings of Shi et al. (1988). There are two results (calculations not shown). (i) If the position of the maximum of the stability function is used to predict the locus of termination, then large variations in the DNA bubble size will not significantly affect one's predictions. This indicates that the sites of termination depend very strongly upon the displacement of the RNA–DNA hybrid by the RNA hairpin, but not strongly upon the size or stability of the DNA bubble. (ii) The value of the stability function depends linearly on the $\Delta G_f^\circ{}_{\text{DNA bubble}}$ term as indicated by eq 5B. If this term is decreased in magnitude by assuming a smaller DNA bubble size, then the stability function will be lowered by a relatively constant amount across all template positions. Above we have used the statement "$\Delta G_f^\circ{}_{\text{termination complex}} \approx 0$ kcal/mol at the site of efficient termination" to constrain the sum of the values of $\Delta G_f^\circ{}_{\text{DNA bubble}}$, $\Delta G_f^\circ{}_{\text{RNA–DNA hybrid}}$, and $\Delta G_f^\circ{}_{\text{pol binding}}$ at that site. This is the basis of our estimate of $\Delta G_f^\circ{}_{\text{pol binding}}$ (see Figure 4 and related discussion). Thus the estimate of $\Delta G_f^\circ{}_{\text{pol binding}}$ depends critically on the size of the DNA bubble that we assume.

(*C*) *The Assumption That* $\Delta G_f^\circ{}_{\text{pol binding}}$ *Is Constant.* This assumption cannot be true in general, for the following reasons.

(i) Some subset of the contacts between the polymerase and the nucleic acids in the transcription bubble must involve positively charged residues on the polymerase and negatively charged phosphates on the nucleic acids. This subset of interactions will be subject to the polyelectrolyte effect (Record et al., 1976) and thus will become thermodynamically less favorable at high salt concentrations.

(ii) There is a structure called the "RNA–DNA separator" on the core polymerase (Figure 1). This has the function of "splaying apart" the RNA transcript and the template DNA strand at a position 12 nucleotides upstream of 3′-OH terminus of the RNA (Richardson, 1975). To perform this function, the "RNA–DNA separator" probably must contact both the RNA strand and the template DNA strand. Thus during intrinsic termination (when the RNA hairpin forms), the polymerase–nucleic acid interactions at this locus must change. But this is equivalent to stating that $\Delta G_f^\circ{}_{\text{pol binding}}$ is not constant during a simple termination event.

Although the assumption of constant $\Delta G_f^\circ{}_{\text{pol binding}}$ cannot be generally true, the analysis is not necessarily made invalid. Changes in the value of $\Delta G_f^\circ{}_{\text{pol binding}}$ could perhaps be small in many intrinsic termination events. We recall that the *location* of a termination event is specified by a very large (~13 kcal/mol) change in the value of the stability function due to sudden displacement of the RNA–DNA hybrid by the RNA hairpin (Figure 6). We would expect the magnitude of this change to overpower any small shifts in $\Delta G_f^\circ{}_{\text{pol binding}}$.

(*D*) *Can Intrinsic Termination Really Be Described as an Equilibrium Process*? Our thermodynamic model is based on an idealized transcription process in which, at each template position, a population of polymerases displays a Boltzmann distribution between states E and T. This idealized model is used to predict sites at which termination is thermodynamically allowed. However, it says nothing about kinetics. Perhaps a termination event could be thermodynamically allowed but kinetically forbidden. This case would arise if translocation of the polymerase were fast, relative to the rate at which energetically comparable RNA and DNA conformations could equilibrate within the transcription bubble. Then the polymerase would translocate out of a terminator region before an RNA hairpin had a chance to form and displace part or all of the RNA–DNA hybrid.

Our equilibrium model accurately predicts the positions of termination for the *thr* and *trp* attenuators of *E. coli* (Figure 5, panels A and B, respectively). It also accurately predicts the termination positions of many other attenuators and terminators in *E. coli* and its phages, and in *S. typhimurium* (see below). This suggests that, in fact, a rapid equilibration may occur between alternative nucleic acid conformations within the transcription bubble while the transcribing polymerase dwells transiently at each template position.

It is worthwhile to examine the question of relative rates more closely. Consider transcription along a poly[d(A-T)]·poly[d(A-T)] template, which probably is a template having a minimal free energy barrier to polymerase movement. The elongation rate on this template has been estimated as 40 ± 5 nucleotides/s at 37 °C (Rhodes & Chamberlin, 1974). Thus an interconversion between alternate nucleic acid conformations in the transcription bubble must be faster than about 25 ms (the residence time of the polymerase at each template position) if our equilibrium model of intrinsic termination is to apply in real time.

Let us consider the interconversion between a state in which the 12-base-pair RNA–DNA hybrid is intact and a state in which the hybrid has been partially or completely replaced by an RNA hairpin:

$$RD + R' \rightleftharpoons RR' + D \qquad (8)$$

Here R is some portion of the terminal 12 nucleotides of the RNA transcript, D is the portion of the DNA template strand that is hybridized to R, and R′ is the upstream complement to R on the same RNA strand. R and R′ constitute the two halves of the stem of a terminator hairpin.

To our knowledge, the kinetics of the displacement reaction described by eq 8 have not been measured. Some insight, however, might be gained from two related studies. (i) For the self-complementary RNA oligomer $A_4GC_5U_5$, the rate of intramolcular hairpin formation has been reported as $5 \times 10^4$ $s^{-1}$ [Cantor and Schimmel (1980), p 1220]. This measurement might be roughly equated to the rate of nuleation of a hairpin in the RNA transcript. (ii) The rate of single-strand branch migration in DNA recombination experiments in vitro has been estimated to be one nucleotide position per 12 μs at 37 °C (Radding et al., 1977). This measurement might be roughly equated to the rate of growth of the hairpin via a mechanism whereby the growing RNA hairpin displaces the RNA–DNA hybrid one position at a time.

Given that the polymerase must have a minimum "dwell time" of at least 25 ms at each template position, these data suggest that a hairpin may have up to 1200 chances to nucleate, and a 10-base-pair stem may have up to 200 chances to grow to completion, during this 25-ms time interval. Thus our treatment, based on equilibrium thermodynamics, may apply in real time to a subset of intrinsic termination events that are not kinetically limited. The predictive success of our model (see below) suggests that this subset may be fairly large.

We are currently examining the question of kinetic constraints on intrinsic termination, using arguments based on elementary transition-state theory (von Hippel & Yager, 1991). Our kinetic model predicts that the efficiency of an intrinsic terminator should be greatly affected when the stability function fluctuates away from 0 kcal/mol. Thus conditions that change any of the terms in the stability function (e.g., $\Delta G_f^\circ{}_{\text{pol binding}}$) by even a few kilocalories per mole should greatly affect the efficiency of termination. We hypothesize that the *efficiencies* of certain termination events may be critical to the physiology of the cell and may be subject to a large degree of regulation through small strategic changes in the magnitude of the $\Delta G_f^\circ{}_{\text{pol binding}}$ term (von Hippel & Yager, 1991).

(*E*) *Alternative Models for Termination.* Our termination model is just one of three that have been proposed in the literature. The two alternatives can be viewed as variants of our model, in which certain of the terms in the stability function are either eliminated or made more complex. We now review the two alternative models.

(*1*) "*RNA Release*" *Model.* In this model (Yager & von Hippel, 1987) the $\Delta G_f^\circ{}_{\text{DNA bubble}}$ and $\Delta G_f^\circ{}_{\text{pol binding}}$ terms are essentially ignored. Only the competition between forming the full RNA hairpin and maintaining the full 12-base-pair RNA–DNA hybrid is viewed as important.

If the RNA release model could predict a relative thermodynamic stability minimum for the transcriptional complex at the right place along a transcribed sequence, it might be useful for predicting the location of a termination event. We have examined the question of how well this model works as a predictor of termination loci for the *thr* attenuator and the *trp* attenuator of *E. coli*. For each attenuator, we have calculated the cost of disrupting only that *portion* of the RNA–DNA hybrid that remains intact after RNA hairpin formation is complete. It appears, for either attenuator, that in vitro termination positions can be predicted relatively well by this type of approach (calculations not shown).

A fundamental shortcoming, however, is that no use is made of the ability to compute $\Delta G_f^\circ{}_{\text{DNA bubble}}$ from the DNA sequence. Thus in principle the model cannot allow an estimation of the $\Delta G_f^\circ{}_{\text{pol binding}}$ term nor, consequently, of the stability function. Thus the model must be inadequate as a basis for quantitative prediction of the *efficiency* of termination (von Hippel & Yager, 1991).

(*2*) "*RNA Hairpin Binding*" *Model.* This model (Farnham & Platt, 1980; Arndt & Chamberlin, 1990) hypothesizes that, during intrinsic termination, the RNA hairpin fits into a "binding site" on the polymerase, which causes a polymerase conformational change to occur. This in turn causes full release of the RNA transcript, closure of the DNA bubble, and dissociation of the polymerase.

The RNA hairpin binding model can be viewed as a more elaborate version of our own model, in which the $\Delta G_f^\circ{}_{\text{pol binding}}$ term changes in value during an intrinsic termination event. The crucial question is whether one needs to postulate a shift in the value of $\Delta G_f^\circ{}_{\text{pol binding}}$ to explain all the available data on intrinsic termination. Our model holds the $\Delta G_f^\circ{}_{\text{pol binding}}$ term constant. It can predict the exact positions of many intrinsic termination events (this paper) and can also rank many intrinsic terminators with respect to relative efficiencies (von Hippel & Yager, 1991).

However, there exist a number of "anomalous" terminators for which our model fails to yield the correct predictions. Also, some intrinsic termination events may display dependencies on downstream DNA sequence (Jin et al., 1988; Telesnitsky & Chamberlin, 1989b) or on specific cations (Gentz et al., 1981; Reynolds, 1988; Arndt & Chamberlin, 1990) that our model, in its simplest form, cannot explain. Thus a more refined or elaborate version of our model may ultimately be required, in which the $\Delta G_f^\circ{}_{\text{pol binding}}$ term is not treated as constant but rather is defined as a function of downstream sequence, salt conditions, etc.

We argue that restraint should be exercised in creating a more elaborate version of the stability function model. There are two criteria to keep in mind: parsimony and falsifiability. It is always possible to "explain" data by introducing extra adjustable parameters into a model. However, each additional parameter makes Occam's razor a little "duller". Also, it is not clear whether a model with extra parameters is scientific in the sense of being falsifiable (disprovable).

In the RNA hairpin binding model, both a hairpin binding site and a polymerase conformational change are postulated. It may be quite hard to experimentally demonstrate the existence of either of these things. A hairpin binding site could perhaps be demonstrated by titrating an artificially stalled elongation complex with an exogenous RNA hairpin. A polymerase conformational change could perhaps be demonstrated with a pulse footprinting technique that uses a brief exposure to a footprinting reagent (Metzger et al., 1989) or with a protein–nucleic acid cross-linking technique that uses short laser pulses (Hockensmith et al., 1990). Such approaches have not yet been explored.

(*VI*) *Further Applications of the Stability Function Model.* (*A*) *Analysis of Many Canonical Terminators.* The terminators of classes I–IV each consist of a GC-rich stem–loop structure that is followed by a largely unbroken run of U residues. Each of these terminators displays a moderate to high efficiency in a minimal in vitro system. None (except possibly the "intergenic attenuators" of class IV) are greatly affected by protein factors. Tables V–VIII list a number of these canonical structures for which the exact sites of termination have been mapped (mostly in vitro). The predicted and observed sites of termination are in very good agreement. Our model disagrees seriously with observations in just two out of fifteen cases for which exact site data are presented. These are the *rpo*H terminator of Table VII and the *tet*A-*orf*L intergenic attenuator of Table VIII. The *rpo*H transcript was mapped in vitro. Possibly an S1 "nibbling" artifact occurred during the mapping procedure. The *tet*A transcript was mapped in vivo, so the transcript may have been "nibbled" by cellular exonucleases. Also, protein termination factor(s) may have shifted the 3′-OH terminus. [See Briat et al. (1987) for an example of a factor-induced shift.]

(*B*) *Bidirectional Terminators.* The experimental studies of Farnham and Platt (1980) and Lynn et al. (1988), as well as the computer simulation of Figure 7, reveal the importance of the $(rU{\cdot}dA)_n$ stretch downstream of the hairpin. Certain simple terminators also contain an oligo(A) stretch just upstream of the hairpin stem. This finding has given rise to the hypothesis that such terminators might function in both directions [see Platt (1986)]. As another application of our model, we attempt to predict whether indeed such terminators should function in both directions.

One of the best documented bidirectional terminators is the *ton*b terminator of *E. coli* (Postle & Good, 1985). This terminator functions in a minimal in vitro system with high (71–72%) efficiency in either direction (Postle & Good, 1985). When the stability function is computed for both orientations of this terminator, a striking symmetry is found (Figure 9A). Thus, according to our thermodynamic model, the two ori-

Table V: Class I: Attenuators of Amino Acid Biosynthesis Operons

| attenuator[a] | obsd sites relative to predicted site[b] | obsd efficiency[c] | max value of stability function (kcal/mol) | ref[d] | data quality |
|---|---|---|---|---|---|
| *leu* | −1, 0, +1, +2 | very high (98%) | +3.1 | Wessler & Calvo, 1981; Searles et al., 1983 | VG |
| *thr* | 0, +1 | high (90%) | +3.0 | Lynn et al., 1985, 1988 | E |
| *trp* | −1, 0 | high (93 ± 5%) | −0.4 | Bertrand et al., 1977; Stauffer et al., 1978 | E |
| *trp* (*S. typhimurium*) | nd[e] | high (70%) | −1.0 | Lee & Yanofsky, 1977; Yanofsky & van Cleemput, 1982; Winkler et al., 1982 | G |
| *trp* (*Serratia marcescens*) | nd | high (70%) | −2.9 | Miozzari & Yanofsky, 1978 | G |
| *his* | +2 | high (>90%) | +6.0 | DiNocera et al., 1978; Verde et al., 1981; Frunzio et al., 1981 | VG |
| *his* (*S. typhimurium*) | −3, −2, −1, 0, +1 | very high (~100%) | +5.6 | Barnes, 1978; Freedman & Schimmel, 1981 | VG |
| *ilv*BN | nd | high (95%) | +4.6 | Hauser et al., 1985 | G |
| *ilv*GMEDA | nd | very high (~100%) | +3.2 | Lawther & Hatfield, 1980 | G |
| *phe*ST | −7, −6, −5, −4 | high (90%) | −5.6 | Fayat et al., 1983 | G |
| *phe*A | nd | medium (60%)[f] | +0.7 | Zurawski et al., 1978; Hudson & Davidson, 1984 | G |

[a] From *E. coli* unless indicated otherwise. [b] Predicted site = position at which the stability function comes closest to zero, under the assumption that $\Delta G_f^\circ{}_{\text{pol binding}} = -30$ kcal/mol. [c] Very high = ⩾98%; high = 67–97%; medium = 34–66%; low = 3–33%; very low = 1–2%; "zero" = <1% (absolute efficiencies of termination). [d] We list only those papers that present the DNA sequence, or the sites or efficiencies of in vitro transcription termination. We have not attempted a comprehensive listing. [e] nd = in vitro termination sites or efficiency not determined exactly. [f] In vivo determination.

Table VI: Class II: Attenuators of Nucleotide Biosynthesis Operons[a]

| attenuator | obsd sites relative to predicted site | obsd efficiency | max value of stability function (kcal/mol) | ref | data quality |
|---|---|---|---|---|---|
| *pyr*BI | nd | very high (98%) | −0.5 | Turnbough et al., 1983; Levin et al., 1989 | G |
| *orf*E-*pyr*E | nd | high (nd) | −3.9 | Poulson et al., 1984 | G |

[a] Footnotes to Table V apply.

Table VII: Class III: Intrinsic Mono- or Bidirectional Operon Terminators[a]

| terminator | obsd sites relative to predicted site | obsd efficiency | max value of stability function (kcal/mol) | ref | data quality |
|---|---|---|---|---|---|
| *sup*B-*sup*E | −2, −1, 0 | high (⩾95%) | −0.5 | Nakajima et al., 1982 | VG |
| *tonb*[b] | −3, −2 | high (71%) | +0.9 | Postle & Good, 1985 | VG |
| *P14*[b] | −3, −2 | high (72%) | +0.7 | Postle & Good, 1985 | VG |
| *rpo*H | −9, −8, −7, −6, −5, −4[c] | high (nd) | −0.7 | Yura et al., 1984; Erickson et al., 1987 | G |
| central (phages M13, f1, fd) | −2, −1 | very high (~100%) | −1.9 | Sugimoto et al., 1977;[d] Beck et al., 1978; Gentz et al., 1981 | VG |
| *trp t* | nd | low (25–30%) | −4.0 | Wu & Platt, 1978; Wu et al., 1981; Christie et al., 1981 | G |
| "oop" (phage λ)[e] | −2 | high (~80%) | +1.8 | Dahlberg & Blattner, 1973; Rosenberg et al., 1975, 1976; Kleid et al., 1976; Howard et al., 1977; Schwarz et al., 1978; Rosenberg & Court, 1979 | VG |
| M1 RNA (RNA component of RNase P) | −3, −2 | high (nd) | +3.3 | Sakamoto et al., 1983 | G |

[a] Footnotes to Table V apply. [b] *P14* terminator ≡ reverse orientation of *tonb* terminator (Postle & Good, 1985). [c] Possible S1 "nibbling" artifact. [d] See especially Figure 2, lane d, of the paper by Sugimoto et al. (1977). [e] Synonyms: "oop" terminator ≡ $t_L'$ terminator ≡ *4S* terminator ≡ $t_0$ terminator.

Table VIII: Class IV: Intergenic Attenuators[a,b]

| attenuator[b] | obsd sites relative to predicted site | obsd efficiency | max value of stability function (kcal/mol) | ref | data quality |
|---|---|---|---|---|---|
| *rps*O-*pnp* | nd | medium (~50%) | −5.7 | Régnier & Portier, 1986 | G |
| *frd*-*amp*C | −3, −1 | high (93%) | −3.1 | Jaurin et al., 1981; Grundstrom & Jaurin, 1982 | VG |
| *rpl*KAJL-*rpo*BC β | −3, −4[c] | high (80%)[d] | −6.1 | Dennis, 1977, 1984; Barry et al., 1980; Downing & Dennis, 1987 | VG |
| *tet*A-*orf*L (transposon Tn10) | −6[c] | high (96–99%)[d] | −11.2 | Hillen & Schollmeier, 1983; Schollmeier et al., 1985 | G |
| *nus*A-*inf*B[e] | −7, −6, −5, −4, −3, −2 (−8, −7) | high (75–90%) | −10.1 (−9.2) | Ishii et al., 1984a,b; Plumbridge et al., 1985; Ishihama et al., 1987; Régnier & Grunberg-Manago, 1989 | G |
| *lac*I-*lac*Z | nd[f] | low (20–25%)[g] | −14.5 | Horowitz & Platt, 1982 | G |
| *lac*Z-*lac*Y | nd | medium? (nd) | −12.2 | Büchel et al., 1980; Kwan et al., 1988 | G |

[a] Footnotes to Table V apply. [b] The attenuator is located between the two genes that are listed. [c] In vivo RNA characterized by S1 or ribonuclease T1 mapping. Transcript 3′-OH termini not determined in vitro. [d] In vivo determination. [e] Two stem–loop structures in tandem. [f] Two sites mapped exactly (Horowitz & Platt, 1982), but we calculate that $\Delta G_f^\circ{}_{\text{RNA hairpin}} \simeq +0.8$ kcal/mol. [g] Heparin increases efficiency to 90% (Horowitz & Platt, 1982).

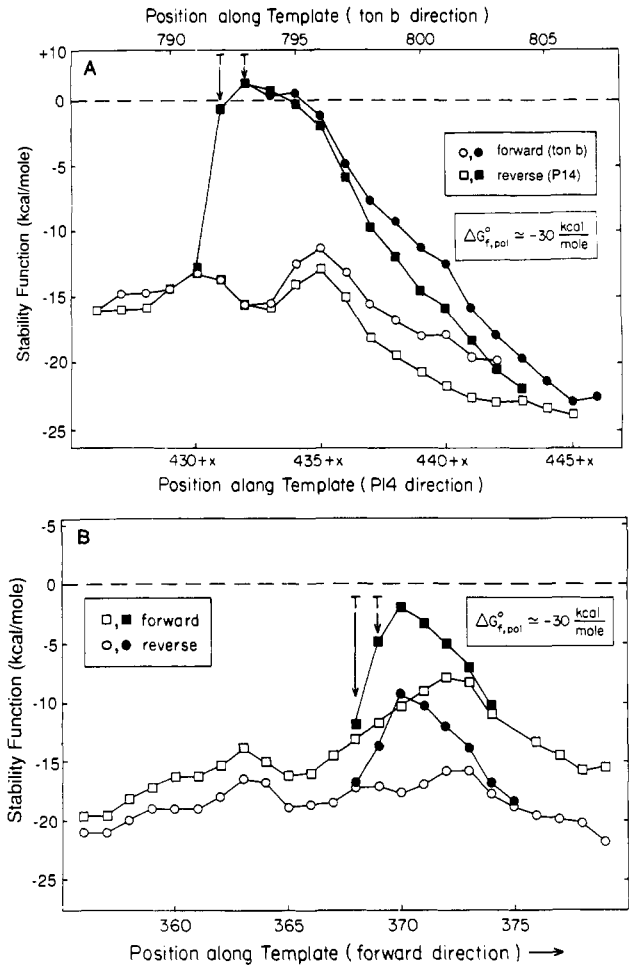



FIGURE 9: Forward and reverse profiles of the stability function for two terminators. (A) *ton*b terminator. Calculations were performed for the forward and reverse orientations of the *ton*b terminator (circles and squares, respectively). Vertical axes: defined as in Figure 5. Horizontal axis: position along template. Open points: $\{\Delta G_f^\circ{}_{\text{DNA bubble}} + \Delta G_f^\circ{}_{\text{RNA-DNA (12-mer)}}\}$. Solid points: stability function. The sites of in vitro termination are marked by vertical arrows and "T" symbols. Measured termination efficiencies in vitro (Postle & Good, 1985) are 71% in the forward (*ton*b) direction, and 72 ± 12% in the reverse (P14) direction. In the reverse orientation the start point of transcription has not yet been precisely mapped (Postle & Good, 1985). The symbol "x" (lower abscissa) denotes the difference in the number of nucleotides between the start point of transcription and the start point of translation (currently an unknown parameter). (B) Central terminator of phages M13, f1, and fd. Calculations were performed for the forward and reverse orientations of the central terminator (squares and circles, respectively). Axes and symbols defined as in (A). The observed termination sites in the forward orientation in vitro are marked with arrows and "T" symbols. In the reverse orientation, the exact termination sites have not been mapped. Forward efficiency ≃ 100% in vitro (Sugimoto et al., 1977); reverse efficiency ≃ 25–50% in vitro (Gentz et al., 1981).

entations of the *ton*b terminator are functionally identical.

The *ton*b terminator should be compared to a terminator that works efficiently in only one direction. In vitro, the "central" terminator of phages M13, f1, and fd has a very high efficiency in the forward direction, but only a low to medium efficiency in the reverse direction (Sugimoto et al., 1977; Beck et al., 1978; Gentz et al., 1981). Figure 9B displays our computation of the stability function for both orientations of the central terminator. A large asymmetry is apparent. The forward orientation gives the higher maximum and has been shown to work with ~100% efficiency in vitro (Sugimoto et al., 1977). From Figure 9B we might expect a low termination efficiency for the reverse orientation, and in fact a 25–50% efficiency is observed (Gentz et al., 1981). A possible reason for the asymmetry in Figure 9B is apparent. In the forward orientation the RNA–DNA hybrid downstream of the stem of the hairpin contains many dA·rU base pairs of low thermodynamic stability. In contrast, in the reverse orientation, the RNA–DNA hybrid downstream of the stem of the hairpin is relatively poor in dA·rU base pairs and thus is relatively stable.

We have extended this analysis to a number of other terminators that have been tested in vitro or in vivo for bidirectional function. In Table IX we compare the results of these experiments and our predictions. The agreement is reasonably good, with three anomalies being observed. One anomaly is the reverse orientation of the *nut*L structure, which is predicted to have a very low efficiency but is observed to have a high efficiency (95%) in vitro (Luk, 1982). We note that the reverse orientation of the *nut*L structure has no known physiological function. The other two anomalies involve the two orientations of the *tet*A-*orf*L intergenic attenuator of transposon Tn10. We predict an efficiency close to zero for both the forward and reverse directions of this attenuator. A high termination efficiency in vivo is observed in each orientation. We note that this intergenic attenuator has not been studied in vitro.

*(C) Finding Terminators within Natural Genetic Sequences.* To our knowledge, there have been two attempts to use a theoretical model as the basis of a method for locating terminators within natural DNA sequence. (1) McMahon and Tinoco (1978) noted that, in an intrinsic terminator, the RNA stem–loop structure tends to be followed by an oligo(rU) sequence. An $(rU{\cdot}dA)_n$ hybrid is relatively unstable (Chamberlin, 1965; Riley et al., 1966; Martin & Tinoco, 1980). Natural DNA sequence was searched for runs of dT, which would correspond to runs of rU in the RNA. In essence, this approach utilized the $\Delta G_f^\circ{}_{\text{RNA-DNA }(n\text{-mer})}$ term of our eqs 5 and 6, but *none* of the structural constraints that the "transcription bubble" paradigm imposes. This approach was able to locate several terminators within the $\phi$X174 genome but was not very discriminating. (2) More recently, Trifonov and co-workers have used a dinucleotide correlation analysis in an empirical attempt to identify "signature" for an intrinsic terminator (Brendel & Trifonov, 1984; Brendel et al., 1986). This approach was able to locate a number of terminators within natural genetic sequence. The discriminating power of this approach is unknown.

We have reexamined the problem of locating intrinsic terminators in light of our thermodynamic model.

*(1) Preselection of Test Sequences.* A search routine could in principle be tested against any subset of known prokaryotic genetic sequence. We apply two criteria to select a subset of sequence data. (1) The transcription bubble paradigm has been rigorously demonstrated only for RNA polymerase from *E. coli*. Therefore, we restrict attention to genetic sequence from *E. coli*, its close taxonomic relatives, and its phages, plasmids, and transposons. (2) An intrinsic terminator could exhibit quite different behaviors in vitro and in vivo, due to the influence of protein factors. Therefore, we choose to examine only those operons that have been transcriptionally well characterized both in vitro and in vivo.

*(2) Outputs of a Search Algorithm.* Over any segment of a genetic sequence, a search algorithm can have five possible outputs. (i) A "true positive" is the correct assertion that a terminator exists. (ii) A "false positive" is the assertion that a terminator exists when in fact it does not. (iii) A "true negative" is the correct assertion that a terminator does not exist. (iv) A "false negative" is the incorrect assertion that a terminator does not exist (i.e., it represents the failure to

Table IX: Terminators That Have Been Tested for Bidirectionality[a]

| terminator | obsd forward efficiency | forward orientation: max value of stability function (kcal/mol) | obsd reverse efficiency | reverse orientation: max value of stability function (kcal/mol) | ref | data quality |
|---|---|---|---|---|---|---|
| *ton*B | high (71%) | +0.9 | high (72%) | +0.7 | Postle & Good, 1985 | VG |
| histidine operon (*E. coli* and *S. typhimurium*) | high (93%)[b] | +0.3 | high (96%)[b] | −0.5 | Carlomagno et al., 1985 | G |
| *rrn*B $t_1$ | high (80–95%) | +1.6 | medium to high?[c] | +0.8 | Brosius et al., 1981; Sarmientos et al., 1983; Glaser et al., 1983; Li and Squires, unpublished, cited in Platt (1986) | G |
| *tet*A-*orf*L intergenic attenuator (transposon Tn10) | high (96–99%)[b] | −11.2 | high (78–96%)[b] | −14.1 | Hillen & Schollmeier, 1983; Schollmeier et al., 1985 | G |
| central (phages M13, f1, fd) | very high (~100%) | −1.9 | low to medium (25–50%) | −9.8 | Sugimoto et al., 1977; Beck et al., 1978; Gentz et al., 1981 | G |
| *DeoCABD* | high (~95%)[b] | −3.9 | low (nd)[b] | −4.1 | Larsen et al., 1987 | G |
| *trp t* | low to medium (25–30%) | −4.0 | zero | −10.6 | Christie et al., 1981 | VG |
| Te (phage T3) | low (5–20%) | −7.3 | zero | −17.0 | Chamberlin et al., 1986; Briat et al., 1987 | VG |
| *lac*Z-*lac*Y intergenic attenuator | medium? (nd) | −12.2 | zero | −11.1 | Kwan et al., 1988[d] | G |
| *nut*L (phage λ) | zero | −13.4 | high (95%) | −9.4 | Luk, 1982 | VG |

[a] Footnotes to Table V apply. [b] In vivo determination. [c] Efficiency not specified explicitly in the publication. [d] Additional information provided by poster at the UCLA Symposium "DNA–Protein Interactions in Transcription", April 4–10, 1988, Keystone, CO.

detect a terminator that is present). (v) The algorithm may make a definite prediction, but there may be insufficient experimental evidence to evaluate it. We must determine the performance of our search algorithm with respect to each of these possible outcomes.

(*3*) *Results of a Search.* Table X summarizes the results of a search for terminators in four *E. coli* operons. The search was conducted according to the algorithm described under Methods. Very good results were obtained in three cases.

(*a*) *The trp Operon of E. coli.* This operon contains five genes. The *trp* attenuator is located immediately downstream of the promoter, in the leader peptide region of the first gene. The intrinsic *trp t* terminator and the ρ-dependent *trp t'* terminator are located (in tandem) downstream of the last gene. Our algorithm correctly identifies the *trp* attenuator and the *trp t* terminator. It misses the ρ-dependent *trp t'* terminator but makes no other false predictions.

(*b*) *The rplKAJL-rpoBC Operon of E. coli.* This operon contains six genes and two unidentified open reading frames. The U terminator is located before the first gene and appears to define the end of an upstream operon. The β attenuator is located between the fourth and fifth genes (*rpl*L and *rpo*B), and the β' terminator is located downstream of the sixth gene (*rpo*C). Our model correctly identifies the U terminator, the β attenuator, and the β' terminator. It makes two minor false predictions of very weak termination events but otherwise appears error-free.

(*c*) *The rpsU-dnaG-rpoD Operon of E. coli.* This operon contains three genes. The $t_1$ terminator (actually an intergenic attenuator) is located between *dna*G and *rpo*D, and the $t_2$ terminator is located downstream of *rpo*D. Our model correctly identifies these two structures and makes no false predictions.

(*d*) *The lacIZYA Operon of E. coli.* This operon contains four genes and an unidentified open reading frame (after *lac*A, in the opposing orientation). The known termination structures include an intergenic attenuator between *lac*I and *lac*Z, one intrinsic and four ρ-dependent termination sites within the *lac*Z gene, and an intergenic attenuator between *lac*Z and *lac*Y. A termination site has also been tentatively mapped at the second hairpin following the *lacA* gene [Figure 5 of Hediger et al. (1985)].

Our search algorithm cannot adequately analyze this operon. There are five problems. (i) The *lac*I-*lac*Z intergenic attenuator is not detected. We note that this structure overlaps with the binding site for CAP protein (Horowitz & Platt, 1982). (ii) The *lac*Z-*lac*Y intergenic attenuator is barely detected, even though it functions with a medium efficiency in vitro (Kwan et al., 1988). (iii) None of the termination sites within *lac*Z are detected. We note that the intrinsic terminator in *lac*Z appears to consist of two widely separated RNA segments (Ruteshouser & Richardson, 1989). (iv) We predict termination to occur at the third hairpin after the *lac*A gene, but not at the second hairpin after this gene. However, the converse is observed [Figure 5 of Hediger et al. (1985)]. We note that the *lac*A gene (and its terminator) may be of plasmid or transposon origin (Hediger et al., 1985). (v) We predict a very weak termination efficiency for the first hairpin following the *lac*A gene, when transcribed in the opposing direction (into the *lac*A gene). This hairpin is located between the *lac*A gene and the unidentified open reading frame and thus might be a terminator for the open reading frame. To our knowledge, no experimental studies have been published on the transcription of this hairpin in the ORF → *lac*A direction.

It therefore seems clear that the *lac* operon is transcriptionally quite complicated, and our model will need refinement before predictions can be reconciled with observation.

## Conclusions

In this paper we have described a thermodynamic model of transcriptional elongation and intrinsic termination in *E. coli*. Our approach is based on the transcription bubble paradigm that applies to *E. coli* RNA polymerase in the elongation mode. The model postulates that intrinsic termination occurs when the RNA–DNA hybrid within the transcription bubble is displaced by a hairpin in the RNA.

We have tested our thermodynamic model against a large number of canonical intrinsic terminators, including the *thr* and *trp* attenuators and the bidirectional *ton*b terminator of *E. coli*. In many cases there is excellent agreement between the predicted and observed sites of termination. In a preliminary search of natural genetic sequence the model was usually highly discriminating in locating intrinsic terminators (Table X).

Table X: Results of a Search of Genetic Sequence[a]

| operon | no. of RNA hairpins: low cutoff[b] | no. of RNA hairpins: high cutoff[b] | true positives[c] | true negatives (bp) | % base pairs correctly analyzed[d,e] | false positives | false negatives | predictions, with insufficient exptl data | ref |
|---|---|---|---|---|---|---|---|---|---|
| *trp* (5 genes; 7335 bp) | 38 | 12 | *trp* attenuator (S), *trp t* terminator (M) | 7100 | 97.3 | none | *trp t'* terminator[f] | none | Wu & Platt, 1978; Wu et al., 1981 |
| *rplKAJL-rpoBC* (6 genes; 12337 bp) | 95 | 29 | U terminator (S), $\beta$ attenuator (S), $\beta'$ terminator (S) | 12230 | 99.6 | $\beta$–$\beta'$ intergenic hairpin (VW) *orfA-orfB* intergenic hairpin (VW) | none | none | Squires et al., 1981; Dennis 1977, 1984; Downing & Dennis, 1987; Morgan & Hayward, 1987 |
| *rpsU-dnaG-rpoD* (3 genes; 5059 bp) | 30 | 5 | $t_1$ terminator (M), $t_2$ terminator (S) | 5020 | 99.2 | none | none | none | Burton et al., 1983 |
| *lacIZYA* (4 genes; 5800 bp) | 39 | 21 | *lacZ-lacY* intergenic attenuator (VW) | 6900 | 92.5–92.8 | third hairpin downstream of *lacA* | *lacI-lacZ* intergenic attenuator; *lacZ* internal terminators; second hairpin downstream of *lacA* (tentative *lacA* terminator) | *orf* terminator (VW) | de Crombrugghe et al., 1971; Kung et al., 1975; Greenblatt et al., 1980; Horowitz & Platt, 1982; Hediger et al., 1985; Nakamura et al., 1986; Kwan et al., 1988; Ruteshouser & Richardson, 1989 |

[a] Particular *E. coli* operons were scanned and potential hairpins were noted. In the region of each hairpin, the stability function was calculated (eq 5B). Termination efficiencies were predicted with an empirical correlation of *max*(stability function) versus efficiency (von Hippel & Yager, 1991). The number of base pairs of sequence falling within all true positives and true negatives was added up and was divided by the total number of base pairs analyzed. This ratio is the "% correct" parameter given in colum 6 of the table. [b] Low cutoff criterion: 3 bp $\leq$ loop length $\leq$ 8 bp; 6 bp $\leq$ stem length; 15 $\leq$ "quality score". High cutoff criterion: 3 bp $\leq$ loop length $\leq$ 8 bp; 8 bp $\leq$ stem length; 18 $\leq$ "quality score" (see Methods). [c] Parenthetical symbols indicate predicted efficiencies (von Hippel & Yager, 1991): VS = very strong ($\geq$98%), S = strong (67–97%), M = medium (34–66%), W = weak (2–33%), VW = very weak (1–2%). [d] (% base pairs correctly analyzed) = (no. of base pairs in true positives and true negatives)/(total no. of base pairs in operon). The total number of base pairs in the operon is provisional, is based on published sequence data, and is listed in the first colum of the table. [e] A putative rho recognition site is taken to be 200 base pairs long (Morgan et al., 1985).

Our model does *not* successfully predict all loci of intrinsic termination. It fails to predict the termination sites in the complex *lac* operon (Table X). It also fails in cases where the sequence that flanks the transcription bubble is important [see Goliger et al. (1989) and Telesnitsky and Chamberlin (1989a,b)].

We restrict attention in this paper to a minimal model that is based on the principles of equilibrium thermodynamics. Obviously this model is simplistic because real termination events must occur in a *kinetic* context. This context is defined by the rate of polymerase movement between adjacent template positions during RNA transcript polymerization. We hypothesize a kinetic competition between elongation and termination at each position along the template, with the relative probability of entry into either path specified by an activation energy barrier of the Eyering type. This extension of our model will be presented elsewhere (von Hippel & Yager, 1991). Here we merely note that the outcome of such a kinetic competition should be decided by a free energy term that could be quite small. This term is the *difference* in the heights of the activation energy barriers for entry into the two kinetic pathways.

Figure 2 defines a state T which must be traversed en route to a final, irreversible dissolution of the transcription complex. Thus state T appears to fit the definition of a transition state in the Eyring formalism. Furthermore, the free energy of state T depends on the value of $\Delta G_f°_{\text{pol binding}}$. This indicates that slight alterations in the value of $\Delta G_f°_{\text{pol binding}}$ could modulate the choice between elongation and termination at an intrinsic terminator and thus could control the efficiency, though not the position, of intrinsic termination events (von Hippel & Yager, 1991).

## Acknowledgments

We thank many of our colleagues, both at the University of Oregon and elsewhere, for criticism and advice. These include, in particular, Otto Berg, Michael Chamberlin, Johannes Geiselmann, Stanley C. Gill, Robert Landick, Jeffrey Roberts, and Charles Yanofsky (several of whom provided experimental results and papers in advance of publication) and the two anonymous referees of the first version of this paper. We also thank Jean Parker for carefully typing the many drafts of the manuscript.

**Registry No.** RNA polymerase, 9014-24-8.

## References

Adhya, S., & Gottesman, M. (1978) *Annu. Rev. Biochem. 47*, 967–996.

Albrechtsen, B., Squires, C. L., Li, S., & Squires, C. (1990) *J. Mol. Biol. 213*, 123–134.

Andrews, C., & Richardson, J. P. (1985) *J. Biol. Chem. 260*, 5826–5831.

Arndt, K., & Chamberlin, M. J. (1988) *J. Mol. Biol. 202*, 271–285.

Arndt, K. M., & Chamberlin, M. J. (1990) *J. Mol. Biol. 213*, 79–108.

Barnes, W. M. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4281–4285.

Barry, G., Squires, C., & Squires, C. L. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3331–3335.

Beal, R. B., Pillai, R. P., Chuknyisky, P. P., Levy, A., Tarien, E., & Eichhorn, G. L. (1990) *Biochemistry 29*, 5994–6002.

Beck, E., Sommer, R., Averswald, E. A., Kurz, Ch., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamota, T., & Takanami, M. (1978) *Nucleic Acids Res. 5*, 4495–4503.

Berg, K. L., Squires, C., & Squires, C. L. (1989) *J. Mol. Biol. 209*, 345–358.
Bertrand, K., Korn, L., Lee, F., & Yanofsky, C. (1977) *J. Mol. Biol. 117*, 227–247.
Brendel, V., & Trifonov, E. N. (1984) *Nucleic Acids Res. 12*, 4411–4427.
Brendel, V., Hamm, G. H., & Trifonov, E. N. (1986) *J. Biomol. Struct. Dyn. 3*, 705–723.
Brennan, C. A., Dombrowski, A. J., & Platt, T. (1987) *Cell 48*, 945–952.
Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746–3750.
Briat, J.-F., Bollag, G., Kearney, C. A., Molineux, I., & Chamberlin, M. J. (1987) *J. Mol. Biol. 198*, 43–49.
Brosius, J., Dull, T. J., Sleeter, D. D., & Noller, H. F. (1981) *J. Mol. Biol. 148*, 107–127.
Büchel, D. E., Gronenborn, B., & Müller-Hill, B. (1980) *Nature 283*, 541–545.
Burton, Z. F., Gross, C. A., Watanabe, K. K., & Burgess, R. R. (1983) *Cell 32*, 335–349.
Cantor, C., & Schimmel, P. (1980) *Biophysical Chemistry*, W. H. Freeman, San Francisco.
Carlomagno, M. S., Riccio, A., & Bruni, C. B. (1985) *J. Bacteriol. 163*, 362–368.
Carpousis, A. J., & Gralla, J. D. (1980) *Biochemistry 19*, 3245–3253.
Chamberlin, M. J. (1965) *Fed. Proc. 24*, 1446–1457.
Chamberlin, M. J., Arndt, K. M., Briat, J.-F., Helmann, J., Reynolds, R. R., Schmidt, M. C., & Singer, V. L. (1986) in *Genetic Chemistry: The Molecular Basis of Heredity*, R. A. Welch Foundation, Houston.
Christie, G. E., Farnham, P. J., & Platt, T. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4180–4184.
Chuknyisky, P. P., Rifkind, J. M., Tarien, E., Beal, R. B., & Eichhorn, G. L. (1990) *Biochemistry 29*, 5987–5994.
Dahlberg, J. E., & Blattner, F. R. (1973) in *Virus Research* (Fox, C. F., & Robinson, W. S., Eds.) pp 533–543, Academic Press, New York.
de Crombrugghe, B., Adhya, S., Gottesman, M., & Pastan, I. (1971) *Nature (New Biol.) 241*, 260–264.
Dennis, P. (1977) *J. Mol. Biol. 115*, 603–625.
Dennis, P. (1984) *J. Biol. Chem. 259*, 3202–3209.
DiNocera, P. P., Blasi, F., DiLauro, F., Frunzio, R., & Bruni, C. B. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4276–4280.
Dove, W. F., & Davidson, N. (1962) *J. Mol. Biol. 5*, 467–478.
Downing, W. L., & Dennis, P. P. (1987) *J. Mol. Biol. 194*, 609–620.
Erickson, J. W., Vaughn, V., Walter, W. A., Neidhardt, F. C., & Gross, C. A. (1987) *Genes Dev. 1*, 419–432.
Farnham, P. J., & Platt, T. (1980) *Cell 20*, 739–748.
Fayat, G., Mayaux, J. F., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M., & Blanquet, S. (1983) *J. Mol. Biol. 171*, 239–261.
Freedman, R., & Schimmel, P. (1981) *J. Biol. Chem. 256*, 10747–10750.
Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986a) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373–9377.
Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986b) *Biochemistry 25*, 3209–3213.
Friedman, D. I., Imperiale, M. J., & Adhya, S. L. (1987) *Annu. Rev. Genet. 21*, 453–488.
Frunzio, R., Bruni, C. B., & Blasi, F. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2767–2771.
Gamper, H. B., & Hearst, J. E. (1982) *Cell 29*, 81–90.
Gardner, J. F. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1706–1710.
Gardner, J. F. (1982) *J. Biol. Chem. 257*, 3896–3904.
Gentz, R., Langner, A., Chang, A. C. Y., Cohen, S. N., & Bujard, H. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4936–4940.
Gill, S., Yager, T. D., & von Hippel, P. H. (1990) *Biophys. Chem. 37*, 239–250.
Glaser, G., Sarmientos, P., & Cashel, M. (1983) *Nature 302*, 74–76.
Goliger, J. A., Yang, X., Guo, H. C., & Roberts, J. W. (1989) *J. Mol. Biol. 205*, 331–341.
Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301–319.
Grayhack, E. J., Yang, X., Lau, L. F., & Roberts, J. W. (1985) *Cell 42*, 259–269.
Greenblatt, J., Li, J., Adhya, S., Friedman, D. I., Baron, L. S., Redfield, B., King, H.-F., & Weissbach, H. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1991–1994.
Grundstrom, T., & Jaurin, B. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1111–1115.
Hanna, M. M., & Meares, C. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4238–4242.
Hansen, U. M., & McClure, W. R. (1980) *J. Biol. Chem. 255*, 9564–9570.
Hauser, C. A., Sharp, J. A., Hatfield, L. K., & Hatfield, G. W. (1985) *J. Biol. Chem. 260*, 1765–1770.
Hediger, M. A., Johnson, D. F., Nierlich, D. P., & Zabin, I. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6414–6418.
Hillen, W., & Schollmeier, K. (1983) *Nucleic Acids Res. 11*, 525–539.
Hockensmith, J. W., Kubasek, W. L., Vorachek, W. R., Evertsz, E. M., & von Hippel, P. H. (1990) *Methods Enzymol.* (in press).
Horowitz, H., & Platt, T. (1982) *J. Biol. Chem. 257*, 11740–11746.
Howard, B. H., de Crombrugghe, B., & Rosenberg, M. (1977) *Nucleic Acids Res. 4*, 827–842.
Hudson, G. S., & Davidson, B. E. (1984) *J. Mol. Biol. 180*, 1023–1051.
Ishihama, A., Honda, A., Nagasawa-Fujimori, H., Glass, R. E., Maekawa, T., & Imamoto, F. (1987) *Mol. Gen. Genet. 206*, 185–191.
Ishii, S., Kuroki, K., & Imamoto, F. (1984a) *Proc. Natl. Acad. Sci. U.S.A. 81*, 409–413.
Ishii, S., Hatada, E., Maekawa, T., & Imamoto, F. (1984b) *Nucleic Acids Res. 12*, 4987–4995.
Jaeger, J. A., Turner, D. H., & Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706–7710.
Jalajakumari, M. B., Guidolin, A., Buhk, H. J., Manning, P. A., Ham, L. M., Hodgson, A. L. M., Cheah, K. C., & Skurray, R. A. (1987) *J. Mol. Biol. 198*, 1–11.
Jaurin, B., Grundström, T., Edlund, T., & Normark, S. (1981) *Nature 290*, 221–225.
Jin, D. J., Walter, W. A., & Gross, C. A. (1988) *J. Mol. Biol. 202*, 245–253.
Kierkegaard, K., Buc, H., Spassky, A., & Wang, J. C. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2544–2548.
Kleid, D., Humayan, Z., Jeffery, A., & Ptashne, M. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 293–297.
Kolter, R., & Yanofsky, C. (1984) *J. Mol. Biol. 175*, 299–312.
Krakow, J. S., & Fronk, E. (1969) *J. Biol. Chem. 244*, 5988–5993.
Kumar, S. A., & Krakow, J. S. (1975) *J. Biol. Chem. 250*, 2878–2884.

Kung, H.-F., Spears, C., & Weissbach, H. (1975) *J. Biol. Chem. 250*, 1556–1562.
Kwan, C., Murakawa, G. J., & Nierlich, D. P. (1988) *J. Cell. Biochem. Suppl. 12D*, 141.
Landick, R., & Yanofsky, C. (1987) in *E. coli and S. typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., et al., Eds.) pp 1276–1301, American Society for Microbiology Publications, Washington, DC.
Larsen, J. E. L., Albrechtson, B., & Valentin-Hansen, P. (1987) *Nucleic Acids Res. 15*, 5125–5140.
Lawther, R. P., & Hatfield, G. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1862–1866.
Lee, F., & Yanofsky, C. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4365–4369.
Levin, H. L., Park, K., & Schachman, H. K. (1989) *J. Biol. Chem. 264*, 14638–14645.
Lindahl, L., & Zengel, J. M. (1986) *Annu. Rev. Genet. 20*, 297–326.
Luk, K.-C. (1982) *Mol. Gen. Genet. 187*, 320–325.
Lynn, S. P., Gardner, J. F., & Reznikoff, W. S. (1982) *J. Bacteriol. 152*, 363–371.
Lynn, S. P., Bauer, C. E., Chapman, K., & Gardner, J. F. (1985) *J. Mol. Biol. 183*, 529–541.
Lynn, S. P., Kasper, L. M., & Gardner, J. F. (1988) *J. Biol. Chem. 263*, 472–479.
Martin, F. H., & Tinoco, I., Jr. (1980) *Nucl. Acids Res. 8*, 2295–2299.
Masukata, H., & Tomizawa, J. (1990) *Cell 62*, 331–338.
McClure, W. R. (1985) *Annu. Rev. Biochem. 54*, 171–204.
McMahon, J. E., & Tinoco, I., Jr. (1978) *Nature 271*, 275–277.
Metzger, W., Schickor, P., & Heumann, H. (1989) *EMBO J. 8*, 2745–2754.
Minton, A. P. (1981) *Biopolymers 20*, 2093–2120.
Miozzari, G. F., & Yanofsky, C. (1978) *Nature (London) 276*, 684–689.
Monforte, J. A., Kahn, J. D., & Hearst, J. E. (1990) *Biochemistry 29*, 7882–7890.
Morgan, B. A., & Hayward, R. S. (1987) *Mol. Gen. Genet. 210*, 358–363.
Morgan, E. A. (1986) *J. Bacteriol. 168*, 1–5.
Morgan, W. D., Bear, D. G., & von Hippel, P. H. (1983a) *J. Biol. Chem. 258*, 9553–9564.
Morgan, W. D., Bear, D. G., & von Hippel, P. H. (1983b) *J. Biol. Chem. 258*, 9565–9574.
Morgan, W. D., Bear, D. G., & von Hippel, P. H. (1984) *J. Biol. Chem. 259*, 8664–8671.
Morgan, W. D., Bear, D. G., Litchman, B. L., & von Hippel, P. H. (1985) *Nucleic Acids Res. 13*, 3739–3754.
Mott, J. E., Grant, R. A., Ho, Y. S., & Putt, T. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 88–92.
Nakajima, N., Ozeki, H., & Shimura, Y. (1982) *J. Biol. Chem. 257*, 11113–11120.
Nakamura, Y., Mizusawa, S., Court, D. L., & Tsugawa, A. (1986) *J. Mol. Biol. 189*, 103–111.
Oppenheim, A. B., Gottesman, S., & Gottesman, M. (1982) *J. Mol. Biol. 158*, 327–346.
Phillips, G. J., Arnold, J., & Ivarie, R. (1987) *Nucleic Acids Res. 15*, 2611–2626.
Pinkham, J., & Platt, T. (1983) *Nucleic Acids Res. 11*, 3531–3545.
Platt, T. (1986) *Annu. Rev. Biochem. 55*, 339–372.
Plumbridge, J. A., Dondon, J., Nakamura, J., & Grunberg-Manago, M. (1985) *Nucleic Acids Res. 13*, 3371–3388.
Postle, K., & Good, R. F. (1985) *Cell 41*, 577–585.
Poulsen, P., Bonekamp, F., & Jensen, K. F. (1984) *EMBO J. 3*, 1783–1790.
Radding, C. M., Beattie, K. L., Holloman, W. K., & Wiegand, R. C. (1977) *J. Mol. Biol. 116*, 825–839.
Record, M. T., Jr., Lohman, T. M., & deHaseth, P. (1976) *J. Mol. Biol. 107*, 145–158.
Record, M. T., Jr., deHaseth, P. L., & Lohman, T. M. (1977) *Biochemistry 16*, 4791–4796.
Régnier, P., & Portier, C. (1986) *J. Mol. Biol. 187*, 23–32.
Régnier, P., & Grunberg-Manago, M. (1989) *J. Mol. Biol. 210*, 293–302.
Reynolds, R. (1988) Ph.D. Thesis in Biochemistry, University of California, Berkeley.
Reznikoff, W. S., Maquat, L. E., Munson, L. M., Johnson, R. C., & Mandecki, W. (1982) in *Promoters: Structure and Function* (Rodriquez, R. L., & Chamberlin, M. J., Eds.) pp 80–95, Praeger, New York.
Rhodes, G., & Chamberlin, M. J. (1974) *J. Biol. Chem. 249*, 6675–6683.
Richardson, J. P. (1975) *J. Mol. Biol. 98*, 565–579.
Riley, M., Maling, B., & Chamberlin, M. J. (1966) *J. Mol. Biol. 20*, 359–389.
Roberts, J. W. (1969) *Nature 224*, 1168–1174.
Rosenberg, M., & Court, D. (1979) *Annu. Rev. Genet. 13*, 319–353.
Rosenberg, M., de Crombrugghe, B., & Weissman, S. M. (1975) *J. Biol. Chem. 250*, 4755–4764.
Rosenberg, M., de Crombrugghe, B., & Musso, R. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 717–721.
Ruteshouser, E. C., & Richardson, J. P. (1989) *J. Mol. Biol. 208*, 23–43.
Sakamoto, H., Kimura, N., & Shimura, Y. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6187–6191.
Sarmientos, P., Sylvester, J. E., Contente, S., & Cashel, M. (1983) *Cell 32*, 1337–1346.
Schildkraut, C., & Lifson, S. (1965) *Biopolymers 3*, 195–208.
Schmeissner, U., McKenney, K., Rosenberg, M., & Court, D. (1984) *Gene 28*, 343–350.
Schmidt, M. C., & Chamberlin, M. J. (1987) *J. Mol. Biol. 195*, 809–818.
Schollmeier, K., Gartner, D., & Hillen, W. (1985) *Nucleic Acids Res. 13*, 4227–4237.
Schwarz, E., Scherer, G., Hobom, G., & Kossel, H. (1978) *Nature 272*, 410.
Searles, L. L., Wessler, S. R., & Calvo, J. M. (1983) *J. Mol. Biol. 162*, 377–394.
Shi, Y., Gamper, H., Van Houten, B., & Hearst, J. E. (1988) *J. Mol. Biol. 199*, 277–293.
Siebenlist, U., Simpson, R. B., & Gilbert, W. (1980) *Cell 20*, 269–281.
Squires, C., Krainer, A., Barry, G., Shen, W.-F., & Squires, C. L. (1981) *Nucleic Acids Res. 9*, 6827–6840.
Stauffer, G. V., Zurawski, G., & Yanofsky, C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4833–4837.
Straney, D. C., & Crothers, D. M. (1985) *Cell 43*, 449–459.
Sugimoto, K., Sugisaki, H., Okamoto, T., & Takanami, M. (1977) *J. Mol. Biol. 111*, 487–507.
Telesnitsky, A. P. W., & Chamberlin, M. J. (1989a) *J. Mol. Biol. 205*, 315–330.
Telesnitsky, A., & Chamberlin, M. J. (1989b) *Biochemistry 28*, 5210–5218.
Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., & Gralla, J. (1973) *Nature (New Biol.) 246*, 40–41.
Tomizawa, J., & Masukata, H. (1987) *Cell 51*, 623–630.

Turnbough, C. L., Jr., Hicks, K. L., & Donahue, J. P. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 368–372.
Verde, P., Frunzio, R., Di Nocera, P. P., Blasi, F., & Bruni, C. B. (1981) *Nucleic Acids Res. 9*, 2075–2086.
von Hippel, P. H., & Yager, T. D. (1991) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
von Hippel, P. H., Bear, D. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389–446.
von Hippel, P. H., Yager, T. D., Bear, D. G., McSwiggen, J. A., Geiselman, J., Gill, S. C., Linn, J. D., & Morgan, W. D. (1987) in *RNA Polymerase and the Regulation of Transcription* (Reznikoff, W. S., et al., Eds.) pp 325–334, Elsevier, New York.
Wessler, S. R., & Calvo, J. M. (1981) *J. Mol. Biol. 149*, 579–597.
Whalen, W., Ghosh, B., & Das, A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2494–2498.
Winkler, M. E., Mullis, K., Barnett, J., Stroynowski, I., & Yanofsky, C. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2181–2185.
Wu, A., & Platt, T. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5442–5446.
Wu, A. M., Christie, G. E., & Platt, T. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2913–2917.
Yager, T. D., & von Hippel, P. H. (1987) in *E. coli and S. typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., et al., Eds.) pp 1241–1275, Americal Society for Microbiology Publications, Washington, DC.
Yang, M.-T., & Gardner, J. F. (1989) *J. Biol. Chem. 264*, 2634–2639.
Yanofsky, C., & van Cleemput, J. (1982) *J. Mol. Biol. 155*, 235–246.
Yura, T., Tobe, Y., Ito, K., & Osawa, T. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6803–6807.
Zurawski, G., Brown, K., Killingly, D., & Yanofsky, C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4271–4275.