
Understanding RNA chain elongation and its regulation requires knowledge of the structure of the transcription complex, in particular, the location, contacts, and movement of the RNA and DNA chains and how these features change at pause and termination sites. In the 1980s, we pictured a transcription complex structure of generally fixed dimensions: an ~35-bp footprint on the DNA, a 17-nucleotide-stranded "bubble" of separated DNA strands, and a 12-bp heteroduplex between the 3′-proximal region of transcript and the DNA template strand (1). RNA chain elongation was considered a repeating cycle in which nucleotide addition was tightly coupled to movement in single-base increments of the RNA and DNA chains through this fixed structure. Explanations for termination and pausing focused on how this stable complex could be altered, principally by changes in RNA–DNA base pairing (1,2).

However, recent observations suggest that chain elongation may be a more complex process that involves multiple transcription complex conformations and movement of the transcription complex or its constituent parts in cycles greater than single-nucleotide addition. We aim here to describe some of these findings and to suggest a framework within which they can be evaluated. We shall then consider the implications of this more elaborate view of transcription for mechanisms that regulate transcript elongation: pausing, termination, and arrest. To develop these ideas, we will describe current views about these mechanisms and highlight two important points: (a) the multipartite architecture of elongation control signals, using for illustration the pause sites in the his and trp operon leader regions, and (b) the recent discovery that cellular factors can promote transcript cleavage. Finally, we will suggest several directions for future studies of RNA chain elongation.
most termination and pause sites it subsequently encounters (5) (see below). Antitermination allows efficient transcription of ribosomal RNA genes and some bacteriophage operons.

Thus, a successful model of transcription complex structure and action must explain six distinct behaviors: remarkable stability at most template positions, termination, pausing, arrest, transcript cleavage, and antitermination.

**Transcription Complex Structure**

**Distinct Sites Within the Transcription Complex**

Several features of transcription complex structure are apparent from theoretical requirements for chain elongation and findings to date (Fig. 1). However, as noted recently by Krummel and Chamberlin (3), this picture is derived largely from studies of halted transcription complexes and may not accurately represent the structure of transcription complexes actively engaged in chain elongation. Further, as we explain below, transcription complexes may not assume an identical conformation at different template positions. In elongation complexes halted at single-template positions, RNA polymerase protects ~25 to ~40 bp of DNA from digestion by various endo- and exonucleases (6 and references therein). Together, topological studies and chemical reactivity studies suggest that the single-stranded bubble var-

ies from ~14 to ~18 nt (7,8, and references therein). Thus, RNA polymerase appears to contact the DNA tightly at sites (A,B, Fig. 1) downstream and upstream from the melted region. Formation of the bubble requires a strand separation site (C, Fig. 1). The duplex must reform either in a second upstream site (D, Fig. 1) or spontaneously as the strands exit the complex. Within the melted region, the template strand presumably occupies a channel through RNA polymerase (E, Fig. 1) that prevents it from pairing with the non-template strand. Several studies of defined complexes reveal an ~8-nt protected region of template DNA within the transcription bubble that could result from these interactions (8 and references therein).

Nucleotide addition presumably occurs in a single active site (F, Fig. 1). Although models of transcription complex structure involving multiple active sites have been proposed and remain formal possibilities, available evidence clearly favors a single-site model. Extensive enzymologic studies, recently reviewed by Erie et al. (9), suggest that the active site contains two subsites: one for the 3'-hydroxyl terminus of the RNA product and a second for the NTP substrate. These studies reveal an $S_2^*$ mechanism for phosphodiester bond formation, which requires that the $\beta\gamma$PP, of the NTP substrate and the 3'-OH on the RNA chain be located at the apical positions of a trigonal bipyramid centered on the substrate $\alpha$-phosphate. A complete cycle of nucleotide addition requires release of pyrophosphate and translocation of the new 3' end to the product terminus site.

**The Nascent Transcript Path**

How extensive is RNA–DNA base pairing in the transcription complex? Although a subject of current debate and vitally important to understanding the mechanism of chain elongation, the extent to which the nascent transcript remains paired to the DNA template after exiting the active site is unresolved. Two alternative views are that it forms an 8- to 12-bp RNA:DNA hybrid (1,2,7,8) or that the hybrid is 3 or fewer bp (10–12, and older papers from Mildvan, Krakow, or Kumar cited therein). Protection of 8–12 nt of template DNA from single-strand–specific chemical modification in many halted elongation complexes (7,8) can be explained by, but does not require, the more extensive hybrid; sensitivity of the nascent transcript to ribonucleases (8,13; S. Milan and M. J. Chamberlin, personal communication) favors the shorter hybrid. Several other findings support one or the other hypothesis (see 2,11, and references therein) and a definitive rejection of either is currently impossible. It is also possible that the transcript alternates between an extensive RNA:DNA hybrid and a single-stranded RNA-binding pocket in different transcription cycle intermediates.

Regardless of which model is correct, sites in the complex must direct the transcript (either as single-stranded RNA or as RNA:DNA hybrid) away from the active site (product-binding site; G, Fig. 1) and, at some position, separate it from the DNA template (H, Fig. 1). Several results favor an ~8-nt product-binding site: (a) ribonuclease V1, which is specific for RNA segments containing stacked bases, cleaves ~8 nt of 3'-proximal transcript in at least some halted complexes (D. N. Lee and R. Landick, unpublished observations); (b) a corresponding ~8-nt region of DNA template is protected from modification by single-strand–specific reagents; and (c) 8 nt can be added to a transcription-priming nucleotide analog when it is cross-linked to His-1237 in the $\beta$ subunit (14; see below). However, 16–24 nt of transcript are protected from cleavage by low levels of ribonucleases (8,13, and older papers from Krakow cited therein) and high concentrations of micrococcal nuclease (D. N. Lee and R. Landick, unpublished). Thus, 8–16 nt of transcript 5' to an ~8-nt product-binding site appears to remain within the transcription complex in some type of transcript exit channel (I, Fig. 1).

**RECENT STUDIES SUGGEST THAT THE TRANSCRIPTION COMPLEX IS A DYNAMIC STRUCTURE THAT MOVES DISCONTINUOUSLY ON THE DNA**

Until recently, the prevailing view held that each nucleotide addition is accompanied by synchronous, single-step movement along the DNA template of a static transcription complex structure (such as depicted in Fig. 1). However, recent reports reveal differences in the footprint and transcription bubble among different halted elongation complexes that are more easily explained by assuming that at least some sites within the complex are not locked to one another during elongation, but may move in cycles greater than single-nucleotide addition. In this view, the paradigm diagrammed in Fig. 1 serves to illustrate potential “moving parts” in a transcription complex “machine.”

**RNA Polymerase-DNA Contacts**

Both the overall length of DNA protected from cleavage by DNase I, hydroxyl radical, or exonucleases and the location of the RNA chain terminus within the footprint vary substantially in different complexes examined to date (6,15–17, and references therein). Transcription complexes halted by selective nucleotide deprivation at several different sites within the initially transcribed 35 bp of two different transcription units exhibit discontinuous shifts in the RNA polymerase footprint on the DNA (6). For at least seven to eight rounds of nucleotide addition the downstream edge of the footprint (A, Fig. 1) remains relatively constant, while the upstream edge of the footprint (B, Fig. 1) moves in rough synchrony with the position of nucleotide addition. This results in a net shrinkage of the footprint from ~40 to ~30
Further elongation results in a large shift in the downstream edge of the footprint (~10 bp). Although additional studies are needed to verify that discontinuous translocation occurs well removed from the events of initiation, these findings are consistent with significant differences in the size and location of the footprint found for isolated trp and thr leader region paused transcription complexes, in which the nascent transcripts are >90 nt (15, 17).

Recently, D. N. Lee in our lab found a remarkably similar discontinuity in RNA polymerase movement after GreA-induced transcript cleavage of both the trpL paused transcription complex and a promoter-proximal halted complex. Although the transcription bubble and upstream DNA contacts (B, Fig. 1) relocate in approximate register with the new transcript 3' end, the downstream DNA contacts (A, Fig. 1) remain nearly unchanged when up to 10 nt is removed from the nascent transcript (D. N. Lee, G. Feng, and R. Landick, unpublished observations).

Transcription Bubble

The patterns of DNA base modification by single-strand-specific reagents in different halted elongation complexes suggest that movement of the transcription bubble also is not locked stepwise to nucleotide addition. In eight different complexes located more than 75 nt from the transcription start, the size of the bubble varied from no more than 14 nt in some complexes to at least 18 nt in others (8). Furthermore, the downstream edge of the bubble (C, Fig. 1) did not move in precise register with the transcript 3' end, but was separated from it by a distance of 0–4 nt. These variations in dimension, although significant, are smaller than those observed for the RNA polymerase footprint on the DNA.

Product Binding Site

Another indication of flexibility within the transcription complex is the recent observation that 8 nt can be added to the priming nucleotide after crosslinking the γ-phosphate to His-1237 in the β subunit of RNA polymerase (14). The most obvious explanations are (a) the RNA product binding site is filled by movement of the active site without significant movement at the 5' end or (b) the 3'-proximal nascent transcript loops out from the site of crosslinking for up to 8 nt before it must rearrange to allow translocation.

The action of transcript-cleavage factors GreA and GreB also suggests that the product binding site may be filled to a variable extent in different transcription complexes. Transcript cleavage rates vary widely among different complexes; further, for any given complex, GreA and GreB yield distinct patterns of cleavage (4). Typically, GreA-induced cleavage removes 2- to 3-nt fragments from the nascent transcript of halted elongation complexes, whereas GreB stimulates removal 2- to 10-nt fragments. Neither of these findings is easily reconciled with a view that the transcription complex maintains a constant structure during elongation.

The Translocation Cycle

Taken together these results suggest that different parts of the transcription complex may move with different step sizes. Whereas the active site (F, Fig. 1) must move on the DNA template in single-nucleotide steps, the downstream edge of the transcription bubble may move in steps of 4 nt, the product binding site by as much as 8 nt, and the downstream edge of the footprint in steps of 10 bp or more. Thus the transcription complex may be thought of as a molecular machine with multiple moving parts whose individual cycles are distinct yet coupled to each other in some as yet unknown fashion. For purposes of this chapter we define a complete translocation cycle as the interval between successive movements of the least frequently moving component of the transcription complex. The available data suggest this is the downstream RNA polymerase–DNA contact (A, Fig. 1) and that a translocation cycle occurs perhaps every 10 bp, within which shorter discontinuous movements of other transcription complex parts (substeps) may occur multiple times. As a hypothetical illustration of this idea, we depict in Fig. 2 how three rounds of ~4-nt jumps in the transcription bubble might be coupled to both nucleotide addition in a single active center and a 10- to 12-nt translocation of the downstream DNA contact. It is also possible that other parts of the transcription complex, such as the transcript exit channel, move discontinuously relative to the nucleic acid chains or to other domains of RNA polymerase. This type of dynamic transcription complex model was first described by Chamberlin, who now has proposed an explanation for how discontinuous transcription occurs without dissociation of the transcription complex (12). His model relies on two separate contacts to the upstream and downstream segments of DNA, two contacts to the RNA transcript, and alternate locking and sliding of the contacts to produce an inchworm-like movement.

Although this view differs from synchronous, single-step RNA chain elongation, it is similar to events that occur during transcription initiation, where significant movement of the complex occurs only after several rounds of polymerization (see Chapter 12). At the extreme, chain elongation can become completely uncoupled from translocation in certain initiation intermediates, producing transcriptional stuttering. Given this behavior of RNA polymerase during initiation, perhaps we should not be surprised to find equivalent complexity during elongation (6).
Discontinuous Translocation as an Explanation of Observed Heterogeneity in Biochemical Properties of Isolated Transcription Complexes

If translocation is discontinuous, the precise structure of the transcription complex will vary from one template position to the next. Therefore, one might then expect transcription complexes at different positions to possess different biochemical properties. In fact, such variability in behavior occurs not just at pause, arrest, and termination sites, but among most complexes that have been studied in any detail. Isolated transcription complexes show pronounced differences in stability (18) and significant variation in electrophoretic mobility (3). In general, the apparent substrate $K_s$ for both elongation and pyrophosphorolysis varies over at least a 500-fold range on natural DNA templates (19 and references therein). Although the dramatic heterogeneity in these rates could result from variations in sequence-dependent interactions in an otherwise static structure, a discontinuous model for transcription complex movement directly predicts it.

Implications of a Translocation Cycle with Multiple Intermediates

Several important questions arise from a conclusion that components of the transcription complex move in steps greater than one nucleotide. Although all of these questions require much further investigation, we attempt to delineate them here.

Determinants of Translocation

First, what determines the positions where discontinuous movements of transcription complex components occur? Are regular jumps begun when polymerase escapes the promoter and continued in uniform intervals, perhaps somehow metered by linkages internal to RNA polymerase? In the two sets of complexes halted along different transcriptional units studied by Krummel and Chamberlin, movement of the downstream RNA polymerase-DNA contact appears to occur at two distinct positions, between positions 23 and 25 on one template and between positions 28 and 34 on the other (6). Further, if the translocation jump constitutes a barrier to rapid elongation, as one might expect, periodic translocation would produce pausing at evenly spaced sites during transcript elongation. Although transcription clearly is discontinuous, no such simple, uniform pattern is apparent in studies to date. Finally, Metzger et al. found that a set of complexes halted by NTP deprivation 20–39 nt after initiation, but upstream from an identical sequence, all show essentially the same upstream and downstream boundaries to exonIII digestion, relative to the site of nucleotide addition (16). Thus, transcription complex movement appears more complex than a regular se-
ries of events begun during initiation. Apparently some feature of nucleic acid sequence determines when RNA polymerase moves on the DNA.

**Step Sizes**

Second, what are the step sizes for translocation and the various translocation substeps? Are they constant or do they vary from one movement to the next during elongation?

**Coupling of Translocation Substeps**

Third, if the size of translocation steps and substeps varies, are they coupled to each other or do they cycle semiindependently in response to different aspects of nucleic acid structure? If the linkage among different cycles is not tight, then we must ask the foregoing questions separately for each of the different components of the transcription complex.

**Translocation Phasing**

Are the movements of parts within the transcription complex completely determined by the sequence of DNA template and RNA transcript or are some movements random? If they are completely determined by interactions between sites on the polymerase and the DNA and RNA chains, then all RNA polymerase molecules should translocate identically along a transcriptional unit and remain completely in-phase with each other regardless of how much the step sizes vary. We define completely phased translocation to mean that all transcription complexes at a given template position will be at equivalent positions in the cycles of footprint and transcription bubble movement.

However, it is possible that translocation becomes unphased because the movement of one or more components of the transcription complex is not tightly coupled to the sequences of the DNA and RNA chains within it. If this occurs, not all polymerases will be in the same step of the translocation cycle when they encounter a given template position and complexes occupying a given template position will be in different conformational states. Unphased translocation predicts that multiple conformations of the transcription complex occur at a single template position.

**Possibility of Different Translocation Cycles**

We noted above two key questions for models of discontinuous translocation: (a) are substeps tightly coupled or somewhat independent and (b) are movements completely determined by the nucleic acid chains or do stochas-

...
A similar slow transition at a single template position was observed by D. Erie and P. H. von Hippel (20a) during studies of misincorporation. The rate of misincorporation is initially rapid when halted complexes are formed by selective NTP deprivation and then incubated with the wrong nucleoside triphosphate substrate. However, again with a half-time of several minutes, these complexes isomerize to a new conformation that misincorporates NTPs at a much slower rate.

A third case most easily explained by conformational heterogeneity of transcription complexes is phage T4 Ale protein-directed transcription termination. Ale selectively terminates transcription of bacterial DNA by E. coli RNA polymerase but does not affect the same enzyme molecules transcribing phage DNA containing 5-hydroxymethylcytosine. Kashlev and coworkers recently found that in vitro Ale directs termination at consensus sequences located frequently in the bacterial DNA, but only when chain elongation occurs at high NTP concentrations. Ale has no effect on RNA polymerase traversing the same sequences at low NTP concentrations (20b). This suggests that some aspect of transcription complex structure differs during rapid elongation. Perhaps Ale recognizes this conformation. Alternatively, the stability of the complex may be lower in this state, allowing Ale to effect termination or RNA polymerase may recognize the consensus sequence only at low levels of NTPs. When RNA polymerase is re-started from a halt site upstream from an Ale termination sequence, Ale directs termination only when the distance between the halt site and the termination sequence is greater than ~20 nt. This result could be explained by a slow conformational transition during elongation that is roughly the reverse of that observed for transcriptional arrest.

TRANSCRIPTIONAL PAUSING

How might a dynamic model of the transcription complex alter our thinking about transcriptional pausing? We will address this question after describing current knowledge about pausing, using the extensively studied his leader pause site as an archetype.

The his leader pause site is located immediately after the DNA segment that encodes the first significant secondary structure in his leader transcript (A:B, Fig. 3a), a position precisely analogous to pause sites found in the leader regions of several other amino acid biosynthetic operons regulated by attenuation (21). These pause sites are thought to halt transcription in the leader region until a ribosome loads on the nascent transcript, thereby synchronizing translation of the leader peptide coding region to transcription of the attenuator. A role for RNA secondary structure in pausing at these sites was apparent from several early findings (see Refs. 1,19,22, and references therein). These findings suggested that pausing might be triggered by...
formation of the complete his A:B RNA hairpin or homologous structures at the other pause sites. This would require base pairing to within 3 nt of the transcript 3' end and could block elongation either by disrupting an RNA:DNA hybrid or by removing the transcript from the active site.

Pausing also was observed at sites where potential for nascent transcript secondary structure was absent, leading to the idea that two classes of pause sites, hairpin-induced and sequence-dependent, existed (see 1,19, and references therein). However, recent findings reveal that pausing at both the his and trp leader pause sites has multiple determinants. Four pause site components are now apparent (22) (Fig. 3c): (a) a pause RNA hairpin that forms 10 or 11 nt upstream from the pause site, (b) the 3'-proximal region of transcript or DNA template in which RNA structure is not important, (c) the 3'-terminal nucleotide, and (d) the DNA sequence downstream from the site of pausing.

Pause Half-life Versus Pausing Efficiency

Two distinct characteristics determine pausing by RNA polymerase: (a) the pause half-life, defined as the time it takes half the paused complexes to escape the pause site (0.693/pseudo-first-order rate constant for nucleotide addition in the paused transcription complex) and (b) pausing efficiency, which is the fraction of transcribing RNA polymerase molecules that recognizes the pause site. Pause half-lives can be measured when the majority of elongating RNA polymerase molecules reach the site synchronously, so that the rate of escape from the pause site is not obscured by continuing arrival of additional complexes. Use of halted elongation complexes formed by nucleotide deprivation greatly facilitates the measurement of half-lives at downstream pause sites (19,22,23) (Fig. 3c).

However, at most sites pausing efficiency is less than 100% (19,22-24). As noted above, a simple explanation is that isomerization to a paused conformation competes with transcript elongation. Understanding of pausing efficiency is limited by the absence of a good method to measure it and most work to date, including that described below, has relied primarily on measurements of pause half-life, although several alternatives have been considered (23 and references therein).

The Pause RNA Hairpin

Examination of base substitutions throughout the region upstream from the his pause site revealed that substitutions at potential base pairs between -12 and -29 reduce pausing in direct proportion to their effect on the predicted stability of the upper portion of the A:B secondary structure (22). Substitutions further upstream, which would affect the 5' side of the lower portion of the A:B structure, had no effect on pausing. Substitutions between -1 and -11 either increase pausing, decrease pausing, or have no effect. Further, compensatory base substitutions that restore the potential for base pairing in the upper segment of A:B, but not elsewhere, return pausing to near wild-type levels. Thus, the true pause RNA hairpin is a 5-bp stem, 8-nt loop structure that forms 11 nt upstream from the transcript 3' end (Fig. 3c).

How does the pause hairpin influence transcription? Since the RNA hairpin does not form within the 3'-proximal region, it is unlikely to remove the transcript from the product binding site (G, Fig. 1). Two remaining possibilities are that hairpin formation blocks elongation by removing the transcript from an RNA-binding channel (1, Fig. 1) or that the hairpin interacts directly with RNA polymerase. These two possibilities are not mutually exclusive.

We favor a model in which the negatively charged phosphate backbone of the pause hairpin interacts electrostatically with a positively charged region of RNA polymerase. Such a charge-charge interaction would be independent of the sequence of the pause hairpin and could account for both the remarkable stability of the pause complex and the effect of the hairpin on pausing. The hairpin interaction may cause an allosteric change that slows elongation, blocks translocation, or shifts the alignment of the transcript 3' end in the active site.

This model can explain the effect of elevated salt concentrations on pausing. At most sites along the DNA template, transcript elongation is inhibited at Cl⁻ ion concentrations above 250 mM. In contrast, the rate of escape from the pause site is increased (half-life decreased). This effect is anion-specific since high concentrations of either glutamate or acetate similarly reduce both the overall elongation rate and the rate of pause site escape. One explanation is that Cl⁻ ions compete with the pause hairpin for binding to RNA polymerase, a view consistent with the Cl⁻ ion–dependent reduction in stability of hairpin-containing transcription complexes observed by Arndt and Chamberlin (18). This model predicts that when formation of the pause hairpin is inhibited, elevated Cl⁻ concentrations should not accelerate elongation at the pause site. In fact, multiple-base substitutions that eliminate base pairing in the his pause RNA hairpin drastically reduce the pause half-life, but no further reduction occurs in the presence of elevated Cl⁻ ions (C. L. Chan, unpublished observations).

The 3'-Proximal Region

Some base substitutions within 11 nt of the transcript 3' end affect pausing, even though they are outside the pause hairpin region. In principle, they could alter critical interactions between RNA polymerase and either the DNA template (E, Fig. 1), the nascent transcript (G, Fig. 1), or the nontranscribed strand alone or in combination. Two possibilities are that (a) an
RNA:DNA hybrid in the 3’-proximal region is important and base substitutions alter its stability or shape and (b) RNA polymerase contacts to the RNA and DNA strands (either as a hybrid or in separate sites) influence the configuration of the RNA or DNA bases in the active site, thus influencing the rate of nucleotide addition. No simple pattern of effects on pausing that might narrow these possibilities has emerged from the substitutions studied to date.

**Downstream DNA Sequence**

A third important determinant of pausing is the untranscribed DNA downstream from the pause site. Effects on pausing of changing this sequence have been observed at the his leader pause site, some early T7 DNA pause sites, at pause sites in a DNA copy of phage QB, and at the trp leader pause site (19,24,25, and references therein). A study of substitutions between 3 and 17 nt downstream of the trp leader pause site reveals that the effect extends to 14 nt downstream from the pause site and that both AT-rich and GC-rich downstream sequences reduce pausing. Thus, although the structural features of the downstream DNA sequence that influence elongation are unclear, they appear independent of the ease of unwinding the DNA helix. The DNA sequences downstream from the wild-type his and trp pause sites reveal no obvious potential for bending or similarity to sequences downstream from other pause sites.

**3’-Terminal Nucleotide**

At all six pause sites in leader regions of amino acid biosynthetic operons regulated by attenuation, the 3’-terminal nucleotide is a pyrimidine (U for trp, his, and leu; C for thr, ilvGMEDA, and ilvBN) and the next nucleotide to be added is a G. At the his pause site, changing the 3’ end of the nascent transcript alters both the position of pausing and its duration. This confirms that the 3’-terminal nucleotide is an important determinant of pausing. The preference for pausing at positions where GTP adds to pyrimidine transcript termini suggests that alignment of the two bases in the active site influences the spatial orientation of the 3’-hydroxyl nucleophile and α-phosphodiester acceptor and thereby affects the rate of phosphodiester bond formation.

**The Nontranscribed DNA Strand—A Key Determinant of Pausing in the Phage λ Late Operon**

A fifth determinant of pausing recently was identified at the site 16 nt downstream from the transcription start site for the phage λ late gene promoter by J. Roberts and coworkers. Q-protein interaction with the +16 paused complex allows antitermination during late gene transcription. Base substitutions at positions +2 and +6 greatly reduce pausing, but in heteroduplex templates only when the substitutions are present in the nontranscribed strand (B. Ring and J. Roberts, personal communication). Thus, it appears that interaction between RNA polymerase and the displaced DNA strand near the upstream edge of the single-stranded transcription bubble inhibits elongation. Although homologous sites in other lambdoid phages are up to 25 nt from the transcription start, pausing is not observed unless the sites are within the promoter-proximal region (26 and references therein).

Thus, an appealing explanation for the role of the nontranscribed strand is that it somehow resists reentry into duplex DNA, either by direct interaction with RNA polymerase or perhaps by inducing mistrreading of the strand at the template-reannealing site (D, Fig. 1). It remains to be determined if the nontemplate strand functions as part of the pause signal at other pause sites.

**Implications of Discontinuous Translocation for Pause Site Recognition**

If multiple translocation intermediates exist during elongation, several important and difficult questions arise about pause site recognition. Presumably, interactions between RNA polymerase and the different pause site components (RNA hairpin, downstream DNA sequence, etc.) would differ among different translocation intermediates. For instance, the intermediate in Fig. 2a might optimally bind the pause hairpin to a positively charged region of the enzyme, whereas the intermediate in Fig. 2f might be rearranged so that the positively charged residues are less available for hairpin interaction. If translocation cycle substeps are not coupled, the number of available states increases.

**How Does the Transcription Complex Achieve the Proper Conformation for Pause Site Recognition?**

If translocation is always phased, the key question for pausing is how the transcription complex arrives at the site in the optimal translocation state for pause signal recognition. Thus, altering the distance between the promoter site and a pause site could cause RNA polymerase to reach the site in the wrong translocation state. However, studies in our lab have used several templates with different promoters and distances to the his or trp pause sites and all exhibit roughly equivalent pausing behavior. Although a systematic study of the effect of pause site position relative to the transcription start or an upstream halt site should be made and might reveal effects that have so far gone undetected, the ~60-bp DNA fragment from 45 bp upstream from the site to 15 bp downstream appears sufficient to direct pause site recognition. If we assume transcription complexes may enter this DNA
segment in different translocation states, then its sequence must dictate the
movements of transcription complex components so the proper configura-
tion for signal recognition is achieved upon arrival at the pause site. This
requires that the translocation cycle be flexible and is consistent with the
conclusion that it is controlled by nucleic acid sequences and not by distance
from the transcription start (see above).

Unphased Translocation Could Reduce Pausing Efficiency:
 Might Pause Sites Be Phasing Signals?

If translocation can be unphased, then multiple transcription complex
conformations may advance to or exist at the pause site. Failure of some
conformations to recognize the pause site could explain less than 100% paus-
ing efficiency. However, at the wild-type his and trp pause sites, pausing is
at least 80% efficient. How could this be achieved if an unphased population
of transcription complexes is distributed among ten or more states? We can
envisage two possibilities. First, the population of translocation intermedi-
ates might become restricted as they pass over the upstream region of the
pause signal, so that most or all complexes reach the site in the state optimal
for isomerization to the paused conformation. Such “phasing” properties for
this region were inferred above from the ability of the pause signal to func-
tion in various locations and different transcriptional units. Second, a het-
erogeneous population of translocation states may reach the pause site, each
with a slightly different energetic route of isomerization to the paused con-
formation. Pausing efficiency then would be determined by the sum of in-
dividual partitions between isomerization to the paused conformation and
elongation past the pause site. At a strong pause signal such as the his and
trp leader sites, the energetic barrier to isomerization for most intermediates
must be lower than the barrier to elongation, ensuring that most complexes
isomerize.

If translocation can become unphased, then either view suggests that
pause signals require some “phasing” properties to work efficiently. If a
method for determining pausing efficiency can be devised and applied to the
many base substitutions available in the his pause signal, we might learn if
a particular pause signal component, such as the pause hairpin or the up-
stream region, phases translocation and ensures efficient pausing. One ap-
pealing idea is that the pause hairpin, by binding to a site on polymerase,
could phase transcription complexes by inhibiting translocation until the
product binding site is full.

A further possibility is that pausing might enforce phasing in certain seg-
ments of DNA. For instance, the pause sites in the his and trp leader regions
could phase translocation of the transcription complex to ensure that it en-
counters the attenuator in a termination-proficient state. This idea might be
tested by varying the distance between the pause and termination sites, but
it will be difficult to control for simultaneous effects on alternative transcript
folding.

TERMINATION

If translocation is discontinuous, then models for termination also require
revision. We will restrict discussion here to ρ-independent termination.
Pausing appears to be an integral component of ρ-dependent termination
(27), suggesting that the above-described considerations could also apply
to it.

ρ-Independent Terminators

Like pause signals, ρ-independent terminators appear to be multipartite.
Traditionally, two components have been considered: (a) a stable RNA hair-
pin with a G+C-rich stem and (b) a uridine-rich sequence, immediately af-
after the hairpin, in which termination occurs. Early models emphasized de-
stabilization of the transcription complex, begun by hairpin-induced partial
disruption of a 12-bp RNA:DNA hybrid and completed by dissociation of
the remaining weak rU-dA base pairs (for a more complete account, see
Refs. 2 and 10). More recently, Yager and von Hippel used a thermodynamic
accounting of the energy required to maintain the transcription bubble ver-
sus the residual stability of an RNA:DNA hybrid after hairpin formation to
predict the positions at which termination can occur (2), assuming that ther-
modynamic equilibrium of the complex is attained at each template position.
These workers then developed a competitive kinetic model to account for
termination efficiency at these sites and found that even small perturbations
in the activation barriers to elongation and termination can shift the
termination efficiency from near 0% to 100% (28). In the Yager–von Hippel
model, the positions of termination are determined by the energetic conse-
quences of altered base pairing, whereas termination efficiency may also
depend on changes in polymerase–nucleic acid contacts (28).

Chamberlin and coworkers defined two additional components of termi-
nation sites in which base substitutions could affect termination efficiency
(10,29, and references therein): (a) the DNA or RNA sequence upstream
from the terminator hairpin and (b) the DNA sequence downstream from the
release site. Swapping elements among different terminators altered termi-
nation efficiencies in unexpected fashion, explainable only if efficiency de-
pears on synergy among the different terminator components. They also
found that increasing concentrations of neutral salts (up to 1 M) increase
termination efficiency much more at some terminators than at others,
whereas decreasing Mg$^{2+}$ to less than 1 mM rendered all terminators nearly 100% efficient. From these findings, they argued that the mechanism of termination is (a) a multistep process, (b) affected by events that occur before RNA polymerase reaches the site of transcript release, and (c) involves direct catalysis of some steps by RNA polymerase (10).

**Terminator Hairpins**

Although there is universal agreement that the terminator hairpin is required for $\rho$-independent termination, how we think about its effect on the transcription complex depends on the size of the RNA:DNA hybrid in an elongation complex (see above). Thus, incomplete understanding of terminator hairpin action stems in part from our lack of knowledge about transcription complex structure.

Contrary to early suggestions that the terminator hairpin functioned to pause RNA polymerase at the terminator, pause and terminator hairpins probably affect the transcription complex differently. Neither the his nor trp pause RNA hairpin extends to within 10 nt of the transcript 3' end, whereas $\rho$-independent terminator hairpins invariably reach to within 7-9 nt of the release site (22 and references therein). Thus, a terminator hairpin may destabilize the transcription complex by disrupting key contacts in the complex that are unaffected by pause hairpin formation, perhaps within the transcript separation site (H, Fig. 1).

Do terminator hairpins make charge-charge contacts with RNA polymerase, as we suggested for pause hairpins? Increased termination at elevated Cl$^-$ ion concentrations (10,18) and low Mg$^{2+}$ ion concentrations (10) suggest such an interaction (Mg$^{2+}$ could mediate hairpin binding). Reduced terminator activity produced in vivo by multiple changes in the stem sequence of the $\lambda$ t22 terminator hairpin that should either maintain or increase its stability led Cheng et al. to suggest that the terminator hairpin could make sequence-specific contact with RNA polymerase (30). K. Wilson in the von Hippel lab observed similar effects with different substitutions at t22 in vitro, particularly at low NTP concentration (K. Wilson and P. H. von Hippel, personal communication). Conceivably these findings reflect changes in hairpin-RNA polymerase interaction caused by an altered hairpin shape. However, base-substitution experiments must be interpreted cautiously, since any change may affect interactions other than formation of the hairpin or its interaction with polymerase (such as alternative transcript folding or interactions between RNA polymerase and single-stranded transcript, DNA template, or nontranscribed DNA strand). Although many studies have shown that substitutions in the terminator hairpin stem reduce termination efficiency, no systematic study of the effect of compensatory substitutions at individual base pairs has been reported.

**Implications of Discontinuous Translocation for Termination and Antitermination**

As for pause site recognition (see above), it seems likely termination would be more efficient from some translocation intermediates than others and that the extent of translocation phasing might in part determine termination efficiency. If so, one or more components of the multipartite termination signal might act to rephase transcription complexes. Obvious candidates are the region upstream from the terminator hairpin identified by Reynolds et al. (10) and the terminator hairpin itself (see above).

The possibility that two or more types of translocation cycles could be related by allosteric transitions in the transcription complex (see above) raises interesting implications for the significant effect of reactions conditions (NTP or Mg$^{2+}$ concentrations, for instance) on termination efficiency and for the mechanism of antitermination generally. If termination occurs most efficiently from a particular translocation intermediate that is accessible only when large discontinuous jumps occur during transcription (e.g., that in Figs. 2a and 2g), an allosteric switch in the translocation cycle could exclude the termination-proficient conformation from the set of conformations available to the transcribing RNA polymerase. Such a switch could be triggered by effectors, transcription factors, or hysteretically. The converse might allow formation of a conformation that factors such as Alc or Nun recognize to direct termination. Although speculative, such a model deserves serious consideration if future work upholds the discontinuous translocation model of transcription.

**Transcriptional Arrest**

The third class of elongation control, transcriptional arrest, is by far the least well understood. Although arrest sites have been found in several transcriptional units in vitro (3,4,10,18), the nature of the signal is completely unknown. The possibility that arrest results from an inappropriate translocation step (4,12) offers an experimental avenue to investigate translocation. What makes arrest particularly interesting is the finding that transcript cleavage can reactivate arrested complexes.

A transcript cleavage activity that removes 2-10 nt from the 3' end of the nascent transcript in elongation complexes halted by NTP deprivation at certain template positions was discovered by Surratt et al. (11). After cleavage, the complexes resume elongation from the new transcript 3'-OH group upon addition of NTPs. Borukhov et al. (4) subsequently showed that transcript cleavage was greatly stimulated by either of two similar but distinct ~18-kDa polypeptides, GreA and GreB, which presumably contaminate purified RNA polymerase. However, even reconstitution of RNA polymerase...
from individually denatured and purified subunits does not completely eliminate the cleavage activity (4), suggesting that cleavage may be an intrinsic activity of RNA polymerase that is greatly stimulated by GreA or GreB. Either factor prevents formation of arrested complexes when added to in vitro transcription reactions, but only GreB efficiently stimulates transcript cleavage and reactivation of arrested complexes once they form (4). A similar process, mediated by TFIIIS, occurs at RNA polymerase II arrest sites (see Chapter 15). GreA has no effect on human RNA polymerase II complexes and, conversely, human SII does not induce cleavage in E. coli complexes (R. Keene and Robert Landick, unpublished); there is no sequence similarity between SII and the E. coli cleavage factors.

G. Feng in our lab has overproduced and purified GreA. Studies with the purified factor show the GreA-mediated cleavage is accompanied by backward translocation of the RNA polymerase on the DNA template (see above). Significant blocks to further cleavage were found after removal of 10 nt from the trp paused transcription complexes and 6 nt from complexes halted 16 nt from the transcription start. However, elongation complexes halted ≥ 140 nt after initiation undergo successive rounds of GreA-stimulated cleavage and back up at least to position +50. This creates a discontinuous ladder of transcripts that is unrelated to stop sites observed during elongation (D. N. Lee, G. Feng, and R. Landick, unpublished observations). During in vitro transcription, recombinant GreA suppresses arrest at some sites, but strongly stimulates pausing at other sites, perhaps by very rapid transcript cleavage in certain transcription intermediates. Remarkably, GreA has no effect on pausing in either the his or trp leader regions or on termination at several different ρ-independent terminators (G. Feng, D. N. Lee, C. L. Chan, and R. Landick, unpublished observations).

CONCLUSIONS AND FUTURE DIRECTIONS

Our consideration of the transcription complex as a machine with multiple moving parts and review of the mechanisms that control transcript elongation lead to the following conclusions:

1. The transcription complex may contain several different components (e.g., polymerase-DNA contacts, DNA strand separation and reannealing sites, product binding sites; Fig. 1) that move in steps greater than single-nucleotide addition during elongation. Movement of these parts may either be directly coupled to each other or occur semiindependently.

2. A transcription complex with parts that move in steps greater than one nucleotide predicts that different conformations exist at different template positions and could account for wide variations in the biochemical properties of transcription complexes, such as apparent substrate Kₐ, stability, and transcript cleavage.

3. A transcription elongation complex with parts that move in greater than single nucleotide steps may display either phased or unphased movement on the DNA template.

4. Variations in the degree of phasing would allow some transcription complexes to encounter pause and termination sites in suboptimal conformations for pausing or termination to occur, thus affecting the response to these elongation control signals.

5. If unphased translocation can occur, we might expect that phasing signals exist in transcription units to ensure that regulatory sites are recognized properly. Transcriptional pause sites are candidates for phasing signals.

6. Both pause and termination sites are multipartite and can include an RNA hairpin element in the nascent transcript, a 3'proximal region of nascent transcript or DNA, the 3'-terminal nucleotide, the DNA immediately downstream from the site, and the nontranscribed DNA strand. It seems likely that different combinations of these components may create different types of pause and termination signals.

Several approaches for future studies deserve mention. Genetic analysis of RNA polymerase subunits has confirmed that several regions conserved between prokaryotic and eukaryotic homologs are important for transcript elongation (31). Further studies of amino acid substitutions in these regions, both biochemically and by isolation of second-site revertants, may reveal if particular genetically defined regions correspond to any of the functional sites depicted in Fig. 1. The definition of multiple components for pause and termination sites allows us to ask if their roles in pausing and termination are similar or different by swapping them between pause and termination sites. Comparable definition of arrest signal components also should be revealing. The recent demonstration by Daube and von Hippel (32) that active elongation complexes can be assembled directly from RNA polymerase and synthetic RNA-DNA bubble duplexes now makes it possible to determine how different types and configurations of RNA and DNA chains affect the final structure and activity of reconstituted complexes. Finally, recent progress in isolating milligram quantities of defined transcription complexes may allow direct electron microscopic study of two-dimensional crystals of transcription complexes and better insight into the structural features of the complex such as the presence or absence of a significant RNA:DNA hybrid (S. Darst and R. Landick, unpublished studies).

In this chapter we have explored ideas provoked by numerous unexpected findings but that currently lack widespread acceptance. We believe this is a useful exercise and will help define experiments that can test these ideas critically. However, we also believe it appropriate to conclude with a caution against accepting new models as established rather than using them as hypotheses that can be tested experimentally to improve our understanding of RNA chain elongation and termination.
We thank M. Chamberlin, J. Roberts, P. von Hippel, and members of our lab for helpful suggestions during preparation of this chapter and M. Chamberlin, J. Roberts, P. von Hippel, and K. Wilson for permission to quote unpublished work. In particular, we acknowledge M. Chamberlin and co-workers both for stimulating a reconsideration of transcription complex models and for first formulating many of the ideas discussed herein. We apologize to all our colleagues for the many instances in which space limitations prevented proper citation of their work.

REFERENCES


