An Integrated Model of the Transcription Complex in Elongation, Termination, and Editing

Peter H. von Hippel

Recent findings now allow the development of an integrated model of the thermodynamic, kinetic, and structural properties of the transcription complex in the elongation, termination, and editing phases of transcript formation. This model provides an operational framework for placing known facts and can be extended and modified to incorporate new advances. The most complete information about transcriptional mechanisms and their control continues to come from the *Escherichia coli* system, upon which most of the explicit descriptions provided here are based. The transcriptional machinery of higher organisms, despite its greater inherent complexity, appears to use many of the same general principles. Thus, the lessons of *E. coli* continue to be relevant.

The central event in transcription is the RNA polymerase–catalyzed "copying" of the sequence of the template strand of a gene into a complementary RNA transcript. This transcript may serve as a message for translation into protein, it may comprise structural RNA that forms the framework of a ribosome or of a transfer RNA "adaptor" molecule in protein synthesis, it may form the genome of an RNA virus, or it may itself serve a regulatory function. The chemistry of transcript formation is straightforward, but the regulatory mechanisms that have been developed by evolving organisms to control this synthetic process appear almost infinite in number, although the basic principles on which they operate are likely to be relatively few. This complexity may reflect the fact that transcription is the primary event of gene expression, which is defined at the molecular level as the transformation of genes (or operons) into the functional proteins and enzymes that direct and catalyze the events of cellular metabolism and differentiation. As a consequence, transcription comprises the first, and thus the most effective, level at which the "reading of the DNA cookbook" can be regulated.

Control at the Gene Level

Formation of a transcript has traditionally been divided into three sequential stages, called initiation, elongation, and termination. All are subject to regulatory control. It may now be more appropriate to divide the overall process into two major phases: (i) activation and transcript initiation and (ii) transcript elongation, with the latter including termination (transcript release) and editing (transcript shortening and resynthesis with increased fidelity), because both can be viewed as alternative pathways that, in principle, compete with elongation at every template position. At the activation-initiation stage, transcription control processes regulate intergene (or interoperon) competition for a limiting amount of RNA polymerase. Each gene, to be competitive, is activated by multiple factors that increase the relative ability of its promoter to recognize and bind polymerase and then to facilitate promoter opening, transcript initiation, and promoter clearance. These activation procedures, sometimes coupled with specific repression events, involve interaction of the promoter-bound polymerase with protein subassemblies that may bind at adjacent or more distant DNA sites and be brought to the promoter by controlled DNA looping (*1*, *2*).

The activation-initiation process is complete when the nascent transcript becomes sufficiently long to stabilize the transcription complex against dissociation from the DNA template. Conformational changes permit the "core" RNA polymerase (defined as the basic enzymatic unit required for the template-directed synthesis of the transcript; the $\alpha_2\beta\beta'$ complex in *Escherichia coli*) to free itself from most of the factors and regulatory subassemblies involved in the activation-initiation process (*1*, *3*). These components are either released into solution or left behind at the promoter, and the core polymerase moves into the elongation phase, which is characterized by multiple and specific pausing events. The duration of these pauses depends on (i) the specific DNA sequence being transcribed, (ii) interactions with regulatory proteins that either bind directly to the elongating transcription complex or are brought to it from protein-binding sites located upstream on the nascent transcript (below), (iii) the concentration of the next required (by the template) nucleotide triphosphate (NTP), and (iv) whether a misincorporation event has occurred at the 3′ end of the chain. These interactions may also change the stability of the complex and thus make dissociation of the polymerase from the DNA template, with concomitant release of the nascent transcript, either more or less likely (*4*).

The final step in transcript formation is termination, which occurs when the elongating transcription complex moves into (or, in some cases, beyond) one or more terminator sequences along the DNA template that may serve as transcription regulators within genes or mark the end of a gene or operon. In *E. coli*, terminators are either "intrinsic," meaning that release of a nascent transcript can be brought about at these terminators without the involvement of protein factors, or "rho-dependent," in that release requires the participation of the *E. coli* transcription termination factor, rho. The efficiency with which transcripts are released at either type of terminator is generally regulated by additional factors (*1*, *4*).

Control at a Specific Template Position

After transcript synthesis has been initiated, the transcription complex at any given template position can, in principle, proceed by several alternative reaction pathways (Fig. 1). Each of these potential pathways is characterized by a particular reaction rate, which may vary from one template position to the next. These alternative reactions are in kinetic competition with one another (*5*); thus, changes in relative rates can "switch" the transcription complex between pathways.

This kinetic competition between potential reaction pathways can be quantified by formulating it in terms of free energy of activation barriers for the competing reactions (*6*). Differences in barrier heights can

The author is at the Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, OR 97403, USA. E-mail: petevh@molbio.uoregon.edu

be used to predict the rate or stability changes within the transcription complex that are required to bring about any particular regulatory consequence. This approach for the competition between elongation and termination at a given template position is shown in Fig. 2. Each pathway consists of a series of steps, and thus the activation barriers shown are oversimplified and represent only the rate-limiting step for each reaction under the conditions of the experiment. Changing conditions can change the relative heights of these barriers, and Fig. 2C shows the termination efficiency (TE; defined as the fraction of the total transcripts reaching a given template position that terminate there) as a function of the difference in barrier heights ($\Delta\Delta G^{\ddagger}$) for this two-pathway competition. Clearly, the transition from domination of the overall reaction by elongation to domination by termination is very abrupt with respect to $\Delta\Delta G^{\ddagger}$, meaning that a very small change in the relative rates of the competing processes, under conditions where the activation free energy barriers are of comparable height (at terminator positions; Fig. 2B), can result in an effective switch from one reaction pathway to the other. In contrast, an equivalent change under conditions of substantially different barrier heights (at elongation positions; Fig. 2B) has virtually no effect on TE (Fig. 2C), meaning that the rate of the elongation process and the stability (and thus the dissociation rate) of the elongation complex can be regulated over a wide range at these positions without risk of transcript release. This quantitative formulation of the kinetic competition model will be exploited in describing transcriptional control mechanisms as regulatable switches in what follows.

Stability of the Elongation Complex

A crucial mechanistic feature of the transcription elongation complex is its extreme stability. Thus, elongation complexes can be halted at elongation positions (most easily by omitting the next required NTP from the transcription mix) and can remain bound to the DNA template and the nascent transcript for long periods without dissociation (*7*). This stability is dynamic, as well as static, because the complex also cannot dissociate while moving from one template position to the next. Control of the "processivity" of the elongation process, defined in terms of the relative propensities of the polymerase to extend the transcript or to dissociate, is central to the regulation of all polymerases. RNA polymerases control processivity by forming elongation complexes that are massively stable, perhaps in part as a consequence of the "closure" of structural components of the polymerase around the DNA and the nascent transcript to limit dissociation (*8*, *9*). The problem then becomes one of bringing about sequence-specific destabilization of the elongation complex when terminator sites are reached, so that the nascent transcript (and the core polymerase) can be released at these template positions.

A Structural Model of the Elongation Complex

The two mechanisms used by *E. coli* to achieve the destabilization necessary for termination are most easily described in the context of a overall structural model of the protein and nucleic acid components of the elongation complex (Fig. 3). A central feature is a transiently open "transcription bubble" [~18 base pairs (bp) in length], which moves with the core RNA polymerase through the otherwise double-stranded DNA while the polymerase catalyzes template-directed transcript elongation. The moving polymerase protects a "footprint" of about 30 bp along the DNA against nuclease digestion, and this footprint includes the transcription bubble as well as some double-stranded DNA on either side. Chemical footprinting of the transcription bubble, as well as direct binding measure-

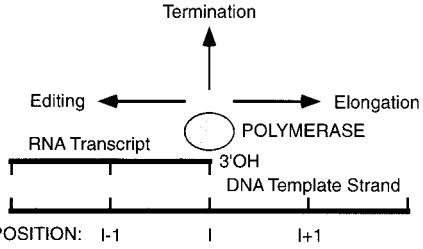

Fig. 1. Three competing pathways for the transcription complex at a given template position. The options are (i) to move forward along the template with the concomitant elongation of the RNA chain by one nucleotide residue, (ii) to move backward along the template (with or without shortening the transcript by one or more residues), or (iii) to dissociate from the template, resulting in release of the nascent RNA.

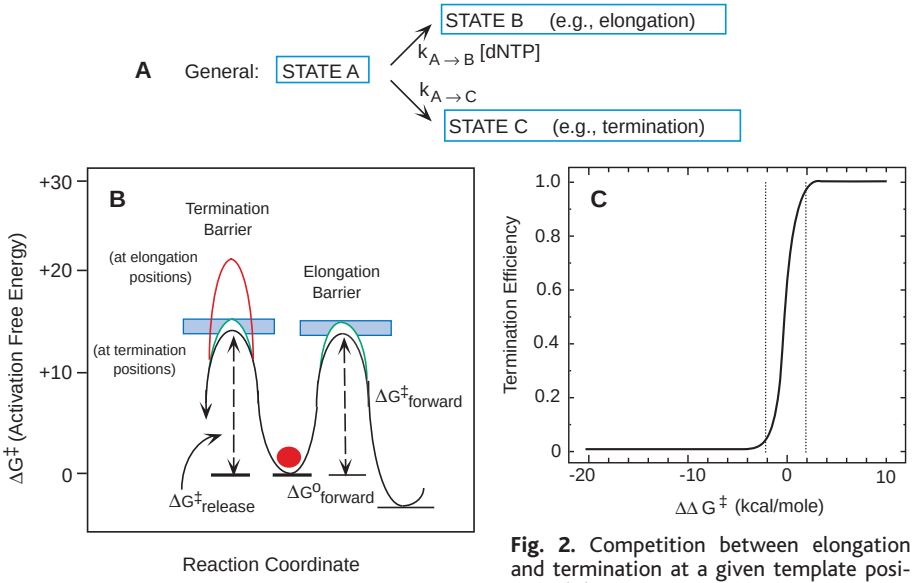

Fig. 2. Competition between elongation and termination at a given template position. (**A**) General formulation of the rate competition process in terms of apparent reaction rate constants. The rate of the transcript elongation process (state A to state B) may be dependent on the concentration of the next required NTP (as shown), or it may be pseudo–first order in the presence of excess NTP. Both rates may also depend on the concentrations of regulatory factors if these are present at less than saturating amounts. $k_{A\rightarrow B}$ and $k_{A\rightarrow C}$ are the rate constants for going from state A to states B and C, respectively. dNTP, deoxynucleotide triphosphate. (**B**) The elongation versus termination rate competition is described as free energy of activation barriers at terminator positions (barriers of about equal height) and elongation positions [termination barrier (red curve) much higher than elongation barrier]. The red circle represents the transcription complex at the ground state. $\Delta G^{\ddagger}_{forward}$ and $\Delta G^{\ddagger}_{release}$ indicate the heights of the free energy of activation barriers to termination or elongation, respectively, at termination positions, and $\Delta G^{o}_{forward}$ represents the standard free energy change of the elongation reaction. The blue bars correspond to the range of differences in barrier heights over which the termination efficiency (TE) ranges from 0.01 (higher termination peak within the blue bar versus the lower elongation peak) to 0.99 (lower termination peak within the blue bar versus the higher elongation peak). (**C**) TE, defined as $k_{term}/(k_{elong} + k_{term})$, where k_{elong} and k_{term} are the rate constants for the single nucleotide elongation and termination reactions, respectively, is plotted here as a function of the difference in the heights of the competing free energy of activation barriers ($\Delta\Delta G^{\ddagger}$) according to the following equation: $TE = [1 + \exp(\Delta\Delta G^{\ddagger}/RT)]^{-1}$, where R is the gas constant and T is absolute temperature. The region of the graph between the two vertical lines corresponds to the blue bars in (B).

ments, suggest that specific sections (Fig. 3) of both the transiently single-stranded and the double-stranded DNA, as well as portions of the RNA-DNA hybrid and a few RNA residues at the 5′ end of the hybrid, make stabilizing interactions (or at least contact) with the polymerase (*10*). The template side of the transcription bubble is hybridized to the 3′ end of the nascent RNA, with the terminal 3′-OH located very close to the downstream edge of the bubble (*11*). Current estimates of the average length of the transient RNA-DNA hybrid range from 9 to 12 bp (*8*, *9*, *12*–*14*). Upstream of the hybrid, the polymerase actively displaces the RNA from the template strand of the bubble (*15*), sending the nascent transcript across a polymerase-binding site for single-stranded RNA (*8*, *9*) and then on into solution, which permits the upstream end of the DNA bubble to reclose.

The catalytic site of the polymerase (inset, Fig. 3) contains both a substrate-binding subsite, at which the incoming NTP is bound to the polymerase and to the complementary nucleotide residue of the template, and a product-binding subsite, at which the 3′ terminus of the growing RNA chain is positioned (*16*). Elongation of the chain by a single nucleotide residue results in phosphodiester bond formation between the NTP bound at the substrate-binding subsite and the 3′ end of the RNA chain bound at the product-binding subsite. This positions the 3′ end of the newly extended chain in the substrate-binding subsite. Completion of the single nucleotide addition cycle thus requires that the chain be released from the substrate subsite, accompanied by a shift of the active site of the polymerase by one position along the template. As a consequence, the RNA-DNA hybrid becomes 1 bp longer at the downstream end, the separation mechanism operates to make the hybrid 1 bp shorter at the upstream end, and the polymerase moves along the template by one position while the hybrid retains a constant length.

A Thermodynamic Model of the Elongation Complex

A structural model of this general sort (*17*) has been used to formulate a quantitative hypothesis for intrinsic termination (*18*) that has withstood the tests of time and experimentation with reasonable success. This hypothesis, which is generally called the "thermodynamic model," posits that the interactions between the protein and nucleic acid components within an elongation complex rearrange rapidly as the complex moves from one template position to the next, meaning that the complex can be considered to be at equilibrium during its "dwell time" at each template position. If this is the case, the free energy of formation [from free core polymerase and a closed (double-stranded) DNA genome] of a stable elongation complex at a particular template position can be written as

$$\Delta G^{o}_{complex} = \Delta G^{o}_{DNA\text{-}DNA} + \Delta G^{o}_{RNA\text{-}DNA} + \Delta G^{o}_{NA\text{-}polymerase} \quad (1)$$

where $\Delta G^{o}_{complex}$ is the net free energy that stabilizes the elongation complex against dissociation, $\Delta G^{o}_{DNA\text{-}DNA}$ is the (unfavorable for elongation complex formation) free energy of opening the DNA-DNA base pairs of the transcription bubble, $\Delta G^{o}_{RNA\text{-}DNA}$ is the (smaller but favorable) free energy of forming the base pairs of the RNA-DNA hybrid within the unpaired transcription bubble, and $\Delta G^{o}_{NA\text{-}polymerase}$ is the (also favorable) net free energy of interaction of the polymerase with the various parts of the nucleic acid (NA) framework of the complex.

The magnitudes of two of these thermodynamic terms ($\Delta G^{o}_{DNA\text{-}DNA}$ and $\Delta G^{o}_{RNA\text{-}DNA}$) can be calculated from known nucleic acid stability basis sets (*18*) at any particular template position, if we assume equilibrium and the above (or any other structurally defined) model for the nucleic acid components of the complex at a given template position. Because the net free energy of complex formation ($\Delta G^{o}_{complex}$) can be estimated independently (*18*), the net (favorable) magnitude of the multipartite $\Delta G^{o}_{NA\text{-}polymerase}$ term can be calculated by difference. The experimentally demonstrated stability (with respect to dissociation) of the transcription complex at elongation positions along the template indicates that the unfavorable $\Delta G^{o}_{DNA\text{-}DNA}$ (transcription bubble formation) term is more than offset by the sum of the favorable $\Delta G^{o}_{RNA\text{-}DNA}$ and $\Delta G^{o}_{NA\text{-}polymerase}$ terms.

Destabilization Is Required for Termination

To make the termination pathway kinetically accessible (Fig. 2, B and C), any model that purports to explain intrinsic termination must provide a mechanism for the massive destabilization of the transcription complex that occurs as it moves through these terminators along the DNA template (*19*). The extent to which termination actually does occur at any particular template position within an intrinsic terminator depends on the interactions of the transcription complex with the local DNA sequence and with factors bound to the polymerase and may also be regulated by decreasing the concentration of the next required NTP. Complexes that emerge from the terminator sequence without having dissociated resume their original stability characteristics (*19*).

Intrinsic terminators, which represent about one-half of the termination sites of *E. coli*, are defined by a template DNA sequence that codes for a stable termination hairpin in the nascent RNA, followed by a run of ribouridylate (rU) residues at the 3′ terminus of the transcript (*17*, *19*). A plausible quantitative model (*18*) for intrinsic termination has been developed on the basis that the RNA-DNA hybrid is substantially destabilized as the complex passes through an intrinsic terminator sequence, thus also destabilizing the elongation complex by reducing the (favorable) contribution of the $\Delta G^{o}_{RNA\text{-}DNA}$ term to $\Delta G^{o}_{complex}$. The run of deoxyadenylate (dA) · rU base pairs that is positioned at the downstream end of the hybrid when the complex is located at template positions within an intrinsic terminator (Fig. 3, region I) is particularly unstable relative to its DNA-DNA or RNA-RNA cognates (*20*). In addition, the formation (in competition with the upstream portion of the RNA-DNA hybrid; Fig. 3, region

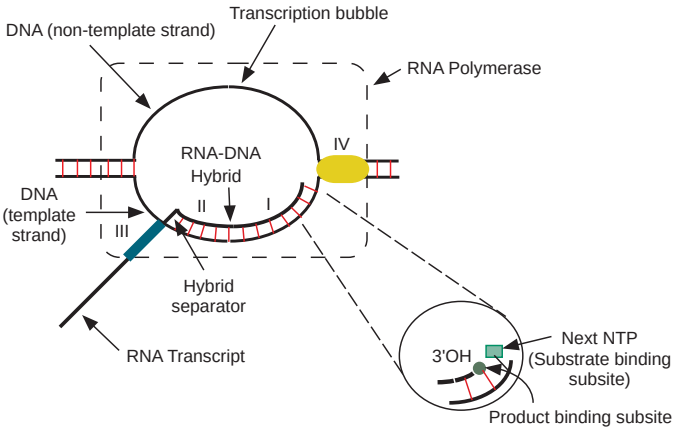

Fig. 3. A schematic representation of essential structural features of the elongation complex. The downstream portion of the RNA-DNA hybrid (I) is the region in which the rU · dA sequence is positioned at intrinsic terminators (see text). The upstream portion of the hybrid (II), as well as the RNA sequence at III, indicate the position of the RNA chain within which the termination hairpin forms at intrinsic terminators. The single-stranded RNA-binding site of the polymerase is located at III, and IV indicates the position of the "thumb" that closes around the double-stranded DNA and may be involved in maintaining the processivity of the polymerase (*21*). Polymerase contacts (and perhaps interactions), defined by footprinting reactions, also occur at other positions along the nontemplate strand of the transcription bubble and elsewhere. The inset shows the active (catalytic) site of the polymerase at the downstream end of the transcription complex, including the substrate-binding subsite for the next NTP (green square) and the product-binding subsite for the 3′ end of the nascent transcript (black circle).

II) of the stable terminator hairpin within the RNA at intrinsic terminators also destabilizes the RNA-DNA hybrid [see also (*21*)]. The sum of these sequence-dependent destabilization events suffices to lower the height of the activation barrier to termination to about the same level as the barrier to elongation (Fig. 2B), primarily by decreasing the (favorable) magnitude of the $\Delta G^{o}_{\text{RNA-DNA}}$ term of Eq. 1. This general model is consistent with the results of experiments on appropriately mutated intrinsic terminators (*19*).

Rho-dependent terminators, which are responsible for the other half of the transcription termination events in *E. coli*, do not share the destabilizing sequence features of intrinsic terminators. Instead, the required destabilization of the transcription complex at these terminators is provided by the RNA-DNA helicase activity of the rho protein (*22*). There are two important sequence requirements for rho-dependent terminators. First, the (rather variable) DNA sequences at the terminators must produce substantial pausing of the transcription complex within these sites. In addition, DNA sequences that code for an extended (~70 nucleotides) site along the transcript that is essentially devoid of RNA secondary structure must be present upstream of these terminators. This unstructured RNA sequence serves as a binding site that permits the hexameric rho helicase to be "loaded onto" the transcript, thus activating the cryptic RNA-dependent adenosine triphosphatase of the protein to provide the chemical free energy needed to drive the bound rho processively and directionally (5′ → 3′) along the nascent RNA. When the translocating rho helicase "catches up with" a paused transcription complex at a rho-dependent terminator, it triggers the release of the nascent transcript, presumably by separating the RNA-DNA hybrid within the transcription bubble, thus again destabilizing the complex sufficiently to permit termination (*22*).

"Fine-Tuning" of Termination Efficiency by Regulatory Factors

A gross destabilization of transcription complexes at intrinsic or rho-dependent terminators is required to make the heights of the competitive free energy of activation barriers for elongation and termination comparable, thus making termination possible. The actual termination efficiencies that are observed at individual template positions within terminators are often fine-tuned by termination or antitermination factors that bind to the transcription complex, either directly or as a consequence of RNA looping (*23*). These factors can function by altering the rate at which the transcribing elongation complex moves along the template (raising or lowering the height of the elongation barrier), by changing the rate of dissociation of the transcription complex (raising or lowering the height of the termination barrier), or by a combination of both mechanisms (*4*, *6*). As Fig. 2C shows, only relatively small changes in these parameters are required to "adjust" TE across its entire range.

Translocation of the Transcription Complex Along the Template

In addition to describing elongation and termination in the context of the equilibrium properties of the transcription complex at a particular template position, we must also consider the kinetic mechanisms responsible for the translocation of the complex along the template. How does forward translocation occur within the single nucleotide addition cycle (*24*)? There is no direct evidence with respect to RNA polymerases on this point, but a minimal proposal can be put forward that is compatible with ideas generated for DNA polymerases (*24*) and can also accommodate the backward movement of the transcription complex (below). This proposal suggests that phosphodiester bond formation [together with inorganic pyrophosphate (PP_i) release; Eq. 2] "triggers" the release of the 3′ terminus of the newly extended RNA from the substrate-binding subsite of the polymerase (Fig. 3). This puts the polymerase into a "sliding mode" (Fig. 4 and below) relative to the nucleic acid framework of the transcription complex, permitting the next required NTP to bind to both the substrate-binding subsite of the polymerase and, by complementary hydrogen-bonding, the next position on the DNA template. This dual binding of the next NTP serves to "relock" the polymerase to the nucleic acid framework, with the 3′ terminus of the extended RNA chain now positioned in the product-binding subsite of the polymerase in preparation for the next phosphodiester bond formation event (*25*).

Regardless of detailed mechanism, this repeating single-nucleotide addition cycle drives the polymerase directionally along the template DNA in much the same fashion as a cytoplasmic "molecular motor" protein (dynein, kinesin, or myosin) moves directionally along a microtubule or actin filament or a nucleic acid helicase moves directionally along DNA or RNA (*26*). Of course, the movement of an elongating polymerase differs from these other molecular motors in that polymerase translocation results in the concomitant extension of the nascent transcript, whereas cytoplasmic motors and helicases generate only heat, adenosine diphosphate, and inorganic phosphate to mark their passage. The elongation-dependent movement of RNA polymerase along the template has recently been observed directly in single-molecule experiments with a laser-trapping procedure, and the force generated by this mo-

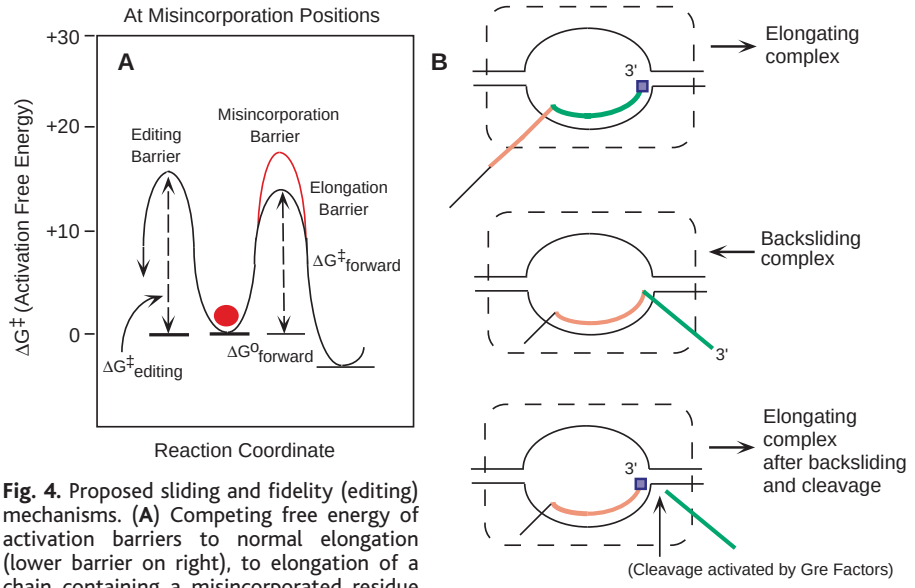

Fig. 4. Proposed sliding and fidelity (editing) mechanisms. (**A**) Competing free energy of activation barriers to normal elongation (lower barrier on right), to elongation of a chain containing a misincorporated residue (higher barrier on right), and to the editing (backward) reaction. Under normal synthesis conditions, the forward pathway is favored, whereas under misincorporation conditions the backward pathway is favored. The red circle represents the transcription complex in the ground state. (**B**) (Top) The transcription complex in the elongation mode. The square marked 3′ represents the substrate-binding subsite containing the next required NTP. (Middle) The transcription complex in the backsliding mode. The green transcript sequence shown within the RNA-DNA hybrid in the top panel has been "extruded" from the front of the complex by "backsliding" (see text). (Bottom) The transcription complex again in the elongation mode after backsliding and Gre factor–activated cleavage of the section of transcript extruded in the middle panel. The polymerase and the transcription bubble move along the transcript RNA and through the double-stranded DNA in register, with the RNA-DNA hybrid "rolling" (or "zippering") along the transcript while maintaining base pair complementarity and constant hybrid length (see text).

lecular motor has been measured (*27*).

The chemical reaction that underlies transcript elongation may be written as

$$RNA_n + NTP \Leftrightarrow RNA_{n+1} + PP_i \quad (2)$$

with incorporation of the next NTP extending the RNA chain by one residue and releasing one molecule of PP_i. This chemical reaction is readily reversible, with a measured equilibrium constant ($K_a = [PP_i]/[NTP]$) ≈100 for template-directed RNA synthesis catalyzed by *E. coli* RNA polymerase (*28*). Although the free energy balance in the presence of 1 mM concentrations of NTPs and PP_i [the estimated physiological concentrations of these species (*28*)] favors elongation, the chemical reaction can be driven backward in the presence of excess of pyrophosphate, resulting in both the shortening the nascent transcript and the movement of the transcription complex backward along the template (Fig. 1) (*28*, *29*).

RNA polymerase can also move backward along the template by a process that is akin to that driven by the 3′ → 5′ exonuclease editing reaction used by DNA polymerases (*30*). Misincorporation at the 3′ terminus of an RNA chain is favored by omitting the next required NTP. This greatly slows the elongation reaction, because both the addition of an incorrect nucleotide residue and the subsequent extension of the RNA chain beyond such a misincorporated residue are slow relative to normal elongation (*31*). During such "enforced" pausing at misincorporation sites, the transcription complex can go into an "unactivated" state from which elongation is not readily resumed when the next required NTP is again added. However, this recalcitrant state can be overcome and elongation resumed, if factors that activate transcript editing are added (*31*, *32*).

Stalled transcription complexes occasionally appear to suffer "spontaneous" transcript cleavage reactions in which a 3′ terminal oligonucleotide of variable length is removed from the transcript, followed by resumption of elongation from the newly created 3′ end of the shortened RNA (*33*). Subsequent work has shown that transcription factors GreA and GreB in *E. coli,* and the equivalent factor SII in eukaryotes, activate RNA polymerase to cleave a nascent transcript at variable (1 to 11 nt) distances back along the chain (*12*, *32*, *34*), suggesting that these processes could account for such spontaneous chain cleavage. What is likely to be happening in such paused or "arrested" complexes is that the 3′ end of the chain is occasionally released from the product-binding subsite of the polymerase, presumably by much the same mechanism that applies under active elongation conditions. At this point, if a terminal misincorporation has occurred or if (perhaps at an "arrest site") the polymerase has adopted an altered conformation that does not immediately permit the next required NTP to "relock" the polymerase at the next template position, the complex may continue to slide backward along the template DNA in a one-dimensional diffusion (random walk) process. This sliding is accompanied by the extrusion of the 3′ end of the nascent RNA from the transcription complex, and when GreA or GreB (or SII) factors are available to activate chain cleavage, a new 3′ terminus is formed that can rebind the shortened transcript to the polymerase at the appropriate template position and permit renewed synthesis in the presence of the next required NTP (Fig. 4) (*12*, *34*, *35*).

This backward sliding pathway results in the editing of the nascent RNA and an increase in the fidelity of transcription (*35*). Transcripts that cannot be extended, either because the next required NTP is missing from the transcription mix or as a consequence of the misincorporation of an incorrect terminal nucleotide residue, are "stalled" on the template. As a result, the probability of release of the 3′ terminus from the polymerase active site at such template positions is increased, favoring the "backsliding" mode of the polymerase with subsequent GreA- or GreB-dependent chain cleavage and release of the 3′-oligonucleotide containing the "incorrect" base, followed by resynthesis. In contrast, chain shortening by pyrophosphorolysis involves the direct reversal of the normal elongation mechanism (Eq. 2). Because a misincorporation event that inhibits elongation would also inhibit transcript shortening by this means, pyrophosphorolysis should not result in effective RNA editing.

How Does Backward Sliding Occur?

The definitive answer to this question is not known, but a plausible mechanism can be formulated in the context of the structural model for the transcription complex presented in Fig. 3. The sliding process appears to be diffusional in nature, in that it does not require the hydrolysis of adenosine triphosphate or other sources of chemical free energy. This suggests that the interactions that hold the template and the nontemplate DNA strands of the transcription bubble and the single-stranded section of the nascent RNA to the polymerase are likely to be non–sequence-specific (electrostatic?) in nature, permitting the polymerase to diffuse along these strands as along an "isopotential surface," with the upstream end of the bubble opening and the downstream end closing to maintain constant bubble size as the complex slides backward and the reverse process occurring when the complex again slides forward (*12*, *34–36*).

Figure 3 suggests that the polymerase is held to the nucleic acid framework of the transcription complex not only by non–sequence-specific protein-DNA and protein-RNA interactions that can permit sliding as described above but also by complementary Watson-Crick hydrogen-bonding between the nascent RNA and the template DNA in the transient RNA-DNA hybrid. This latter specific and complementary interaction certainly cannot slide in the same way. However, if the normal processes that maintain the hybrid at a constant (9 to 12 bp) length in elongation also operate in the backward sliding process, then the hybrid can essentially "roll" along the DNA template, shifting the sequence that is base-paired within the hybrid backward or forward along the template and the transcript in concert with the one-dimensional diffusion of the polymerase. Thus, the maintenance of the transient hybrid at a fixed length also serves to make it a part of the isopotential one-dimensional diffusion surface, in that whenever a base pair at one end of the hybrid closes, a compensating base pair opens at the other end. Furthermore, as shown experimentally (*12*, *34*) and as expected in terms of this model, the polymerase cannot diffuse to within less than ~12 bp of the 5′ terminus of the transcript (or to within less than ~12 bp of the end of a "blocking" complementary oligonucleotide that has been hybridized to the nascent RNA), nor can it diffuse beyond the 3′ end of the transcript, because either of these processes would result in a net opening (shortening) of the RNA-DNA hybrid and therefore cost more free energy than is available through a diffusional mechanism. This proposed mechanism for the movement of the polymerase along the nucleic acid framework of the transcription complex is illustrated in Fig. 4.

Conclusions and Perspectives

Although the integrated model of the elongation complex that has been presented here can rationalize most of what we know about the behavior of this entity to date, it is still short of both structural and mechanistic detail. Thus, we do not yet know the molecular structure of any multisubunit RNA polymerase, although the common features of the various DNA polymerase and reverse transcriptase structures that have been solved provide useful hints (*37*) and available low-resolution images of transcription complexes and biochemical cross-linking studies provide insight into some details of structure and topology (*8*, *9*). However, we do not know the exact path taken through the polymerase by the DNA and RNA framework of the transcription complex, how various transcription factors change the rates of movement or the stability properties of the transcription complex, or how these changes are further

modulated by the local sequences of the template and nontemplate DNA and the nascent transcript. Additional facts, as they come in, will further define or modify the conceptual framework that has been presented here.

References and Notes

1. S. L. McKnight and K. R. Yamamoto, *Transcriptional Regulation* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992).
2. D. Herschlag and F. B. Johnson, *Genes Dev.* **7**, 173 (1993); K. Rippe, P. H. von Hippel, J. Langowski, *Trends Biochem. Sci.* **20**, 500 (1995).
3. A. A. Travers and R. R. Burgess, *Nature* **222**, 537 (1969); J. Greenblatt and J. Li, *Cell* **24**, 421 (1981); S. C. Gill, S. E. Weitzel, P. H. von Hippel, *J. Mol. Biol.* **220**, 307 (1991).
4. P. H. von Hippel, W. A. Rees, K. Rippe, K. S. Wilson, *Biophys. Chem.* **59**, 231 (1996); M. R. Van Gilst, W. A. Rees, A. Das, P. H. von Hippel, *Biochemistry* **36**, 1514 (1997).
5. K. M. Arndt and M. J. Chamberlin, *J. Mol. Biol.* **213**, 79 (1990); R. R. Reynolds, C. R. M. Bermudez, M. J. Chamberlin, *ibid.* **224**, 31 (1992); K. S. Wilson and P. H. von Hippel, *ibid.* **244**, 36 (1994); W. A. Rees, S. E. Weitzel, A. Das, P. H. von Hippel, *ibid.* **273**, 797 (1997).
6. P. H. von Hippel and T. D. Yager, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2307 (1991); *Science* **255**, 809 (1992).
7. Although elongation complexes are stable with respect to dissociation under such lengthy incubation conditions, slow reactions may alter the kinetic or conformational properties of such halted complexes during extended "dwell times" at a given template position (see discussion of editing and "sliding" reactions in text).
8. A. Polyakov, E. Severinova, S. A. Darst, *Cell* **83**, 365 (1995); E. Nudler, E. Avetissova, V. Markovtsov, A. Goldfarb, *Science* **273**, 211 (1996); F. J. Asturias, G. D. Meredith, C. L. Poglitsch, R. D. Kornberg, *J. Mol. Biol.* **272**, 536 (1997); for earlier references, see (*18*).
9. E. Nudler, I. Gusarov, E. Avetissova, M. Kozlov, A. Goldfarb, *Science* **281**, 424 (1998).
10. M. Kainz and J. Roberts, *Science* **255**, 838 (1992); see also references in (*19*) and (*21*). A more three-dimensionally correct and complete model of the elongation complex [for example, Fig. 1 of (*9*)] would show that the double-stranded DNA is sharply bent in passing through the transcription complex [W. A. Rees, R. W. Keller, J. P. Vesenka, G. Yang, C. Bustamante, *Science* **260**, 1646 (1993)]. This bending could facilitate the opening of the transcription bubble and the displacement of the nascent RNA and perhaps also improve access for regulatory factors to specific portions of the elongation complex.
11. Y. Shi, H. Gamper, B. Van Houten, J. E. Hearst, *J. Mol. Biol.* **199**, 277 (1988).
12. N. Komissarova and M. Kashlev, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1755 (1997).
13. I. Sidorenkov, N. Kommissarova, M. Kashlev, *Mol. Cell*, in press.
14. K. S. Wilson, C. Conant, P. H. von Hippel, unpublished data.
15. J. P. Richardson, *J. Mol. Biol.* **98**, 565 (1975).
16. J. S. Krakow and E. Fronk, *J. Biol. Chem.* **244**, 5988 (1969).
17. H. B. Gamper and J. E. Hearst, *Cell* **29**, 81 (1982); T. D. Yager and P. H. von Hippel, in Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology*, F. Neidhardt, Ed. (American Society of Microbiology, Washington, DC, 1987), pp. 1241–1275.
18. T. D. Yager and P. H. von Hippel, *Biochemistry* **30**, 1097 (1991).
19. K. S. Wilson and P. H. von Hippel, *J. Mol. Biol.* **244**, 36 (1994); *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8793 (1995).
20. F. H. Martin and I. Tinoco Jr., *Nucleic Acids Res.* **8**, 2295 (1980).
21. Both the run of rU resides and the contiguous terminator hairpin within the RNA are required for effective intrinsic termination (*19*). The formation of the termination hairpin could also partially destabilize the elongation complex by competing with the binding of a single-stranded RNA segment of the nascent transcript to the RNA-binding site (Fig. 3, region III) located just upstream of the hybrid (*19*) [see also C. R. Altmann, D. E. Solow-Cordero, M. J. Chamberlin, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3784 (1994); (*9, 14*)], resulting in an additional partial destabilization of the complex by decreasing the (net favorable) contribution of the $\Delta G^{\circ}_{\text{NA-polymerase}}$ term (Eq. 1) to the stability of the complex.
22. W. D. Morgan, D. G. Bear, B. L. Litchman, P. H. von Hippel, *Nucleic Acids Res.* **13**, 3739 (1985); C. A. Brennan, A. J. Dombroski, T. Platt, *Cell* **48**, 945 (1987); D. Jin, R. R. Burgess, J. P. Richardson, C. A. Gross, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1453 (1992); J. Geiselmann, Y. Wang, S. E. Seifried, P. H. von Hippel, *ibid.* **90**, 7754 (1993); J. P. Richardson, *J. Biol. Chem.* **271**, 1251 (1996); K. M. Walstrom, J. M. Dozono, S. Robic, P. H. von Hippel, *Biochemistry* **36**, 7980 (1997); A. Q. Zhu and P. H. von Hippel, *ibid.*, in press.
23. Tethering the regulatory proteins to the nascent RNA, and thus controlling the effective concentration of these proteins at the transcription complex by RNA looping, permit the introduction of terminator specificity into the regulatory process (*4*). Thus, only terminators located fairly closely downstream of the sequence that codes for the specific site on the transcript that binds the regulatory protein will be affected. In the same way, the requirement for an unstructured rho-loading site on the nascent RNA transcript introduces terminator specificity into rho-dependent termination (*22*).
24. D. Erie, T. D. Yager, P. H. von Hippel, *Annu. Rev. Biophys. Biophys. Chem.* **21**, 379 (1992); F. R. Fairfield, J. W. Newport, M. K. Dolejsi, P. H. von Hippel, *J. Biomol. Struct. Dyn.* **1**, 715 (1983); T. A. Steitz, *Curr. Opin. Struct. Biol.* **3**, 31 (1993).
25. The simplest version of this model suggests that the transcription complex as a whole moves "monotonically" along the template, maintaining register between the 3′ end of the elongating transcript and the active site of the polymerase. A more elaborate scheme, based largely on nuclease footprinting observations of halted transcription complexes, suggested that the polymerase might move (at least in part) by an "inchworming" mechanism [M. J. Chamberlin, *Harvey Lect.* **88**, 1 (1995)]. In this scheme, the polymerase was thought to be "flexible," and although the active site moved monotonically with the elongating transcript, it was suggested that other parts of the polymerase might remain fixed and then move along the template in larger, multi–nucleotide-residue "jumps," reflecting a cycling between "strained" and "relaxed" polymerase conformations. This alternative hypothesis has generated much valuable experimentation that has advanced the field, but it now appears (*12*, *34*) that some of the evidence on which it was based may have involved dynamic footprinting artifacts associated with movement (during footprinting) of the halted and "back-sliding" transcription complexes.
26. T. M. Lohman, K. Thorn, R. D. Vale, *Cell* **93**, 9 (1998).
27. H. Yin *et al.*, *Science* **270**, 1653 (1995); H.-Y. Wang, T. Elston, A. Mogilner, G. Oster, *Biophys. J.* **74**, 1186 (1998); J. Gelles and R. Landick, *Cell* **93**, 13 (1998).
28. T. D. Yager, J. Korte, and P. H. von Hippel, unpublished, cited in D. A. Erie *et al.* (*31*); see also K. I. Kato, J. M. Goncalves, G. E. Houts, F. J. Bollum, *J. Biol. Chem.* **242**, 2780 (1967).
29. T. A. Rozoskaya, A. A. Chenchik, R. S. Bibilashvilli, *FEBS Lett.* **137**, 100 (1982); G. A. Kassevetis, P. G. Zenner, E. P Geiduschek, *J. Biol. Chem.* **261**, 14256 (1986).
30. H. Echols and M. F. Goodman, *Annu. Rev. Biochem.* **60**, 477 (1991).
31. D. A. Erie, O. Hajiseyedjavadi, M. C. Young, P. H. von Hippel, *Science* **262**, 867 (1993).
32. M. G. Izban and D. S. Luse, *Genes Dev.* **6**, 1342 (1992); S. Borukhov, V. Sagitov, A. Goldfarb, *Cell* **72**, 459 (1993). The analogy to editing by DNA polymerase is not complete, because these transcription factors do not themselves seem to have any exonucleolytic or endonucleolytic activity. Rather, they appear to activate chain cleavage by an as yet undefined mechanism, probably involving the same catalytic subsites of the polymerase that function in transcript elongation.
33. C. K. Surratt, S. C. Milan, M. J. Chamberlin, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7983 (1991).
34. T. Reeder and D. K. Hawley, *Cell* **87**, 767 (1996); N. Komissarova and M. Kashlev, *J. Biol. Chem.* **272**, 15329 (1997); E. Nudler, A. Mustaev, E. Lukjtanov, A. Goldfarb, *Cell* **89**, 33 (1997); R. Landick, *ibid.* **88**, 741 (1997).
35. M. J. Thomas, A. A. Platas, D. K. Hawley, *Cell* **93**, 627 (1998).
36. This proposed sliding mechanism for polymerase relative to the nucleic acid framework of the transcription complex is similar to the one-dimensional diffusion of electrostatically (and nonspecifically) bound proteins along double-stranded DNA in locating their regulatory targets [see P. H. von Hippel and O. G. Berg, *J. Biol. Chem.* **264**, 675 (1989)]. However, the "zippering" movement of the transcription bubble and the RNA-DNA hybrid along template DNA is not entirely "isopotential," because translocation of the complex by one base pair may result in the opening (for example) of two more stable G · C base pairs and the closing of two less stable A · T (or A · U) base pairs. The resulting small differences in the net stability of the transcription complex as a function of template position may help to define the pausing positions of elongation complexes moving both backward and forward along the template.
37. For example, see H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, J. Kraut, *Science* **264**, 1891 (1994); S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **391**, 251 (1998); J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *ibid.*, p. 304.

38. I thank my laboratory colleagues, whose experimental and theoretical work has led to many of the ideas developed here, as well as many colleagues at the University of Oregon and elsewhere for stimulating discussions. References included in this article can only "scratch the surface" in acknowledging the work of colleagues in other laboratories and of those who came before. The preparation of this article was supported in part by U.S. Public Health Service research grants GM-15792 and GM-29158. The author is an American Cancer Society Research Professor of Chemistry.