ELSEVIER

Biophysical Chemistry 101–102 (2002) 401–423

Biophysical Chemistry

www.elsevier.com/locate/bpc

Reaction pathways in transcript elongation☆

Peter H. von Hippel*, Zvi Pasman

Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, OR 97403, USA

Abstract

Transcription of DNA into RNA is a central part of gene expression, and is highly regulated in all organisms. In order to approach transcription control systems on a molecular basis we must understand the mechanisms used by the transcription complex to discharge its various functions, which include transcript initiation, elongation, editing, and termination. In this article we describe recent progress in sorting out the multiple reaction pathways that are, at least in principle, available to the transcription complex at each DNA template position, and show how transcription control systems partition active complexes into these pathways. Understanding these regulatory processes requires an elucidation of the molecular details of how sequence- and factor-dependent changes in the conformations, stabilities, and reaction rates of the complexes determine function. Recent progress in unraveling these issues is summarized in this article and emerging principles that govern the regulation of the elongation phase of transcription are discussed.

Keywords: RNA polymerase; Transcription regulation; Transcription factors; Transcript elongation; Termination; Editing

☆ This article is dedicated to John Schellman by PHvH, in recognition of all that he has taught me in the course of our long personal and scientific friendship. A lot of what I understand about the physical chemical aspects of biology has grown out of countless luncheon conversations with John over the years. In addition, many of the ideas that underlie our present quantitative models of how the macromolecular machinery of transcription might work and be controlled have been strengthened and refined by countless diagrams scribbled on paper napkins in the course of those same lunches. As indicated in this article, recent developments in structural and kinetic analyses have sharpened our view of how many of these regulatory interactions might actually operate in detail. In that sense this paper might be viewed as comprising a summary 'report to John' on these recent developments, as well, of course, as a report to others who might be interested.

*Corresponding author.

E-mail address: petevh@molbio.uoregon.edu (P.H. von Hippel).

0301-4622/02/$ - see front matter
PII: S0301-4622(02)00160-6

1. Overview

Our understanding of the regulation of transcription has developed on a number of parallel fronts. Largely separate efforts have been made to describe the physiological processes of transcript initiation, elongation, editing, and termination, although these processes have many features in common that can, in principle, be described by the same formalism.

1.1. Initiation

The regulation of initiation is conceptually more complicated than that of the other processes listed above. This follows because initiation involves both the physical location of the promoter (the transcription start point) of the selected gene by the appropriate RNA polymerase, and the setting of the overall level of initiation of expression of that gene relative to those of the other genes of the organism. These processes require the participation of a whole ensemble of regulatory subunits. Some serve as DNA sequence recognition factors that also bind to the relevant 'core' RNA polymerase, while others serve as recognition, activation, or repression subunits by binding to the DNA at or near the relevant promoter(s). Still others bind to more distant enhancer sites on the genome and interact with the promoter and the polymerase by looping, twisting, or otherwise distorting the intervening double-stranded DNA.

This macromolecular machinery serves to regulate the distribution of active RNA polymerase over the available transcription start sites on the genome, and this machinery is further controlled by different levels of expression of the regulatory proteins themselves at different stages of the life cycle of the cell. The process of initiation is completed by the local opening ('melting') of the promoter, which is a complex process that involves bending the DNA by $\sim 90°$, together with substantial conformational rearrangements of the protein subunits of the polymerase [1]. In functional terms the opening of the promoter region is accompanied by the polar binding of the active site of the core polymerase at a particular locus on one of the strands of the resultant DNA 'bubble', which serves then to define both the template strand itself and the starting position of the RNA transcript on that template. From this point the elongation of the nascent transcript ($5' \rightarrow 3'$) is processive and directional ($3' \rightarrow 5'$) along the coding DNA template strand, with elongation proceeding concomitantly with the movement of the polymerase through the DNA of the gene that is being transcribed.

Despite the fact that the first extension steps of transcript synthesis comprise an elongation reaction, this initial process of transcript extension is considered a part of the initiation phase of transcription because it is quantitatively, although neither qualitatively nor chemically, distinct from the elongation steps that occur in transcription further downstream. This follows because in these very early stages of transcript extension the RNA–DNA hybrid duplex, and perhaps the transcription bubble, have not yet attained their mature size. As a consequence the transcription complex has not reached its full stability and, at least in *E. coli*, the sigma (specificity) factor of the holoenzyme has not yet been released from its interactions with the DNA of the promoter and the core subunits of the polymerase. These aspects together result in a less stable transcription complex during initiation, and at some promoters also leads to the release of short (2–10 nt) nascent RNA oligomers. For promoters at which such RNA release occurs, the polymerase is able to return to the transcription start point of the gene while maintaining its 'hold' on the template and also maintaining an open transcription bubble. This 'short RNA oligomer release' process, which may have regulatory significance, is called 'abortive initiation' in prokaryotes and probably has a counterpart in eukaryotic transcript initiation as well.

We will not consider the complexities of initiation further in this article. Rather we will focus on the actions and interactions of the core polymerase in the course of processive transcript elongation, where many additional aspects of transcription regulation are manifested.

1.2. Elongation

In terms of the above definition, the initiation phase of transcription ends in prokaryotes when

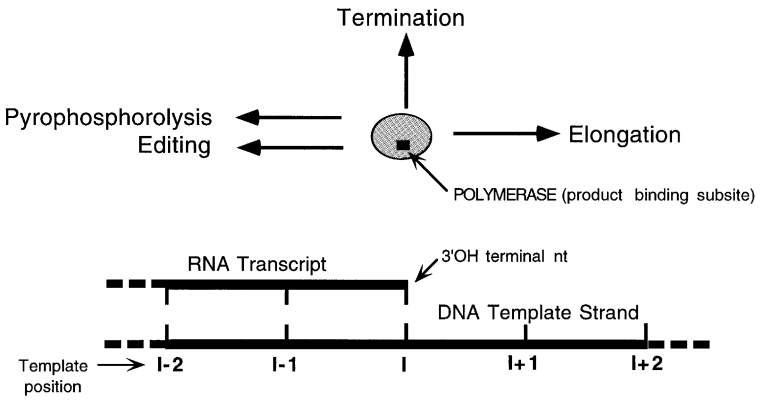

Fig. 1. Competing pathways for extension, shortening or release of the transcript at template position (I). The shaded sphere represents the polymerase and the square denotes the product binding sub-site of the active site (see also Fig. 3a). The RNA transcript and the DNA template strand are shown below, and the 3′-terminus of the nascent transcript is considered to be bound at the product binding sub-site. Shortening the transcript by pyrophosphorolysis or editing moves the bound 3′-terminus of the transcript (and the polymerase) to the left (to positions I−1, I−2,…). Elongating the transcript moves the 3′-terminus (and the polymerase) to the right (to positions I+1, I+2,…). Paused and arrested states represent intermediates in some of these physiological processes (see text).

the RNA–DNA hybrid is extended to its mature length (8–9 bp), the binding affinity of the sigma specificity factor for the promoter DNA and the core subunits of the polymerase is reduced, and sigma is generally (but not always) released from the transcription complex. In eukaryotes the equivalent point occurs when the polymerase has moved far enough from the promoter to leave behind (probably at the promoter) at least a portion of the complex of regulatory proteins involved in activation, and the transcription complex has reached its mature level of stability. By definition the transcription process moves into the elongation phase at this point in all organisms.[1]

In a mechanistic context, it is now more straightforward to include the processes of transcript extension, termination, and shortening (by pyrophosphorolysis and RNA editing), as well as intermediate states that may include transcription complex sliding, pausing, and arrest, as comprising alternative pathways for the polymerase within the elongation phase of transcription. This notion of alternative reaction pathways that can (at least potentially) engage in kinetic competition with elongation at every template position is illustrated schematically in Fig. 1. The binding of polymerase to the DNA, once transcription is safely beyond the initiation phase, is very stable. As a consequence, short ‘excursions’ by the complexes down the alternative pathways of Fig. 1 will eventually lead back to elongation. Therefore, at least until a termination site is reached, the active RNA polymerase remains bound to the template and non-template strands of the DNA and to the nascent RNA transcript within the transcription bubble, and the overall process of transcript elongation is fully processive.

At least in *E. coli*, the elongation complex becomes sufficiently destabilized to permit a measurable (and regulated, see below) probability of irreversible dissociation as it moves through intrinsic termination sites along the DNA template. The complexes that terminate at a given intrinsic terminator release the nascent RNA and polymerase from the template and non-template DNA, and the

[1] The core RNA polymerase of the transcription complex comprises a single subunit in many bacteriophage systems, four subunits (2α, β and β′) in *E. coli* and related prokaryotes, and ten or more subunits in typical eukaryotes. The core polymerase of all these organisms, however, seems to operate within a DNA transcription bubble, and with an RNA–DNA hybrid, of about the same size [2].

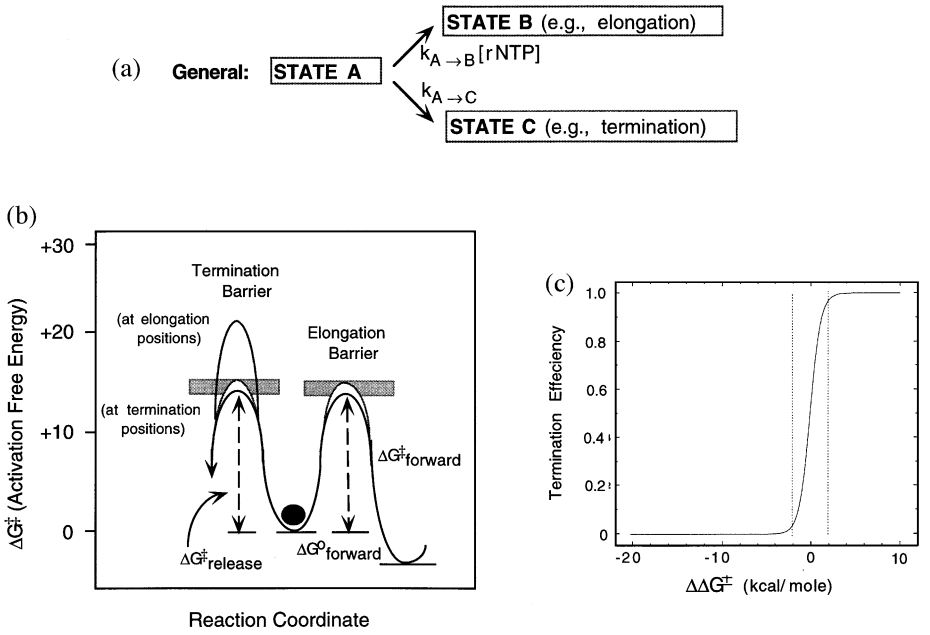

Fig. 2. Partitioning of the transcription complex between two potential reaction pathways (here elongation and termination). (a) State A here represents the active elongation complex with its 3′-terminus located at position (I) on the template (and presumably in the product binding sub-site of the polymerase; see text). From here the transcript can either be extended to form State B by adding one residue (placing the new 3′-terminus at template position (I+1) and again in the product binding sub-site of the polymerase), or released to form State C by dissociation of the transcription complex. These reactions are characterized by rate constants k_{A-B} and k_{A-C}, respectively. The elongation rate also depends on the concentration of the next-required NTP, if this concentration is sub-saturating. (b) The two free energy of activation barriers that the polymerase (black circle) faces in 'deciding between' the elongation and the termination pathway (i.e. in deciding whether to move to State B or State C). The elongation pathway is the only one that is effectively available at elongation (non-terminator) template positions. At terminator positions both barriers are of comparable height, and thus both pathways are available and in effective competition. (c) A plot of the TE of the complex as a function of the difference in barrier heights ($\Delta\Delta G^{\ddagger}$) between the two pathways. Note that the transition is sharp, and that the TE changes from <0.01 to >0.99 over a very narrow range of $\Delta\Delta G^{\ddagger}$ values. Such a system of competitive pathways is very stable and favors elongation only at non-terminator positions, but can, at a 'window of opportunity' provided by a terminator sequence, operate as a regulatory 'switch' that can be triggered by very small changes in the stability or synthesis rate of the complex. (Taken with permission from Ref. [5].)

transcription bubble closes to reform the original duplex DNA. In contrast, the complexes that succeed in 'reading-through' each terminator continue to elongate in a fully processive manner until again subjected to a finite probability of dissociation at the next terminator encountered in moving down the template.

The processes of editing and pyrophosphorolysis represent additional competing pathways for a transcription complex at any particular template position, and result in shortening the transcript from the 3′-end. These pathways are shown schematically in Fig. 1. In contrast to termination, the transcription complex remains bound to the DNA (and the shortened RNA) while involved in these pathways, and eventually elongation is resumed at the new 3′-terminus of the shortened transcript.

A central question that is somewhat concealed in the rendering of Fig. 1 is the nature of the mechanisms that partition the elongation complex that arrives at each template position into these different pathways, and how this partitioning is controlled. In the balance of this article we will describe what is known about each of the competing pathways shown in Fig. 1. Each of these descriptions starts by asking 'what happens' to the

transcript elongation complex in that pathway, including reaction components and stoichiometries. We then describe what is known about the thermodynamics of each reaction, defining, in free energy terms, how far each reaction could potentially go in isolation. A transition state description of each process will be used to define the kinetics of the reaction; i.e. to ask how fast each reaction goes. Finally, we will introduce new findings about the structure of the transcription complex in each of these competing pathways. We conclude with a discussion of the control mechanisms that regulate the relative rates of each pathway, and thus determine, at each template position, which course a transcription complex will actually take, to what extent each pathway is utilized, and how this utilization is controlled.

2. Mechanisms of transcription elongation and competing pathways

2.1. A probabilistic view of the branching pathways of transcription

Each reaction pathway shown in Fig. 1 can be considered to be involved in a potential kinetic competition at each template position, meaning that the transcription complexes that arrive at each position are partitioned into the various available reaction pathways with probabilities that depend on the relative heights of the transition state barriers that define the 'entrances' to those pathways. Implicit in this picture is the assumption that the form (or forms) of a complex located at any given template position that is active in terms of synthesis (i.e. 'elongation-competent') reaches energetic and conformational equilibrium during its 'dwell-time' at that position prior to the next elongation step. This definition of elongation-competent complexes explicitly excludes forms of the transcription complex that do not readily interconvert with the elongation-competent form(s), such as arrested and paused complexes that cannot directly add the next-required NTP to the elongating transcript.[2] In this view the complex must be 'reset' into a form (or forms) that is active for synthesis before elongation to the next template position can occur. Further details of this model are developed in Section 3, below.

It is important to be clear about this, because it has been known for some years that transcription complexes can exist at any particular template position in more than one functional state [3], and that these different states favor, define, and lead into the different reaction pathways of Fig. 1. What we know of these states, and how they might interconvert, will be discussed in the balance of this article. With respect to the elongation pathway, the central import of the equilibrium assumption stated above is that only one form (or one population of rapidly interconverting forms) of the elongation complex is active in synthesis.[3]

An alternative view to the equilibrium or rapid interconversion model suggests that individual, long-lasting states of the transcription complex can be specified for particular genes, perhaps as early as at the promoter. These putative multiple states of the elongation complex are considered to have distinct functional properties, and to be stable over many steps of elongation and synthesis. Such 'stable heterogeneity' models, featuring multiple forms of non-equilibrating polymerases, would seriously complicate our views of regulatory mechanisms, since these mechanisms would then have to deal with an undefined multiplicity of stable elongation-competent transcription complexes. We

[2] As will be detailed below, the editing pathway that is triggered by misincorporation, and perhaps some complexes in an arrested state, require the intervention of a covalent bond cleavage reaction before the complex can be reset into an activated state, and this cannot be viewed as corresponding to equilibration of the forms of the transcription complex in a simple sense. Clearly elongation at any particular template position cannot proceed until the elongation-competent state is re-established, either by conformational equilibration to the active state or by chain cleavage followed by such conformational equilibration.

[3] To put this more generally, two or more synthetically active states of the elongation complex could exist in alternation at a given template position, but the equilibrium model (which we have elsewhere called the 'rapid interconversion' model [4]) requires that these putative active states equilibrate within the 'dwell-time' of the transcription complex at each template position.

will consider such hypotheses, and describe experiments that render them unlikely, in the sections that follow.

To avoid confusion, we emphasize once more that there is no question that elongation complexes can exist in several forms at each template position, that these different forms interconvert at very different rates, and that these different forms represent intermediates on the various competing reaction pathways of Fig. 1. In addition, the changes that occur in the complexes in passing through some of these synthetically inactive pathways can be reversed and can, with characteristic kinetics, regenerate an elongation-competent state of the transcription complex. However, the basic hypothesis that we are putting forward here, and which underlies any simple version of regulatory control, is that there is only one state—or a rapidly (on the time scale of the dwell-time of the elongation-competent form of the complex at each template position) interconverting set of states—of the core polymerase that is active in elongation at any template position. In this view controlling the fraction of complexes that can surmount the transition state barrier to elongation at each template position represents the ultimate target of transcriptional regulation.

Fig. 2a shows the kinetic partitioning between the elongation and termination pathways at one template position to illustrate this idea, although, of course, a real picture of what is going on at any given template position (I) would allow for the simultaneous partitioning of the transcription complex among the transition states of all the potential pathways that apply at that point (i.e. elongation, pausing, pyrophosphorolysis, editing, and termination) in a multi-dimensional plot of competing reaction pathways. Fig. 2b illustrates the actual barriers to elongation and termination that apply at both a terminator and a non-terminator template position, and Fig. 2c plots the 'sharpness' of the switch between elongation and termination as the difference in barrier heights ($\Delta\Delta G^{\ddagger}$) between these two pathways is changed by regulatory interactions (see Refs. [5,6] for a more extended treatment of these ideas).

Clearly if more than one stable (heterogeneous) form of the elongation-competent transcription complex could be present at a given template position, regulation mechanisms would have to deal with such multi-dimensional transition states and the control of their utilization for each elongation-competent form of the complex, leading to a continuously bifurcating control scheme. We note that while this is conceptually possible it is also unlikely, since it would render control at each template position 'hostage' to the sequence- and environment-dependent experiences that each transcription complex has experienced in arriving at that template position. We describe below some recent experiments involving transcription through multiple terminators that we have used to test the competitive equilibrium (or rapid interconversion) models of the synthetically active transcription complex against stable heterogeneity models [4].

2.2. Structure and properties of the active elongation complex

Fig. 3a shows a schematic model of the elongation complex that illustrates features of the complex that are particularly important to the regulatory process. This is contrasted, in Fig. 3b, with a structural view of the same complex that is based on recent X-ray crystallographic and fluorescent energy transfer studies (see Refs. [7,8]). Fig. 3a shows that ~14 bp of duplex DNA are separated at the polymerase to form a 'transcription bubble' that moves along the genomic DNA with the polymerase. The formation of such a bubble is obviously destabilizing, but this unfavorable contribution to the free energy of formation of the transcription complex is offset by favorable interactions that the existence of the bubble make possible. These interactions include the formation of an ~8–9 bp RNA–DNA hybrid duplex, comprising the 3′-terminal residues of the nascent RNA and the complementary sequence of the template strand of the DNA at the downstream end of the transcription bubble.

The RNA–DNA hybrid is maintained at an approximately constant length by an active separation or unwinding process at its upstream end, which separates the nascent transcript from the template DNA. This permits the 5′-end of the released ssRNA to pass through a single-stranded

RNA binding site (Fig. 3a) and the RNA 'exit tunnel' (Fig. 3b), ultimately leading to the extrusion of the nascent ssRNA transcript from the complex. In addition, this continuous separation of the hybrid at a position located ~8–9 bp from the 3′-end of the elongating transcript also permits the reformation of the complementary duplex DNA upstream of the bubble, and maintains the transcription bubble, as well as the hybrid, at an approximately constant size as the elongation complex moves along the template.

Thermodynamically, the formation of the RNA–DNA hybrid partially offsets the destabilizing free energy that is contributed by the transcription bubble. Favorable interactions between the polymerase and the nucleic acid (NA) framework provide the additional stabilizing free energy to the overall complex that renders the net free energy of the complex favorable for its formation. These interactions make the elongation complex very stable and completely processive at non-terminator positions along the DNA (for a more extensive discussion of these thermodynamic aspects see [5,6]). The physical nature of these stabilizing polymerase–NA interactions are still being elucidated, and the details of the partitioning of the total binding free energy of the polymerase to the NA framework of the complex into specific binding interactions are not well-understood. It is known, however, that these interactions include a specific binding site for the RNA–DNA hybrid within the β and β′ subunits of the *E. coli* core polymerase, an ssRNA binding site located 'behind' the moving RNA–DNA hybrid and a dsDNA binding site (comprising a set of 'jaws' between the β and the β′ subunits) that close around the dsDNA just ahead of the front end of the moving transcription bubble (Fig. 3a and b). An important feature of the structure of the elongation complex that is not shown in Fig. 3a, but can be seen in Fig. 3b, is that the projected overall path of the DNA duplex through the polymerase is bent by ~90° within the complex. This bending is energetically costly and may have important consequences for the structure (and the stability) of the NA features of the elongation complex, as will be discussed further below.

Finally, a central feature of the structure of the transcription complex, which defines the enzymology of the polymerase as well as its interactions with the nascent RNA and the template DNA, is the polymerase active site. Some important features of the active site, including the substrate and product sub-sites and their relationship to the template DNA and the nascent RNA, are shown schematically in the insert in Fig. 3a (see also Fig. 1).

2.3. Competing reaction pathways for the transcription complex

As shown schematically in Fig. 1 and Fig. 2, there are a number of reaction pathways that are, in principle, available to the transcription complex at each template position. The heights of the transition state barriers that control the entrances to these pathways, and thus their relative probabilities of utilization, differ at each template position. Elongation is the most probable pathway at virtually all template positions in the presence of saturating concentrations of all four NTPs, and thus the elongation complex remains predominantly in its 'active' elongation state as it moves from one template position to the next. This follows because the activation barriers to the other pathways are significantly higher than the barrier to elongation at most template positions.

2.3.1. Elongation

An important feature of the active elongation state of the polymerase is the 'cycling' of the growing 3′-end of the nascent transcript between the substrate and product binding sub-sites of the active site of the enzyme (Fig. 3a). The single nucleotide addition cycle begins with the 'templated' binding of the next-required NTP in the substrate binding sub-site of the polymerase; i.e. the NTP in this site is bound both to the polymerase and also directly to the next position of the DNA template by complementary Watson-Crick hydrogen bonding. NTP binding is followed by the chemical step of the elongation reaction, which leads to the formation of a phosphodiester bond between the 3′-end of the RNA chain that lies in the product binding sub-site and the nucleotide

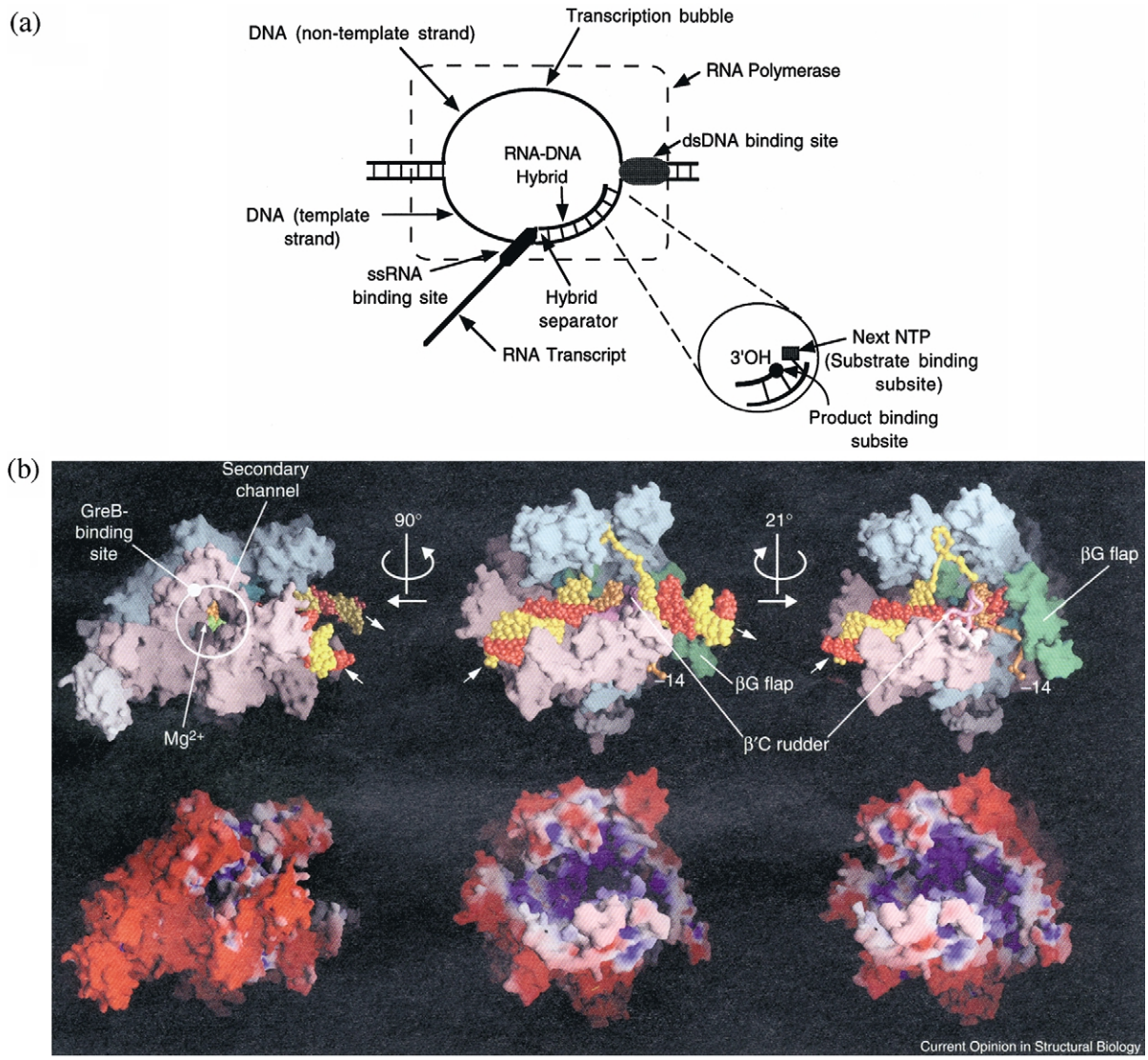

Fig. 3.

that is bound in the substrate binding sub-site. This reaction is accompanied by the release of the pyrophosphate (PP_i), which is the other product of phosphodiester bond formation in the RNA elongation reaction.

These events are thought to induce a conformational change in the polymerase that triggers the release of the newly added 3′-terminal NMP residue of the transcript from the substrate binding sub-site. As a consequence the nascent RNA is no longer fixed in position relative to the polymerase and goes into a one-dimensional diffusion (sliding) mode, eventually resulting in the diffusion of the new 3′-terminus of the transcript ‘back’ into the product binding sub-site of the active site of the polymerase. This completes the elongation reaction cycle by making available the substrate binding sub-site of the polymerase for the templated bind-

ing of the next-required NTP. The transiently released transcript remains electrostatically bound to the polymerase during this sliding reaction, and also to the NA framework of the complex via the RNA–DNA hybrid duplex. The rebinding of the 3′-end of the RNA into the product sub-site of the polymerase is then further stabilized by the binding of the properly hydrogen bonded next-required NTP into the substrate binding sub-site. (For a further discussion of the elongation reaction cycle see Refs. [5,9,10].)

As indicated above, the elongation reaction predominates at most non-terminator positions on the DNA template in the presence of saturating NTP concentrations (~0.1–0.5 mM). However, other reaction pathways can also be accessed, especially if NTP concentrations are low. Omission of the next-required NTP from the reaction mix can 'stall' transcript elongation at defined template positions. Sequential omissions and additions of the next-required NTP, combined with knowledge of the template sequence, have been exploited in in vitro experiments designed to stall complexes at template positions of functional interest [11]. This methodology has also been used as an experimental tool to 'walk' the 3′-terminus of the elongating chain from one defined template position to the next [12]. The stalling of the transcript elongation reaction increases the overall 'dwell-time' of the complex at a given template position, and thus increases the probability that other reaction pathways (Fig. 1) will be explored.

2.3.2. Pyrophosphorolysis

The reversible elongation reaction for adding the next-required NTP is

$$\mathrm{RNA} + \mathrm{NTP} \Leftrightarrow \mathrm{RNA} - \mathrm{NMP} + \mathrm{PP_i} \quad (1)$$

Here the forward reaction corresponds to 5′→3′ elongation of the RNA transcript and the reverse reaction results in 3′→5′ transcript shortening by pyrophosphorolysis. The forward nucleotide addition reaction is favored over pyrophosphorolysis by a factor of ~100 (Yager et al., unpublished, reviewed in Ref. [10]). However, elevated concentrations of PP_i can be used at low NTP concentrations to drive the transcript elongation reaction backwards over many template positions [13,14].

This process represents a direct reversal of the elongation reaction of Eq. (1), meaning that the terminal NMP of the RNA is attacked by PP_i while the 3′-end of the nascent chain is in the substrate binding sub-site, resulting in the release of the 3′-terminal residue of the RNA as NTP. Since this reaction can be repeated multiple times in the presence of significant excesses of PP_i, it is clear that the individual steps of the elongation reaction are also reversible, as required by the principle of microscopic reversibility. The existence of the facile pyrophosphorolysis reaction further demonstrates that the 3′-end of the nascent chain can 'slide back' into the substrate binding sub-site when this sub-site is not occupied by the next-required NTP. Finally, we note that the reversibility of the elongation reaction means that chain shortening by pyrophosphorolysis cannot serve as an 'editing' pathway to improve the fidelity of the nucleotide incorporation process, since effective editing requires the use of a transcript shortening reaction that is not the reverse of the forward reaction. Thus, a different reaction pathway with

Fig. 3. Structural representations of the transcription complex. (a) A schematic view of the some of the central features of the transcription complex, focussing, in particular, on the NA framework of the complex and its relation to the polymerase. The relative positions and features of the transcription bubble, the RNA–DNA hybrid, the ssRNA binding site, the dsDNA binding site, and the polymerase active site are shown. The insert shows a 'blown-up' view of the sub-sites of the polymerase active site. (b) A structural view of the bacterial transcription complex. The three upper views represent three different rotations of the crystal structure of the core RNA polymerase of *Thermus aquaticus* (which closely resembles that of *E. coli*). The NA framework of the elongation complex has been superimposed on the polymerase by modeling techniques. The DNA template strand is shown in red, the non-template strand in yellow, and the nascent RNA transcript in gold. The β subunit of the RNAP is cyan, the β′ subunit is pink, and the α subunits are white. The catalytic Mg^{++} cation is shown as a magenta sphere, and the incoming NTP is green. Other features of the structure are indicated, including the secondary channel for the incoming NTP and the pathway (and exit tunnel) for the synthesized transcript. White arrows indicate the path of the incoming and exiting dsDNA (note that the trajectory of the dsDNA bends by approximately 90° within the polymerase). The lower three views are the same as the upper, but with the surfaces shaded to indicate the local electrostatic potential (red for acidic; white for neutral; and blue for basic). Note that the polymerase surfaces on which the NA framework must slide are largely basic, as expected. (Taken with permission from [8].)

a different transition state must be used for this purpose [15].

2.3.3. Pausing, arrest, editing, and the 'slid-back' state

As indicated in Fig. 1, the editing pathway, like pyrophosphorolysis, results in transcript shortening from the 3′-end, although by a different reaction pathway. In general the transcript elongation reaction pauses (stalls) when the next-required NTP is present only at very low concentrations (or is missing altogether), and the PP_i concentration is also low. Under these conditions there is a high probability that the substrate sub-site of the polymerase will be empty and thus that binding of the 3′-end of the nascent chain in the product sub-site will be destabilized. As a consequence the 3′-end of the RNA has a significant probability of 'escaping' from both binding sub-sites of the active site of the polymerase, although without dissociation of the polymerase from the NA framework. Because the primary interaction that 'holds' the polymerase at the next template position is the proper binding of the next-required NTP [5,9], there is a significant probability (in the absence of this NTP) that the bound polymerase will 'slide back' along the NA framework of the complex.

As visualized in Fig. 4, this sliding-back reaction maintains a constant transcription bubble size and a constant length of base-paired RNA–DNA hybrid. In addition, the nucleotide residues of the nascent RNA that are actually involved in hybrid formation will shift towards the 5′-end of the transcript as the hybrid 'zippers' open from its 3′-(RNA) end and zippers closed at its 5′-end. Simultaneously, the corresponding residues of the single-stranded DNA (ssDNA) 'arms' of the transcription bubble maintain a constant bubble size while sliding through the non-specific binding sites of the polymerase that hold the bubble open.

As a consequence the polymerase moves along (and 'through') the NA framework of the transcription complex in this process, with the base pairs at the ends of the bubble and those of the corresponding RNA–DNA hybrid zippering open and closed as the polymerase passes. This movement of the polymerase backwards (and forwards) along the RNA and DNA of the complex corresponds to a one-dimensional diffusion reaction along an approximately isopotential surface, although the diffusion process cannot be considered to be entirely random because the stability of the complex varies somewhat with template position. This follows because of the small differences in the stability of the RNA–DNA hybrid that reflect its changing net content of A·T and G·C bp as it 'zippers' along the NA framework of the complex.

This back-sliding process has been demonstrated experimentally for *E. coli* RNA polymerase [16,17] and eukaryotic Pol II complexes [18]. In principle this sliding reaction can carry the polymerase 'backwards' along the NA framework of the complex to within ~8–9 bp of the 5′-end of the nascent RNA, with concomitant extrusion of the 3′-end of the RNA from the complex (Fig. 4). This limit, which represents the extreme position to which 'backward' sliding towards the 5′-end of the nascent RNA can go without shortening the fixed length of the RNA–DNA hybrid, has been demonstrated experimentally (Kashlev, personal communication). Thus, no further sliding of the polymerase in the backward direction should be possible, because continued movement towards the 5′-end of the transcript would be opposed by the free energy cost of shortening the RNA–DNA hybrid that such further sliding would entail. Similarly the sliding or diffusion reaction cannot carry the polymerase forward beyond the template position that corresponds to the 3′-end of the transcript, since further sliding in that direction would also shorten the RNA–DNA hybrid and thus be thermodynamically disfavored.

It has been shown [19] that transcription complexes located at template positions involving relatively unstable RNA–DNA hybrids (i.e. hybrids rich in rU·dA and rA·dT bp) are particularly prone to this sliding reaction. This seems to result in preferential pausing at template positions that correspond to such hybrids and, at some positions, to dead-end complex formation and transcription arrest. Thus paused complexes may often represent complexes that have, in fact, slid-back as described above, although elongation can resume from such paused positions if the hybrid 'slides back' to the 3′-end of the transcript and the 3′-terminal residue

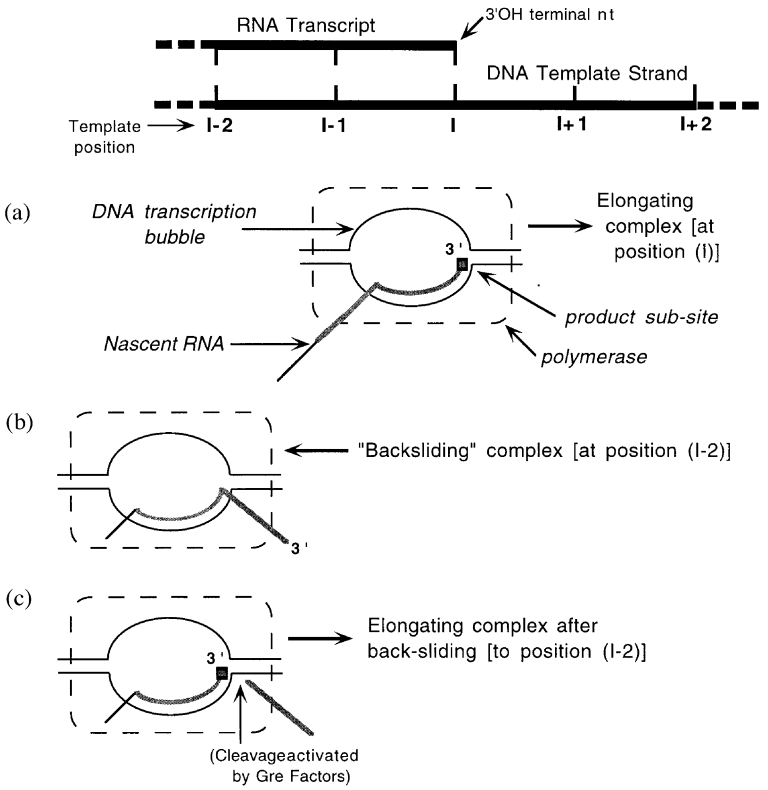

Fig. 4. 'Sliding-back' pathways of the transcription complex and editing. The elongating RNA transcript and the DNA template strand are shown at the top, with template positions numbered. The product binding sub-sites of the slid-back complexes are aligned with the template positions of the upper schematic. (a) The complex is shown in the actively elongating state at template position (I). (b) The complex is shown in the slid-back state at template position (I−2), with extrusion of the 3′-end of the transcript. (c) The previous complex in template position (I−2), after the extruded 3′-end of the transcript has been cleaved off in a reaction catalyzed by the GreA and B RNA editing factors of *E. coli*. The new 3′-end of the transcript is positioned for renewed synthesis in the product binding sub-site of the polymerase at position (I−2).

is properly re-bound in the product binding sub-site of the polymerase. Such spontaneous elongation 'reactivation' reactions seem to be less likely at arrest sites, perhaps because hairpins or other elements of secondary structure have formed in the nascent RNA to prevent or inhibit downstream sliding of the polymerase along the NA framework of the complex.

2.3.4. Misincorporation

Low concentrations of the next-required NTP can favor incorporation of an 'incorrect' (mispaired) NMP residue at the 3′-end of the transcript [3]. Such misincorporation results in unproductive positioning of the end of the transcript in the active site of the polymerase [9,10] and prevents, or significantly inhibits, the addition of the next-required NTP. This inhibition, in turn, favors backsliding of the complex into a paused or arrested state. Both prokaryotic and eukaryotic organisms contain transcript editing factors, called GreA and B in *E. coli* [20,21] and SII in higher organisms [22]. These editing factors work by stimulating an RNA exo- or endo-nuclease activity of the polymerase, which cleaves off one or more nts of the extruded 3′-end of the RNA from the slid-back complex, thereby removing the misincorporated base (and often several additional residues) from the nascent RNA (Fig. 4). This activity permits the newly formed 3′-end of the transcript to slide

back into position in the product binding sub-site at the appropriate template position, accompanied by resumption of transcript elongation. These editing factors are not themselves RNA exo- or endonucleases, and the mechanisms that they use to activate the polymerases to catalyze cleavage are not understood.

2.3.5. Termination

The final reaction pathway shown in Fig. 1 is termination (transcript release), which normally occurs only at specific termination sites that mark the ends of genes on the DNA template.[4] Termination is defined as the (irreversible) dissociation of the protein and RNA components of the elongation complex from the DNA of the genome. Mechanisms that are used to achieve termination differ more widely between organisms than do the other reaction pathways discussed above. Termination is probably best understood in *E. coli*, and we will limit our discussion of these pathways to those of this organism.

Termination in *E. coli* proceeds by two distinct classes of mechanisms, called intrinsic and Rho-dependent termination, respectively. Intrinsic termination occurs at defined template sequences that code for a stable stem-loop (terminator) hairpin in the RNA, followed directly by a run of uridine residues. In general transcript release occurs after the termination hairpin has formed and 7–8 rU residues have been synthesized. Intrinsic termination proceeds with a characteristic termination efficiency (TE) that depends on terminator sequence, NTP concentration, transcription rate, and the presence of transcription factors. Values of TE for intrinsic terminators range from 0.10 to 0.90, meaning that these terminators are between 10 and 90% efficient at bringing about transcript release (see Ref. [23] for a detailed discussion of TE measurements).

[4] Termination sites are also sometimes located within genes, or between the genes of polycistronic operons. In these positions such terminators participate in transcription control by serving as sites of 'premature' termination or as regulators of transcription polarity, respectively. The efficiency of these regulatory terminators are often controlled by antitermination mechanisms, as in the N- and Q-dependent antitermination systems of phage λ.

These TE levels mean that a substantial fraction of the elongation complexes that arrive at a given intrinsic terminator is released, while a substantial fraction also transcribes through. Assuming that the active elongation complexes that arrive at a given terminator represent a functionally homogeneous population, the relative probabilities of elongation and termination at these template positions are comparable. In terms of Fig. 2b this means that the heights of the free energy of activation barriers to elongation and to termination are about the same at terminator template positions. In contrast, this is certainly not true at non-terminator positions; here termination is very improbable and the barrier that leads to termination is much higher than that leading to elongation.

Rho-dependent termination (reviewed in Refs. [24,25]) is quite different. Rho-dependent terminators are bipartite and less well defined in terms of sequence than intrinsic terminators. Furthermore, Rho-dependent termination requires an upstream sequence in the nascent RNA that is both largely unencumbered by secondary structure and relatively rich in rC residues. This sequence comprises the Rho 'loading site' at which transcription termination factor Rho (a hexameric RNA–DNA helicase [26]) is assembled onto the transcript. This loading activates the cryptic ATPase activity of the Rho hexamer, driving Rho (in an ATP-dependent process) in a $5' \rightarrow 3'$ direction along the nascent RNA, 'in pursuit' of the transcribing RNA polymerase that is moving along the DNA template in the same (downstream) direction. The RNA polymerase tends to move more slowly at template positions within Rho terminator sequences. As a consequence, the Rho helicase is likely to 'catch-up' with the transcription complex at these terminator positions.

Depending on the relative translocation rates of Rho and RNA polymerase along their respective NA lattices [27], as well as on a variety of other non-relative-rate-dependent factors [28,29], Rho helicase then unwinds the RNA–DNA hybrid within the transcription bubble to facilitate transcript release with characteristic TE values at template positions within the Rho-dependent terminator sequence. The details of the mechanisms whereby Rho induces termination, to the extent to

which these processes are understood, are described elsewhere [30–35].

In contrast to Rho-dependent termination, intrinsic termination can be considered completely within the context of the interactions of the RNA polymerase with the NA framework of the transcription complex, meaning that no extraneous transcription factors are required. As stated above, intrinsic terminators are characterized by two distinct sequence features within the template DNA strand. One of these features encodes a stable stem-loop hairpin in the RNA. Of course this sequence also specifies a potential stem-loop structure in the DNA, but this hairpin does not form within the template DNA itself for several reasons. First, the 'window' of ssDNA exposed within the transcription bubble is too short to contain the entire stem-loop sequence at any one template position. In addition, a significant portion of the template DNA strand within the transcription bubble is 'tied up' as a part of the double-stranded RNA–DNA hybrid. Finally, the base-paired hybrid, and as well as other portions of the transiently ssDNA of the transcription bubble, are non-sequence-specifically (electrostatically) bound at sites within the β- and β′-subunits of the core polymerase. No such constraints inhibit stem-loop formation within the nascent RNA, and in principle the termination hairpin can form within the transcription complex as soon as the second half of the hairpin stem has been transcribed [23,36]. We note that the rate of formation of this hairpin relative to the rate of continued elongation may be involved in regulating TE at intrinsic terminators.

The second sequence feature of the DNA template that is crucial for intrinsic termination is a long run of dA residues, which code for an equivalent run of rU residues within the nascent RNA. This rU sequence is located directly downstream of the termination hairpin in the nascent RNA, and both features must be present (and in direct proximity) for the termination process to occur [23]. The role of the rU sequence, as previously described (see Ref. [6]), is to form a significantly unstable rU·dA RNA–DNA hybrid sequence within the transcription complex. This relatively unstable hybrid sequence, in turn, partially destabilizes the elongation complex and thus makes the transition state barrier leading to termination more accessible (Fig. 2). The termination hairpin may serve primarily to interact with the polymerase and thus to force the moving elongation complex to pause significantly (1 s or more) [36,37].[5] Such pauses raise the barrier to elongation and therefore also favors termination.

It appears that the formation of the termination hairpin also destabilizes the upstream end of the directly adjacent RNA–DNA hybrid, thus further destabilizing the elongation complex and favoring termination. It had been proposed earlier that the RNA–DNA hybrid might be longer than the 8–9 bp we now believe it to be, and this led to the proposal that hairpin formation might directly destabilize the adjacent bp of the RNA–DNA hybrid by invading and shortening the hybrid from the upstream end as a consequence of the competitive use of the same nucleotide residues in the stem of the hairpin [6]. With the shorter hybrid this cannot occur in the direct form originally envisioned, but it appears that the formation of the hairpin, possibly in conjunction with the overall bending of the DNA within the transcription complex, sufficiently distorts the neighboring bp of the hybrid to achieve significant destabilization of, at least, the upstream end of the RNA–DNA duplex.

The RNA termination hairpin could also bind to a site on the polymerase and thus 'lock' the transcription complex into a fixed position along the template, inhibiting both backsliding and the release of the 3′-terminus of the transcript from the active site of the polymerase [36]. Thus, complex dissociation and transcript release probably occur directly from the terminator sequence itself, rather than after an intervening sliding event. This proposal is consistent with the notion that dissociation of the complex should occur while it is at its most destabilized position along the template since back-sliding, which would be accompanied by the shifting of the hybrid away from the maximally unstable rU·dA sequence, would serve to return the complex to a more stable state.

[5] Such hairpins in the RNA, in the absence of the vicinal rU sequence, are called 'pausing hairpins', since they also cause the complex to pause, although not to terminate [38]. Hairpin-induced pausing also plays an important role in the regulation of transcript elongation.

2.3.6. Conclusions

In this section we have summarized the various possibilities and reaction pathways that are, in principle, open to the transcription complex at each template position. We have suggested that the synthesis pathway is favored elongation proceeding at an average rate of ~30–35 nt/s (i.e. with an average dwell-time of approximately 30 ms) at most template positions under conditions of NTP saturation. Under these circumstances the next NTP binds to the substrate binding sub-site of the polymerase as soon as that sub-site is vacated by the 3′-terminus of the newly extended RNA chain. The efficient transfer of the 3′-terminus of the chain to the product binding sub-site permits the nucleotide addition cycle to continue rapidly and effectively. Normally, if an incorrect NTP binds at the substrate binding sub-site, the resulting lack of hydrogen bonding complementarity with the underlying template position destabilizes the NTP binding interaction and leads to prompt ejection of the incorrect NTP from the substrate sub-site. Furthermore, binding of the 'incorrect' NTP results in 'mispositioning' of the NTP relative to the end of the growing chain within the product binding sub-site [10], further disfavoring incorrect nucleotide incorporation.

If misincorporation does occur the result is to further disfavor elongation and concomitantly (in part because of incorrect hydrogen bonding to the template) to favor release of the 3′-end of the transcript from the active site of the polymerase. This leads to stalling of transcript elongation and favors formation of arrested (dead-end or slid-back) complexes. In turn, this results in extrusion of the 3′-end of the RNA from the transcription complex, thus leading into the editing pathway. The actual editing process involves the cleavage of one or more residues from the 3′-end of the chain by the polymerase, with this cleavage process being activated by the GreA or B transcription factors in *E. coli* or by the SII factor in eukaryotes. The new 3′-end of the shortened chain then rebinds to the active site of the polymerase at the slid-back template position, and elongation resumes.

Pausing of the elongation reaction can be achieved by reducing the concentration of the next-required NTP, which also leads to an increased probability of a slid-back state of the transcription complex, although here this occurs without introducing a misincorporated residue. Subsequent provision of the next-required NTP permits elongation, although paused complexes have an increased probability of following the arrest pathway as the length of the pause is increased. The addition of high concentrations of PP_i to such a paused complex can drive chain shortening by pyrophosphorolysis. We note that this requires that the 3′-end of the chain be positioned in the substrate binding sub-site of the polymerase, since this process represents the chemical reverse of the elongation reaction (Eq. (1)). Therefore, chain shortening by pyrophosphorolysis, like elongation, is disfavored for a chain containing a 3′-terminal misincorporated nucleotide residue.

Finally, the termination pathway is favored at intrinsic terminator sequences at which the elongation complex is destabilized by the transient presence of an RNA–DNA duplex containing a run of rU·dA bp. The complex pauses at this point, probably due in part to the formation of a termination hairpin within the nascent RNA. Since the heights of the free energy barriers to elongation and termination at terminator positions are comparable, the relative usage of these competing pathways is easily regulated by small changes in the rate of elongation or the stability of the elongation complex (Fig. 2).

3. Is there more than one functional form of the active elongation complex?

In concluding this overview of the reaction pathways that are potentially available to a transcription complex at any particular template position, we now ask explicitly if more than one functional form of the complex can be involved in transcript elongation. This question is timely because the recent literature contains numerous allusions to more than one functional form of the elongation complex (e.g. see Ref. [38]), with the definite suggestion that these forms may retain their functionally distinct characteristics (altered rates, inability to recognize termination and pausing sites, etc.) while transcribing for great distances along the DNA template [39–41]. More

fundamentally, however, this question is of major importance because the demonstrated existence of such states would force us to adopt a very different and much more complicated view of the regulatory mechanisms that control the process of transcript elongation.

It is certainly clear from the information summarized above that a number of functionally and conformationally distinct states of the transcription complex can exist at a given template position[6], and that these complexes differ from one another in at least some important molecular details. Thus the existence of this type of transcription complex heterogeneity is not at issue. Rather, we ask here whether all these forms, in becoming activated to resume transcription at any given template position, are 'reset' into one synthetically active form that is conformationally and functionally defined only by interactions that are actually 'present' at that template position.

Given the fact (e.g. see Ref. [3]) that there can be more than one form of the transcription complex at a given template position, we have assumed for some time [5,6,42] (based in part on the 'competing-transition-states' model for the various reaction pathways) that these forms 'equilibrate' at each template position before further synthesis can occur. An important consequence of this assumption is that the functional properties of the transcriptionally active complex will depend only on the interactions with components (i.e. the template, the nascent RNA, and any bound protein factors) that are actually present and in contact with the transcribing polymerase at the template position at issue. This assumption underlies most of our ideas of transcription regulation, including the notion that the elongation properties of a transcription complex at a particular position on the DNA template depend only on the local DNA sequence and are not changed by moving this DNA within the genome (e.g. by recombination).[7]

These ideas have lead to a class of models that we have termed 'equilibrium', or more precisely, 'rapid interconversion' models of the transcription elongation process [4]. These models state that the synthetic reaction at a given template position occurs from a single conformationally defined state (or rapidly interconverting set of states) that is regulated by the interactions that are actually present at that position (see above). Of course the length of time that a complex remains at a given template position will determine whether some fraction of the complex population has further equilibrated into less probable (e.g. arrested or slid-back) states. If this has occurred, the equilibrium model posits that these forms must re-equilibrate with the elongation-competent form(s) before exiting (via synthesis) from that template position. We note that this notion is consistent with (e.g.) the results of Artsimovitch and Landick [19], who have shown that transcription complexes elongate from a pause position at two different rates. The equilibrium model would suggest that the slow rate reflects the rate-limiting step for the 'reactivation' of complexes that have partitioned into slid-back or other inactive forms, and must be reset into a synthetically active form before elongation can proceed. The rapidly synthesizing fraction would then represent the population that remains in the elongation-competent form at the end of the pause.

[6] It is important to define what we mean when we say that a transcription complex is 'located' at a given template position. In Fig. 1 and Fig. 4 we mean the DNA template position with which the product binding sub-site of the polymerase is physically aligned. However, this cannot always be ascertained. The alternative is to define the position of the complex as that to which the nascent RNA has been extended. This latter definition, of course, is independent of whether the polymerase has slid-back from, or remained at, that position.

[7] The fragment of DNA that is moved in this 'thought experiment' must be long enough to include any upstream sequence that may be encoded into the nascent transcript and, for example, binds a regulatory protein that then interacts with the polymerase at a downstream template position RNA looping. The phage λ N-protein-dependent antitermination system provides a case in point [42]. The critical distinction is that 'rapid interconversion' models posit that the functional and conformational properties of synthetically active elongation complexes depend only on actual interactions at the template position of interest, while 'stable heterogeneity' models propose that 'upstream experiences' may result in changes that remain functionally stable within the transcription complex and may be retained through many single nucleotide addition cycles, despite the fact that the interactions or sequences that produced them are no longer present.

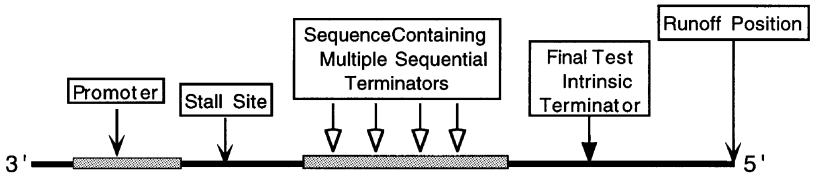

Fig. 5. Schematic of termination experiments designed to discriminate between 'equilibrium' ('rapid interconversion') and 'stably heterogeneous' models of active elongation complexes. The horizontal line represents the template strand of the DNA construct used in these experiments. The 3′- and 5′-ends of the template strand are indicated, as are the locations along the template of the promoter, the stall site (position to which elongation could be extended with 3 NTPs and at which elongation complexes were incubated), the sequence within which the multiple sequential upstream intrinsic or Rho-dependent terminators were placed, the positions of the final downstream 'test' terminator, and the runoff position.

The alternative notion leads to the 'stable heterogeneity' class of models described above. These models posit that stably heterogeneous populations of complexes are formed at upstream template positions, perhaps even at the promoter, and retain their distinct functional characteristics in the form of a 'molecular memory' while transcribing across many template positions. Stable heterogeneity models have been proposed explicitly [39] or considered implicitly [38,40,41] in many recent reports in the literature.

3.1. Experimental tests of stable heterogeneity models of active elongation complexes

We have recently conducted a series of quantitative experiments, using measured intrinsic TE to assay function, that were designed to discriminate between these classes of models [4]. We constructed DNA templates containing a common promoter, followed by a sequence that permitted the nascent RNA transcript to be extended to a defined stall site located at transcript position +24 (i.e., well into the elongation phase) by using only three of the four required NTPs. Here the transcription complexes were initially accumulated and then the fourth required NTP was added, permitting transcription to proceed to the end of the template. The general form of the template constructs used in these experiments is outlined in Fig. 5.

In these experiments elongation complexes were required to transcribe through a series of identical intrinsic terminators to reach the runoff position at the end of the linear template, and the TE at each terminator was measured. If the initial population of transcription complexes at the stall site contained functionally distinct and stable populations that differed in their ability to read through a terminator, we would expect the initial complexes to be fractionated into 'termination-resistant' and 'termination-prone' populations as a consequence of transcription, with the presence of such functionally diverse populations being manifested by a progressively decreasing apparent TE as the complexes transcribed through a series of identical terminators. This follows because the termination-prone complexes would be expected to dissociate preferentially at each intrinsic terminator, thus the fraction of the remaining population consisting of termination-resistant complexes should increase.

These experiments showed unequivocally that the measured TEs of the complexes were the same (within the standard error of the measurements) at all the identical terminators, regardless of the fraction of the initial complexes that actually reached a given terminator. This result is, of course, inconsistent with the expectations of stable heterogeneity models, but is completely consistent with equilibrium or rapid interconversion models. These results were extended by attempting to fractionate the stalled elongation complexes into functionally distinct classes by other means, such as inactivating a large proportion of the stalled complexes by lengthy incubation at room temperature, or by shorter incubations at elevated temperatures. Templates were also constructed that required complexes to transcribe through different intrinsic or Rho-dependent terminators before reaching a downstream (test) intrinsic terminator

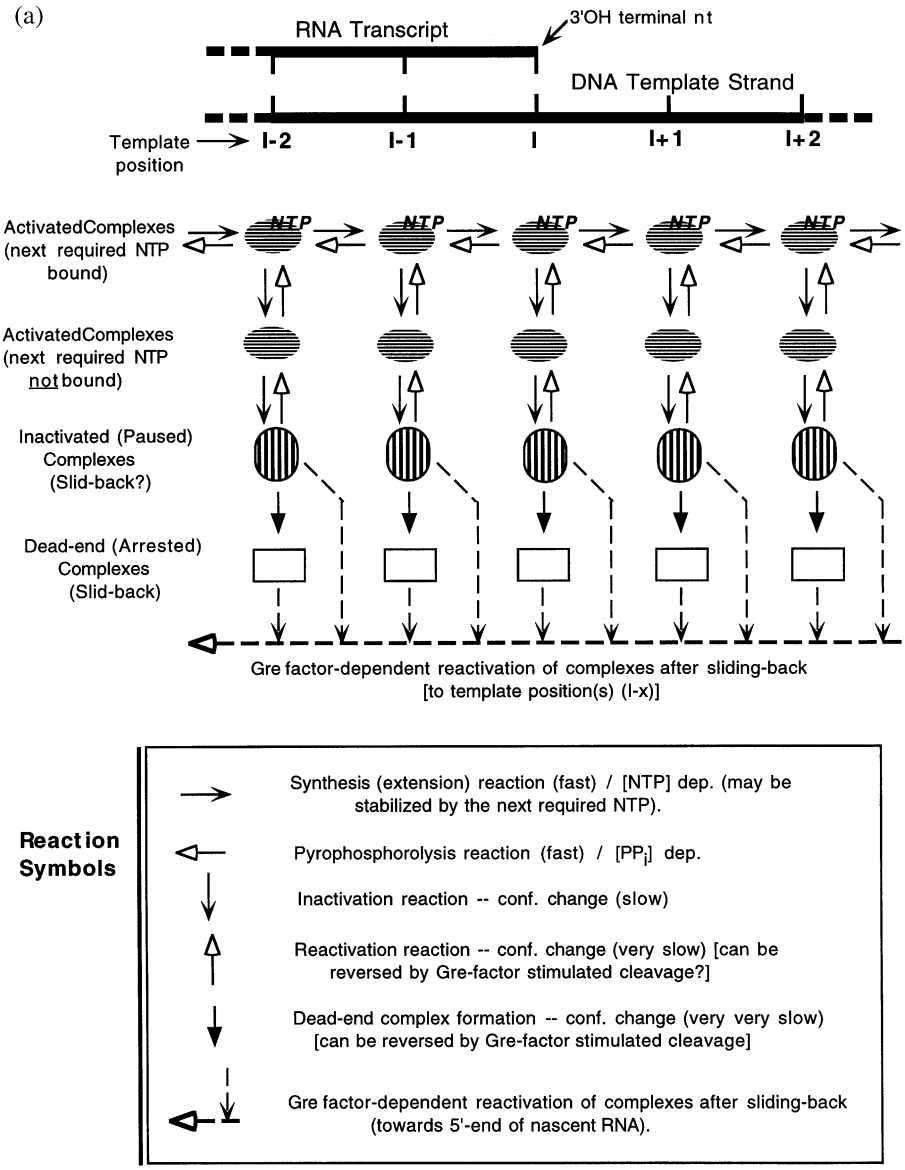

Fig. 6. Models for the multiple possible forms of the transcription complex and their interconversion interactions at defined template positions. Schematic of multiple possible forms of the transcription complexes located at template positions I−2, I−1, I, I+1, and I+2. (a) Model with a single active elongation-competent form of the transcription complex (top row, horizontal arrows). This form can interconvert (vertical arrows) into various forms that are synthetically inactive, proceeding through an (otherwise active) conformation without an NTP bound in the substrate binding sub-site to a reversibly inactivated form (represented as a slid-back complex with the 3′-terminus of the transcript no longer in the active site), and finally to a dead-end complex. Reaction symbols are defined in the box at the bottom of the figure. (b) The non-essential allosteric model of Foster et al. [41], showing two synthetically active forms of the elongation complex at each template position. Reaction symbols are defined in the box at the bottom of the figure. The same progression of inactive forms is shown as in a. (See text for further details and comparison of models.)

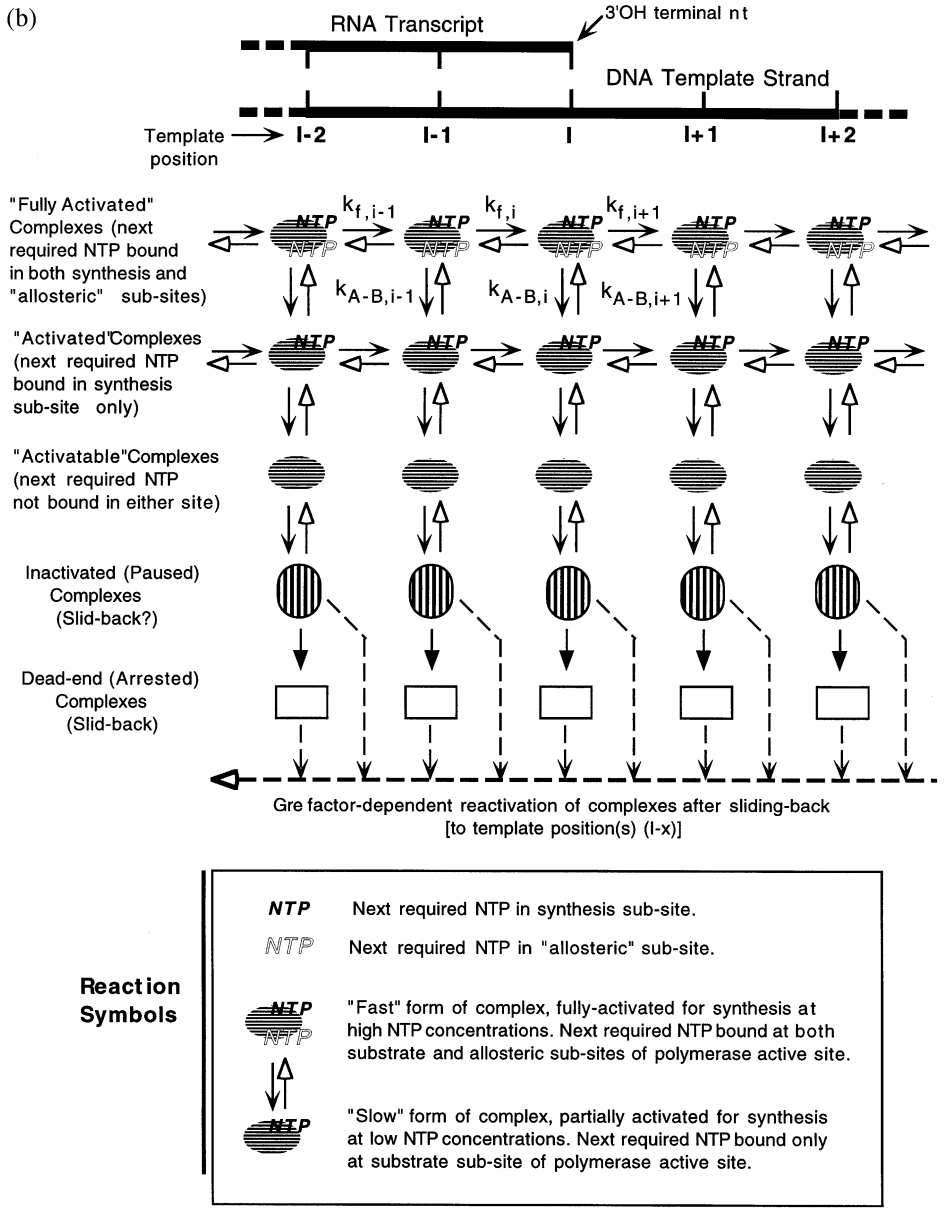

Fig. 6. (Continued).

(Fig. 5). In no case were we able to fractionate the complexes into functionally stable heterogeneous populations.

The details of these experiments have been presented and discussed elsewhere [4], and show that actively synthesizing transcription complexes are 'reset' into a single active form (or a rapidly interconverting population of forms) at each sequential terminator along a given template. We note, however, that these terminators were separated by an average of 40 bp in our template constructs, and what we really wanted to know is

whether the actively synthesizing transcription complexes reached conformational and functional equilibrium at each template position located between the test terminators. By examining the kinetics of our experiments and considering the results of earlier experiments in our laboratory and elsewhere (for details see Ref. [4]), we have been able to extend the above conclusion to argue that equilibrium or rapid interconversion models for the active complexes apply not only at sequential terminators, but are also likely to apply at most or all of the template positions that separate the terminators.

3.2. Further consideration of rapid interconversion models

The results of the experiments described above, together with reconsideration of the explanations of the experiments cited by others to support stable heterogeneity models (see Ref. [4]), have permitted us to conclude that actively synthesizing elongation complexes do indeed equilibrate to a single form (or to two or more rapidly interconverting forms), and that the outcome of synthesis at a given template position depends only on interactions with regulatory factors and ligands that are actually present at that template position. The simplest model of such a mechanism, illustrated schematically for several adjacent template positions, is presented in Fig. 6a. As set up this model shows that while different forms of the transcription complex that lead into different reaction pathways at a given template position can coexist at each position, extension of the transcript to the next template position can only occur via a single form of the complex to which the other forms present must revert before elongation can proceed.

An interesting model of elongation that introduces an additional role for the next NTP as a non-essential allosteric regulatory component has recently been put forward by Foster et al. [41] (see also comments on this model by Sousa [43]). In an important experiment these workers showed that the kinetics of transcript elongation at a specific template position seemed to display a higher order dependence on the concentration of the next-required NTP than the first order dependence that would be expected from conventional Michaelis-Menten kinetics. An even more striking observation made in this study was that, at low concentrations of the next-required NTP, the addition of a non-hydrolyzable analogue of the next-required NTP actually increased, rather than decreased, the rate of transcript elongation. Moreover, this effect of the nucleotide analogue was template-position-specific, meaning that the effect of the ATP analogue used in these studies was observed only at positions at which ATP was the next-required NTP.

Foster et al. [41] interpreted these results by suggesting that (in addition to the substrate and product sub-sites of the active site of the polymerase shown in Fig. 3a) the polymerase may contain an additional allosteric regulatory site that can bind an additional molecule of the next-required NTP and further activate the elongation complex, although filling this site is not essential for synthesis. They proposed that elongation complexes might synthesize RNA either as a fast 'activated' form with a filled allosteric site, or as a slower 'unactivated' form with an empty allosteric site. A representation of this model is shown in Fig. 6b, in which we have termed the fast form of RNAP 'fully activated' and the slower form 'activated', in keeping with the nomenclature that we have used in this article.

Foster et al. [41] suggested that the fully activated and activated forms of the elongation complex might interconvert slowly on the time scale of the single nucleotide addition cycle, thus providing a possible explanation for the stably heterogeneous populations of polymerases that had been proposed by others. Our results [4] cited above seem to rule out this proposal, along with other stable heterogeneity models, but do not rule out the proposed non-essential allostery model [41] if the two putative forms of the synthetic complex interconvert rapidly. In terms of the model shown in Fig. 6b, this means that the forward rate constants for the (horizontal) synthesis reactions (k_f) at the various positions, for both proposed forms of the elongation-competent complex, must be significantly larger (faster) than the rate constants for the (vertical) reactions of the complexes locat-

ed at a defined template position that convert these active complexes to inactive forms.

We believe that there may be a simpler explanation for the Foster et al. results [41]. Instead of an actual non-essential allosteric NTP binding site, for which, as the authors admit, there is presently no structural evidence, we propose a 'competitive inactivation' model that follows from the competing-reaction-pathways ideas discussed above. At low NTP concentrations in such models (see also Ref. [9]), the active elongation complex is destabilized by the protracted absence of the next-required NTP from the NTP binding sub-site. A complex carrying an empty substrate sub-site is much more likely to 'decay' into an inactive state than one with a filled site, which is why, in Fig. 6a and b, we show such 'empty' complexes as lying on the path to reversibly inactivated, and finally to dead-end, forms of the transcription complex.

Complexes with an empty NTP binding site, but otherwise still in the synthetically active form, are thus likely to exist as precursor forms of the various paused (or inactivated) states described in Section 2. In contrast, at saturating concentrations of NTPs the active elongation complex is stabilized by the rapid binding of the next-required NTP at the substrate binding sub-site of the polymerase. Fast exchange of the analogue with the next-required NTP would then result in the reasonably fast elongation rates observed, even at low NTP concentrations. At the same low NTP concentrations, but in the absence of the relevant analogue, elongation complexes are much more likely to decay into inactive states from which return is slow (Fig. 6a and b), resulting in a decreased rate of apparent transcript elongation.

This explanation might also serve to rationalize the much earlier observation of Chamberlin et al. [44] that the apparent K_m for the insertion of the next-required NTP appeared to vary widely as a function of template sequence and position. This result has always seemed perplexing because, in the absence of other complications, the rate at which polymerases insert NTPs into elongating NA chains at sub-saturating NTP concentrations should depend solely on whether or not the candidate NTP exhibits Watson-Crick complementarity with the template DNA strand. Thus a dependence of the apparent K_m on template sequence seemed hard to explain. The 'competitive inactivation' model suggested here can provide a simple explanation for these results.

This follows because values of K_m are measured in experiments in which the apparent rate of the reaction is determined as a function of decreasing NTP concentration, and at substrate concentrations in the vicinity of K_m (for systems where K_m is approximately equal to K_d) the apparent rate of the reaction will decrease because the substrate binding sub-site is only partially filled. Under these conditions of lower substrate binding sub-site occupancy, active transcription complexes (unlike most enzymes) become more prone to conversion into the other inactive forms that are favored at pausing and arrest positions, and this tendency will, of course, depend on template and transcript sequence. Therefore, the wide variations of the apparent K_m with template sequence observed by Kingston et al. [44] may actually reflect the fraction of transcription complexes that have (e.g., at pausing sites) assumed an inactive form, rather than serving as a direct measure of K_m (i.e., K_d or site occupancy) in the conventional sense. This proposal also explains, of course, how a non-hydrolyzable analogue of the next-required NTP might activate transcription at low NTP concentrations, since by increasing the occupancy of the substrate sub-site this analogue would also stabilize the active form of the elongation complex.

4. Final comments on regulation and evolution

The above discussion has dealt primarily with the transcription complexes of *E. coli*—a relatively obscure bacterium (except to molecular biologists)—and with the interactions of these complexes with regulatory factors, including both those of the host and others that are encoded by certain bacteriophage sequences that can become stably incorporated into the *E. coli* genome. The initial rationale for this focus, of course, was that *E. coli* was a tractable organism, and at the same time the ideas of evolution provided faith that its macromolecular components and organelles, as well as many aspects of its operating principles, would

turn out to be carried over relatively intact into higher organisms when these became experimentally accessible.

The recent explosive progress in eukaryotic molecular and cell biology has strikingly vindicated this view. Clearly eukaryotes have developed more elaborate complexes and mechanisms to deal with the fact that they are larger than *E. coli* and have more genes. They also engage in development and need added levels of inter-cellular and inter-tissue regulation to provide levels of control of gene expression that are commensurate with their larger genomes and developmental requirements. One striking example of such added levels of complexity follows from the appearance of the nuclear membrane. This entity separates the nuclear machinery of DNA replication and rearrangement, as well as that of RNA transcription, editing, and splicing, from the protein synthesis machinery of the cytoplasm in eukaryotic cells, thus providing opportunities for the incorporation of additional layers of gene expression control.

Nevertheless, all organisms must handle and utilize DNA, RNA, and proteins. From this perspective it may not be surprising that the polymerases of prokaryotes and eukaryotes are virtually identical in their central structural aspects, including features alluded to in the preceding discussion such as the size of the transcription bubble and the length of RNA–DNA hybrid within the transcription complex [2]. This follows, of course, because DNA, RNA, proteins, and all other cellular and extra-cellular macromolecules impose their own structural, dynamic, and stability limitations on how they can be manipulated and regulated in aqueous solution and at physiological temperatures, regardless of the size and complexity of the organism within which they are constrained to operate.

These molecular limitations, together with the evolutionary skein that connects all organisms, also constrain the regulatory principles that can be applied to the control of gene expression. An overriding theme is the cooperative self-assembly of complicated macromolecular machines. The controlling principles here are: (i) templating mechanisms, including the involvement of 'codes' and adaptors to permit 'communication' between sets of molecules with incompatible recognition systems; (ii) intra- and inter-molecular cooperativity, to permit macromolecules and their complexes to develop stable and defined conformations that are, nevertheless, easily driven into other forms across sharp transition boundaries by small changes in the interacting components or the environment; and (iii) the central role of water and the aqueous ionic environment in setting and controlling the interaction potentials of the component macromolecules in folding and assembly. A related requirement is that the stability of biological complexes must be limited in order to permit biological molecules and their complexes to find (within the times permitted by the cell cycle) their free energy minima and escape from metastable 'trapped' states, either by spontaneous rearrangement or with the assistance of various types of chaperone complexes.

These limitations on stability provide opportunities for the evolution of biological regulatory pathways as well. Thus the delicate balance of the regulatory machinery permits the insertion, by evolution, of additional control pathways into an organism without destroying the fundamental mechanism(s) of the processes involved. The reaction pathways available to the transcription complex, which are shown schematically in Fig. 1 and have been described in this article, provide a good case in point. We note that the balanced nature of these pathways permits the rapid and stable synthesis of RNA transcripts on the DNA template, and at the same time permits easy switching to other pathways when, for example, a nucleotide residue has been misincorporated or a termination signal is encountered.

One might argue that it would be 'safer' to make functional macromolecules or macromolecular complexes more stable, and thus avoid the dangers inherent in such delicately balanced states (e.g. see Ref. [39]). However, this scenario also results in the loss of the very flexibility that the evolutionary development of new pathways and new regulatory and signaling systems requires, and eventually leads organisms into evolutionary 'dead-ends'.

Finally, returning to the central theme of transcription regulation via kinetically competing reac-

tion pathways, the discussion in Section 3 (in which we ask whether there is more than one stable actively synthesizing form of the transcription complex) provides another example of the advantages of marginal stability for regulation and evolution. If synthetically active transcription complexes could display 'macromolecular memories' that transcend the dwell-time at a given template position, this would require the simultaneous development of endlessly branching cascades of regulatory mechanisms to control the function of these stably heterogeneous populations of transcription complexes as they move from one template position (and regulatory environment) to the next. Clearly a population of transcription complexes that must equilibrate within their dwell-time at a given template position comprises the simpler alternative, since then a single DNA-sequence-dependent set of regulatory signals can suffice for the control of all the active elongation complexes involved in gene expression.

Acknowledgments

In addition to the many valuable discussions with John Schellman that have been acknowledged in the dedication, the authors would also like to thank their laboratory colleagues, and friends and colleagues elsewhere, for many valuable discussions and insights that have helped to frame the ideas presented here. The work of our laboratory that is described in this article has been supported over the years by NIH grants GM-15792 and 29158 (to PHvH). The authors are also grateful to the American Cancer Society for a postdoctoral fellowship (ZP) and for long term support via an American Cancer Society Research Professorship (PHvH).

References

[1] K.S. Murakami, S. Masuda, E.A. Campbell, O. Muzzin, S. Darst, Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex, Science 296 (2002) 1285–1290.

[2] R.H. Ebright, RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II, J. Mol. Biol. 304 (2000) 687–698.

[3] D.A. Erie, O. Hajiseyedjavadi, M.C. Young, P.H. von Hippel, Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription, Science 262 (1993) 867–873.

[4] Z. Pasman, P.H. von Hippel, Active *E. coli* transcription elongation complexes are functionally homogeneous, (submitted for publication).

[5] P.H. von Hippel, An integrated model of the transcription complex in elongation, termination, and editing, Science 281 (1998) 660–665.

[6] P.H. von Hippel, T.D. Yager, Transcript elongation and termination are competitive kinetic processes, Proc. Natl. Acad. Sci. USA 88 (1991) 2307–2311.

[7] V. Mekler, E. Kortkhonjia, J. Mukhopadhyay, et al., Structural organization of RNA polymerase holoenzyme and the RNA polymerase-promoter open complex, Cell 108 (2002) 599–604.

[8] S.A. Darst, Bacterial RNA polymerase, Curr. Opin. Struct. Biol. 11 (2001) 155–162.

[9] R. Guajardo, R. Sousa, A model for the mechanism of polymerase translocation, J. Mol. Biol. 265 (1997) 8–19.

[10] D.A. Erie, T.D. Yager, P.H. von Hippel, The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective, Annu. Rev. Biophys. Biomol. Struct. 21 (1992) 379–415.

[11] J.R. Levin, B. Krummel, M.J. Chamberlin, Isolation and properties of transcribing ternary complexes of *Escherichia coli* RNA polymerase positioned at a single template base, J. Mol. Biol. 196 (1987) 85–100.

[12] M. Kashlev, E. Martin, A. Polyakov, K. Severinov, V. Nikiforov, A. Goldfarb, Histidine-tagged RNA polymerase: dissection of the transcription cycles using immobilized enzyme, Gene 130 (1993) 9–14.

[13] G.A. Kassavetis, P.G. Zentner, E.P. Geiduschek, Transcription at bacteriophage T4 variant late promoters. An application of a newly devised promoter-mapping method involving RNA chain retraction, J. Biol. Chem. 261 (1986) 14256–14265.

[14] T.A. Rozovskaya, A.A. Chenchik, R.S. Beabealashvilli, Processive pyrophosphorolysis of RNA by *Escherichia coli* RNA polymerase, FEBS Lett. 137 (1982) 100–104.

[15] J.J. Hopfield, Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity, Proc. Natl. Acad. Sci. 71 (1974) 4135–4139.

[16] N. Komissarova, M. Kashlev, Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded, Proc. Natl. Acad. Sci. USA 94 (1997) 1755–1760.

[17] E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, The RNA–DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase, Cell 89 (1997) 33–41.

[18] T.C. Reeder, D.K. Hawley, Promoter proximal sequences modulate RNA polymerase II elongation by a novel mechanism, Cell 87 (1996) 767–777.

[19] I. Artsimovitch, R. Landick, Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals, Proc. Natl. Acad. Sci. USA 97 (2000) 7090–7095.

[20] S. Borukhov, A. Polyakov, V. Nikiforov, A. Goldfarb, GreA protein: a transcription elongation factor from *Escherichia coli*, Proc. Natl. Acad. Sci. USA 89 (1992) 8899–8902.

[21] S. Borukhov, V. Sagitov, A. Goldfarb, Transcript cleavage factors from *E. coli*, Cell 7 (1993) 459–466.

[22] D. Reines, M.J. Chamberlin, C.M. Kane, Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro, J. Biol. Chem. 264 (1989) 10799–10809.

[23] K.S. Wilson, P.H. von Hippel, Stability of *Escherichia coli* transcription complexes near an intrinsic terminator, J. Mol. Biol. 244 (1994) 36–51.

[24] T. Platt, Rho and RNA: models for recognition and response, Mol. Microbiol. 11 (1994) 983–990.

[25] J.P. Richardson, Structural organization of transcription termination factor Rho, J. Biol. Chem. 271 (1996) 1251–1254.

[26] C.A. Brennan, A.J. Dombroski, T. Platt, Transcription termination factor Rho is an RNA–DNA helicase, Cell 48 (1987) 945–952.

[27] D.J. Jin, R.R. Burgess, J.P. Richardson, C.A. Gross, Termination efficiency at Rho-dependent terminators depends on kinetic coupling between RNA polymerase and Rho, Proc. Natl. Acad. Sci. USA 89 (1992) 1453–1457.

[28] A.Q. Zhu, P.H. von Hippel, Rho-dependent termination within the *trp t′* terminator. I. Effects of Rho loading and template sequence, Biochemistry 37 (1998) 11202–11214.

[29] A.Q. Zhu, P.H. von Hippel, Rho-dependent termination within the *trp t′* terminator. II. Effects of kinetic competition and processivity, Biochemistry 37 (1998) 11215–11222.

[30] K.M. Walstrom, J.M. Dozono, S. Robic, P.H. von Hippel, Kinetics of the RNA–DNA helicase activity of *Escherichia coli* transcription termination factor. 1. Characterization and analysis of the reaction, Biochemistry 36 (1997) 7980–7992.

[31] K.M. Walstrom, J.M. Dozono, P.H. von Hippel, Kinetics of the RNA–DNA helicase activity of *Escherichia coli* transcription termination factor. 2. Processivity, ATP consumption, and RNA binding, Biochemistry 36 (1997) 7993–8004.

[32] T.J. Allison, T.C. Wood, D.M. Briercheck, F. Rastinejad, J.P. Richardson, G.S. Rule, Crystal structure of the RNA-binding domain from transcription termination factor, Nat. Struct. Biol. 5 (1998) 352–356.

[33] D.M. Briercheck, T.C. Wood, T.J. Allison, J.P. Richardson, G.S. Rule, The NMR structure of the RNA binding domain of *E. coli* factor suggests possible RNA–protein interactions, Nat. Struct. Biol. 5 (1998) 393–399.

[34] C.E. Bogden, D. Fass, N. Bergman, M.D. Nichols, J.M. Berger, The structural basis for terminator recognition by the transcription termination factor, Mol. Cell 3 (1999) 487–493.

[35] R.R. Wei, J.P. Richardson, Identification of an RNA-binding site in the ATP binding domain of *Escherichia coli* by H_2O_2/Fe-EDTA cleavage protection studies, J. Biol. Chem. 276 (2001) 28380–28387.

[36] I. Gusarov, E. Nudler, The mechanism of transcription termination, Mol. Cell 3 (1999) 495–504.

[37] H. Yin, I. Artsimovitch, R. Landick, J. Gelles, Nonequilibrium mechanism of transcription termination form observations of single RNA polymerase molecules, Proc. Natl. Acad. Sci. USA 96 (1999) 13124–13129.

[38] R. Landick, Shifting RNA polymerase into overdrive, Science 284 (1999) 598–599.

[39] H.J. Harrington, R.B. Laughlin, S. Liang, Balanced branching in transcription termination, Proc. Natl. Acad. Sci. USA 98 (2001) 5019–5024.

[40] R.J. Davenport, G.J.L. Wuite, R. Landick, C. Bustamante, Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase, Science 287 (2000) 2497–2500.

[41] J.E. Foster, S.F. Holmes, D.A. Erie, Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation, Cell 106 (2001) 243–252.

[42] M.R. Van Gilst, P.H. von Hippel, Quantitative dissection of transcription control systems: N-dependent antitermination complex of phage lambda as regulatory paradigm, Methods Enzymol. 323 (2000) 1–3231.

[43] R. Sousa, A new level of regulation in transcription elongation?, Trends Biochem. Sci. 26 (2001) 695–697.

[44] R.E. Kingston, W.C. Nierman, M.J. Chamberlin, A direct effect of guanosine tetraphosphate on pausing of *Escherichia coli* RNA polymerase during RNA chain elongation, J. Biol. Chem. 256 (1981) 2787–2797.