Specificity mechanisms in the control of transcription

Peter H. von Hippel [*], William A. Rees [1], Karsten Rippe [2], Kevin S. Wilson [3]

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

Abstract

In this overview we analyze and illustrate the principles underlying some of the specificity mechanisms that control the initiation, elongation, and termination phases of transcription. Thermodynamic mechanisms dominate in the first steps of initiation, where promoters at various levels of activation can be considered to be in competition for a limiting supply of core RNA polymerase. In the later stages of initiation, as well as in elongation and termination, the regulatory mechanisms that control specificity are largely kinetic, involving rate competition between branching reaction pathways where the outcome depends on the rates (and equilibria) of reaction and interconversion of different forms of the transcription complex. Elongation complexes are very stable at most positions along the DNA template, meaning that only RNA chain elongation (and editing) can occur at these positions. However, the stability of transcription complexes decreases abruptly when termination sequences are encountered, and here the outcome can be easily switched between elongation and termination (RNA release) by minor changes in the relative rates of these competing processes. *Cis* effectors, defined as sites at which regulatory proteins bind to upstream activation loci on either the DNA or the nascent RNA, play important roles in the control of both initiation and of the elongation–termination decision. Examples, drawn from studies of phage λ N-dependent antitermination and *E. coli* rho-dependent termination processes, illustrate the flexibility and additivity of regulatory components within control mechanisms in transcription that involve multiple determinants. The generality of such regulatory principles are stressed.

Keywords: RNA polymerase; Looping; DNA looping; RNA looping; Nucleotide addition; Transcript initiation; Elongation; Termination

1. Introduction

Bill Harrington loved a good specificity story, and whenever we met we used to regale one another with such tales [4]. Bill particularly enjoyed examples of what one might call 'physical biochemical leverage',

---

* Corresponding author.
[1] Present address: Howard Hughes Medical Institute, National Jewish Center for Immunology and Respiratory Medicine, 1400 Jackson Street, Denver CO 80206, USA.
[2] Present address: German Cancer Research Center, Department: Biophysics of Macromolecules, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.
[3] Present address: Department of Biology, Sinsheimer Laboratories, University of California, Santa Cruz, CA 95064, USA.

---

[4] The comments and reminiscences concerning Bill Harrington in this section are by PHvH; for more context for these comments see "Remembrances of Bill Harrington" (the first article in this issue).

in which a minor change in a thermodynamic or kinetic parameter pushes a reaction into a different free energy minimum or down a different reaction pathway and often, as a result of an amplification or cascade mechanism, leads to an entirely different biological outcome. One such specificity mechanism that Bill and I studied together in the late 1950s had to do with the central role of the proline residue (as a consequence of its constraining steric effects on neighboring residues of the polypeptide chain) in controlling the formation of the collagen helix and regulating many of the resultant interactions of collagen molecules in forming connective tissue and initiating bone mineralization (some aspects of these studies are reviewed [1]). Later research into force generation models involving helix–coil transitions in the 'hinge' region of the myosin molecule by Bill's laboratory seemed to illustrate another such mechanism [2,3].

The regulation of transcription by protein–nucleic acid interactions is rife with examples of such systems, and thus in memory of Bill (and because he would have enjoyed discussing them), we here describe some recent developments in the regulation of transcription that can be considered from this point-of-view [5].

2. Regulation in transcription

Molecular mechanisms in the regulation of transcription can be formulated in terms of the sequence of events involved in the passage of the core DNA-dependent RNA polymerase through an operon (defined as the stretch of DNA containing the regulatory and coding sequences of a particular gene) as the polymerase catalyses the template-directed transcription of DNA into RNA. This process is traditionally divided into three phases, corresponding to initiation, elongation, and termination of the nascent transcript. Some features of the overall process are outlined in Fig. 1.

Different overall regulatory mechanisms dominate the various phases of transcription. Regulation in initiation is basically inter-operon in nature, in the sense that the promoters of the various operons are in direct competition with one another for the limiting amount of free polymerase available. This competition is both thermodynamic and kinetic, since the ultimate objective is to control the rate of formation of specific transcripts. The thermodynamic part involves specific and competitive binding of the polymerase to promoter sequences, and the resulting equilibrium distribution of polymerases on the available promoters is determined both by intrinsic (sequence-dependent) affinities of the competing promoters and by the levels to which these binding affinities are enhanced by general and specific transcription activation factors. The kinetic component involves the rate at which the bound polymerases can 'melt-in' to form open promoter complexes, and then initiate RNA formation and 'clear' the promoter by moving into the elongation phase of transcription [6].

Transcription regulation becomes intra-operon in nature within the elongation phase, since control of the relative rates of elongation and termination here determines the rate of completion of a specific transcript. After the transcription complex has cleared the promoter and crossed the initiation–elongation boundary the competition for transcript initiation between different operons can begin again at the promoter. These global regulatory issues have been discussed in detail elsewhere [4,5].

3. Direct recognition and binding interactions

In the first step of transcription at any particular operon, the central DNA-dependent RNA polymerase and the associated activator (or repressor) proteins must find, recognize, and bind to the pro-

---

[5] The reader should appreciate that space limitations prevent us from providing a fully referenced overview of transcriptional regulation here. Thus only selected references are cited to document each point, and these should be consulted to locate the essential background literature.

---

[6] As shown in Fig. 1, prokaryote and eukaryotic promoters utilize somewhat different initiation processes, since eukaryotic promoters require protein activation factors to function at all in this inter-operon competition, while prokaryotic promoters can function without factors to some extent (except, of course, for the specificity subunit sigma, which is required as a part of the prokaryotic holopolymerase for effective initiation).
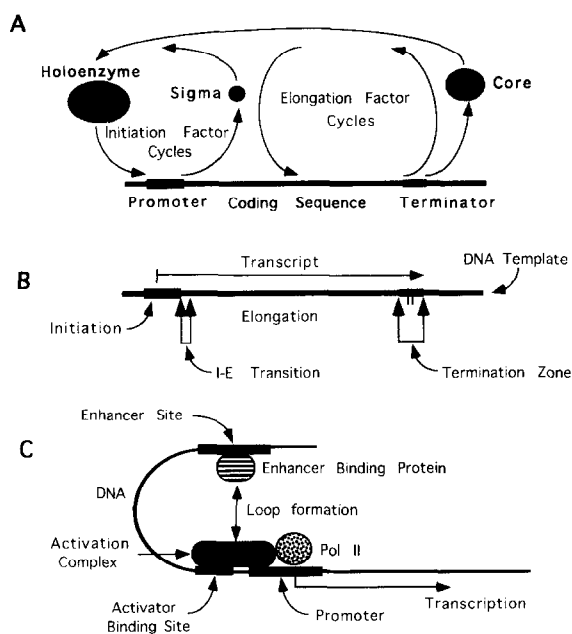
A

B

C

Fig. 1. Molecular events involved in the various phases of transcription in prokaryotes and eukaryotes. (A) In prokaryotes the 'core' polymerase, defined as the minimal enzyme capable of carrying out template-directed RNA synthesis, is directed to the regulatory DNA sequences located at the beginning of the gene (the promoter) by the presence of a specificity subunit (sigma) that binds to the core enzyme in solution to form the polymerase holoenzyme. The promoter is 'closed' (fully base-paired) during initial recognition by the holoenzyme. The holoenzyme then 'melts-in' to the promoter to form the 'open' promoter complex, in which the polymerase binds in a polar fashion to the template strand and transcript synthesis begins. The sigma subunit is released from the transcription complex at the end of the initiation phase, leaving the core polymerase (together with regulatory factors that cycle on and off and, with the local template sequence, control the rate of synthesis, stability, and conformation of the transcription complex) to synthesize along the template through the coding sequence until a terminator sequence is reached. At this point the complex is destabilized sufficiently so that termination (defined as release of the nascent RNA and the core polymerase) can occur and the overall transcription cycle can begin anew. (B) Shows the initiation, elongation, and termination phases of transcript formation, as well as the template positions of the initiation–elongation (I–E) transition (at which the sigma subunit dissociates) and the 'zone of opportunity' for termination (see text). (C) Activation of the initiation of messenger RNA synthesis by 'core' RNA polymerase II (pol II) in eukaryotes involves the binding of the regulatory subunits of the promoter activation complex at upstream activator sites on the DNA, as well as at more distant DNA sites called enhancers. The activation proteins that bind to both of these classes of sites are thought to contact the 'core' pol II subassembly located at the promoter by DNA looping (see text and Fig. 3).

moter and to specific DNA regulatory sequences located nearby. At this level a protein must be able to recognize a specific sequence of DNA base pairs, and to discriminate this sequence from all others. As a consequence these sequences must be long enough so that the probability of the random reoccurrence of the same sequence within the genome is significantly less than unity. In *E. coli* this sequence length is 12 base-pairs; in higher organisms (with larger genomes) it can be as long as 16–17 base pairs [6]. The information content of specific DNA sequences that bind regulatory proteins has been thoroughly studied [7,8].

How might a protein recognize a specific DNA sequence? Linear (one to one) amino acid–base pair recognition codes were rejected early in the development of field, and have been replaced by recognition mechanisms based on the favorable interactions of 'complementary' DNA–protein surfaces. These schemes were derived (at least in spirit) from the ideas of the multipartite recognition of substrates by the active sites of enzymes. Initial proposals in this area were thus framed by asking how the binding site of a protein, consisting of a number of specifically positioned amino acid side-chains and peptide backbone functional groups, might 'recognize' and bind to the functional groups that define a specific sequence of DNA base pairs [9]. Complementary matrices of hydrogen bond acceptors and donors, typically involving the functional groups of the major or minor grooves of double-stranded DNA [10], seemed best suited to provide the primary specificity of the recognition interaction, with stability (and perhaps a little additional specificity) being imparted by favorably placed charge–charge and 'hydrophobic' (water release) interactions [11,12].

Since the net free energy changes corresponding to the formation of such multiple (cooperative) recognition complexes are sensitive to the relative positioning of the interacting functional groups of the binding partners, it soon became clear that secondary interactions could improve and extend this primary recognition specificity. For example, the introduction of some flexibility into either the protein or the nucleic acid (or both) could lead to improved relative positioning of the functional groups involved, though, of course, such distortion from its unliganded free energy minimum conformation of

either partner must be 'paid for' in the coin of reduced binding free energy.

In addition, site-specific mutational alterations of defined base pairs within the regulatory target DNA sequences soon showed that not every base pair interacts directly with protein functional groups; clearly to some extent intervening (non-contacted) base pairs must be recognized indirectly as a consequence of the relative positions into which they place the functional groups of the base pairs that do interact directly with the protein [6,13]. Finally, it was also shown that hydrogen-bonded acceptors and donors can be geometrically 'extended' by bridging them with specifically placed water molecules [14].

These views, and their development as more protein–DNA complexes have been defined structurally, have led to a reasonably coherent picture of how individual proteins and DNA sequences might recognize one another (for recent structural overviews see [15–17]).

4. The problem of the other sites

Specificity in such interactions implies more than just recognition (and binding) of the protein to the correct base-pair sequence in isolation. In protein–DNA interactions in particular, where each base pair defines the beginning of an overlapping competitive ('wrong') DNA binding site, discrimination is perhaps more important than direct recognition of the correct site per se. A fortuitous property of recognition via complementary hydrogen-bonding matrices in water makes the necessary discrimination possible.

Since the surface hydrogen-bond donors and acceptors of the unliganded DNA and protein recognition sites are bound to water molecules, it was early recognized (e.g., see [18]) that such matrices cannot gain much stability by interacting with their correct partners, since such interactions would merely replace the free energies of protein–water and DNA–water hydrogen bonds for each unliganded partner with the free energies of an equivalent number of protein–nucleic acid and water–water hydrogen bonds after the complex had been formed. Thus relatively little net free energy is likely to be gained in the recognition reaction, except for that produced by the increased entropy of the released water

molecules (though see [19]). As a consequence, the stability of specific protein–DNA complexes must be largely ascribed to other sources of binding free energy, including the relatively non-sequence-specific charge–charge and water displacement reactions that occur when complementary protein and DNA surfaces come together in specific complex formation.

Nevertheless, such recognition mechanisms can still lead to functional specificity because incorrect interactions (between protein binding sites and 'wrong' DNA base-pair sequences) result in the apposition of non-complementary hydrogen-bonding interactions within the recognition matrices. Thus in wrong interactions the water molecules that provide hydrogen-bonding partners for the protein and DNA hydrogen-bond donors and acceptors within the recognition sites of the unliganded species are displaced without replacement by an appropriate protein or DNA acceptor or donor, resulting in the potential 'burial' (away from solvent water) of protein hydrogen-bond donors without a suitable DNA acceptor partner, and protein acceptors without a DNA donor.

Since each of these 'unmade' hydrogen bonds can provide an unfavorable free energy contribution of up to +5 kcal/mol (the free energy of a hydrogen bond formed in vacuo) to a wrong interaction, such contacts will destabilize the 'wrong' interactions relative to the 'right' ones, especially since the total stability ($\Delta\Delta G^\circ$) of a specific protein–nucleic acid complex (relative to its separated components free in solution) may be as little as -5 to -10 kcal/mol [12]. If too many 'unsatisfied' hydrogen-bonding interactions result, the protein may undergo a conformational change to a more 'general' non-specific binding form in which specific hydrogen-bonding groups are withdrawn from the protein–DNA interface, resulting in a non-sequence-specific complex that may be held together totally by electrostatic interactions [6,13].

Thus binding specificity is achieved in large measure not because the 'right' interactions are particularly stable, but because the 'wrong' interactions are (relatively) unstable. We note that most aspects of biological specificity, including protein folding, enzyme–substrate recognition, and organelle assembly, are ultimately likely to be based on the application of this water-dependent specificity principle.

5. How do proteins find their specific regulatory targets on the DNA?

In addition to the equilibrium discrimination conundrum posed above as "the problem of the other (DNA) sites", this issue also has kinetic implications, in that the multistep binding and rejection processes that must be involved in the finding and identifying of specific DNA target sites by regulatory proteins could make this process inordinately slow. In fact it is easy to imagine that this 'trial-and-error' process could take much longer than the minutes to hours available for an entire cell cycle in an actively dividing organism or tissue.

Early in vitro studies with the *lac* repressor of *E. coli* suggested that a mechanism exists to speed up this process, since the *lac* repressor seemed to find its operator target on the DNA genome at rates in excess of those predicted if the reaction were fully diffusion-controlled [20]. This idea is, of course, mechanistically impossible as stated, and was eventually interpreted by showing that specific DNA-binding proteins, such as *lac* repressor, can equilibrate between, and interact with, DNA binding sites in both a specific and a non-specific binding mode (and perhaps a pseudo-specific binding mode as well; see [6,13,21]), as described above.

Because of the purely electrostatic character of the non-specific binding interaction and the polyelectrolyte nature of DNA [11], such non-specifically bound (and positively charged, at least within the DNA binding site) proteins are held to the DNA largely by the mixing entropy of the counterions that are displaced from the 'condensed ion layer' that surrounds the DNA as a consequence of charge–charge interactions that form between the DNA phosphates and basic side-chains of the protein. As long as the rate of relaxation of the ion atmosphere exceeds the rate of linear diffusion, the DNA can be viewed as an 'isopotential surface' on which the protein can diffuse, exploring the DNA by one-dimensional 'sliding' and 'intersegment transfer' (see Fig. 2), permitting many translocations of the protein between binding sites on the DNA without dissociation (see [22]).

Both of these mechanisms can greatly increase the rate of regulatory target location on the DNA, since the DNA itself serves as a 'diffusion guide' for the
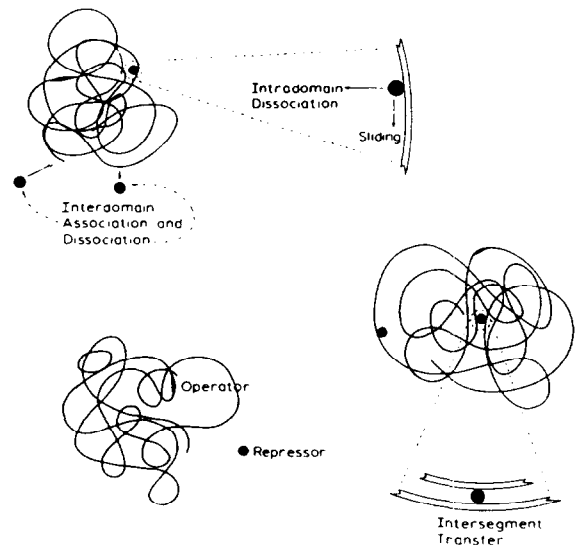

Fig. 2. Schematic view of the *E. coli lac* repressor interacting with a large operator-containing DNA molecule in dilute solution. The upper expanded view shows repressor bound to a segment of non-specific DNA, on which it can either 'slide' or engage in intradomain dissociation–association processes in seeking its specific (operator) target site. The lower expanded view shows a repressor molecule doubly bound to two DNA segments; this corresponds to the intermediate state in the intersegment transfer process. The DNA molecules are well separated into 'domains' under these conditions. (Modified with permission from [22].)

protein in its non-specifically bound state. Perhaps because of microscopic viscosity barriers encountered in moving along DNA (or between DNA strands), the calculated linear diffusion coefficient is smaller [23] than the one-dimensional coefficient of diffusion in free solution, but such a restricted random walk process can still significantly increase the overall rate of specific target location because such restricted diffusion along and within the domain of the double-helical DNA random coil so greatly decreases the volume of solution that needs to be searched for the regulatory target [24]. These mechanisms, and the experiments that were performed to sort them out, have been reviewed elsewhere ([22]; see also [25,26]).

6. Regulatory protein complexes extend specificity and DNA sequence recognition possibilities

Most transcriptional regulatory proteins, especially those involved in promoter activation, function

as parts of complexes that, in aggregate, bind to longer DNA sequences than do single regulatory molecules. However, these activation complexes often include 30–50 or more individual protein subunits [27,28], and if these proteins were all required to recognize their target DNA sequences individually, this would impose an insurmountable recognition burden on the system, given the limited specificity determinants included in even these longer DNA base-pair sequences. How then is the binding specificity of these complexes controlled by the DNA sequence?

It appears that many of these regulatory protein subunits may bind only to other protein components of the activation complex and not touch the DNA directly at all. Clearly the 'primary' regulatory proteins or protein sub-assemblies (including the RNA polymerase) that do bind directly to the DNA are central to 'nucleating' binding at a larger set of DNA sequences that are not completely specific. The binding sites for subsequent 'secondary' proteins may then depend on the formation of new multipartite binding sites involving both protein–protein and protein–DNA interactions, which further increases the binding specificity of the overall complex. Finally 'tertiary' proteins may use binding (recognition) sites that involve only protein–protein interactions to stabilize the entire activation assembly. Thus the overall binding specificity of a transcription activation complex can arise from a 'cascade' of recognition events, with the primary proteins binding relatively weakly to a larger set of partially specific DNA sequences, followed by 'coupled' and perhaps 'cooperative' binding of the secondary and tertiary proteins to add elements of both specificity and stability to increasingly focus the binding of the activation complex at the correct DNA sites [29] [7].

Such cascade mechanisms have multiple advantages. (i) They reduce the DNA determinants relative

to those that would be required if the DNA had to recognize each protein component of the activation complex individually. In such situations the specific DNA target recognition problem (the problem of the other sites; see above) applies not to the individual protein components, but to the specific activation complex as a whole. (ii) They also facilitate correct complex assembly in that the primary recognition proteins (or protein sub-assemblies) need bind only weakly to the initial DNA target sites (e.g., with K_a values of approximately 10^7 M^{-1}). Such weak binding assures relatively rapid dissociation (in milliseconds) from targets that do not include the combined protein–DNA determinants for secondary protein binding. Binding of secondary proteins will then stabilize nascent complexes at a subset of 'more correct' DNA target sites, and even further specificity and stabilization will occur as tertiary proteins interact with the partially assembled activation complex. (iii) Finally, interactions with secondary and tertiary proteins can be 'coupled' or 'cooperative', thus facilitating assembly as described above.

More detailed and fully worked out examples of such assembly processes for regulatory complexes have been presented elsewhere in terms of: (i) the assembly of a homo-protein regulatory system at a messenger RNA 'translational operator' that controls the transcription of the single-stranded DNA binding protein (gp32) of T4 bacteriophage [30]; and (ii) the hetero-protein assembly of the prokaryotic ribosome on its ribosomal RNA framework (see [29]).

7. Cis interactions contribute to recognition specificity via DNA looping

As implied in Fig. 1 and developed above, the burden of protein recognition and discrimination on the base-pair sequences at and immediately adjacent to the promoter can become immense in the formation of eukaryotic promoter activation complexes, with 30 or more activation subunits being involved [27,28]. Some of these regulatory subunits bind directly to the polymerase at the promoter, but many also bind at nearby (within 100–200 base pairs of the promoter) sites of defined sequence called upstream activation elements, or further afield along the DNA (several thousand base pairs) at binding

---

[7] ·Coupled' binding means that the binding of a secondary protein increases the apparent binding constant of the primary protein without an increase in the cooperativity of binding with respect to protein concentration. 'Cooperative' binding means that the apparent binding isotherm of the primary protein is sharpened as a consequence of secondary protein binding, indicating a higher than first order dependence on free protein concentration (see [29]).

RNA polymerase II
and associated factors

promoter

upstream element with
transcription factors
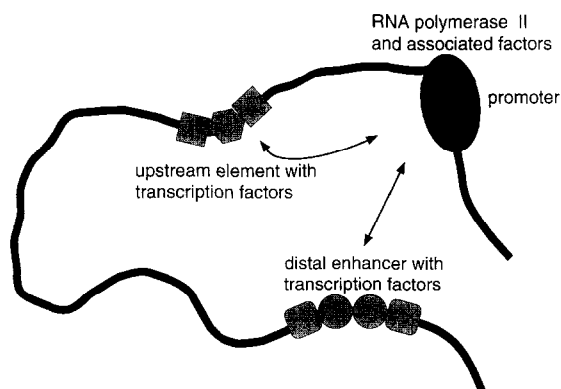
distal enhancer with
transcription factors

Fig. 3. DNA looping brings proteins bound at upstream activation and enhancer sites to the promoter. In this simplified view the many protein subunits that comprise the RNA polymerase II (pol II) transcription complex are not shown separately. Upstream activation elements are typically found at distances of 100–200 base pairs from the transcription start site at the promoter, whereas enhancers can be located up to several thousand base pairs upstream or downstream of their target promoters. (Reproduced with permission from [31].)

sites called enhancers. These DNA sites serve to tether and assemble parts of the activation complex and then utilize DNA looping to bring these components into contact with the portions of the activation complex located at the promoter (see Fig. 3). The quantitative consequences of such looping interactions on activation complex assembly (called *cis* activation effects since they involve interactions of proteins bound to DNA sites on the *same* DNA molecule [8]) will depend, at least in part, on the contour distance along the DNA molecule at which these regulatory sites are located relative to the promoter.

As described in a recent quantitative study of such *cis* tethering and looping effects on the local concentrations of protein components in the vicinity of the promoter [31], these mechanisms have numerous advantages over the assembly and binding of activator proteins directly from solution. Due to the stiffness of double-stranded DNA, such tethering and unconstrained looping may not significantly increase the concentration of individual activator proteins (or ac-

tivator subassemblies) above the levels at which they are present in solution. We calculate that the local concentration in the vicinity of promoters of proteins bound either to upstream activation sites or enhancer sites on the DNA is on the order of 10^{-8}–10^{-9} M, which is close to the estimated concentration of these proteins in solution within the nucleus.

How then might such looping result in *cis* activation by proteins bound to upstream activator or enhancer sites? Looping can provide several thermodynamic advantages in the assembly of an activation complex at the promoter. (i) It permits other specific DNA sequences to participate in, and to provide additional DNA sequence-dependent 'information' (in the form of protein binding sites) for, the assembly of a three-dimensional activation complex. Furthermore, if several protein subunits bind to these sites as organized sub-assemblies, this effectively reduces the entropy of mixing of the protein subunits and increases their binding affinities for the complex relative to the situation that would apply if all the subunits were free in solution within the nucleus. (ii) Binding of (e.g.) the TATA-binding protein (TBP) or specifically placed nucleosomes at sites near the promoter also provides an opportunity for sequence-specific DNA bending that can significantly increase (to 10^{-5} M or more; see [31]) the concentration at the promoter of regulatory proteins bound at upstream activator sites. (iii) In contrast, because several 'persistence lengths' of DNA typically intervene (see [31]), such bending may not greatly change the local concentrations at the promoter of proteins bound to the more distant enhancer sites. However, these enhancer sites do provide the opportunity for activator proteins to assemble as sub-complexes on specific DNA sequences and then to participate in the assembly of the final activation complex in a manner that is more flexible and less geometrically constrained than for proteins bound to the closer upstream activation elements [9].

The quantitative details of these attributes of *cis* binding to the formation of eukaryotic activator complexes have been presented elsewhere [31], but in the

---

[8] In these terms a *trans* interaction involves two DNA sites that are *not* on the same DNA molecule.

[9] Of course, this situation may be further constrained if the DNA between the enhancer and the promoter is encumbered with nucleosomes.

present context it is clear that such *cis* binding does provide the opportunity to introduce significantly increased amounts of DNA sequence information into the transcription initiation process, and also replaces the limited organizational flexibility of transcription complexes that are constrained by a strictly linear arrangement of proteins along the DNA near the promoter with the much greater variety of opportunities afforded by a defined three-dimensional assembly of protein and DNA components.

8. The role of kinetic competition in determining regulatory specificity

Except for certain simple repressor interactions that may decrease polymerase function at promoters by competitive occlusion of promoter sites, specific binding *per se* is not the ultimate goal of protein–protein and protein–DNA interactions in transcription. In general, specific binding merely serves as a necessary prelude to a series of other reaction steps that lead eventually to the production of the RNA transcripts in amounts and at times required by the regulatory context. Clearly initial binding must be tight enough to achieve recognition specificity as outlined above, but it must not be so tight as to preclude subsequent necessary events, such as the 'melting-in' of RNA polymerase to form the open-promoter complex, the specific (and polar, relative to the DNA) binding of the polymerase to the template strand, and the manifold other steps involved in transcription initiation [10].

Furthermore, these subsequent events must be achievable on the biological time scale under in vivo conditions, in that each reaction step that follows initial binding must be sufficiently accessible from the preceding state (i.e., sufficiently probable in a Boltzmann sense) to permit the overall reaction to go to completion within the time available. This again introduces kinetics into the problem, but in an interesting and powerful way, since it is generally not the
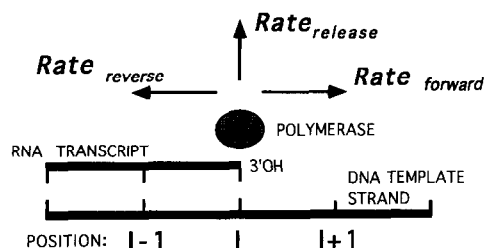

Fig. 4. Competitive reaction pathways potentially available to the transcription complex at each template position. $Rate_{forward}$ is a function of the intrinsic rate of elongation at each template position ($V_{max,NTP,I}$) and of the concentration of the next required nucleotide when [NTP] is rate-limiting. $Rate_{release}$ is a function of the intrinsic stability of the transcription complex at each template position. $Rate_{reverse}$ is a function of the intrinsic rate of pyrophosphorolysis at each template position ($V_{max,PP,I}$) and of pyrophosphate concentration when [PP$_i$] is rate-limiting. This rate is also a function of GreA, GreB, and other intrinsic nuclease activation factors.

absolute rate of the required subsequent process that is controlling, but rather the relative rates of all possible subsequent pathways. As a consequence, as we will show, kinetic control of biological processes can be very directive and can play a key role in the regulation of transcriptional specificity.

The powerful role of such competitive kinetic control in transcription can be illustrated by considering the reaction possibilities available to an elongating transcription complex at each template position once the initiation–elongation interface has been passed and the elongating polymerase is moving through the coding region of the gene (see Fig. 1). The ternary elongation complex (polymerase [11], template DNA and nascent RNA) is very stable and moves along the DNA template processively and without dissociation until a defined termination site is reached.

Fig. 4 shows that at each template position (I)

---

[10] Even repressors must bind weakly enough, on an absolute scale, to permit dissociation (free or induced) on the time scale of subsequent events in the life cycle cell (e.g., changes in nutrient concentration, cell division, etc.).

[11] By polymerase we mean here the core polymerase itself, together with any activation factors that may be tightly bound to it and move with it. For *E. coli* RNA polymerase this could correspond simply to the four subunit ($\alpha_2\beta\beta'$) core complex that comprises the minimal functional enzyme required for elongation, although the *E. coli* transcription elongation complex often carries regulatory elongation and termination factors as well. The eukaryotic 'core' complex carries both 'elongation' and regulatory subunits.

there exist three competing reaction pathways that are, in principle, available to the transcription complex [32]. Thus the polymerase can: (i) move forward along the template (to position I + 1) by elongating the nascent transcript by template-directed addition of the next required nucleotide at the 3′-end; (ii) terminate transcription (at position I) by releasing the nascent RNA into solution; or (iii) move backward (to position I − 1 or beyond) by shortening (and perhaps editing) the nascent RNA by one (or more) nucleotide residues from the 3′-end using pyrophosphorolysis or an intrinsic nucleolytic cleavage reaction that is activated by transcription factors such as GreA and GreB (in *E. coli*) or by TFIIS (in eukaryotes). We note that elongation (and shortening) of the transcript is not uniform with template position; rather the rates of these processes are very dependent on template sequence and also on the concentration of the next-required NTP. Thus the transcription complex can pause in moving down the template and the 'dwell-time' at any particular template position can range from a few ms up to minutes or more.

One aspect of transcript elongation in which such competitive kinetic processes play a critical role is in the regulation of fidelity. Thus both misincorporation of a nucleotide at the 3′-end of the nascent transcript, and the addition of further nucleotide residues to a 3′-terminus carrying an incorrect residue, are much slower than the regular chain elongation process. Furthermore, these delays appear to trigger changes in the 'stalled' transcription complexes, in which they decay in a time-dependent manner into inactivated forms in which residues located at or near the 3′-end of the nascent chain are subject to nucleolytic editing. These editing processes involve shortening the chain by a number of residues from the 3′-end, resulting both in the removal of misincorporated residues and the avoidance of the formation of 'dead-end' (permanently stalled) complexes that might otherwise block, and thus inactivate, the operon. (For further details see [33,34].)

9. Kinetic competition controls the elongation–termination decision

What factors are important in deciding whether termination is a viable alternative at any particular
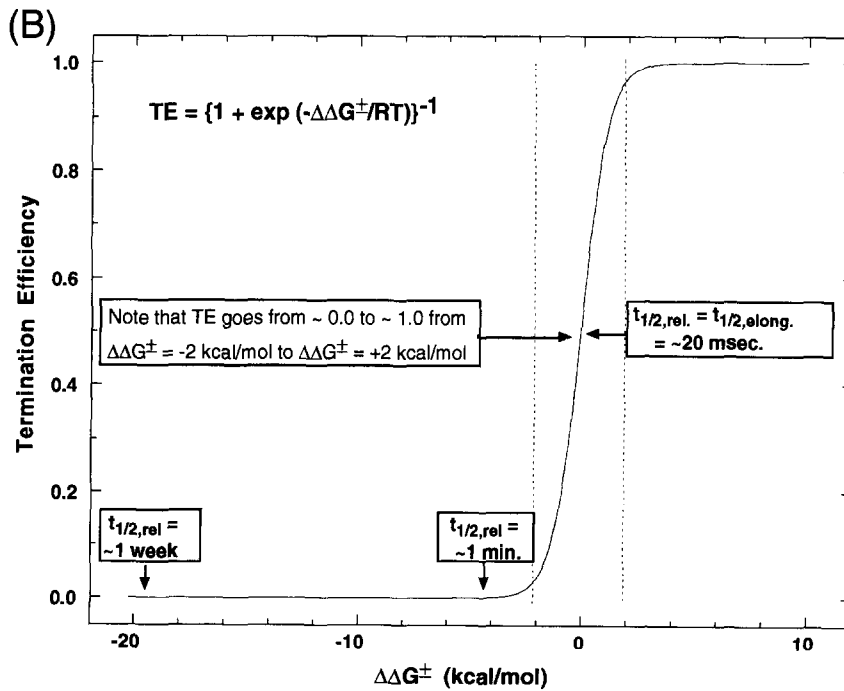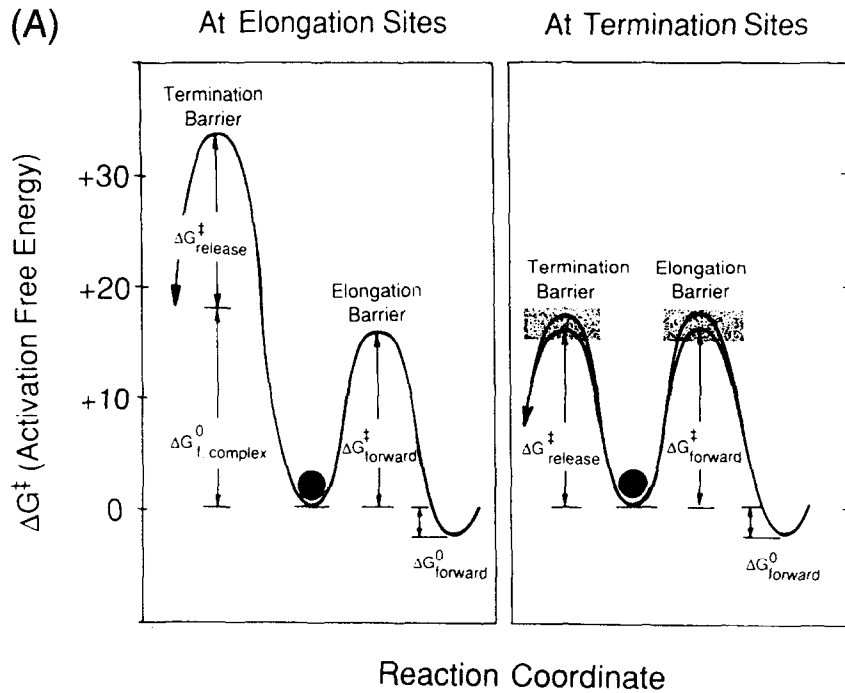
template position? As indicated above, the transcription complex is stable and processive at most template positions in the elongation phase of transcription (Fig. 1), and termination only becomes possible at specific terminator sites along the template at which elongation complexes are destabilized so that elongation and termination become kinetically competitive possibilities. In *E. coli*, the only organism for which we currently have much information about termination mechanisms, these terminators are of two types, intrinsic and rho-dependent.

An intrinsic terminator codes for two crucial sequence elements at the 3′-end of the terminated transcript. These elements are a penultimate termination hairpin with a stable (G · C-rich) stem six to eight base-pairs in length, followed by a run of six to eight rU residues at the end of the RNA chain. Transcription into and through this sequence makes termination possible by destabilizing the previously very stable and processive elongation complex (possible mechanisms that explain this destabilization are discussed in [35] and references cited therein). In contrast, *E. coli* transcription termination factor rho is thought to destabilize the elongation complex by means of its intrinsic ATP-dependent RNA–DNA helicase activity (see below and [36]).

A useful representation of this kinetic competition between pathways, leading either to continued transcript elongation or RNA release and termination, is shown in Fig. 5A [37]. The left-hand panel of this figure represents the competitive situation at nonterminator positions along the template. Here the height of the free energy of activation barrier to elongation is approximately 16 kcal/mol, corresponding (via the Eyring relation) to a forward rate constant of about 30–50 s^{-1}; i.e., to a 'dwell-time' at a typical elongation position of 20–40 ms. In contrast, the height of the free energy of activation barrier to termination is typically greater than +30 kcal/mol, corresponding to a characteristic dissociation time measured in hours to days. This simply shows in activation barrier terms that termination is not a significant possibility at most template positions, and that significant changes in the rate of elongation (or the stability) of the complex as a consequence of sequence-specific pauses can be used to control elongation without any risk of RNA release.

In contrast, the competitive situation shown in the right-hand panel of Fig. 5A represents events at terminators where the free energy of activation bar-

rier to RNA release has been significantly lowered, either in a sequence-dependent fashion at intrinsic terminators or by the action of transcription termina-

(A)

At Elongation Sites   At Termination Sites

Reaction Coordinate

(B)

$$TE = \{1 + \exp(-\Delta\Delta G^{\pm}/RT)\}^{-1}$$

Note that TE goes from ~ 0.0 to ~ 1.0 from $\Delta\Delta G^{\pm}$ = -2 kcal/mol to $\Delta\Delta G^{\pm}$ = +2 kcal/mol

$t_{1/2,rel.}$ = $t_{1/2,elong.}$ = ~20 msec.

$t_{1/2,rel}$ = ~1 week

$t_{1/2,rel}$ = ~1 min.

$\Delta\Delta G^{\pm}$ (kcal/mol)

tion protein rho. Here the rates of elongation and termination are comparable, as shown by the fact that the free energy of activation barriers are of comparable height and small changes in the relative heights of the barriers (within the stippled zones shown) can tip the balance almost entirely either toward elongation or toward termination. In Fig. 5B these differences are illustrated quantitatively by plotting termination efficiency (TE) as a function of the difference in height of the competing free energy of activation barriers ($\Delta\Delta G^*$). Clearly the entire range of TE is covered by a change in $\Delta\Delta G^*$ from -2 to $+2$ kcal/mol, meaning that at termination sites a very small change in either the rate of elongation of the transcript or the stability (with respect to dissociation and RNA release) of the complex, or a combination of both, can effectively move the termination efficiency across its entire regulatory range (for further details see [37,38]). This principle, together with cis effects of the type described for transcript initiation above, is applied in controlling termination efficiency at both intrinsic and rho-dependent terminators.

10. Cis effects as regulatory elements in transcript elongation and termination

How can the ideas of processivity, stability, and kinetically competing pathways be combined to regulate transcription as a function of template position? Obviously, if we continue our focus on termination as an example, this process is only possible at positions at which the elongation complex has been destabilized. However the degree to which such destabilization within a template 'zone of opportunity' for termination [39] actually results in termination and RNA release, as opposed to continued transcript elongation, depends on the fine-tuning of

the termination efficiency by small changes in the rates of elongation and/or RNA release, as shown in Fig. 5B. Furthermore, the specificity with which this termination control is exerted at some template positions and not at others may depend on additional regulatory signals that are coded into the template at effector sites located upstream of the terminators to be controlled.

There are a number of ways in which a processive process, such as the directional movement of a transcription complex along the template, can be used to control downstream events. One possibility is that the complex is somehow transiently altered on transcribing through a regulatory sequence of base pairs along the template, with this alteration decaying with time or with distance moved along the template. Such processes have been described as time-dependent (or distance-dependent) 'molecular clocks', reflecting, for example, the slow (relative to the time required for adding a nucleotide to the 3'-end of the transcript) decay of a particularly stable polymerase conformation. Though such ideas have often been proposed in considering transcriptional control in elongation and termination, no convincing evidence for the actual existence of such mechanisms has yet been put forward.

The other possibility for transiently altering a transcription complex at a regulatory locus located upstream of a functional template site (such as a terminator) is to postulate that the altered state of the complex is maintained by 'tethering' the polymerase to the regulatory site by nucleic acid looping while the complex continues downstream to the terminator. In such a scenario the altered state of the complex would persist only as long as it remains tethered to the regulatory site. In principle this tethering could involve binding of the regulatory DNA sequence itself to the moving transcription complex, thus forming a DNA loop much like that described above

---

Fig. 5. (A) Schematic diagrams of the relative heights of the free energy of activation barriers to elongation and to termination at a typical elongation position (left) and at a typical terminator position (right). The zero-point on the y-axis is set equal to the free energy of the reactant state of the transcription complex at template position I. The total height of each barrier is the sum of a thermodynamic (stability) component ($\Delta G^\circ_{f,complex}$) and a kinetic component ($\Delta G^*_{forward}$ or $\Delta G^*_{release}$). Barrier heights corresponding to 10-fold changes in rate are shown as alternative heavy lines at the tops of the peaks in the right-hand panel; the stippled areas corresponding to peak height variations of ± 1.4 kcal/mol over which range the termination efficiency (TE) goes from ca. 0.01 to ca. 0.99 (see B). (B) Termination efficiency (TE) as a function of the difference in the heights of the free energy of activation barriers to termination and to elongation at template position I ($\Delta\Delta G^*$). Some important features of this relationship are indicated in the figure.

in discussing cis-effects in the control of transcript initiation. This mechanism also has not been observed in elongation control.

A more attractive alternative (because single-stranded RNA is more flexible than double-stranded DNA) is that tethering involves binding to the *product* of the regulatory DNA sequence; i.e., to a particular sequence or secondary structure within the nascent RNA. This RNA sequence or structure could then interact (either directly or indirectly through bound protein factors) with the moving polymerase complex to regulate its properties at downstream sites, presumably by altering either its rate of movement along the template or its stability with respect to dissociation. As examples of systems in which interactions with the nascent RNA result in such

cis-modulation at downstream regulatory sites, we will discuss the control of the lysis–lysogeny decision in *E. coli* bacteriophage λ by the antitermination transcription factor N, and the control of termination at rho-dependent terminator sites by *E. coli* factor rho.

Cis-regulation by phage λ antitermination protein N involves RNA looping. Phage λ can exist as a dormant prophage in the so-called lysogenic state through many cycles of cell division within the genome of *E. coli*. However, this situation can be upset at any time by the onset of unfavorable environmental conditions. Such conditions trigger the excision of the phage DNA from the bacterial chromosome, followed by massive phage production and lysis of the bacterial cell. One of the initial biochem-
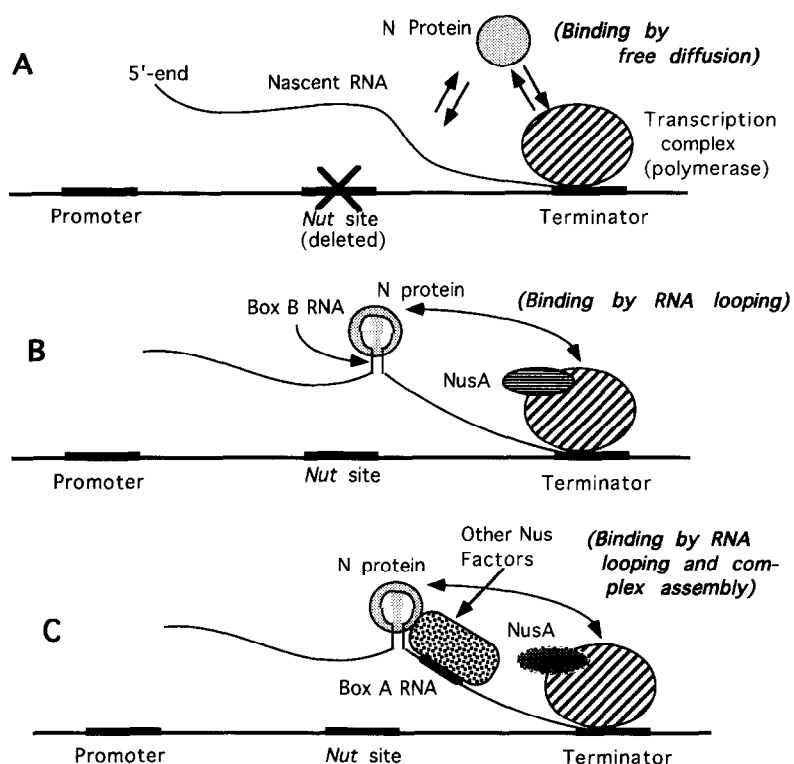

Fig. 6. Three models of N-dependent antitermination. (A) The non-processive (non-terminator-specific) antitermination induced by phage λ N protein alone in the absence of a *nut* site. (B) The minimally processive (minimally terminator specific) N-dependent antitermination system, with N binding tightly to the boxB RNA hairpin coded by the *nut* site and interacting with the polymerase and NusA by RNA looping to form an antitermination complex that is stable over several hundred nucleotide residues of RNA chain elongation. (C) The fully processive (fully terminator specific) N-dependent antitermination system, with N binding tightly to the boxB RNA hairpin and NusA and the other Nus factors binding to the boxA sequence of the *nut* RNA to make an antitermination complex that is stable over thousands of nucleotide residues of RNA chain elongation. (Taken with permission from [56].)

ical events in this switch from the lysogenic to the lytic state is the increased production of phage-coded N protein, which binds to host transcription complexes and permits them to read through strong termination signals to transcribe the genes responsible for phage replication, maturation, and coat protein production. Thus the induction of N-dependent antitermination comprises a crucial step in activating this switch in life style of phage λ [40].

Mechanistically, N-dependent antitermination involves a number of host (Nus) proteins in addition to the λ-coded N protein itself (for recent reviews see [41,42]). These proteins assemble with the moving polymerase to form a stable termination-resistant transcription complex that can processively read through a series of widely spaced terminator sites, as illustrated in Fig. 6C. A reduced set of protein components (Fig. 6B) can induce a less stable and processive antitermination complex [43], and N alone can also induce antitermination (Fig. 6A), though only in a non-processive, and thus non-terminator specific form [44].

From the point-of-view of this discussion, the essential question is how the system works to establish the specificity of the antitermination process; i.e., how does it decide which terminators are to be read through? Here cis-dependent effector specificity comes into play, because physiological N-dependent antitermination is totally dependent on the presence of a specific template sequence called the nut site upstream of the terminators to be regulated. In this system the nut site is expressed through the product RNA, with cis RNA looping bringing nut-site-coded RNA features (called boxA and boxB; see Fig. 6B and C) to the polymerase in order to participate, with the Nus proteins and N, in the formation of a protein–RNA complex that stabilizes a termination-resistant form of the transcription elongation complex [41,42].

Studies of in vitro transcription termination in the presence of N and in the absence of a nut site (Fig. 6A), have shown that N can bind weakly to the elongation complex and to non-specific sites on the nascent RNA to form an unstable regulatory complex with antitermination properties at N concentrations significantly higher (and salt concentrations somewhat lower) than those characteristic of physiological N-dependent antitermination [44]. Using these

data and assuming an equilibrium model, it has been proposed that the rate of dissociation of N from this complex is comparable to the dwell-time of the moving transcription complex at a single template position, thus rendering such antitermination induced by N protein alone to be effectively non-specific (i.e., there is no regulation by upstream sites).

In separate experiments it has been shown that N binding both increases the rate of transcription of the elongation complex 2- to 5-fold [56], and that the N–boxB RNA–NusA subassembly (Fig. 6B) stabilizes the elongation complex against RNA release (Rees et al., [56], see also [45]). It is this increase in the rate of elongation, combined with the stabilization of the transcription complex to decrease the rate of RNA release, that changes the relative heights of the free energy of activation barriers to elongation and termination (see Fig. 5) sufficiently to bring about the observed antitermination.

This transcription complex stabilization (and perhaps also, in part, the rate increase effect) depend critically on RNA looping as a cis-effector to increase the concentration of boxB RNA (and thus of the stable antitermination subassembly) sufficiently at the transcription complex to permit read-through of the target terminators. As a consequence we obtain cis-dependent specificity regulation for several hundred base pairs downstream of the nut site for the minimally processive antitermination system (Fig. 6B), and for thousands of base pairs downstream of the nut site for the fully processive and stable antitermination assembly (Fig. 6C).

Rho-dependent termination also depends on cis-specificity effects involving the nascent RNA. The control of termination at E. coli rho-dependent terminators also involves specific protein interactions with the nascent RNA, as well as competing kinetic pathways, but in quite a different way from that manifested by the N-dependent antitermination system (above).

Rho is a hexameric protein that binds single-stranded RNA [46]. This binding activates an RNA-dependent ATPase of rho that fuels the directional translocation ($5' \rightarrow 3'$) of the rho hexamer along the nascent RNA ([47,48]; Katherine Walstrom, unpublished data). Termination is thought to be triggered by the intrinsic ATP-dependent RNA–DNA helicase activity of rho [49] when it reaches the transcription

complex paused at putative rho-dependent terminators along the template [50–52]. Thus the specific positions of rho-dependent terminators along the template (as well as the efficiency of the rho-dependent termination process) are controlled by the kinetic coupling of the relative rates of the translocation of rho along the nascent RNA and of polymerase along DNA template ([53]; see also [37] and Fig. 5A and B).

Rho-dependent termination specificity also has a cis-specificity element, in that rho will only bind to a subset of nascent transcripts that carry a specific 'rho loading site', defined as a stretch of RNA that is effectively free of secondary structure over 70–80 nucleotide residues [54]. This requirement follows from the fact that rho is a single-stranded RNA binding protein with a hexameric RNA site size of 70–80 nucleotide residues. As a consequence, rho-dependent termination also displays both binding and kinetic specificity determinants. These features of the regulation of rho-dependent termination, and the relevant thermodynamic and kinetic parameters that apply, are illustrated schematically in Fig. 7. We note again that the regulatory specificity of this system, both in terms of the positions of rho-dependent termination sites and in terms of the efficiency at which rho-dependent termination occurs at these sites (Qin Zhu, unpublished data), depends on the relative rates of two competing processes as described in Fig. 5. In this case these processes are the rate of translocation (and helicase activity) of the rho along the nascent RNA chain, relative to the rate at which the transcription complex extends the nascent RNA along the DNA template.

11. Specificity control involves multiple determinants

In summary, certain general principles of specificity regulation seem to emerge from this overview of the mechanisms of the many regulatory elements that are encountered by the DNA-dependent RNA polymerase in its journey through an operon along the DNA template. First, of course, one is struck by the relative simplicity and generality of the thermodynamic and kinetic principles that underlie the function of these 'switches'. However, one is also struck
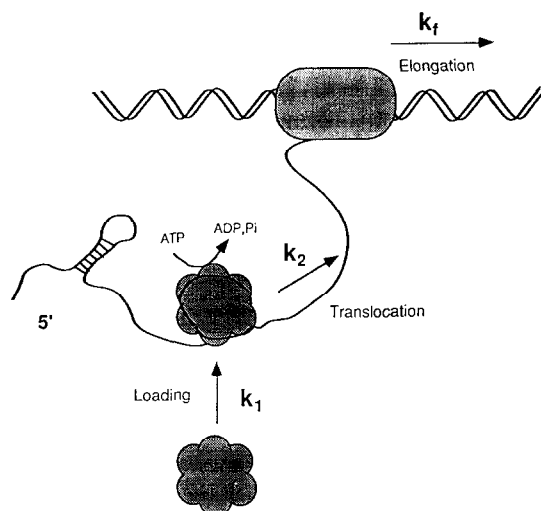

Fig. 7. A model of rho-dependent termination, showing 'kinetic coupling' between the rate of RNA polymerase elongation on the DNA template and the overall rate of rho moving along the RNA chain to induce termination. The *E. coli* rho hexamer binds at the unstructured 'rho loading site' on the nascent RNA with a rate constant k_1, from which it translocates $5' \to 3'$ along the RNA (driven by ATP hydrolysis) with a rate constant k_2, to 'catch up' (at a pause site) with the transcription complex, which elongates the RNA with a rate constant k_f at that template position [53]. (Taken with permission from [55].)

by the complexity and overlapping redundancy (and sometimes even by the apparent clumsiness relative to modern engineering standards) with which these principles have been deployed and combined in the course of evolution. Of course this follows because, in the design of a new regulatory mechanism to control an additional facet of transcription, nature (unlike engineers) is never able to 'start over'. Rather these mechanisms have to be 'added on' to the existing regulatory systems without perturbing those that are already in place and playing central roles in the regulation of the cell or organism.

Nevertheless, it seems to us (and it certainly seemed to Bill Harrington) that understanding the underlying principles of these switches from a quantitative and physical chemical point-of-view is worthwhile, since this approach makes it possible to dissect biological control mechanisms and often to establish at least the *magnitudes* of the changes in the relative kinetic and stability parameters of the switch that permit it to function in the physiological milieu. It is our hope and expectation that such

knowledge will further focus the search for, and the actual analysis of, the components of the real regulatory elements that control and direct the processes of replication, growth, and differentiation in all biological systems.

References

[1] W.F. Harrington and P.H. von Hippel, Adv. Protein Chem., 16 (1961) 1–138.

[2] W.F. Harrington, Proc. Natl. Acad. Sci. USA, 68 (1971) 685–689.

[3] M.E. Rodgers, J.J. Englander, S.W. Englander and W.F. Harrington, Biophys. Chem., 59 (1996) (this issue).

[4] T.D. Yager and P.H. von Hippel, in F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter and H.E. Umbarger (Eds.), *E. coli* and *S. typhimurium*: Cellular and Molecular Biology, Vol. 1, American Society for Microbiology, Washington, DC, 1987, pp. 1241–1275.

[5] P.H. von Hippel, T.D. Yager and S.C. Gill, in S. McKnight and K. Yamamoto (Eds.), Transcription Regulation, Cold Spring Harbor Laboratories, 1992, pp. 179–201.

[6] P.H. von Hippel, in R.F. Goldberger (Ed.), Biological Regulation and Development, Vol. 1, Plenum Press, New York, 1979, pp. 279–347.

[7] T.D. Schneider, G.D. Stormo, L. Gold and A. Ehrenfeucht, J. Mol. Biol., 188 (1986) 415–431.

[8] O.G. Berg and P.H. von Hippel, J. Mol. Biol., 193 (1987) 723–750.

[9] P.H. von Hippel and J.D. McGhee, Ann. Rev. Biochem., 41 (1972) 231–300.

[10] N.C. Seeman, J.M. Rosenberg and A. Rich, Proc. Natl. Acad. Sci. USA, 73 (1976) 804–809.

[11] M.T. Record, Jr., T.M. Lohman and P.L. deHaseth, J. Mol. Biol., 107 (1976) 145–158.

[12] R. Spolar and M.T. Record, Jr., Science, 263 (1994) 777–784.

[13] P.H. von Hippel and O.G. Berg, Proc. Natl. Acad. Sci. USA, 83 (1986) 1608–1612.

[14] Z. Otwinoski, R.W. Schevitz, R.-G. Zhang, C.L. Lawson, A. Joachimiak, R.Q. Marmostein, B.F. Luisi and P.B. Sigler, Nature, 335 (1988) 321–329.

[15] C.O. Pabo and R.T. Sauer, Ann. Rev. Biochem., 61 (1992) 1053–1095.

[16] T.A. Steitz, Quart. Rev. Biophys., 23 (1990) 205–280.

[17] N.D. Arbuckle and B. Luisi, Struct. Biol., 2 (1995) 341–346.

[18] I.M. Klotz and J.S. Franzen, J. Am. Chem. Soc., 84 (1962) 3461–3466.

[19] S.M. Habermann and K.P. Murphy, submitted for publication.

[20] A.D. Riggs, S. Bourgeouis and M. Cohn, J. Mol. Biol., 53 (1970) 401–407.

[21] G.R. Bellomy and M.T. Record, Jr., Prog. Nucleic Acids Res., 39 (1990) 81–128.

[22] P.H. von Hippel and O.G. Berg, J. Biol. Chem., 264 (1989) 675–678.

[23] R.B. Winter, O.G. Berg and P.H. von Hippel, Biochemistry, 20 (1981) 6961–6977.

[24] G. Adam and M. Delbruck, in A. Rich and N. Davidson (Eds.), Structural Chemistry and Molecular Biology, W.H. Freeman, San Francisco, CA, 1968, pp. 198–215.

[25] T. Ruusala and D.M. Crothers, Proc. Natl. Acad. Sci. USA, 89 (1992) 4903–4907.

[26] H. Kubata, O. Kurosawa, I. Arai, M. Washizu, S.A. Margarson, R.E. Glass and N. Shimamoto, Science, 262 (1993) 1561–1563.

[27] R. Hori and M. Carey, Curr. Opin. Genet. Dev., 4 (1994) 236–244.

[28] A.J. Koleske and R.A. Young, Trends Biochem. Sci., 20 (1995) 113–116.

[29] P.H. von Hippel and F.R. Fairfield, Pure Appl. Chem., 57 (1985) 45–56.

[30] P.H. von Hippel, S.C. Kowalczykowski, N. Lonberg, J.W. Newport, L.S. Paul, G.D. Stormo and L. Gold, J. Mol. Biol., 162 (1982) 6795–6818.

[31] K. Rippe, P.H. von Hippel and J. Langowski, Trends Biochem. Sci., 20 (1995) 500–506.

[32] D.A. Erie, T.D. Yager and P.H. von Hippel, Annu. Rev. Biophys. Biophys. Chem., 21 (1992) 379–415.

[33] D.A. Erie, O. Hajiseyedjavadi, M.C. Young and P.H. von Hippel, Science, 262 (1993) 867–873.

[34] S. Borukov, V. Sagitov and A. Goldfarb, Cell, 72 (1993) 459–466.

[35] K.S. Wilson and P.H. von Hippel, Proc. Natl. Acad. Sci. USA, 92 (1995) 8793–8797.

[36] T. Platt and J.P. Richardson, in S. McKnight and K. Yamamoto (Eds.), Transcription Regulation, Cold Spring Harbor Laboratories, 1992, pp. 365–388.

[37] P.H. von Hippel and T.D. Yager, Proc. Natl. Acad. Sci. USA, 88 (1991) 2307–2311.

[38] P.H. von Hippel and T.D. Yager, Science, 255 (1992) 809–812.

[39] K.S. Wilson and P.H. von Hippel, J. Mol. Biol., 244 (1994) 36–51.

[40] D.I. Friedman, in R. Calendar (Ed.), The Bacteriophages, Vol. 2, Plenum Press, New York, 1988, pp. 263–318.

[41] A. Das, Annu. Rev. Biochem., 62 (1993) 893–930.

[42] J. Greenblatt, J.R. Nodwell and S.W. Mason, Nature, 364 (1993) 401–406.

[43] W.A. Whalen and A. Das, New Biol., 2 (1990) 975–91.

[44] W.A. Rees, S.E. Weitzel, T.D. Yager, A. Das and P.H. von Hippel, Proc. Natl. Acad. Sci. USA, 93 (1996) 342–346.

[45] J. DeVito and A. Das, Proc. Natl. Acad. Sci. USA, 91 (1994) 8660–8664.

[46] L.R. Finger and J.P. Richardson, J. Mol. Biol., 156 (1982) 203–219.

[47] J. Geiselmann, Y. Wang, S.E. Seifried and P.H. von Hippel, Proc. Natl. Acad. Sci. USA, 90 (1995) 7754–7778.

[48] E.J. Steinmetz, C.A. Brennan and T. Platt, J. Biol. Chem., 265 (1990) 18408–18413.

[49] C.A. Brennan, A. J. Dombroski and T. Platt, Cell, 48 (1987) 945–952.

[50] R. Reisbig and J.E. Hearst, Biochemistry, 20 (1981) 1907–1918.

[51] L.F. Lau and J.W. Roberts, J. Biol. Chem., 258 (1983) 9391–9397.

[52] W.D. Morgan, D.G. Bear and P.H. von Hippel, J. Biol. Chem., 258 (1983) 9565–9574.

[53] D. Jin, R. Burgess, J.P. Richardson and C.A. Gross, Proc. Natl. Acad. Sci. USA, 89 (1992) 1453–1457.

[54] W.D. Morgan, D.G. Bear, B.L. Litchman and P.H. von Hippel, Nucleic Acids Res., 13 (1985) 3739–3754.

[55] Y. Wang, PhD Thesis, University of Oregon, 1992.

[56] W.A. Rees, S.E. Weitzel, A. Das and P.H. von Hippel, in preparation.