INITIATION OF TRANSCRIPTION BY RNA POLYMERASES OF *E. COLI* AND PHAGE T3

Ekkehard K.F. BAUTZ

Institut für Molekulare Genetik, Universität Heidelberg, 69 Heidelberg 1, Berlinerstrasse 15, GFR

In bacteria, transcription of all the genetic information is mediated by a single species of RNA polymerase. Evidence for a single enzyme comes from studies on the drug rifampicin which interacts with RNA polymerase and, upon addition to a bacterial culture, eliminates the synthesis of ribosomal, transfer and messenger RNA. A single-step mutation to rifampicin resistance restores the ability to synthesize all RNA species in the presence of the drug [1].

A bacterial genome consists of several hundred transcriptional units or operons. To transcribe these, a cell contains about 7000 molecules of RNA polymerase, i.e. about 1.5% of the total cellular protein [2]. Regulation of transcription appears to occur primarily at the level of initiation. We can group the transcriptional units of a bacterial genome roughly into three types: 1) those that are nearly always expressed at optimal rates, e.g., the rRNA genes; 2) those that are always expressed at very low levels, e.g., the *i* gene coding for the *lac* repressor and 3) those being subject to repression and induction, e.g., catabolite operons. The difference between genes of type one and two appears exclusively in the structure of their promotors, the RNA polymerase having a very high affinity for the first and a very low affinity for the second class. An example is the i^Q promotor mutation which results in an increased production of *lac* repressor [3].

Repression and induction in the classical sense via repressor molecules is called negative control; the repressor blocks initiation of transcription by occupying a site near to or overlapping the promotor, whereas the presence of an inducer removes such blocks. Some operons, such as the *lac* operon, can be additionally regulated through positive control elements, auxiliary factors, which increase the affinity of RNA polymerase for the promotor. One example is the catabolite gene activating protein which together with cyclic AMP is required for optimal transcription of the *lac* operon [4].

Since the action of all regulatory elements appears to affect in some way the affinity of RNA polymerase to the start signals of transcription, most questions concerning the regulation of transcription in bacteria center on: 1) the interaction of RNA polymerase with promotor sites; 2) the interaction of auxiliary factors with RNA polymerase and/or promotor sites and 3) the interaction of repressors with operator sites. In order to study the first of these, namely the recognition of the initiation signals by the polymerase, it is necessary to know more about the structure of the enzyme as well as the start signals.

RNA polymerase isolated from *E. coli* cells consists of 5 subunits, $\alpha_2\beta\beta'\sigma$, having a combined mol. wt of approx. 470 000 [5]. Except for the two α polypeptides, the subunits are non-identical. This means that nearly 9×10^6 daltons of DNA is required to code for all the subunits. Thus, elucidation of the structure and detailed information of this enzyme promises to prove rather difficult. The almost prohibitive size of the bacterial enzyme has persuaded us to look for alternative proteins of similar function but less complexity. Recently, such RNA polymerases have been found [6, 7] and one of these, the RNA polymerase produced in T3-infected *E. coli* cells, has not only become a valuable tool for a better understanding of the biochemical mechanism of transcription but it has also made us appreciate even more the complexity of the bacterial enzyme which appears to reflect nothing less but an almost unlimited versatility in being able to cope with many different regulatory elements.

Initiation of transcription by *E. coli* RNA polymerase

To produce an RNA chain the enzyme has to perform a series of different reactions which can be divided into 3 steps: initiation; chain elongation and termination. While the α, β and β' subunits are required for all steps, the σ protein was found to be required for initiation but not for the catalytic function of the polymerization proper [8]. This finding has focused considerable attention on the steps of initiation which, according to our present knowledge, may be outlined as follows:

1) $E\sigma + DNA \rightarrow E\sigma \cdot DNA$.

2) $E\sigma \cdot DNA \underset{< 17°C}{\overset{> 17°C}{\rightleftharpoons}} (E\sigma \cdot DNA)^*$.

3) $(E\sigma \cdot DNA)^* + NTP_1 + NTP_2 \rightarrow (E\sigma \cdot DNA)^* \cdot NTP_1 \cdot NTP_2$.

4) $(E\sigma \cdot DNA)^* \cdot NTP_1 \cdot NTP_2 \rightarrow (E\sigma \cdot DNA)^* \cdot NTP_1 - NMP_2 + PP_i$.

5) Translocation of enzyme on DNA template.

6) $(E\sigma \cdot DNA)^* \cdot NTP_1 - NMP_2 + NTP_3 \rightarrow (E\sigma \cdot DNA)^* \cdot NTP_1 - NMP_2 - NMP_3 + PP_i$.

The initial association (step 1) with DNA is nonspecific and reversible; both holoenzyme ($\alpha_2\beta\beta'\sigma$) and core enzyme ($\alpha_2\beta\beta'$) can form complexes and there only spatial limitations as to the number of enzyme molecules which can associate with DNA. If an enzyme is bound to a promotor site, there is a transition to a highly stable complex (step 2) which forms only in the presence of the σ factor, at temperatures above 17°C and at low ionic strength [9–11]. These complexes are specific as they seem to form only at genuine promotor sites. Formation of a tight complex involves a conformational change which probably affects the structure of both the enzyme [10, 11] and the promotor [12]. The next step requires the addition of the first and second substrate molecules into the initiation and chain elongation sites. Binding to the initiation site appears to be purine-specific with a K_m several-fold higher than that of the second site [13]. PP_i is then split off from the second nucleoside triphosphate to yield the first phosphodiester bond. From here on the enzyme enters the catalytic cycle involving a translocation step (step 5), with the initiation site becoming the product terminus site holding the 3′ terminal nucleotide of the growing chain and the chain elongation site becoming free to accept a new substrate molecule (step 6). Step 2 deserves the greatest attention as it encompasses the actual recognition of the promotor site by the enzyme. In switching from a loose to a tight complex, the enzyme appears to undergo a conformational change for which the σ factor is absolutely required. It can be shown that on preincubation with a native DNA template a limited number of holoenzyme molecules become partially resistant against rifampicin [10, 11] and totally resistant against template competitors such as heparin polyinosinic acid (poly I) [9, 14], or denatured DNA. It seems that in the latter case, all loosely bound enzyme molecules are removed by the competitor except those enzymes that have undergone the conformational change. Under these conditions, core enzyme is removed quantitatively by the competitor [14].

The removal of nonspecifically bound enzymes by polyanions of high affinity for the polymerase has been employed to isolate the tight enzyme-binding sites of DNA. The experimental scheme is analogous to the isolation of ribosome-binding sites as done by Steitz [15], involving digestion of unprotected DNA with DNAase, followed by gel filtration of the enzyme–promotor complex to separate it from the digestion products. Such tight binding sites have been isolated in several laboratories and are being analyzed for their primary sequences. In our own experience, the success of obtaining unique sequences depends upon the selection of a template containing only one or two promotor sites, such as DNA of phage fd [16] or T3 and T7, as well as a reasonably low initial ratio of RNA polymerase molecules to genomes in the incubation mixture. This is necessary since at high enzyme to DNA ratios most promotors seem to bind more than one enzyme rather tightly. This has been correlated with the finding that the site of polymerase recognition (the genetically defined promotor) and the site specifying the start of the RNA chain may be a few hundred base pairs apart [17]. It has been suggested that the region between these two sites can serve as a 'storage stretch' where polymerases can line up to ensure production of multiple messenger copies in short succession. The average length of such storage stretches was found to vary for different species of

DNA [18]. The number of heparin-resistant polymerases per promotor site also varies with ionic strength and it is not necessarily a linear function of enzyme concentration. In the extreme case of phage T5, so many polymerase molecules can become resistant that at least half of the entire early region (if not all of it) serves as storage stretches [19]. This makes one wonder whether, given the right experimental conditions and a sufficient number of enzyme molecules, several not too widely spaced enzyme-binding sites on DNA could be transformed to nucleation sites from where polymerases in a cooperative fashion could alter the tertiary structure of a vast stretch of DNA. In that case, 'storage stretches' could well become synonymous with genes.

Nevertheless, under very controlled conditions it is possible to bind polymerase to, e.g., T3 DNA to such a limited extent that only 0.1 to 0.2% of the total genome becomes resistant to digestion with pancreatic DNAase yielding fragments of unique length and a homogeneous distribution of pyrimidines suggesting that they contain unique sequences [14]. These fragments are double stranded and between 35 and 40 base pairs long and it appears that the strands can be separated by gel electrophoresis [20]. Sequence work on the fragments isolated by us is currently under way by Dr. Donaldson (Cambridge). Sequencing of one of the two promotors of fd DNA has already neared completion (Schaller, personal communication) and it will be of interest to compare sequence data of polymerase binding sites obtained in different laboratories on different species of DNA, especially since it is expected that the many different promotors that can be recognized by a bacterial RNA polymerase will prove to show sequence similarities but not identity.

Initiation of transcription by T3 RNA polymerase

In the steps of the initiation reaction postulated above, the addition of the first and second nucleoside triphosphate is difficult to investigate with the enzyme from *E. coli*, as this enzyme does not start all RNA chains with a unique sequence. In that respect, the enzyme synthesized after infection with phage T3 is a better object to study as it appears to start all RNA chains with the same trinucleotide sequence. Furthermore, the enzyme is less complicated in structure than the host enzyme, consisting of a single polypeptide chain of a molecular weight of 110 000, and it exhibits a very stringent template specificity, vastly preferring T3 DNA over any other DNA template. In this section I am recounting some of the findings which make the comparison between this and the bacterial enzyme in the following section worthwhile.

The dependence of enzyme activity on the concentration of each of the four nucleoside triphosphates which serve as substrates was determined and Lineweaver–Burk plots indicated that varying the concentration of either ATP, CTP or UTP resulted in a linear dependence of activity on substrate concentration whereas variation of the GTP concentration yielded a curvilinear plot which becomes linear when plotted as $^1/_V$ vs $^1/_{S^2}$. This result suggests that the simultaneous addition of two GTP moieties is required at some stage during the synthetic reaction. Studies on the exchange of pyrophosphate (PP_i) which measures the reverse reaction of synthesis, support the contention that this unusual dependence of enzyme activity on the concentration of GTP reflects the initiation with the sequence pppGpG, i.e., the fact that one obtains high levels of PP_i exchange in the presence of only GTP as substrate indicates that the sequence of at least the first two nucleotides is GG [21]. In contrast, with T7 DNA as template, almost no PP_i exchange was observed with GTP alone but only when GTP and ATP were both present, suggesting that, in this case, an adenylic acid residue occupies one of the first two positions at the 5′ end of the message [21].

The latter result corroborates the fact that a Lineweaver–Burk plot with T7 DNA as template gives a curvilinear dependence on substrate concentration only if both GTP and ATP are varied simultaneously. These data indicate that the T3 RNA polymerase is forced to start RNA chains with another sequence on T7 DNA than on the closely related T3 DNA. The T7 RNA polymerase, however, appears to be able to start with the sequence GG on both T7 and T3 DNA templates [21].

The T3 RNA polymerase is sensitive to either the rifamycin derivative AF/013 or to heparin. Binding of enzyme to DNA does not render it resistant to heparin and, unlike *E. coli* RNA polymerase [14], the T3 enzyme remains sensitive to heparin or high

salt concentration even after the formation of the first dinucleotide since preincubation with GTP alone leaves the enzyme sensitive. Upon preincubation with GTP and ATP, however, the enzyme becomes almost fully resistant to heparin. It was therefore of interest to determine the size of the products that accumulate under the two conditions, i.e., incubation with GTP alone or with GTP and ATP. The results were rather clear cut; incubation with GTP alone resulted in the exclusive production of the dinucleotide pppGpG whereas incubation with both GTP and ATP yielded a tetranucleotide as the major product with tri- and pentanucleotides as minor components [21]. These results indicate that most RNA chains start with the sequence $pppGpGpAp\begin{matrix}A\\G\end{matrix}$. Quantitatively, there are many more dinucleotides produced during incubation with GTP alone than tetranucleotides when ATP is also present, indicating that the enzyme recycles much more quickly if it can produce only dinucleotides. We have shown that under these conditions not only is the dinucleotide released but that the enzyme also leaves the template. That a tetranucleotide but not a dinucleotide can stabilize the enzyme–DNA complex suggests that the growing RNA chain serves to stabilize the transcription complex. Thus, during polymerization the T3 RNA polymerase is more tightly bound to DNA than during initiation.

Comparing the two enzymes

In table 1, I have listed those structural and functional properties of the T3 and *E. coli* RNA polymerase which are useful for comparison. The most pronounced differences between the two enzymes are size and complexity of the proteins as well as the fact that the T3 enzyme starts all chains with a unique sequence. This is not because there is only one promotor for the T3 enzyme on T3 DNA; although the exact number is difficult to establish, it is definitely more than one and most likely five or six (Küpper, unpublished). It is also worth noting that the T3 enzyme, in the absence of any additional protein factors, transcribes very efficiently, initiating without a lag phase and synthesizing RNA chains in vitro at a rate which is 3–4 times higher than expected from the known in vivo rates of polypeptide synthesis. Thus, if the same rates were attained in vivo, T3 late transcription would vastly outpace translation. While initial rates as high as 500 nucleotides/sec have been reported for the *E. coli* enzyme [18], I have not used these in the comparison since they are not sustained in prolonged synthesis. Kinetic analyses have shown that on a T3 DNA template the T3 enzyme initiates all RNA chains with the sequence pppGpGpApPu. The apparent K_m for GTP determined from these experiments is about 5 times that of other nucleotides which are involved only in chain elongation. On a T7 DNA template, where the starting sequence is pppGpA, the K_m for ATP was no different than that for CTP (which is involved only in polymerization). Since K_m values are an indication of the affinity of an enzyme for it substrate, these findings suggest that there are only two substrate binding sites on the T3 enzyme. The first, or initiation, site (site I) is filled by GTP, and the K_m for this site is about 5 times that of the second or polymerization site (site II). Such a model assumes that after the formation of a phosphodiester bond the enzyme translocates on the DNA template and that the 3′ terminal nucleotide of the growing product then occupies the initiation site. The polymerization site (site II) thus continues to function in the same manner during the successive addition of all subsequent nucleotides. A similar two site model for the *E. coli* RNA polymerase has been proposed by Goldthwait et al. [13]. It is interesting to note that the initiation site of the host enzyme also exhibits a K_m value that is higher than that of the polymerization site, and that the 5′ terminal nucleotide is always a purine.

In a reaction mixture containing GTP alone as substrate, the T3 RNA polymerase continuously catalyzes the formation of dinucleotides without the formation of a stable transcription complex. In contrast, it has been found that the host enzyme becomes resistant to the effects of high salt concentrations after the formation of the first phosphodiester bond [22]. The T3 polymerase becomes resistant to high salt concentrations (or the polyanion heparin) only after the formation of an RNA chain 3–4 nucleotides long. It is not known whether these differences reflect dissimilarities in the manner in which the enzymes undergo conformational changes that result in a tighter binding to the template. However, in both cases the enzymes are more tightly bound to DNA during polymerization than during initiation.

Table 1
Properties of the T3 and *E. coli* RNA polymerases.

	T3	*E. coli*
Enzyme structure	1) Mol. wt = 110 000 2) Only 1 subunit 3) Probably active as monomer 4) Requires no protein factors for accurate transcription 5) Two substrate binding sites: K_m for initiation site = 5 × K_m for polymerization site	1) Mol. wt = 470 000 2) Five subunits ($\beta\beta'\alpha_2\sigma$) 3) Probably active as monomer 4) Requires σ factor for accurate initiation and in some instances ρ factor for termination 5) Two substrate binding sites: K_m for initiation site = 10 × K_m for polymerization site
Binding to DNA	1) Complex unstable 2) Complex not resistant to heparin or high salt	1) Complex highly stable above 20°C 2) Complex resistant to heparin, but not high salt
Initiation	1) Starts with: pppGpGpApPu on T3 DNA pppGpA on T7 DNA 2) Inhibited at salt concentrations > 0.1 M KCl 3) Formation of dinucleotide does not give resistance against heparin or high salt; formation of tetranucleotide stabilizes 4) No lag phases at salt optimum	1) Depending on start site, starts with either pppA or pppG, usually has pyrimidine in second position 2) Inhibited at salt concentrations > 0.2 M KCl 3) Formation of dinucleotide gives stable complex resistent to high salt 4) Lag phase at salt optimum
Polymerization	1) Chain elongation rate = 170 nucl/sec at 37°C 2) Size of products identical at 0.005 M KCl and 0.15 M KCl; no change with ρ factor	1) Chain elongation rate = 20–50 nucl/sec at 37°C 2) Size of products altered above 0.15 M KCl or with ρ factor

A useful probe to study enzymes of the same function but different structure is their sensitivity to various inhibitors. In table 2, I have listed the sensitivity of *E. coli* and T3 RNA polymerases to a number of drugs which are most commonly used in inhibition studies. I have also included the A and B enzymes from calf thymus for comparison. The T3 enzyme shares the insensitivity against rifampicin, a potent inhibitor of the *E. coli* enzyme, with the mammalian polymerases. Likewise, only the *E. coli* enzyme is sensitive to streptolydigin. All enzymes are inhibited by the rifamycin derivative AF/013. While claims of a selective inhibition of initiation by mammalian polymerases have been challenged for this derivative as it inhibits only in large doses and then rather nonspecifically inactivates many other enzymes [23], it could be shown that AF/013 inhibits the T3 enzyme at lower concentrations and only before the enzyme has initiated [24]. The complete insensitivity of the T3 enzyme to exotoxin from *B. thuringiensis* is unique. So far all RNA polymerases tested have been inhibited by this exotoxin which appears to function as an ATP analogue in substrate binding. This suggests that the T3 enzyme shows the most stringent specificity in substrate binding.

Finally, it is interesting to note that while at low ionic strength heparin binds T3 RNA polymerase more strongly than T3 DNA, it binds the enzyme less strongly at higher ionic strength so that at 0.15 M KCl it is no longer inhibitory. It has also been claimed

Table 2
Sensitivity of RNA polymerases to various inhibitors.

Inhibitor	RNA polymerase			
	E. coli	T3	Calf thymus A	Calf thymus B
Rifampicin	+	–	–	–
Rifamycin AF/013	+	+	+	+
Streptolydigin	+	–	–	–
Exotoxin	+	–	+	+
α Amanitin	–	–	–	+
Heparin at 0.05 M K^+	+	+		
Heparin at 0.15 M K^+	+	–		

+ = Inhibition; – = no inhibition.

that, in low salt, heparin inhibits the related T7 enzyme during polymerization [25], we have shown, however, that the T3 enzyme is not affected by heparin at all during the polymerization reaction [21].

Outlook for future progress

1. *Structure of the initiation complex*

There are two components which can be analyzed separately: the polymerase and the promotor. The primary structure of the latter will be much easier to determine and it is expected that the DNA sequences of several polymerase binding sites will become obtainable within the next one or two years. The DNA region binding the *lac* repressor has been sequenced [26] and there is no reason why the somewhat larger promotors should not be.

To sequence the bacterial enzyme would require an experimental effort of the same magnitude as that currently invested into the structure of ribosome. There are few laboratories in the position to do that. The T3 or T7 polymerase should be easier to sequence although a single polypeptide of a mol. wt of 10^5 presents quite a challenge. There is the additional problem that phage-infected cells are not obtainable in unlimited quantities.

There are a few physical methods used to study directly the interaction of polymerase with DNA, the problem being that the number of base pairs which are possibly in an altered configuration are so small compared with the number of base pairs in nonpromotor regions that reasonable results can only be obtained if a change in the conformation of only a few base pairs can be made to affect the overall structure of the genome [27].

2. *Mechanism of transcription*

Biochemically, the most obvious questions concerning the *E. coli* enzyme is the assignment of individual functions to the subunits. We know that β' has the strongest affinity to DNA [9], that σ affects the binding to promotor sites [10], and there is reason to believe that β contains the substrate binding sites. We are left with the question, what is the function of the α subunits? Do they just have some structural reasons for existing? Or do they serve some regulatory purposes since they are modified after phage T4 infection [28]? It seems to me that *E. coli* mutants producing altered subunits are most badly needed to approach these questions. Fortunately, this need has been recognized in a few laboratories and a considerable effort is being invested in obtaining and analysing such mutants, although it is far more difficult to obtain clear cut temperature sensitive mutants than one might have expected.

A very important problem from the regulatory point of view is the interaction of RNA polymerase with other proteins, as has been found in the case of phage T4 infection [29] and sporulation in *B. subtilis* [30]. It is hoped that analysis of outside supressors of bona fide *ts* mutants of RNA polymerase might uncover additional regulatory proteins.

References

[1] Yura, T., Igarashi, K. and Masukata, K. (1970) in: Lepetit Colloquium on RNA Polymerase and Transcription (Silvestri, ed.), pp. 71–89, North-Holland, Amsterdam.
[2] Mastzura, H., Hansen, B.S. and Zeuthen J. (1973) J. Mol. Biol. 74, 9–20.
[3] Müller-Hill, B., Crapo, L. and Gilbert, W. (1968) Proc. Natl. Acad. Sci. U.S. 59, 1259–1264.
[4] Eron, L., Arditti, R., Zubay, G., Connaway, S. and Beckwith, J.R. (1971) Proc. Natl. Acad. Sci. U.S. 68, 215–218.
[5] Burgess, R.R. (1969) J. Biol. Chem. 244, 6168–6176.
[6] Chamberlin, M., McGrath, J. and Waskell, L. (1970) Nature 228, 227–231.
[7] Dunn, J.J., Bautz, F.A. and Bautz, E.K.F. (1971) Nature New Biol. 230, 94–96.
[8] Burgess, R.R., Travers, A.A., Bautz, E.K.F. and Dunn, J.J. (1969) Nature 221, 43–46.
[9] Zillig, W., Zechel, K., Rabussay, D., Schachner, M., Sethi, V.S., Palm, P., Heil, A. and Seifert, W. (1970) Cold Spring Harbor Symp. Quant. Biol. 35, 47–58.
[10] Bautz, E.K.F. and Bautz, F.A. (1970) Nature 226, 1219–1222.
[11] Sippel, A.E. and Hartmann, G.R. (1970) Eur. J. Biochem. 16, 152–157.
[12] Travers, A., Baillie, D.L. and Pedersen, S. (1973) Nature New Biol. 243, 161–163.
[13] Goldthwait, D.A., Anthony, D.D. and Wu, C.-W. (1970) in: Lepetit Colloquium on RNA polymerase and Transcription (Silvestri, ed.), pp. 10–27, North-Holland, Amsterdam.
[14] Bautz, E.K.F., Bautz, F.A. and Beck, E. (1972) Mol. Gen. Genet. 118, 199–207.
[15] Steitz, J.A. (1969) Nature 224, 957–964.

[16] Heyden, B., Nüsslein, C. and Schaller, H. (1972) Nature New Biol. 240, 9–12.
[17] Blattner, F.R., Dahlberg, J.E., Boettinger, J.K., Fiandt, M. and Szybalski, W. (1972) Nature New Biol. 237, 232–236.
[18] Schäfer, R., Zillig, W. and Zechel, K. (1973) Eur. J. Biochem. 33, 207–214.
[19] Schäfer, R., Krämer, R., Zillig, W. and Cudny, H. (1973) in press.
[20] Beck, E. and Bautz, E.K.F., unpublished.
[21] McAllister, W.T., Küpper, H.A. and Bautz, E.K.F. (1973) Eur. J. Biochem. 34, 489–501.
[22] So, A.G. and Downey, K.M. (1970) Biochemistry 9, 4788–4793.
[23] Juhasz, P.P., Benecke, B.J. and Seifart, K.H. (1972) FEBS Letters 27, 30–34.
[24] Küpper, H.A., McAllister, W.T. and Bautz, E.K.F. (1973) Eur. J. Biochem., in press.
[25] Chamberlin, M. and Ring, J. (1973) J. Biol. Chem. 248, 2245–2250.
[26] Gilbert, W., Maizels, N. and Maxam, A. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, in press.
[27] Saucier, J.-M. and Wang, J.C. (1972) Nature New Biol. 239, 167–170.
[28] Seifert, W., Rabussay, D. and Zillig, W. (1971) FEBS Letters 16, 175–179.
[29] Stevens, A. (1972) Proc. Natl. Acad. Sci. U.S. 69, 603–607.
[30] Greenleaf, A.L., Linn, T. and Losick, R. (1973) Proc. Natl. Acad. Sci. U.S. 70, 490–494.