4. T7 RNA Polymerase

Rui Sousa

Introduction

As part of the mechanism by which cellular resources are subverted to the production of phage particles, a number of bacteriophage encode their own RNA polymerases (RNAPs). Soon after the bacterium is infected by one of these phage, the endogenous bacterial RNAP is inactivated, and the phage encoded RNAPs take over transcription of the phage genes. These phage RNAPs are structurally simpler than the familiar cellular RNAPs (Sousa, 1997). For example, the best characterized of these phage enzymes, that of the T7 bacteriophage, is a single subunit enzyme of 99 kD, which may be contrasted with E. coli RNAP which is a 5 subunit enzyme of approximately half a million Daltons. Despite this relative simplicity, these phage RNAPs are fully competent in carrying out a transcription reaction similar to that carried out by the cellular RNAPs, and they are also subject to regulation during the course of a phage infection. Because of their relative structural simplicity, these enzymes became attractive targets for investigators interested in understanding the fundamental mechanisms of transcription (Mcallister, 1993a).

Several of these RNAP genes were cloned in the early 1980’s. Analysis of their sequences revealed extensive homology between the RNAPs encoded by different phage (Mcallister, 1993b). The cloning of a yeast mitochondrial RNAP (mtRNAP) gene (Masters et al., 1987), and later of mammalian mtRNAP genes (Tiranti et al., 1997), unexpectedly revealed that the mtRNAPs are also homologous to the phage enzymes. This ‘phage RNAP’ family is therefore very widespread. The availability of these cloned genes allowed overexpression systems for the phage RNAPs to be developed. The ability to produce large amounts of these RNAPs has had an enormous impact on the RNA structure–function field, since it has allowed investigators to prepare specific RNAs by in vitro transcription of DNAs linked to phage promoters (Milligan et al., 1987). The phage RNAP genes, particularly that of the T7 phage, have also been used to construct overexpression systems for heterologous proteins in both E. coli and eukaryotic hosts (Tabor and Richardson, 1985; Studier and Moffat, 1986; Fuerst et al., 1986). Use of these overexpression systems is now widespread, and the study of phage RNAPs is therefore justified both from a fundamental viewpoint and because of the technical applications of these exceptional enzymes.

This review focuses almost exclusively on the RNAP from bacteriophage T7. This is the only member of this family of RNAPs whose 3-dimensional structure has been determined, and it has therefore become the object of intensive study. Most of what we have learned about structure–function relationships and mechanisms for this enzyme is directly relevant to other members of the ‘T7–like’ RNAPs (sometimes referred to as the ‘single–subunit RNAPs’), and much of this information can also help us understand fundamental aspects of transcription common to both single–and multi–subunit RNAPs. The first part of this review addresses T7 RNAP mechanism from a structure–function perspective: the roles of individual T7 RNAP domains, subdomains, or single amino acids in executing the transcription reaction are described. The second part of this review addresses aspects of T7 RNAP mechanism which are not amenable to being addressed from in terms of the structure–function relationships of the RNAP.

T7 RNA Polymerase Structure–Function Relationships

T7 RNAP has been crystallized as an apoenzyme (Sousa et al., 1994), in complexes with promoter DNA (Cheetham and Steitz, 1999) or with the transcriptional regulator T7 lysozyme (Jeruzalmi et al., 1998), and as an initial transcription complex carrying a 3 base RNA molecule (Cheetham et al., 1999). The structure of the apoenzyme was described first (Fig. 1a), and revealed a highly α-helical molecule approximately 60 Åon each side. A deep cleft running the length of one side of the molecule was proposed to bind DNA and to contain the active site where RNA synthesis takes place. One of the most impor-
Fig. 1. A: Ribbon diagrams of the polymerase domain of DNA Polymerase I (left; Beese et al., 1993) and of the entire T7 RNA Polymerase molecule (right). The thumb, palm, and fingers subdomains of the DNAP 1 polymerase domain are indicated. Homologous structures occur in the T7 RNAP molecule, but T7 RNAP contains 4 additional structural elements which are indicated as the N-terminal domain, the extra 4-helix bundle, the promoter recognition loop, and the C-terminal loop. The active site in both enzymes is marked by two spheres indicating the position of the catalytic carboxylates. B: The pattern of motif conservation in the nucleic acid polymerases: h, hydrophobic residues; +, positively charged residue; -, any residue; ., a sequence gap. Motifs A, B, and C line the active site and are present in DNAP 1, T7 RNAP and other Pol I family members. Motifs A, B, C, D, D', and E identified by Poch et al., 1989 and Delarue et al., 1990. The T/DxxGR motif was identified by Blasco et al., 1995.

Important findings revealed by this structure was its extensive similarity to the structures of the Klenow fragment of DNA polymerase I and HIV I reverse transcriptase (Sousa, 1996). The possibility of such similarity had been suggested on the basis of the detection of a few short sequence motifs which appeared to be conserved in most nucleic acid polymerases, irrespective of their template or substrate specificity (Fig. 1b; Poch et al., 1989; Delarue et al., 1990; Blasco et al., 1995). However, these sequence similarities were so limited that it was unclear if they reflected genuine and extensive 3-d structural homology, if structural similarity was localized only to the active site, or if, in fact, these sequence similarities were only apparent and did not reflect underlying structural similarity. The X-ray studies confirmed the presence of extensive structural similarity and also showed that the conserved sequence motifs corresponded to the amino acids which lined the active sites of these enzymes.

The structural similarity between DNAPI and T7 RNAP was sufficiently extensive to justify inclusion of T7 RNAP in the Pol I family. Approximately half of the T7 RNAP molecule was seen to be homologous to the polymerase domain of DNAP 1. This homology included identical folding topology and conservation of almost all secondary structure elements. The overall shape of this polymerase domain was similar to a cupped right hand and this domain was further subdivided into three subdomains dubbed ‘thumb’, ‘palm’, and ‘fingers’. Each of these elements has been subject to extensive structure–function analysis in T7 RNAP as well as in related polymerases. This has made it possible to assign distinct functions to each of these domains.

The thumb subdomain: In the Pol I family this domain is N-terminal to the palm and fingers subdomains. In T7 RNAP it is comprised of residues 330–414, and forms a largely α-helical projection on one side of the template binding cleft (Fig. 1). In the apoenzyme the part of the thumb subdomain which projects furthest from the polymerase (residues 345–383) is disordered (Sousa et al., 1994; Jeruzalmi et al., 1998), but assumes α-helical structure upon binding template (Cheetham and Steitz, 1999). Similar template–binding dependent order–disorder transitions or movements of their respective subdomains have been observed in crystal structures of a number of polymerases (reviewed in Sousa, 1996). On the basis of such observations it has been proposed that a primary function of the thumb subdomain is to hold the polymerase on the DNA during transcript or primer elongation, either through direct interactions between thumb subdomain side-chains and the DNA or RNA phosphate–ribose backbone, or by forming a structure which wraps around the DNA and keeps the polymerase from falling off but still allows it to slide along the DNA. Such a function is supported by a number of studies showing that deletions or point mutations in the thumb subdomains of different polymerases decrease their processivity and increase the rates of elongation complex dissociation (Bebenek et al., 1995; Bonner et al., 1994b; Minnick et al., 1996; Gopal et al., 1999). In particular, it has recently been shown that a
cluster of positively charged amino acids (residues 391–395) of the T7 RNAP thumb subdomain contact the phosphate-ribose backbone of the RNA 4–6 nucleotides away from the RNA 3′-end. Point mutations in these amino acids greatly increase the spontaneous dissociation rate of the T7 RNAP elongation complex (Brieba et al., 2001).

The palm subdomain: In the Pol I family the palm subdomain is not contiguous in sequence, but is interrupted by the fingers subdomain (Fig. 1). In T7 RNAP the palm subdomain is comprised of residues 414–448, 535–538, and 787–819. The palm subdomain lines one side of the template binding cleft with 4 or 5 β-strands (4 in T7 RNAP, 5 in DNAP 1). The essential catalytic residues are found in the palm subdomain within motifs A and C (Fig. 1), which form T7 RNAP β-strands 5, 8, and 9. In T7 RNAP these motifs include aspartic acids 537 and 812, both of which are involved in binding the two Mg++ ions which catalyze phoshodiester bond formation. Mutation of either of these aspartate essentially inactivates the polymerase (Bonner et al., 1992; 1994a). It has been proposed that the two Mg++ ions catalyze bond formation by: (1) stabilizing the development of negative charge on the 3′-OH to make it a more effective nucleophile, (2) stabilizing negative charge development on the pyrophosphate so as to facilitate the release of this group and, (3) stabilizing the structure and charge of the pentacovalent transition state centered on the α-phosphosphate (Steitz, 1998). There does not appear to be any acid-base catalysis or a direct role for any amino acid side chain in this reaction; i.e., the role of the polymerase is essentially limited to positioning the 3′− nucleotide, the incoming NTP, and the Mg++ ions in appropriate geometry for catalysis. A similar active site architecture and catalytic mechanism may be common to many enzymes which carry out phosphoryl transfer reactions (Steitz, 1998), though its occurrence may more often reflect convergent evolution rather than true homology.

A β-hairpin formed by strands 3 and 4 does not contain residues directly involved in catalysis. However, this hairpin does include the well-conserved T/DxxGR sequence motif (Fig. 1b), and both mutagenesis and structural studies demonstrate that this element contacts the 3′−segment of the RNA in a T7 RNAP transcription complex (Fig. 2).

The fingers subdomain: The fingers subdomain of T7 RNAP is formed by residues 539−738 and 771−786. In all of the Pol I family members the bulk of the interactions with the template strand are made by the fingers subdomain. Like the thumb subdomains, the most distal elements of the fingers subdomains contain disordered elements which become ordered upon binding template (Jeruzalmi et al., 1998; Cheetham and Steitz, 1999). In addition, the fingers subdomain also participates in a remarkable conformational transition during every cycle of nucleotide addition. This conformational transition was first predicted by Benkovic, Johnson, and co-workers who, on the basis of kinetic studies, proposed that in the absence of bound dNTP, DNAP 1 formed an ‘open’ complex which was catalytically incompetent and able to slide on the DNA (reviewed in Johnson, 1993). Upon binding dNTP, the polymerase was proposed to isomerize to a ‘closed’, catalytically competent form which could no longer slide on the DNA. This conformational isomerization has now been observed directly by comparison of structures of DNAPs with and without bound dNTP. It is seen to involve a rotation of the fingers subdomain that leads to a closing in of the fingers around the template: primer (Doubli et al., 1998; Kiefer et al., 1998; Li et al., 1998). The T7 RNAP initial transcription complex structure which has been described (Fig. 2) does not contain bound NTP, and the conformation of the fingers and template strand in this complex is similar to that seen in the DNAP ‘open’ complexes. Though the ‘closed’ structure of T7 RNAP (with NTP bound) is not yet available, the high degree of homology and conservation of mechanism amongst the Pol I family members and the similarity in the ‘open’ complex structures of T7 RNAP and the DNAPs means that it is almost certain that a similar induced fit mechanism will operate in T7 RNAP. Additional support for such a mechanism has recently been obtained in studies examining the effect of NTP binding on the stability and lateral mobility of T7 RNAP elongation complexes. These studies showed that the binding of NTP resulted in a remarkable stabilization of the elongation complex against dissociation (Brieba et al., 2001; Mentesanas et al., 2000), as well as a restriction in the lateral mobility (sliding) of the complex on the DNA (Huang and Sousa, 2000). The resistance to sliding due to NTP binding was most apparent at the downstream end of the complex on the template strand. In agreement with the transitions observed in the DNAP crystal structures which showed the fingers subdomain clamping down on the downstream end of the template strand upon NTP binding.

The energy driving this induced fit transition comes from interactions with the NTP. In particular, amino acids within T7 RNAP helix Y (amino acids 627−637) and the turn immediately C-terminal to this helix (amino acids 628−641) contact the triphosphate and ribose moieties of the incoming NTP. This part of the polymerase corresponds to conserved motif B (Fig. 1b). Residues within 627−634 of helix Y, including the invariant lys 631, probably interact with the triphosphate moiety of the NTP, while residues within 635−639 interact with the ribose moiety. Mutations which alter the ribose specificity of polymerases are found in the latter segment. For example, mutation of the phe 762 in DNAP 1 (which corresponds to T7 RNAP M 635) to tyrosine greatly enhances the ability of the DNAP to utilize 2′−3′−ddNTPs (Tabor and Richardson, 1995; Astatke et al., 1998a). Mutation of T7 RNAP tyr 639 to phenylalanine enhances the ability of T7 RNAP to use 2′−
dNTPs (Sousa and Padilla, 1995). In fact, this Y639F mutation makes T7 RNAP almost completely indiscriminate with regards to the chemical character of the ribose 2'-substituent on the rNTP (though substituents that are much larger than a 2'-OH group are not effectively utilized), and this mutant enzyme has now become widely used by investigators wishing to incorporate non-canonical 2'-groups into RNAs for structural or functional studies (Padilla and Sousa, 1999; Sousa, 2000). The basis for the effect of this mutation is now understood, since it appears that T7 RNAP discriminates between rNTPs and dNTPs via a hydrogen bond between the Y639 hydroxyl and the 2'-OH of the rNTP (Huang et al., 1997; Brieba and Sousa, 2000). Interestingly, this tyrosine is conserved in DNAPs (in DNAP I it corresponds to Y766, see Fig. 1b), where it must clearly have a different function. Resolution of the apparent paradox that this amino acid may be conserved structurally, but not functionally, emerges from an examination of the mechanism of rNTP/dNTP discrimination in the DNAPs. Such discrimination relies on a steric clash between an active site glutamic acid side-chain and the 2'-OH group of an rNTP (Doublié et al., 1998; Asstatke et al., 1998b). In T7 RNAP, a glycine takes the place of this glutamic acid and the smaller side-chain makes room for the 2'-OH group. It turns out that the glutamic acid in the DNAPs H-bonds to the tyrosine which corresponds to T7 RNAP Y639. Thus, substitution of this glutamate with glycine (in T7 RNAP) not only makes room for a ribose 2'-OH group, but also frees the tyrosine to make an H-bond with the 2'-OH group. This is a particularly clear example of how the function of an amino acid depends not only on what it is and where it is, but also on the amino acids near it.

The induced fit transition from the open to the closed complex represent the rate limiting step in nucleotide ad-
dition, and has been proposed to be critical in differentiating between complementary and non-complementary nucleotides so as to limit base substitution misincorporation (Johnson, 1993). Consistent with this idea, most mutations which increase misincorporation in DNA polymerases map to the fingers subdomain, and, in particular, map to elements which contact the template strand near the 3'-end of the primer and which correspond to DNA I helix O and the turn connecting helices O and P (Minnick et al., 1999; Osheroff et al., 1999). A recent study of the fidelity mechanisms of T7 RNAP similarly found that mutations in g1y 640, phe 644, or g1y 645 (all within the turn connecting the helices which correspond to DNA I helices O and P) increase base misincorporation (Huang et al., 2000). This study also found that the intrinsic fidelity of T7 RNAP was comparable to that of exonuclease deficient Klenow fragment (~1 base substitution error in 2 x 10^4 bases), however, unlike what has been reported for most DNA polymerases or the multisubunit RNA polymerases, it was also found that T7 RNAP displayed no post-misincorporation proofreading.

Accessory Domains: The portion of T7 RNAP which encompasses the thumb, palm, and fingers subdomains corresponds to the polymerase domain and is homologous to the polymerase domains of other Pol I family members. The polymerase domain accounts for roughly half of the T7 RNAP molecule and its function is limited to carrying out template-directed, processive nucleic acid polymerization. In addition to this activity, which is common to all processive nucleic acid polymerases, T7 RNAP displays additional activities. These include promoter, pause site, and terminator site recognition (Pol I family DNA polymerases display none of these), template unwinding activity (DNA polymerases usually work with a single-stranded template, while T7 RNAP uses a duplex template), and RNA displacement activity (DNA I leaves its product hybridized to the template strand, while T7 RNAP actively displaces RNA into solution). These additional functions appear to be conferred on T7 RNAP through addition of novel domains which are not present in the Pol I family DNA polymerases (similarly, the DNA polymerases display domains and activities, such as proofreading, which are not found in T7 RNAP). Thus the structure-function organization of T7 RNAP is modular: common, conserved functions reside in a conserved, core polymerase domain; additional functions are achieved through addition of domains unique to specific classes of polymerases. In T7 RNAP these additional domains include the N-terminal domain (residues 1-311), the central 4-helix bundle domain (residues 449-531), the promoter recognition loop (residues 739-770), and the C-terminal loop (residues 820-883). Structural and structure-function studies over the past decade have made it possible to assign specific functions to each of these accessory domains.

The N-terminal domain: The largest accessory domain in T7 RNAP is the N-terminal domain. This element has roles in promoter recognition and promoter opening, in nascent RNA binding and displacement, and possibly in pause site recognition. A role for this element in promoter recognition was revealed by the crystal structures of a T7 RNAP-promoter complex and an initial transcription complex (Fig. 2). Sequence specific binding of the 23 base-pair T7 promoter involves interactions between the polymerase and the -17 to -6 base-pairs (the -5 to +6 sequence is important for promoter opening and productive initiation, but not for sequence specific binding; Ujvari and Martin, 1997; Chapman and Burgess, 1987). In the crystal structure of T7 RNAP complexed with promoter, contacts with the -6 to -17 base-pairs are made with a single contiguous surface of the polymerase which includes parts of the promoter recognition loop and the N-terminal domain (Fig. 2). In particular, the N-terminal domain contacts and bends the AT-rich -13 to -17 sequence of the promoter.

The T7 RNAP-promoter structure also revealed an 'intercalating β-hairpin' (residues 231-241) which inserted between the template and non-template strands of the promoter and made stacking interactions with the exposed bases (Fig. 2). A role for this hairpin in promoter opening was confirmed by experiments showing that point and deletion mutations in this element reduced the ability of the RNAP to open or initiate on duplex promoters, but did not reduce activity on templates in which the promoter had been artificially opened by removal of part of the non-template strand (Brieba and Sousa, 2001). On the basis of the structural work it had also been suggested that the intercalating hairpin would be involved not only in opening the promoter during initiation, but also in holding the transcription bubble open during transcript elongation. However, this suggestion was not supported by the mutagenesis results. While mutations in the intercalating hairpin clearly reduced promoter opening during initiation, the same mutations did not alter the stability or structure of the transcription bubble in the elongation complex. Mutations in the intercalating hairpin also did not affect the polymerase's ability to displace the RNA during transcript elongation.

While mutations in the intercalating hairpin did not affect RNA displacement activity, a role for the N-terminal domain in nascent RNA binding and displacement is indicated by results from mutagenesis and proteolysis of this domain. T7 RNAP contains a region between amino acids 170 and 180 which is hypersensitive to proteolysis. When the enzyme is cleaved in this region the resulting fragments remain non-covalently associated and the polymerase retains catalytic activity (‘nicked’ RNAP). However, its processivity, RNA binding and displacement activity, and elongation complex stability are all reduced (Ikedo and Richardson, 1987a, 1987b; Muller et al., 1988; Gopal et al., 1999; Huang et al., 1999; Mentesanas et al., 2000).
In addition, protection of the RNA in the elongation complex is reduced. In the normal elongation complex ~ 8 bases of RNA are strongly protected from RNase T1 digestion, probably because they are hybridized to the template. Another ~ 6 bases of RNA are more weakly protected from T1 digestion (Huang and Sousa, 2000). An elongation complex formed with the ‘nicked’ RNAP displays strong protection of ~ 8 bases of RNA, but the weak protection of the 6 additional bases seen in the normal elongation complex is eliminated (Huang and Sousa, 2000). This suggests that protection of these 6 additional bases is due to an interaction with the N-terminal domain which is disrupted in the nicked enzyme. Significantly, residues 170–180 lie on the ‘back’ side of the RNAP, well away from the any likely contacts with RNA or DNA. It therefore appears likely that proteolytic nicking affects the polymerase’s activity by altering the conformation or position of the N-terminal domain. This hypothesis is consistent with the observation that nicking reduces protection over a ~ 10 base-pair region of the footprint of T7 RNAP on its promoter (Ikeda and Richardson, 1987b), a disruption which is too extensive to explain in terms of loss of RNAP : promoter interactions localized to a few amino acids. It is also consistent with the observation that insertions or deletions within amino acids 170–180 cause a phenotype similar to nicking, while point mutations in this region do not (Lyakhov et al., 1997), indicating that the mutant phenotype depends on alteration of the normal path of the polypeptide backbone rather than loss of specific side-chains. A final aspect of the phenotype of the nicked RNAP is loss of recognition of the T7 concatemer junction pause/termination site (Lyakhov et al., 1997). This may mean that the N-terminal domain plays an important role in recognizing or responding to the concatemer junction sequence. Alternatively, it may be that this site is not directly involved in concatemer junction recognition, but that the disruption of the RNA : RNAP interaction in the nicked EC allows formation of alternate RNA secondary structures or interactions which interfere with concatemer junction recognition.

The promoter recognition loop: The promoter recognition loop in T7 RNAP is formed by residues 739–770. Evidence that this element is involved in T7 promoter recognition was first obtained in experiments showing that point mutations in this region could alter promoter specificity (Raskin et al., 1992, 1993). Swapping of the recognition loops between T7 and T3 RNAPs, for example, exchanges the promoter specificities of the resulting chimeras (Raskin et al., 1992). The crystal structure of a T7 RNAP : promoter complex confirmed the role this loop in promoter recognition, revealing extensive interactions between amino acids 746, 748, 756, and 758 and base-pairs −7 to −11. Interestingly, a recent cross-linking study suggested a second role for this element. It revealed the presence of an interaction (or close approach) between the promoter recognition loop and the RNA ~ 8 bases upstream of the 3’-end in an elongation complex (Temniakov et al., 2000). This observation raises the intriguing possibility that the establishment of this interaction may help trigger the release of the RNAP from the promoter. Thus, in the initial transcription complex, the promoter recognition loop makes interactions with the promoter. However, once the RNA reaches 8 bases in length, it may establish an interaction with the promoter recognition loop and this may simultaneously disrupt the promoter interaction so as to free the transcription complex to move away from the promoter.

The C-terminal loop: The C-terminal loop of T7 RNAP corresponds to residues 820–883. Much of this loop is exposed on the surface of the polymerase, though the C-terminus of this loop (residues 882–883) is tucked back into the polymerase in a pocket immediately adjacent to the active site (Fig. 3a). DNAP1 also contains a C-terminal loop, but its structure differs markedly from that of T7 RNAP (Fig. 3b). The structural differences between the C-terminal loops of DNAP1 and T7 RNAP may explain why only the latter enzyme is regulated via an interaction with T7 lysozyme (Moffat and Studier, 1987). This interaction regulates T7 RNAP activity during T7 phage infection. Late in infection T7 lysozyme accumulates and binds to T7 RNAP. The binding of lysozyme to T7 RNAP inhibits transcription initiation, but inhibition is much greater for transcription of the T7 class II promoters. Class II promoters are generally weaker and display various base differences (usually in the −4 to +4 region) relative to the stronger T7 class III promoters (Moffat et al., 1994). Proteins required during the early and middle stages of T7 phage infection are usually expressed from class II promoters, while proteins required late in infection are expressed from class III promoters. The net effect of the greater repression of class II promoters due to T7 lysozyme accumulation late in infection is an increase in the relative synthesis of proteins expressed from class III promoters (Moffat and Studier, 1987). The greater repression of the class II promoters caused by binding of T7 lysozyme to the polymerase generally decreases its affinity for NTPs (Villemin and Sousa, 1998; Huang et al., 1999). While this effect is itself nonspecific, the intrinsically higher K_{NTP} for initiation from class II promoters means that the repressive effect of a decrease in NTP affinity leads to greater repression of transcription from these promoters. The loss in NTP affinity upon lysozyme binding to the polymerase appears to be due to a conformational change in the C-terminus of the RNAP (Huang et al., 1999). Comparison of structures of
the free and lysozyme-complexed RNAPs reveals that lysozyme causes the C-terminus to flip out of the active site (Fig. 3 a). A study in which the exposure of the RNAP C-terminus was characterized by carboxypeptidase sensitivity confirmed that the structural transition seen in these crystal structures is due to lysozyme binding (Huang et al., 1999), and was not an artifact of different crystallization conditions or crystal packing environments. Mutation of the T7 RNAP C-terminal residues or deletion of Ala 883 decreases NTP affinity, so it is easy to understand why lysozyme binding should have similar effects (Patra et al., 1992; Gardner et al., 1997).

Based on this understanding of the lysozyme mechanism, it is possible to understand how the structural differences between the C-terminal loops of T7 RNAP and DNAP 1 underlie the lysozyme sensitivity of the former enzyme. The C-terminal loop of DNAP 1 differs from that of T7 RNAP in two ways. First, the DNAP 1 loop lacks a long, surface exposed extension which forms part of the T7 lysozyme binding site. Second, interactions between the C-terminal residues of the DNAP 1 loop and the DNAP 1 active site are extensive (the C-terminal loop of DNAP 1 forms the 5th strand of an extended β-sheet which lines the palm subdomain). By comparison, the C-terminus of the T7 RNAP loop is truncated, and the interactions with the active site residues are much more limited. Thus the structural differences between the C-terminal loop of DNAP 1 and T7 RNAP: (1) create a binding site for T7 lysozyme in T7 RNAP and (2) facilitate the extraction of the T7 RNAP C-terminus from its position adjacent to the active site.

The 'extra' 4-helix bundle: This structure is formed by residues 449–531 (Fig. 1 a). T7 RNAP mutants from which this element is deleted form inclusion bodies and cannot be resolubilized to active enzyme, and most point mutants in this region either have no apparent phenotype or are generally poorly active (author’s laboratory, unpublished observations). There are therefore no data which reveal a distinct function for this element.

We may now turn to questions of T7 RNAP mechanism which are not readily addressed from the perspective of the structure–function relationships of the RNAP.

Mechanisms of T7 RNAP Transcription

T7 promoter structure and function: The consensus T7 promoter is 23 base-pairs in length and may be divided into 3 functionally distinct segments. Base-pairs from -17 to -6 form a recognition element which is important for
sequence specific binding. Mutations in this region reduce RNAP : promoter affinity, and this part of the promoter must be duplex for maximal binding (Maslak and Martin, 1993). The −4 to −1 TATA element is important for promoter opening. Both crystallographic and solution studies show that, in the open complex, the promoter is melted up-stream to −4 (Brieba and Sousa, 2001 ; Bandwar and Patel, 2001). Mutations in this region reduce initiation rates, but not RNAP : promoter affinity (Chapman and Burgess, 1987). The effects of these mutations can be overcome by using a negatively supercoiled template (Chapman and Burgess, 1987), presumably because supercoiling facilitates opening. The +1 to +6 initially transcribed sequence (‘GGGAGA’) is also not important for RNAP : promoter affinity, but is important for efficient initiation and progression through the initial or ‘abortive’ phase of transcription. In particular, T 7 RNAP has a strong preference for initiating with ‘GG’, and a change in either the +1 or +2 base will reduce the initiation rate (Milligan et al., 1987). Divergence from consensus in the initially transcribed sequence will also increase the rate at which the RNAP aborts (releases the short nascent transcript and then reinitiates) during initial transcription (Lopez et al., 1997 ; Guajardo et al., 1998). The promoter itself can be single-stranded (missing non-template strand) downstream of −5 with no loss of activity (Maslak and Martin, 1993). In fact, RNAP will bind to a partially single-stranded promoter 2−3 orders of magnitude more tightly than to a fully duplex promoter (Diaz et al., 1996 ; Bandwar and Patel, 2001).

Promoter binding and opening: Upon binding promoter T 7 RNAP induces a large bend in the DNA (Ujvari and Martin, 2000). This has been observed both crystallographically and through gel shift analysis. Bending may help nucleate promoter opening. Opening itself is extremely rapid and does not limit the maximal rate of initiation (Bandwar and Patel, 2001). However, until stabilized by binding of the initiating NTPs, the open complex is thermodynamically less stable than the closed complex. In the RNAP : promoter binary complex, closed and open complexes are therefore in rapid equilibrium, and closed complexes are 2−7−fold more abundant than open complexes (Villemain et al., 1997 ; Bandwar and Patel, 2001). T 7 promoter opening kinetics are therefore similar to those of E. coli ribosomal RNA promoters which also display kinetically rapid, but thermodynamically disfavored, open complex formation (Gaal et al., 1997). The net effect of such kinetics is to cause initiation rates to be sensitive to initiating NTP concentrations, because NTP binding is required to stabilize the open complexes. This may explain why T 7 RNAP transcription, like E. coli rnr transcription, is sensitive to physiological conditions that cause NTP levels to drop (Yamagishi et al., 1987). Though it has been suggested that the decrease in T 7 RNAP activity observed under such conditions may be due to the accumula-

Progression through initial transcription and promoter release: During the initial, or abortive, phase of transcription T 7 RNAP retains its interaction with the recognition element of the promoter, as revealed by the observation that the footprint on the promoter lengths via downstream extension while the upstream border of the footprint remains in place (Ikeda and Richardson, 1986 ; Mookhtar et al., 1991). During this initial phase the RNAP frequently aborts transcription (releases the nascent transcript) and reinitiates (Martin et al., 1988). When the RNA reaches a length of ~8 bases the polymerase releases the promoter (the footprint is observed to move downstream ; Ikeda and Richardson, 1986 and Brieba and Sousa, unpublished observation) and transcript release is curtailed. The mechanism by which the RNAP is able to maintain interactions with the upstream elements of the promoter while still transcribing as far downstream as +8 has been unclear, and it has been variously proposed that the RNAP could either stretch, or that the DNA would be looped out or compressed (scrunch) within the RNAP. The crystal structure of an initial transcription complex implies that DNA ‘scrunching’ is the correct explanation because the DNA is observed to become scrunched into a deep pocket in the RNAP as transcription proceeds from +1 to +3 (Cheetham et al., 1999). This study also suggests that scrunching could trigger promoter release: the pocket in T 7 RNAP which is proposed to accommodate the scrunched DNA appears to be large enough to accommodate no more than 6−8 nucleotides of DNA. Therefore, once this pocket is filled, the RNAP might have to release the promoter to transcribe further. Alternate models suggest that when the RNA reaches a certain length and contacts a site on the RNAP it causes releases, either via an allosteric mechanism, or by directly disrupting the promoter interaction (see section on promoter recognition loop, above, and Temiakov et al., 2000). These models are not necessarily exclusive and it is possible that RNA length and scrunched DNA both contribute to releasing the RNAP from the promoter.

Elongation complex structure, stability and movement: Unlike the initial transcription complex, the elongation complex is remarkable stable. The spontaneous dissociation rate of an elongation complex halted on a duplex template is on the order of 0.1min−1 (Gopal et al., 1999 ; Mentesan es et al., 2000 ; Huang et al., 1999 ; Brieba et al., 2001). A crystal structure of a T 7 RNAP elongation complex has yet to be described. In its absence, solution studies have attempted to provide answers to the following questions: What confers such remarkable stability on the
elongation complex and how is such stability consistent with a requirement for RNAP movement on the DNA? What are the lengths and dispositions of the RNA : DNA hybrid and transcription bubble? What is the extent of the RNA : RNAP interaction and how does the nascent RNA exit the elongation complex?

Multiple RNA : RNAP, RNAP : DNA, and RNA : DNA interactions contribute to the stability of the elongation complex. The fingers and palm subdomains interact with the 3'-segment of the RNA : DNA hybrid, and the thumb subdomain makes direct interactions with the RNA phosphate-ribose backbone, and may also contribute to elongation complex stability by wrapping around the DNA (Cheetham et al., 1999; Brieba and Sousa 2001). These interactions are common to both the initial transcription complex and the elongation complex. The greater stability of the elongation complex reflects the establishment of an additional interaction between the N-terminal domain and single-stranded RNA 8-14 bases upstream of the RNA 3'-end (Gopal et al., 1999; Montesanas et al., 2000; Huang et al., 1999; Huang and Sousa 2000). The RNA : DNA hybrid also makes an important contribution to elongation complex stability. On the basis of the structure of an initial transcription complex, it was initially proposed that this hybrid could be no more than 3 bases in length because a larger hybrid modeled into this structure would cause a steric clash with the N-terminal domain (Cheetham et al., 1999). However, the release of the RNAP from the promoter is known to be accompanied by a conformational change, most probably involving the N-terminal domain (Sousa et al., 1992). It therefore appears likely that the hybrid in the elongation complex could be larger than suggested by the structure of the initial transcription complex. Indeed, solution experiments measuring hybrid length by a variety of approaches concur in supporting a hybrid length of ~8 base-pairs (Huang and Sousa 2000; Temiakov et al., 2000; Brieba and Sousa, 2001). The transcription bubble appears to be just large enough to accommodate this hybrid. It is ~9 bases long, with the DNA bases immediately downstream of the RNA 3'-end paired, and only 1 or 2 base-pairs upstream of the hybrid unpaired (Huang and Sousa, 2000). The basic features of T 7 RNAP elongation complex structure are summarized in Fig. 4. The hybrid and transcription bubble sizes are not, however, always constant. Though T 7 RNAP actively displaces the RNA from the template via the interaction between the N-terminal domain and the RNA, reannealing of the template and non-template strands also makes an important contribution to RNA displacement. When the driving force for reannealing is weak, extended hybrids tend to form (Gopal et al., 1999). This occurs, for example, on supercoiled templates. The ability of T 7 RNAP to occasionally form such extended hybrids may be important to its role as a priming RNAP during T 7 DNA replication (Fuller and Richardson, 1985).

Given the extensive RNA : DNA and RNA : RNAP interactions proposed to be important for elongation complex stability, the question of elongation complex movement may appear problematic. However, these RNA : DNA or RNAP : RNA interactions are not sequence specific, and, in the absence of bound NTP, the RNAP appears relatively free to slide on the DNA, with its average position and range of sliding determined by interactions with the RNA and DNA and by the relative energetics of competing positions for the transcription bubble and RNA : DNA hybrid (Huang and Sousa, 2000). Thus elongation complex movement appears to be driven passively, by random fluctuations. When the elongation complex happens to slide into the appropriate (post-translocated) position, the NTP can bind. The NTP-induced 'open' to 'closed' isomerization then restricts the lateral mobility of the complex. This mechanism for polymerization translocation is likely to be general, and may be described as a translocational equilibrium in which multiple elongation complex species (distinguished by their position on the DNA) rapidly exchange with one another by random sliding (Guajardo and Sousa, 1997). It is important to understand that the effective 'concentration' of the elongation complex species which is properly positioned to bind NTP is likely to be strongly affected by local sequence or by opposed DNA binding proteins (Guajardo et al., 1998). As a
consequence the $K_{\text{off}}$ for bond formation may vary almost indefinitely as the polymerase moves along the DNA. Such variation has, in fact, been observed for *E. coli* RNAP (Levin and Chamberlin, 1987). This variation in apparent $K_{\text{off}}$ may be important in modulating RNAP progression, in pausing, and in sequence specific regulation.

**Pausing and termination**: T 7 RNAP terminates specifically at two types of terminators (McAllister, 1997; Hartvig and Christiansen, 1996; Jeng et al., 1990; He et al., 1998; Lyakhov et al., 1997). Class I terminators are similar in structure to the canonical rho-independent or intrinsic terminators of *E. coli* RNAP. They contain a sequence which can form a stable hairpin when transcribed into RNA, and immediately following the hairpin is a stretch of uracil residues. Formation of the hairpin is likely to be important for disrupting an RNA : RNAP interaction normally made with single-stranded RNA, and the stretch of uracils may be important for destabilizing the RNA : DNA hybrid, as well as in slowing the progression of the RNAP. The combination of loss of the RNA : RNAP interaction due to hairpin formation, and the weak U–rich RNA : DNA hybrid may be sufficient to cause elongation complex dissociation (Hartvig and Christiansen, 1986).

Class II terminators contain the sequence AUCUGUU, followed by at least 6 uracils. There is no clear secondary structure in the RNA transcribed at these terminators which can explain how they cause termination, and it is possible that recognition of these terminators involves sequence specific RNAP : RNA or RNAP : DNA interactions (He et al., 1998; Lyakhov et al., 1997, 1998). A class II terminator was first identified in a human PTH gene cloned downstream of a T 7 promoter (Mead et al., 1986). Its physiological significance was initially unclear (in contrast, T 7 DNA contains a number functionally important class I terminators), but a similar sequence was then identified at the T 7 DNA concatemer junction (CJ; Lyakhov et al., 1997, 1998). The CJ contains the same AUCUGUU element but without the extended stretch of uracils. Apparently, the absence of the long uracil segment causes the CJ to act primarily as a pause site rather than a terminator. Binding of T 7 lysozyme to the elongation complex enhances pausing (or termination) at the CJ, even though the elongation complex is resistant to the effects of T 7 lysozyme when transcribing heterologous DNA (Zhang and Studier, 1997; Lyakhov et al., 1997). Pausing of T 7 RNAP at the CJ appears to be important for processing and packaging of T 7 DNA, and it has been suggested that this is because it is important to keep the region of the DNA downstream of the CJ clear of transcription during processing (Zhang and Studier, 1997; Lyakhov et al., 1997), but this question remain unresolved, as does the basic mechanism of how the CJ sequence actually causes pausing.

**Similarities to multisubunit RNAP mechanism**: The similarities between the mechanism of T 7 RNAP transcription and transcription by the multisubunit RNAPs are extensive. It is likely that many of these similarities are the result of convergent evolution. As a result they are perhaps most useful in addressing the question: What is the most efficient catalytic strategy for execution of a transcription reaction?

A promoter with separate recognition elements and a TATA (or similar) sequence used to nucleate opening appears to be common to all RNAP promoters, so this may represent an efficient solution for the problem of how recognition and binding (which must initially involve a duplex DNA) can be followed by DNA melting. Bending of the DNA upon RNAP binding is also seen with both T 7 RNAP and the multisubunit RNAPs and may be important for opening the DNA (Ujvari and Martin, 2000; Korsheva et al., 2000). The opening step itself is a key regulatory step in transcription initiation and opening is characterized by distinct kinetics in different RNAP : promoter systems (i.e., slow, favored and ATP–independent at most σ70 promoters (McClure, 1985); ATP–dependent at σ54 promoters (Tintut et al., 1995); rapid but ATP–independent and disfavored at rrn promoters (Gaal et al., 1997)). One of these classes of E. coli promoters (for ribosomal RNA) displays opening kinetics that are similar to those of T 7 promoters, and in both cases the functional consequences (high sensitivity to NTP concentrations during initiation) are similar (Gaal et al., 1997; Villemain et al., 1997).

An initial abortive phase of transcription during which the polymerase retains its interaction with the promoter and tends to release short nascent transcripts also appears to be an obligatory step in the initiation reaction of all RNAPs (McAllister, 1997; McClure 1985). This may reflect a requirement that the interaction with the 3’–5’ segment of the RNA be weak enough to allow translocation during elongation and transcript release during termination, though it may also be that abortive transcription provides an additional point for regulation and specification of productive initiation. Remarkably, for both the multisubunit RNAPs and T 7 RNAP, promoter release and the end of abortive transcription occur when the RNA reaches a length of 8 – 9 bases. This is similar to the length of the RNA : DNA hybrid in the elongation complexes of both T 7 RNAP (Huang and Sousa, 2000; Temiakov et al., 2000) and the multisubunit enzymes (Kireeva et al., 2000; Komissarova and Kashlev, 1998; Korsheva et al., 2000). It may therefore be important that promoter release occur only when the RNA has become long enough to form the full-length hybrid seen in the elongation complex. However, it is unclear why the hybrid length should be similar in all these complexes. Other aspects of elongation complex structure differ between T 7 RNAP and the multisubunit enzymes. The transcription bubble in the *E. coli* RNAP elongation complex, for example, is twice as long as...
in the T 7 RNAP complex, and the extent and nature of the RNAP : DNA interactions in these complexes also differ greatly (Korsheva et al., 2000).

Certain sequences which cause the single–subunit T 7–like RNAPs to pause or terminate have been found to similarly affect the multisubunit RNAPs (Jeng et al., 1990; Mote and Reines, 1998). The general similarity in the structure of T 7 class I terminators and intrinsic E. coli RNAP terminators has already been mentioned. Many of these similarities may be most easily understood in terms of the intrinsic energetics of nucleic acid secondary structure formation. Thus, RNA hairpin formation may be an especially efficient way to disrupt a single–stranded RNA : protein interaction, and a uracil–rich sequence may be the most effective way to weaken an RNA : DNA hybrid.

**Future directions** The past decade has seen an explosion of structural and structure–function information on T7 RNA polymerase and the complexes it forms during the course of a transcription reaction. The roles of individual domains, subdomains and even individual amino acids in the different steps of this reaction are now well understood. It is now necessary to make the picture of T7 transcription which emerges from these studies more dynamic. The challenge will be to characterize the conformational transitions and changes in DNA, RNA, and RNAP interactions which occur during the successive steps of promoter opening, initial transcription and promoter release, pausing and termination.

**References**


22) Fuller, C. W., and Richardson, C. C. (1985). Initiation of DNA Replication at the Primary Origin of Bacterio-


32. 4281–5.


