bonding environment occur dynamically during the simulations. This approach opens the door for proposing reaction pathways that involve rehybridization of C atoms in molecules. For diamond film formation we have suggested that $-CH_2$ species are highly reactive and may be important for growth. The resulting pictures present a challenge to experimentalists for laboratory confirmation and to theorists for calculating the energetics by first-principle methods (*22*).

REFERENCES AND NOTES

1. *Chem. Eng. News* **68** (no. 23), 32 (4 June 1990).
2. D. W. Brenner, *Phys. Rev. B* **42**, 9458 (1990).
3. G. C. Abell, *ibid.* **31**, 6184 (1983).
4. J. Tersoff, *Phys. Rev. Lett.* **56**, 632 (1986); *Phys. Rev. B* **37**, 6991 (1988).
5. U. Burkert and N. L. Allinger, *Molecular Mechanics* (American Chemical Society, Washington, DC, 1982); S. Weiner, P. Kollman, D. Nguyen, D. Case, *J. Comput. Chem.* **7**, 230 (1986); B. Brooks *et al.*, *ibid.* **4**, 187 (1983); K. D. Gibson and H. A. Scheraga, *ibid.* **8**, 826 (1987); N. L. Allinger, F. Li, L. Yan, J. C. Tai, *ibid.* **11**, 868 (1990); W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.* **110**, 1657 (1988); C. F. Wong and J. A. McCammon, *ibid.* **108**, 3830 (1986).
6. D. W. Brenner, J. A. Harrison, C. T. White, R. J. Colton, *Thin Solid Films*, in press.
7. D. Vanderbilt and S. G. Louie, *Phys. Rev. B* **30**, 6118 (1981).
8. A. V. Hamza, G. D. Kubiak, R. H. Stulen, *Surf. Sci.* **206**, L833 (1988).
9. M. Page and D. W. Brenner, *J. Am. Chem. Soc.* **113**, 3270 (1991).
10. Y. L. Yang and M. P. D'Evelyn, *ibid.*, in press.
11. *J. Mater. Res.* **5** (November 1990).
12. F. G. Celii, P. E. Pehrsson, H.-T. Wang, J. E. Butler, *Appl. Phys. Lett.* **52**, 204 (1988).
13. M. Frenklach and K. E. Spear, *J. Mater. Res.* **3**, 133 (1988); D. Huang, M. Frenklach, M. Maroncelli, *J. Phys. Chem.* **92**, 6379 (1988); M. Tsuda, M. Nakajima, S. Oikawa, *J. Am. Chem. Soc.* **108**, 5780 (1986).
14. W. A. Yarbrough, Diamond Optics IV, *SPIE Proc. 1534*, in press; D. N. Belton and S. J. Harris, *J. Chem. Phys.*, in press.
15. S. J. Harris, *Appl. Phys. Lett. B* **56**, 2298 (1990).
16. T. Tsuno, T. Imhi, Y. Nishibayashi, K. Hamada, N. Fujimore, *Jpn. J. Appl. Phys.* **30**, 1063 (1991).
17. D. W. Brenner and B. J. Garrison, *Surf. Sci.* **198**, 151 (1988); D. Srivastava, B. J. Garrison, D. W. Brenner, *Phys. Rev. Lett.* **63**, 302 (1989); *Langmuir* **7**, 683 (1991).
18. S. A. Adelman and J. D. Doll, *J. Chem. Phys.* **64**, 2375 (1976); R. R. Lucchese and J. C. Tully, *Surf. Sci.* **137**, 570 (1983); H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
19. Absolute temperature calibration is difficult with empirical potentials because most are not fit to the melting temperature. In addition, because of the exponential dependence of most reaction rates on temperature, it is quicker to examine reaction mechanisms at higher temperatures.
20. D. Srivastava and B. J. Garrison, *J. Chem. Phys.* **95**, 6885 (1991).
21. A. F. Voter, *Phys. Rev. B* **34**, 6819 (1986).
22. After submission of this manuscript, a preprint (D. Huang and M. Frenklach, *J. Phys. Chem.*, in press) was called to our attention. Using semiempirical electronic structure calculations, they examined a reaction pathway similar to C to E, although with a different transition state.

23. Supported by the Office of Naval Research, the National Science Foundation, and a Teacher-Scholar Award to B.J.G. from the Camille and Henry Dreyfus Foundation. Pennsylvania State University supplied a generous grant of computer time for these studies. D.W.B. is supported by the Office of Naval Research through the Naval Research Laboratory.

21 October 1991; accepted 3 December 1991

Structure of Transcription Elongation Complexes in Vivo

MARK KAINZ AND JEFFREY ROBERTS

The opening of duplex DNA in the elongation phase of transcription by *Escherichia coli* RNA polymerase in vivo was detected at a regulatory site where a prolonged pause in transcription occurs. Single-stranded DNA in the transcription bubble was identified by its reactivity with potassium permanganate ($KMnO_4$). The elongation structure in vivo was similar to that of transcription complexes made in vitro with some differences. The observed reactivity to $KMnO_4$ of the DNA template strand was consistent with the existence of an RNA-DNA hybrid of about 12 nucleotides.

THE COMPLEX OF RNA POLYMERASE (RNAP), DNA, and the elongating mRNA transcript is an intermediate in the enzymatic reaction and the site of the regulatory processes of termination and antitermination. The elongation complex of *Escherichia coli* RNAP has the following properties in vitro: (i) its footprint is more compact than that of initiation complexes (20 to 30 bp versus 65 to 95 bp) (*1, 2*); (ii) base-specific contacts are not evident (*1*); (iii) the DNA (~17 bp) is unwound in the transcription bubble (*3*); and (iv) the DNA template and the nascent RNA form a hybrid of 10 to 12 bp (*4–6*), although this value has been questioned (*7*).

During transcription in vitro of the *E. coli* bacteriophage λ late gene operon (*8*) and the late gene operons of related phages 82 (*9*) and 21 (*10*), RNA polymerase pauses after making only 15 to 25 nucleotides of RNA, at a position specific to each phage DNA. These paused elongation complexes are the substrates on which each phage gene Q-encoded antiterminator protein acts in vitro to modify RNAP. We have identified the paused elongation complexes in vivo, and have used them to study the structure of the elongation complex in vivo.

We detected paused complexes in cells using $KMnO_4$ to modify thymines and cytosines in denatured DNA (*11*) and thus to mark the transcription bubble. Cells bearing plasmids with the late promoter and initially transcribed region of the late operon of λ or related phages were treated with $KMnO_4$ during growth, and $KMnO_4$-modified pyrimidines were detected by primer extension. In some experiments the drug rifampicin, which prevents initiation but not elongation of transcription by *E. coli* RNA polymerase, was added before $KMnO_4$ treatment. Inhibition of initiation allows elongating and paused RNAP to clear the DNA and thus identifies such enzymes, and also freezes RNAP at the promoter in the open complex.

We analyzed three plasmids containing the λ late operon: (i) pXY306 (*12*) contains the wild-type λ promoter p_R' with its associated regulatory sequence *qut* (for "Q utilization"); (ii) pXY306 (+6C) is a *qut*$^-$ point mutant that does not support the in vitro pause at $^+16$–$^+17$ (and cannot be modified by Q), but still has an active promoter (*13–15*); and (iii) pXY306 (−11G) contains a mutation that nearly abolishes promoter function (*13*). Wild-type pXY306 has reactive residues at positions +1 and +2 of the bottom (template) strand, and +14, +15, and +16 of the top (nontemplate) strand (Fig. 1A) (*16*). Treatment with rifampicin alters this pattern in two ways. Reactivity at +1 and +2 (lane 2) and +14, +15, and +16 (lane 8) is abolished, and the minor reactivity of T at −11 (lane 2) and −2, −6, −7, and −8 (lane 8) increases. We infer that bases at +1 and +2 on the bottom strand and at +14, +15, and +16 on the top strand occur in single-stranded regions of the transcription bubble of RNA polymerase paused after making 16 or 17 nucleotides of the late RNA, whereas T at −11 on the bottom strand and −2, −6, −7, and −8 on the top strand are present in single-strand regions of open complex formed at the promoter. Open and paused complexes coexist in a population, although for steric reasons they probably could not exist on the same DNA molecule. A time course shows that the half-life of the paused complex in vivo is about 30 s (*17*).

This identification of melted DNA in open and paused complexes was supported by analysis of the mutant plasmids. $KMnO_4$ treatment of cells bearing pXY306 (+6C) detected reactivity ascribed to open complex [at −11 (Fig. 1A, lane 3) and weakly at −2,

Section of Biochemistry, Molecular, and Cell Biology, Biotechnology Building, Cornell University, Ithaca, NY 14853.

−6, −7, and −8 (lane 9)], but not reactivity ascribed to the paused complex [at $^{+}1$ and $^{+}2$ (lane 3) and $^{+}14$, $^{+}15$, and $^{+}16$ (lane 9)]; as expected, the open complex signals were increased by prior treatment with rifampicin (lanes 4 and 10). Finally, $KMnO_4$ treatment of cells bearing pXY306 (−11G) gave no modification of pyrimidines ascribed to either paused or open complex (Fig. 1A, lanes 5, 6, 11, and 12).

The mutants and the rifampicin treatment also revealed a general transcription-dependent background reactivity. The reactivity of +12, +5, and −4 of the bottom strand was only slightly affected by the +6C mutation, which eliminated the pause and its associated reactivity at +1 and +2, but was much reduced by rifampicin; this is true even on the −11G promoter mutant, which should not interact with RNAP. This background may be a result of transcription from other plasmid promoters.

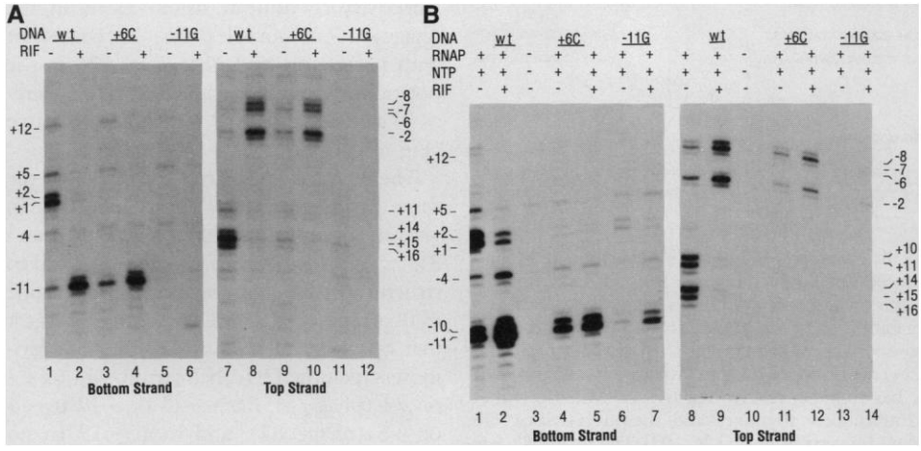

Fig. 1. $KMnO_4$ reactive pyrimidines of transcribed DNA identify paused transcription elongation complexes near the RNA start sites of lambdoid phage late operons, and open complexes at the late promoters. (**A**) Phage λ DNA in vivo. Lanes 1 to 6 show $KMnO_4$ reactive pyrimidines in the bottom (template) strand and lanes 7 to 12 show reactive pyrimidines in the top (nontemplate) strand. Odd-numbered lanes, treatment with $KMnO_4$ alone; even-numbered lanes, treatment with rifampicin (RIF) followed by $KMnO_4$. The "wt" plasmid is pXY306 (*13*); "+6C" and "−11G" are point mutants of pXY306 and are described in the text. (**B**) Phage λ DNA in vitro. Lanes 1 to 7 show reactive pyrimidines in the bottom (template) strand, and lanes 8 to 14 show reactive pyrimidines in the top (nontemplate) strand. DNA of plasmids named in (A) were transcribed in vitro (*18*) and $KMnO_4$ modification and primer extension (*19*) reactions were performed as described.

We next compared the pattern of $KMnO_4$ reactivity on the λ late promoter in vivo with that of in vitro transcription complexes. We formed open complexes with purified RNA polymerase holoenzyme and the NusA protein on supercoiled pXY306, pXY306 (+6C), and pXY306 (−11G) DNA in vitro (*18*) and either initiated transcription by addition of rNTPs or inhibited initiation with rifampicin. Complexes were treated with $KMnO_4$ and modified bases mapped by primer extension (*19*). Reactivity in vitro mirrored the pattern in vivo in almost every detail (see Fig. 1B), except that thymine at +10 and +11 was more reactive in vitro (compare Fig. 1B, lane 8, with Fig. 1A, lane 7) (*20*). In contrast to the total lack of promoter or pause signal on the −11G plasmid in vivo, we saw some reactivity of pyrimidines at −10 and −11 on the bottom strand (Fig. 1B, lane 7). Weak promoter function may be compensated by the long incubation with excess RNAP in vitro. An extremely low background reactivity was observed with supercoiled DNA in the absence of protein (Fig. 1B).

Models for the structure of transcription elongation complexes, based on measurements of complexes made in vitro, predict the promoter distal margin of the transcription bubble to be immediately downstream of the catalytic center of RNA polymerase and the promoter proximal margin of the bubble to remain a constant 17 to 18 bp upstream of the catalytic center (*3, 21*). The $KMnO_4$ reactivity of four independent paused elongation complexes on three different templates supports this model. Pyrimidines in the top strand at or just upstream of the 3′ end of the nascent RNA chain were reactive, whereas pyrimidines downstream of the RNA 3′ end were not. For example, the +16–17 paused complex of λ DNA showed top strand modification of T at positions +14 and +15 and weaker modification of C at +16, whereas C at +18 and +20 and T at +19 were unreactive. Results

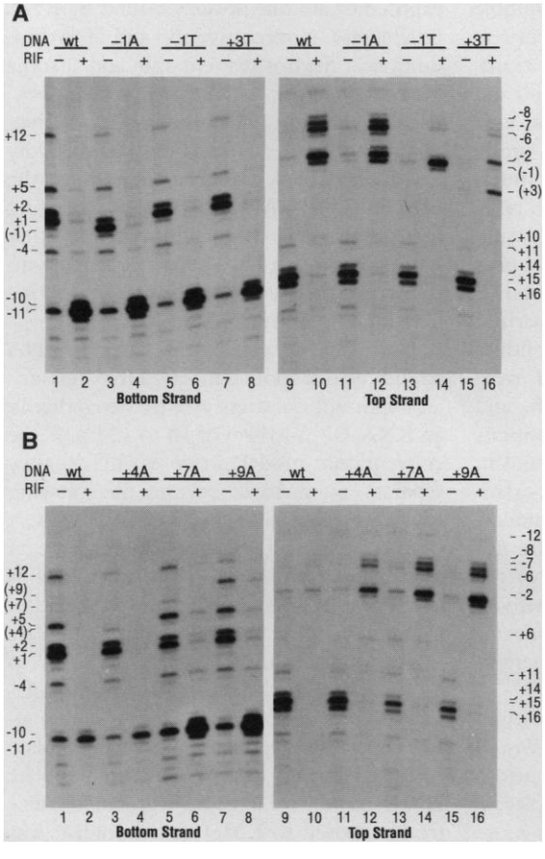

Fig. 2. Point mutations define the margins and internal structures of λ +16–+17 paused complexes and open complexes at the λ late promoter in vivo. Parentheses indicate mutant sites. (**A**) $KMnO_4$ reactivity in vivo of thymines at the margin of the transcription bubble. Lanes 1 to 8: bottom (template) strand. Lanes 9 to 16: top (nontemplate) strand. The "wt" DNA is pXY306; "−1A," "−1T," and "+3T" DNAs are point mutants of pXY306 that do not affect the ability of the template to support the transcriptional pause in vitro (*13*). (**B**) $KMnO_4$ reactivity in vivo of thymines at internal positions in the template strand of the transcription bubble. Lanes 1 to 8: bottom (template) strand; lanes 9 to 16: top (nontemplate) strand. The "wt" DNA is pXY306, and "+4A," "+7A," and "+9A" are point mutants of pXY306 that affect pausing at +16–+17 only slightly (+9A and +7A) or not at all (+4A) (*13*).

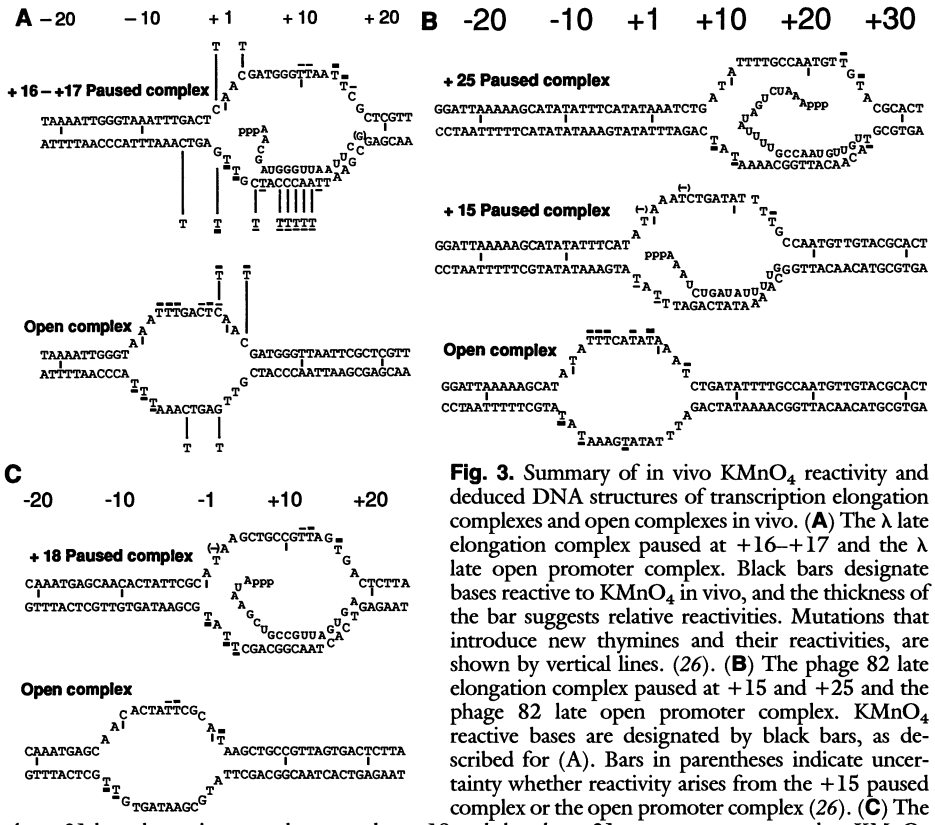

Fig. 3. Summary of in vivo $KMnO_4$ reactivity and deduced DNA structures of transcription elongation complexes and open complexes in vivo. (**A**) The λ late elongation complex paused at +16–+17 and the λ late open promoter complex. Black bars designate bases reactive to $KMnO_4$ in vivo, and the thickness of the bar suggests relative reactivities. Mutations that introduce new thymines and their reactivities, are shown by vertical lines. (*26*). (**B**) The phage 82 late elongation complex paused at +15 and +25 and the phage 82 late open promoter complex. $KMnO_4$ reactive bases are designated by black bars, as described for (A). Bars in parentheses indicate uncertainty whether reactivity arises from the +15 paused complex or the open promoter complex (*26*). (**C**) The phage 21 late elongation complex paused at +18 and the phage 21 open promoter complex. $KMnO_4$ reactive bases are indicated by black bars as described for (A) (*26*).

are similar for the +25 paused complex of phage 82 DNA and the +18 paused complex of phage 21 DNA (see Fig. 3).

The distance between the upstream edge of the bubble and the catalytic center of RNAP was also approximately constant. For the +16–17 paused complex of λ DNA, for example, T at +1 and +2 of the bottom strand was strongly reactive, whereas T at −4 was unreactive. We examined the upstream margin of the transcription bubble in the paused elongation complex of λ DNA in more detail using two point mutations that changed the DNA sequence at position −1 but did not affect the ability of the template to support a pause at +16–17. The reactivity of T at −1 in the bottom strand (mutant −1A) implies that the promoter proximal margin of the transcription bubble extended at least as far as −1 (Fig. 2A, lane 3) (however, it is possible that the mutant and wild type do not have bubbles of exactly the same length). A T placed at −1 in the top strand (mutant −1T) was not reactive in the paused complex in vivo (Fig. 2A, lane 13) or in vitro and was probably protected by RNA polymerase.

The transcription elongation complex is thought to include 10 to 12 bp of RNA-DNA duplex along the template strand of the DNA (*4, 5*), a structure that should protect the template strand from $KMnO_4$ reactivity. We found low reactivity of pyrimidines at internal positions of both strands relative to the strong reactivity at the bubble margins, and this reactivity did not entirely derive from the paused complex. For example, +5 and +12 of the λ bottom strand and +10 and +11 of the λ top strand in vivo show this intermediate reactivity. To probe this entire segment, we used a set of mutations that introduced new thymines in positions +4 to +11 of the λ bottom strand without changing the properties of the promoter; data for three mutants are shown (Fig. 2B) and the results are summarized (Fig. 3A). The newly introduced T residues at +4, +7, +8, +9, +10, or +11 were similar in reactivity to the T at +5, and much less reactive than T at edge positions or T at +10 and +11 in the top strand in vitro. Note that reactivities vary—for example, +4 is weakly reactive and +9 is moderately reactive—and that reactivities are not independent, as shown by the fact that the change of +4 to T suppresses the reactivity of the wild-type +5T. This analysis confirms the intermediate reactivity of this segment of bottom strand DNA.

In two cases there was more reactivity in the top strand than in the bottom, as would be expected if RNA protected the bottom strand. For the phage 21 paused complex, T in the top strand at $^{+}12$ and $^{+}13$ was more reactive compared to the margin base $^{+}16$, but the T at $^{+}14$ of the bottom strand was weakly reactive compared to the margin bases $^{+}1$ and $^{+}3$ (Fig. 3C); for the λ in vitro complexes, T at +10 and +11 of the top strand reacted strongly, but T at +12 and +13 of the bottom strand did not (Fig. 1B). Reactivity of bottom strand DNA in the phage 82 $^{+}25$ complex was also consistent with protection by RNA: T at +26 in the bottom strand is moderately reactive, whereas T at +18 and +19 is nearly unreactive (Fig. 3B).

The patterns of $KMnO_4$ reactivity of the three phage promoter open complexes in vivo are all similar to each other (Figs. 1 and 3), to the *lac L8:UV5* promoter (*11*), and to structures predicted for other prokaryotic σ70 promoters (*22*). These results suggest that open complexes on the three late promoters melt the DNA helix from −11 to +3 or +4 (phage λ), from −11 or −12 to +4 or +5 (phage 82), and from −12 to no more than +5 (phage 21).

Our results are consistent with the following structure of the transcription elongation complex in vivo. The DNA helix is unwound over 17 to 18 bp, with the catalytic center of RNA polymerase located within 2 bp of the downstream margin of the bubble. The top DNA strand near the downstream margin and the bottom DNA strand near the upstream margin of the bubble are $KMnO_4$ reactive and thus are either solvent exposed or are only loosely bound by RNA polymerase. Nonreactive top strand pyrimidines near the upstream margin and at some internal positions may be protected by proteins of the transcription complex in vivo. However, T at +10 and +11 is strongly reactive in the λ complex in vitro in the presence of RNAP and NusA, suggesting that other proteins—for example, the other Nus proteins that have been associated with transcription elongation complexes (*23*)—may interact in this region.

The disposition of the nascent RNA strand relative to its template DNA remains an important question: Are they together in an RNA-DNA hybrid of 10 to 12 bp, as the conventional model proposes, or do they separate a few bases beyond the growing site, as inferred from the sensitivity of most of the putative hybrid to ribonuclease at high concentration (*7*)? The first view is supported by cross-linking of the 5′ end of the emerging transcript to the template DNA strand until its length exceeds 10 to 12 bases (*4*), and by the shortest homopolymeric template segment (10 bases) that allows transcript slippage against the template (*6*). Our results show that the template DNA strand over a 12-bp segment was not freely exposed to $KMnO_4$, although it was

more reactive than duplex DNA. Possibly the ~12-bp hybrid does exist but has a structure sufficiently different from extended hybrids that its DNA component is more reactive than a strand of DNA in a duplex. Alternatively, the ~12-bp hybrid might be in equilibrium with free RNA and DNA and thus somewhat open to both nuclease and $KMnO_4$ attack.

REFERENCES AND NOTES

1. A. J. Carpousis and J. D. Gralla, *J. Mol. Biol.* **183**, 165 (1985).
2. W. Metzger, P. Shickor, H. Heumann, *EMBO J.* **8**, 2745 (1989); B. Krummel and M. J. Chamberlin, *Biochemistry* **28**, 7829 (1989).
3. H. Gamper and J. Hearst, *Cell* **29**, 81 (1982).
4. M. Hanna and C. F. Meares, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4238 (1983).
5. T. Yager and P. H. von Hippel, in Escherichia coli *and* Salmonella typhimurium *Cellular and Molecular Biology*, F. C. Neidhardt, Ed. (American Society for Microbiology, Washington, DC, 1987), pp. 1241–1275.
6. L. A. Wagner, R. B. Weiss, R. Driscoll, D. S. Dunn, R. F. Gesteland, *Nucleic Acids Res.* **18**, 3529 (1990).
7. G. A. Rice, C. M. Kane, M. J. Chamberlin, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4245 (1991).
8. E. J. Grayhack, X. Yang, L. F. Lau, J. W. Roberts, *Cell* **42**, 259 (1985).
9. X. Yang, J. A. Goliger, J. W. Roberts, *J. Mol. Biol.* **210**, 453 (1989).
10. H. C. Guo, M. Kainz, J. W. Roberts, *J. Bacteriol.* **173**, 1554 (1991).
11. S. Sasse-Dwight and J. D. Gralla, *J. Biol. Chem.* **264**, 8074 (1989).
12. X. Yang, C. M. Hart, E. J. Grayhack, J. W. Roberts, *Genes Dev.* **1**, 217 (1987).
13. X. Yang, thesis, Cornell University (1988).
14. Point mutations are named according to position relative to the RNA start site and to the sequence change in the top (RNA-like) strand of DNA.
15. The paused transcription complex of λ DNA characterized in vitro contains RNAs at both +16 and +17 (W. Yarnell and J. Roberts, unpublished results), although the RNA was originally described as "+16" (*8*).
16. Although adenine adjacent to reactive thymine may be identified by primer extension (*11*), we have noted only pyrimidines in Fig. 3.
17. M. Kainz and J. W. Roberts, unpublished data.
18. In vitro transcription reactions were similar to those described (*8*). Open complexes were formed by incubating reactions containing 5 nM supercoiled plasmid, 25 nM *E. coli* RNAP (purified in this laboratory by G. DiCenzo), and 150 nM NusA protein (purified in this laboratory by J. Goliger) in transcription buffer (20 mM tris-HCl, pH 7.5, 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 10 mM KCl, and 4 mM $MgCl_2$), with or without rifampicin at 10 μg/ml, at 37° for 10 min. Transcription was begun by addition of 200 μM adenosine triphosphate (ATP), cytidine triphosphate (CTP), and guanosine triphosphate (GTP) and 50 μM uridine triphosphate (UTP), and incubation was continued at 37°C for 5 min.
19. J. A. Boroweic, L. Zhang, S. Sasse-Dwight, J. D. Gralla, *J. Mol. Biol.* **196**, 101 (1987).
20. Note that there is more signal in lanes 1 and 2 of Fig. 1B than in other lanes; lanes 4 through 7 better display the strong overall similarity of reactivity in vivo and in vitro.
21. Y. B. Shi, H. Gamper, B. Van Houten, J. E. Hearst, *J. Mol. Biol.* **199**, 277 (1988).
22. K. Kirkegaard, H. Buc, A. Spassky, J. C. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2544 (1983); U. Siebenlist, R. B. Simpson, W. Gilbert, *Cell* **20**, 269 (1980).
23. S. Barak, B. Ghosh, W. Whalen, D. Lazinski, A. Das, *Cell* **50**, 885 (1987); R. J. Horwitz, J. Li, J. Greenblatt, *ibid.* **51**, 631 (1987).
24. $KMnO_4$ reactivity in vivo was analyzed essentially as described (*11*). *Escherichia coli* HB101 transformed with plasmid was grown to mid-log phase in M9 glucose supplemented with 0.5% casamino acids and 100 μg of ampicillin per milliliter. Culture aliquots were treated at 37°C with 10 mM $KMnO_4$ for 2 min or with 200 μg of rifampicin per milliliter for 5 min followed by 10 mM $KMnO_4$ for 2 min. Plasmid DNA was extracted by a boiling lysis method as described (*25*), except that 0.06 mg of lysozyme per milliliter was added, and then purified [as described (*11*)]. Plasmid yields were quantified with a slot blot hybridization assay in which aliquots of purified $KMnO_4$-modified plasmid were transferred to nylon hybridization membrane (GeneScreen Plus, New England Nuclear, Boston, MA) and hybridized with ^{32}P-labeled oligonucleotides used as primers in the primer extension assay. The hybridization signal was quantified by densitometry of the autoradiogram of the membrane or by direct counting of the membrane using a Betascope blot analyzer (Betagen). Equal amounts of plasmid were analyzed by primer extension as described (*19*).
25. G. Del Sol, G. Manfioletti, C. Schneider, *Biotechniques* **7**, 514 (1989).
26. Because only pyrimidines react significantly with $KMnO_4$, some boundaries of melted segments are not determined exactly. For open complexes, uncertainties in upstream and downstream boundaries are, respectively: λ, 1 and 2 bp; 82, 1 and 0 bp; and 21, 1 and 3 bp. For paused elongation complexes: the λ +16–+17 complex, 2 and 1 bp; the 82 +15 complex, 1 and 1 bp; the 82 +25 complex, 0 and 0 bp; and the 21 +18 complex, 1 and 0 bp.

27. We thank W. Yarnell and H.-C. Guo for advice and discussion, G. DiCenzo and J. Goliger for materials, J. Gralla for assistance with the $KMnO_4$ footprinting procedure, and J. Lis, J. Helmann, and members of the laboratory for their critique of the manuscript. Oligonucleotides were synthesized by the Cornell University oligonucleotide synthesis facility. This research was supported by NIH grant GM 21941 and a grant from the Cornell Biotechnology Program.

27 August 1991; accepted 3 December 1991

Mechanism of Transduction by Retroviruses

AMANDA SWAIN AND JOHN M. COFFIN*

Retroviruses can capture cellular sequences and express them as oncogenes. Capture has been proposed to be a consequence of the inefficiency of polyadenylation of the viral genome that allows the packaging of cellular sequences flanking the integrated provirus in virions; after transfer into virions, these sequences could be incorporated into the viral genome by illegitimate recombination during reverse transcription. As a test for this hypothesis, a tissue culture system was developed that mimics the transduction process and allows the analysis and quantitation of capture events in a single step. In this model, transduction of sequences adjacent to a provirus depends on the formation of readthrough transcripts and their transmission in virions and leads to various recombinant structures whose formation is independent of sequence similarity at the crossover site. Thus, all events in the transduction process can be attributed to the action of reverse transcriptase on readthrough transcripts without involving deletions of cellular DNA.

RETROVIRUSES HAVE BEEN STUDIED for their ability to acquire and express cellular sequences as oncogenes (*1, 2*). However, the mechanism by which these viruses incorporate cellular genes into their genome is still undefined. Transduction of oncogenes is thought to proceed from a provirus integrated within or near a proto-oncogene (*3–5*). Joining of cell and viral sequences can proceed by deletions that remove portions of viral and cellular sequences (*4*) or by formation of joint "readthrough" transcripts (*5*). Illegitimate recombination of the packaged hybrid transcripts during reverse transcription will incorporate proto-oncogene sequences into the provirus (Fig. 1).

That readthrough transcripts could be intermediates in transduction is implied by the observation that cleavage and polyadenylation of viral transcripts in the 3′ long terminal repeat (LTR) of avian leukosis virus is inefficient (*6*). A significant portion (about 15%) of viral transcripts from infected cells contain neighboring cell sequences, are packaged efficiently, and can serve as templates for reverse transcription. Indeed, a mutation inactivating the polyadenylation signal of Rous-associated virus (RAV-1) does not greatly affect viral replication. Most such mutant viral genomes have cellular sequences appended to their 3′ ends, but these sequences are usually lost during reverse transcription (*7*).

To test the role of readthrough transcripts in the transduction of cellular sequences, we constructed a model resembling a provirus integrated upstream of a proto-oncogene and introduced it into cells in culture. The SV-*neo* cassette (*8*), containing the selectable *neo* marker, flanked by DNA containing promoter and polyadenylation signals from SV40, was placed downstream of an RAV-1 provirus with a mutation in the polyadenylation signal (*7*) (Fig. 2A). QT35 cells [a transformed quail cell line (*9*)] were transfected with this DNA, and *neo*-expressing

Tufts University School of Medicine, Department of Molecular Biology and Microbiology, 136 Harrison Avenue, Boston, MA 02111.

*To whom correspondence should be addressed.