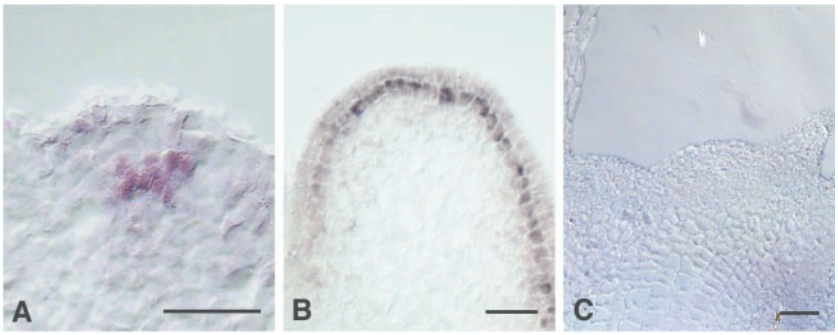

Fig. 4. *CLV3*-dependent regulation of *WUS* mRNA expression. In situ hybridization with *WUS* digoxigenin-labeled antisense probe. (**A**) Longitudinal section through a wild-type shoot meristem. *WUS* is expressed in deep regions of the meristem. (**B**) In the enlarged meristem of a *clv3* mutant, the *WUS* expression domain is expanded. (**C**) Arrested meristem of a *CaMV35S*::*CLV3* plant. *WUS* RNA is not detectable. Bars, 20 μm.

(*17*), the stem cell–promoting pathway may involve additional factors that may overlap in function with *WUS* and be potential targets for regulation by the *CLV* pathway, such as *POLTERGEIST* (*18*).

In wild-type *Arabidopsis* plants, the CLV3 signal is likely to be released from the stem cells at the apex of the meristem and to activate a CLV1/CLV2 receptor complex in underlying cells (*2*). Signaling through the *CLV* pathway limits *WUS* activity by restricting its expression to a narrow domain of cells in deeper layers of the meristem. Constitutive signaling through *CLV3* enhances this negative pathway, causing down-regulation of *WUS* and complete loss of stem cells. When the negative pathway is disrupted in *clv* mutants, the *WUS* expression domain expands laterally and upward, resulting in, or resulting from, an accumulation of stem cells. Activity of the positive pathway promotes the expression of *CLV3* or maintenance of the *CLV3* expression domain. This mutual regulation, involving positive and negative interactions, provides a feedback system for maintaining the delicate balance required for proliferation of stem cells to proceed at the right time and in the right place.

References and Notes

1. T. A. S. Steeves and I. M. Sussex, *Patterns in Plant Development* (Cambridge Univ. Press, New York, 1989); S. E. Clark, *Semin. Cell Dev. Biol.* **7**, 873 (1996); E. M. Meyerowitz, *Cell* **88**, 299 (1997); P. Laufs, O. Grandjean, C. Jonak, K. Kieu, J. Traas, *Plant Cell* **10**, 1375 (1998).
2. J. C. Fletcher, U. Brand, M. P. Running, R. Simon, E. M. Meyerowitz, *Science* **283**, 1911 (1999).
3. J. C. Fletcher and E. M. Meyerowitz, *Curr. Opin. Plant Biol.* **3**, 23 (2000).
4. H. M. O. Leyser and I. J. Furner, *Development* **116**, 397 (1992); S. E. Clark, M. P. Running, E. M. Meyerowitz, *Development* **119**, 397 (1993).
5. S. E. Clark, R. W. Williams, E. M. Meyerowitz, *Cell* **89**, 575 (1997).
6. J. M. Kayes and S. E. Clark, *Development* **125**, 3843 (1998); S. Jeong, A. E. Trotochaud, S. E. Clark, *Plant Cell* **11**, 1925 (1999).
7. S. E. Clark, M. P. Running, E. M. Meyerowitz, *Development* **121**, 2057 (1995).
8. R. W. Williams, J. M. Wilson, E. M. Meyerowitz, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10467 (1997); A. E. Trotochaud, T. Hao, G. Wu, Z. Yang, S. E. Clark, *Plant Cell* **11**, 393 (1999).
9. A *CLV3* full-length cDNA was inserted between the *CaMV35S* promoter and the 3′*NOS* terminator of pRTΩNot I to give pBU4. Clone orientation and fidelity was verified by sequencing. A Pst I restriction fragment of pBU4 was introduced into pGREEN (P. Mullineaux and R. Hellens, John Innes Centre, Norwich, UK) to give pBU6 (*CaMV35S*::*CLV3*). pBU6 was transformed into *Agrobacterium tumefaciens* strain GV3101 and used to transform *Arabidopsis* plants by vacuum infiltration [N. Bechtold, J. Ellis, G. Pelletier, *C. R. Acad. Sci. Paris Sci. Vie* **316**, 1194 (1993)]. Transformation efficiency ranged from 0.5 to 5.0%.
10. Some 2% of all transgenic plants presumably failed to express the transgene at sufficient levels to give a phenotype.
11. In situ hybridizations were performed as described in (*2*).
12. A 5.5-kb Eco RI genomic DNA fragment containing the *CLV3* gene was cloned into pGREEN, giving gE55*CLV3*, and transformed into *Agrobacterium* strain GV3101. Introduction of gE55*CLV3* into *clv3-2* mutants rescued the mutant phenotype. Plants that carried four copies of the *CLV3* gene were obtained by introducing two copies of gE55*CLV3* into the genome of wild-type plants.
13. I. Lee, D. S. Wolfe, O. Nilsson, D. Weigel, *Curr. Biol.* **7**, 95 (1997); J. A. Long and M. K. Barton, *Development* **125**, 3027 (1998).
14. A DNA fragment containing 3.5-kb DNA upstream of the putative transcription start site of the *UFO* gene was used to replace the *CaMV35S* promoter in pRTΩNot I. A *CLV3* cDNA clone was then inserted between the Ω-leader and the *NOS*-terminator giving pBU11. The *UFO::CLV3* gene construct was transferred as a Pst I fragment into a pGREEN T-DNA vector, giving pBU12. Transformation of *Arabidopsis* wild type and *clv3-8* mutants via *Agrobacterium* GV3101 was performed as described above. We analyzed 203 transformants.
15. T. Laux, K. F. Mayer, J. Berger, G. Jürgens, *Development* **122**, 87 (1996); K. F. Mayer *et al.*, *Cell* **95**, 805 (1998).
16. H. Schoof *et al.*, *Cell* **100**, 635 (2000).
17. U. Brand and R. Simon, unpublished data.
18. L. P. Yu, E. J. Simon, A. E. Trotochaud, S. E. Clark, *Development* **127**, 1661 (2000).

19. We thank N. Beinert and J. Wong for technical assistance and members of the Meyerowitz lab and the Simon lab for discussion. Supported by the NIH (J.C.F.), the U.S. Department of Agriculture (J.C.F.), the NSF (E.M.M.), Zeneca Seeds (E.M.M.), and the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich 243 (U.B., M.H., and R.S.).

13 April 2000; accepted 1 June 2000

A Structural Model of Transcription Elongation

Nataliya Korzheva,[1] Arkady Mustaev,[1] Maxim Kozlov,[1] Arun Malhotra,[2] Vadim Nikiforov,[3] Alex Goldfarb,[1] Seth A. Darst[4*]

The path of the nucleic acids through a transcription elongation complex was tracked by mapping cross-links between bacterial RNA polymerase (RNAP) and transcript RNA or template DNA onto the x-ray crystal structure. In the resulting model, the downstream duplex DNA is nestled in a trough formed by the β′ subunit and enclosed on top by the β subunit. In the RNAP channel, the RNA/DNA hybrid extends from the enzyme active site, along a region of the β subunit harboring rifampicin resistance mutations, to the β′ subunit "rudder." The single-stranded RNA is then extruded through another channel formed by the β-subunit flap domain. The model provides insight into the functional properties of the transcription complex.

Transcription is the major control point of gene expression, and RNAP is the central enzyme of transcription. In bacteria, the essential core RNAP (subunit composition $\alpha_2\beta\beta'$) has a molecular mass of around 400 kD and is evolutionarily conserved among all cellular organisms (*1*). The processive ternary elongation complex (TEC) of core RNAP, DNA template, and RNA transcript is extremely stable, yet RNAP can translocate in both directions on the DNA and disassembles rapidly upon encountering specific termination signals. Empirical models of TEC structure and function have attempted to reconcile these properties (*2*–*5*), but were not based on

[1]Public Health Research Institute, 455 First Avenue, New York, NY 10016, USA. [2]Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, FL 33101, USA. [3]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia. [4]Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: darst@rockvax.rockefeller.edu

detailed structural information on RNAP. The x-ray crystal structure of *Thermus aquaticus* (*Taq*) core RNAP (*6*) provides a framework for more detailed modeling. We have used RNA- and DNA-protein cross-linking to define the path of the nucleic acids across the RNAP in the TEC.

For RNA cross-linking, we used two kinds of TECs (*7*). The first type was generated from the T7A1 promoter by using derivatized trinucleotide primers (Fig. 1A). A TEC carrying an 11-nucleotide (nt) RNA transcript [TEC11 (*7*)] was either extended by walking with limited sets of substrates to obtain TECs with RNAs of 12 to 44 nt, or shortened by limited pyrophosphorolysis to TECs with RNAs of 7 to 9 nt (Fig. 1A). The second type of complex was assembled by using preformed "minimal nucleic acid scaffold" of RNA/DNA oligonucleotides (Fig. 1A, bottom) that bound core RNAP to yield functional TECs (*7*). Two kinds of cross-linking groups were used (Fig. 1A): the fluorinated aromatic azide (I), which reacts with any amino acid, and the aldehyde (II), which cross-links exclusively to lysine residues. The β or β′ subunits were the cross-linking

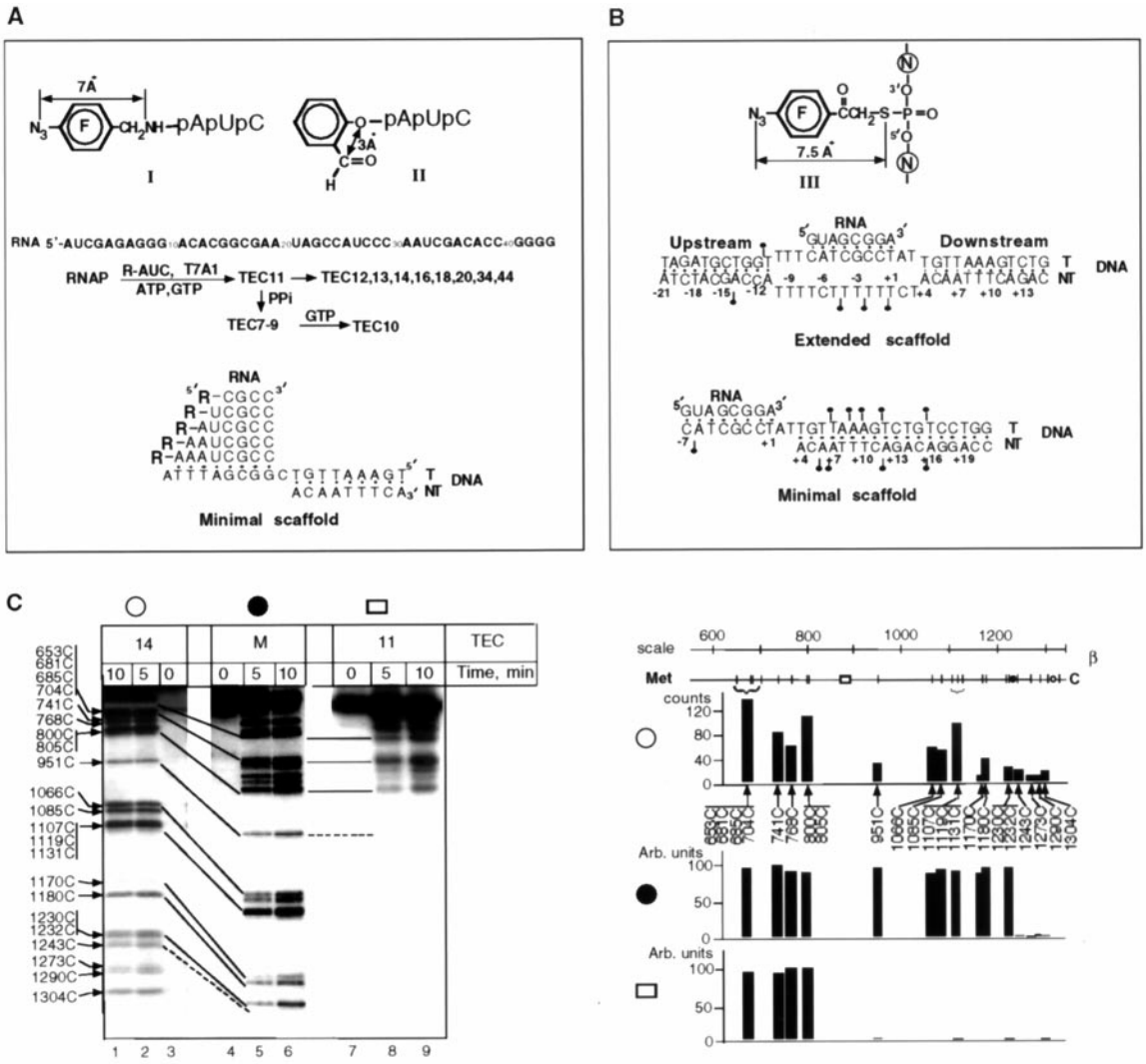

Fig. 1. Systems used for nucleic acid–protein cross-linking in transcription complexes and the principle of cross-link site mapping. (**A**) Systems used for RNA-protein cross-linking. (Top) Reagents used for cross-linking from the RNA 5′-terminus. Photoreactive fluorinated (F) aromatic azide (I) and Lys-specific aldehyde (II) RNA primers used in cross-linking experiments were synthesized as described for reagent IV (*9*), starting from 4-azidotetrafluorobenzylamine (*36*) and 2-hydroxybenzaldehyde. (Middle) The sequence of RNA synthesized from the T7A1 promoter and the route to defined TECs containing the reactive groups. "R" denotes the cross-linking probe attached to the 5′-terminus of the trinucleotide primer (either reagent I or II from the top). (Bottom) The structure of minimal RNA:DNA scaffolds used for RNA-protein cross-linking. "R" denotes the cross-linking probe attached to the 5′-terminus of the RNA oligonucleotides (either reagent I or II from the top). (**B**) Systems used for DNA-protein cross-linking. (Top) Structure of the photoreactive probe, 1-(4-azidotetrafluorophenyl)-ethanone (*36*), attached to internucleotide thiophosphates of the DNA. Circled letters (N) indicate adjacent nucleotide residues in the DNA sequence. (Middle and bottom) Structures of nucleic acid scaffolds used for DNA-protein cross-linking. The positions where reactive groups were introduced are marked by filled circles. (**C**) Principle of cross-link site mapping. (Left) SDS-PAGE and autoradioagraphy of single-hit CNBr cleavage products from cross-linked β subunit. Cleavage products are labeled from the RNA 5′-terminus of TEC14 (lanes 1 to 3) and TEC11 (lanes 7 to 9). The radioactive label must be linked to a residue between the shortest labeled cleavage product and the next CNBr product that is not radioactive and thus not visible on the gel. Lanes 4 to 6 are marker lanes (M), showing a characteristic ladder of CNBr cleavage products for the β subunit labeled at a known site, His^{1237}, near the COOH-terminus (*9*). Continuous lines connect bands corresponding to the same cleavage products used for mapping from each lane, and dashed lines show the first cleavage site that gives nonradioactive product. TEC14 has the most COOH-terminal cross-links and thus has labeled bands extending to 1304C. The position of the radioactive label was localized between the CNBr cleavage sites that gave rise to the shortest labeled cleavage product and the next shortest product, which was not radioactive. (Right) Quantitation of CNBr cleavage products. The top horizontal bar represents the COOH-terminal part of the 1342–amino acid *E. coli* β polypeptide. Met residues (CNBr cleavage sites) are marked on the scale bar directly below. Filled and open circles and the open rectangle indicate the positions of the mapped cross-linking sites shown on the left. (○) Histogram of the intensity of the cleavage products from lane 1. The bars are aligned vertically with the cleavage site that gave rise to the corresponding band, as shown on the scale bar above. (●) Intensity of the cleavage products from lane 6, normalized with respect to lane 1. The highest normalized intensity was set to 100 arbitrary units. The shortest radioactive product resulted from cleavage at 1230C/1232C. Cleavage at 1243C did not give radioactive product, thus localizing the cross-linking site between 1232 and 1243. In this instance, the cross-linking site is known to be 1237 (*9*). (□) Normalized intensity of the cleavage products from lane 9. The loss of radioactivity COOH-terminal of 800C/805C localizes the cross-link between 800C/805C and the next cleavage site at 951C.

targets from all nucleic acid positions (*7*). Nonspecific cross-linking in controls with nonmodified nucleic acids was always less than 2%.

For DNA cross-linking, TECs were assembled from minimal (*3*) and extended (*8*) scaffolds (Fig. 1B). The cross-linking probe (Fig. 1B, top) was attached to DNA through phosphorothioate linkages (*7*) at defined positions (Fig. 1B). Cross-links were to the β or β′ subunits, and strong cross-links were mapped (*7*).

For cross-link mapping (*9*), the subunit carrying the radiolabeled adduct was subjected to single-hit chemical degradation, yielding characteristic NH_2- and COOH-terminal fragmentation ladders on SDS gels. Cross-links were localized by determining the site closest to the NH_2- or COOH-terminus where cleavage resulted in the loss of radioactive bands from the ladder. This is illustrated by examples of RNA cross-link mapping on the β subunit from TEC14 and TEC11 (Fig. 1C). The complete mapping data and methodology are presented as an online supplement (*7*). The mapping results are summarized in Fig. 2, where the register of the incoming nucleotide substrate is denoted "+1." Upstream positions (where RNAP has been) are denoted −1, −2, and so on, while downstream positions (where RNAP is going) are denoted +2, +3, and so on.

The cross-links in Fig. 2 are presented in the context of prior knowledge about the nucleic acids in the TEC. The DNA is melted in a 12- to 14-nt region (the transcription bubble) between +2 to −11 (*10*, *11*). The DNA strands reanneal only one or two bases downstream of the RNA 3′-terminus. Further downstream, both strands of the DNA are completely protected from hydroxyl-radical cleavage, suggesting enclosure within the protein, whereas the upstream duplex DNA is only partially protected (*11*). Within the bubble, the template strand of DNA is completely protected, whereas the upstream part of the nontemplate strand is only partially protected, suggesting exposure of the bases (*12*). The 3′-proximal 8 to 9 nt of the RNA transcript (positions +1 to −8) form a hybrid with the DNA template strand (*4*). RNA folding and protection experiments indicate that the transcript is enclosed from its 3′-end to around −14 (*13*). Finally, RNAP occasionally "backtracks," sliding backwards (in the upstream direction) on the DNA. The bubble translocates with the RNAP, and the transcript reverse-threads through the complex, maintaining the hybrid with the DNA, generating a downstream segment of single-stranded RNA (*14*, *15*), of which about 9 nt are protected from ribonuclease digestion (*15*).

There is an ambiguity of 1-nt position in the interpretation of the cross-linking data because it is not known whether the transcript 3′-terminus occupies the *i* or *i* + 1 site of RNAP in the absence of substrates. We assumed a pretranslocation state, with the 3′-end of the transcript occupying the *i* + 1 site. Thus, the 5′-end of an 11-nt RNA is assumed to be at position −10.

In addition to the data of Fig. 2, the following considerations went into the generation of the TEC model: (i) The duplex DNA outside of the transcription bubble was constrained to B-form. (ii) The RNA/DNA hybrid was constrained to the hybrid crystal structure (*16*). (iii) Analagous to crystal structures of other polymerases (*17*), the 3′-OH of the RNA terminal residue occupying the *i* site (position −1), and the α-phosphate of the incoming nucleotide substrate occupying the *i* + 1 site (position +1), were constrained to chelate the catalytic Mg^{2+} ion. (iv) Changes in the RNAP structure were not allowed during the calculations (*18*), although insights gained from the modeling suggest that the conformation of the RNAP in the TEC is different from that of the crystal structure.

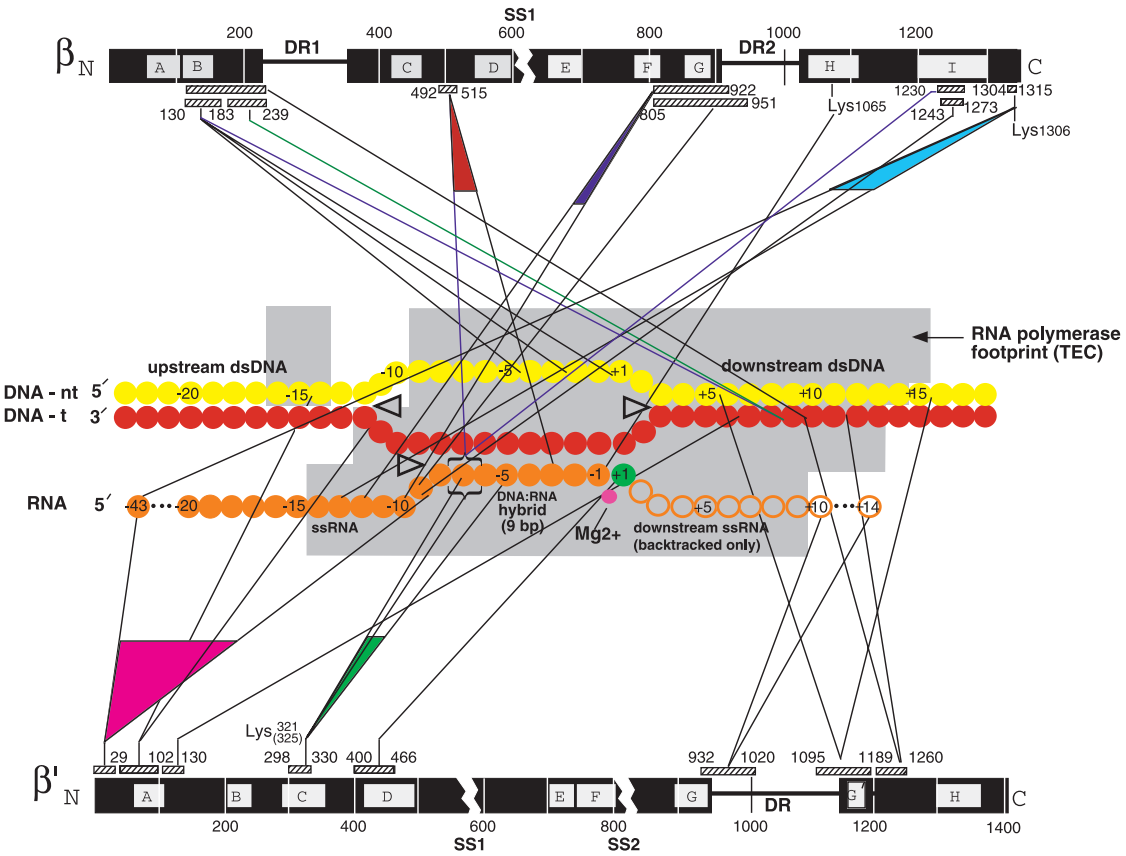

Fig. 2. Nucleic acid scaffold of the transcription complex and its contacts with RNAP subunits. Horizontal bars represent *E. coli* RNAP β (top) and β′ (bottom) subunits. Lettered gray boxes indicate evolutionarily conserved regions (*1*). Dispensable regions [DR (*34*)] and split sites [SS (*37*)] are indicated. Nucleic acid scaffold of the transcription complex is shown in the middle. Filled circles represent the DNA (red, template strand; yellow, nontemplate strand) and RNA (gold) chains. The incoming nucleotide substrate is green. Open circles show the segment of single-strand downstream RNA in the backtracked TEC. The positions of nucleotide residues are numbered relative to the *i* + 1 site of the active center, which is denoted +1. The 12- to 14-base transcription bubble is indicated (*10*, *11*). Results of chemical and enzymatic footprinting of the DNA and RNA are also indicated by gray shading (*11*–*13*). Lines connect contacts identified by cross-links from specific nucleotide positions and protein segments (indicated by cross-hatching) of RNAP subunits (*7*, *30*). Lysine residues that are targets of Lys-specific cross-linking are numbered. Colored triangles show segments of nucleic acid scaffold that cross-link to the same protein regions.

Distances between specific sites in RNAP and the nucleic acids in the model agree well with the cross-linking results [Web figure 9 (*7*)]. One exception is that cross-links from the upstream RNA (positions −9 to −43) to the NH_2-terminus of β′ were observed (Fig. 2), but these sites are up to 40 Å apart in the model. The lack of sequence conservation between *Escherichia coli* and *Taq* in the β′ NH_2-terminal 32 amino acids, as well as the presence of an extra 12 amino acids in *E. coli*, which could easily span the 40 Å distance, may explain this discrepancy.

The positioning of specific nucleic acid elements in the TEC model (Figs. 3 and 4) is discussed below.

At least 9 base pairs (bp) of downstream duplex DNA are required for TEC stability (*5*). This DNA cross-links to regions of β′ that form the side walls of a "trough" that accommodates the downstream DNA [Fig. 4A and Web figure 9A (*7*)]. Additional cross-links from the downstream duplex DNA map to a region of β that forms the roof of the tunnel (Fig. 3D).

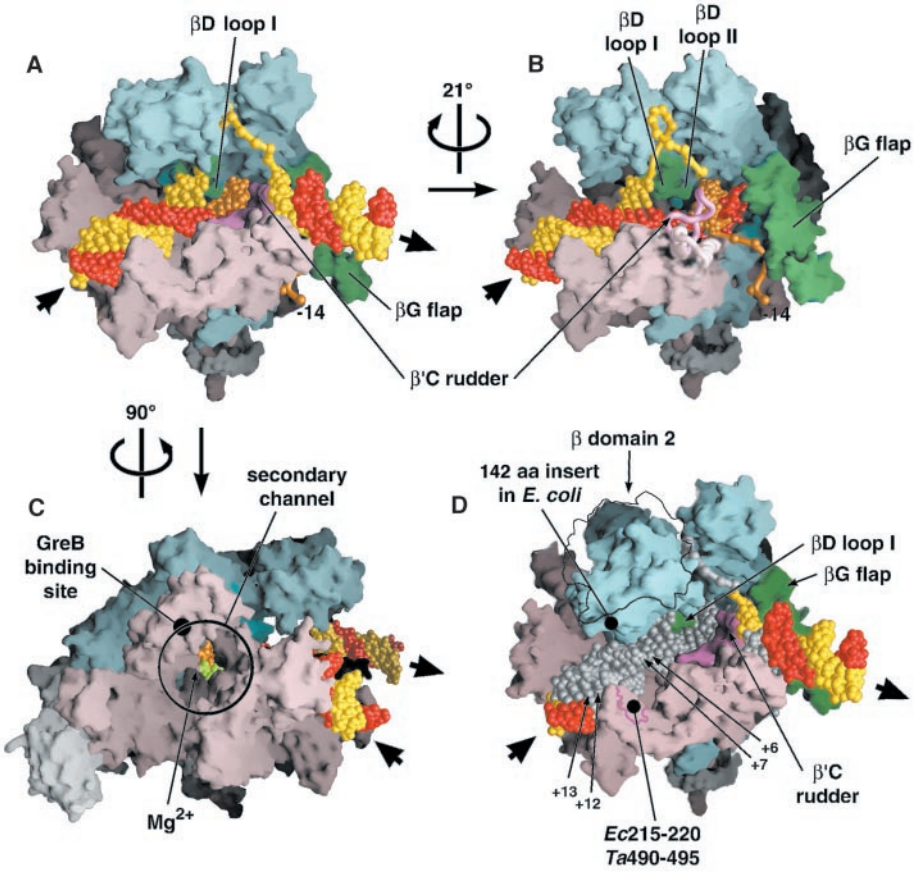

Fig. 3. (A to C) Three views of the TEC model. The model consists of three components: (i) the *Taq* core RNAP crystal structure (*6*), shown as color-coded molecular surfaces (β, cyan; β′, pink; α and ω, white; catalytic Mg^{2+}, magenta sphere); (ii) the DNA template (template strand, red; nontemplate strand, yellow); and (iii) the RNA transcript (gold) plus incoming nucleotide substrate (green). The nucleic acid backbones are shown as worms with CPK phosphate atoms. For base-paired sections of the nucleic acids, the bases are shown as CPK atoms. The directions of the entering downstream DNA and exiting upstream DNA are indicated (large arrows). (**A**) View perpendicular to the main active-site channel, which runs roughly horizontal. Parts of the protein structure discussed in the text are colored (βD loop I and βG flap in green, β′C rudder in rose). (**B**) View roughly along the axis of the exiting upstream duplex DNA. For clarity, parts of the β′ subunit nearest the viewer are shown as a backbone worm without the corresponding molecular surface (the β′C rudder in rose, and its coiled-coil platform in pink), and the upstream duplex DNA has been removed. Parts of the protein structure discussed in the text are colored (βD loops I and II and βG flap in green). (**C**) View down the secondary channel, showing the path for diffusion of the incoming nucleotide substrate (green) into the active site. The catalytic Mg^{2+} ion (magenta sphere) is just visible to the left of the substrate nucleotide. (**D**) Correspondence between the model and RNA and DNA footprinting data. Segments of the RNA and DNA backbone protected from either hydroxyl-radical or enzymatic digestion (*10*–*12*) are shown in gray. The RNAP structure shown is after the modeled conformational change involving rotation of two β segments (β domain 2 and βD loops I and II, see text) toward the DNA. An outline of β domain 2 before the conformational change is shown. The location of the middle of a 27-residue nonconserved segment (β*Ta*208-234) that harbors a 142-residue insertion in *E. coli* (dispensable region I, β*Ec*221-362) is indicated. A six–amino acid segment (outlined in rose), when deleted from *E. coli* RNAP, results in unstable promoter complexes in which the deoxyribonuclease footprint is shortened from the downstream edge (*35*).

In a previous study (*19*), cross-links from the downstream duplex DNA were mapped near the COOH-terminus of β (β*Ec*1230-1273) (*20*). This result is not consistent with our model, because this region of RNAP is far away, near the upstream DNA. These cross-links were likely an artifact of nonspecific photoreaction of the DNA nucleotides rather than of photoreaction with the specific cross-linking probe used in that study.

The RNA/DNA hybrid was placed in the RNAP channel according to the cross-link mapping, resulting in a remarkable complementary fit with the protein structure (Figs. 3 and 4). The hybrid is cradled between a helix comprising conserved region F of β′ (denoted β′F) (*21*) and the catalytic Mg^{2+} at its downstream end, and the β′C rudder upstream (Fig. 4). The transcript within the hybrid from about −3 to −5 passes near a pocket in β that is lined with rifampicin resistance mutations (Fig. 4B) (*22*). Binding of the bulky rifampicin molecule in this pocket would likely block elongation of the transcript beyond 2 to 3 nt, consistent with the mechanism of rifampicin inhibition (*23*).

Cross-links from TEC7-9 (positions −6 to −8) map to the β′C rudder (β′*Ec*298-330), and the Lys-specific group reacts with *Ec*Lys321 and/or *Ec*Lys325, located on the rudder. The 8- to 9-bp length of the RNA/DNA hybrid (*4*), combined with the finding that RNAs longer than 9 nt no longer cross-link to the rudder (Fig. 2), suggests that the rudder separates the exiting RNA from the DNA template strand, which is an active function of RNAP (*3*, *24*). Residues at the downstream edge of the rudder (facing the RNA/DNA hybrid) are highly conserved (*7*). Two basic amino acids, *Ec*Arg322 (*Ta*Arg598) and *Ec*Lys325 (*Ta*Arg601) (the only side chains shown in Fig. 4), are nearly invariant (Arg or Lys in 94% of all β′ homologs) and are positioned such that they may play a role in the proposed rudder function (Fig. 4).

The transcript is enclosed within the TEC from the 3′-end to around position −14 (*13*), but only the first 9 nt of the RNA are enclosed in the main channel of RNAP as part of the RNA/DNA hybrid. Previously, Zhang *et al.* (*6*) suggested that the single-stranded RNA between −9 to −14 extends through the groove underneath the βG flap, and this is supported by cross-links from TEC10-13 (positions −9 to −12) that map to the βG flap [cyan, Web figure 9C (*7*)].

The RNA/DNA hybrid is 8 to 9 bp (*4*), whereas the bubble spans 12 to 14 nt, leaving 3 to 6 nt of single-stranded template DNA upstream of the hybrid, enough to reach the up-

stream side of the rudder to reanneal with the nontemplate DNA. According to the model, the rudder may participate in two important functions, maintaining the upstream edge of the RNA/DNA hybrid by removing the RNA transcript from the DNA template strand (*3*, *24*), while also maintaining the upstream edge of the bubble. Cross-linking data do not precisely position the upstream duplex DNA [it cross-links to β′*Ec*29-102, a region disordered in the crystal structure (*6*)], but this DNA may be flexible in the TEC because it is only partially protected from hydroxyl-radical cleavage (*11*). In the model, the upstream duplex DNA lies on a platform, formed by an α-helical coiled-coil domain of β′. This is consistent with hydroxyl-radical footprinting results (Fig. 3D) and gives a distance that can be bridged by the single-strand nontemplate DNA between the upstream and downstream duplex DNA. The incoming downstream duplex DNA and the exiting upstream duplex DNA form a bend angle (*25*) of about 90°, arising mainly from a kink between the downstream duplex and the RNA/DNA hybrid (Fig. 4A). A bend angle of 92° was experimentally determined by atomic force microscopy of active TECs (*25*).

The nontemplate strand DNA within the bubble cross-links exclusively to β*Ec*130-183. This region forms part of a groove along the inner roof of the RNAP main channel between two structural domains of β [domains 2 and 3, as defined previously (*6*)] that appear situated to accommodate the single-strand DNA from about +2 to −6/−7. The single-strand nontemplate DNA from −8 to −11 is solvent exposed, consistent with hydroxyl-radical and nuclease digestion studies (*11*, *12*) (Fig. 3D).

The model positions the two principal elements of the TEC, the downstream duplex DNA and the RNA/DNA hybrid, within RNAP with high confidence. It is consistent with nearly all available data, including cross-linking and footprinting data. Because the duplex DNA and RNA/DNA hybrid were restricted to known structures (*18*) and protein conformational changes were not allowed, the model may not accurately describe protein–nucleic acid interactions between specific bases and side chains, nor the interactions between substrate nucleotides, the catalytic Mg^{2+} ion, and protein side chains near the catalytic center. The primary goal of this modeling was to indicate the paths of RNA and DNA through the RNAP, to provide insight into the functional properties of the TEC, and to guide further experiments. In this spirit, some possible functional consequences of the model are discussed below.

The "canonical" bacterial intrinsic termination signal in DNA comprises a GC-rich dyad symmetry element followed by an oligo(T) sequence (*26*), and formation of a hairpin in the RNA is essential for termination (*27*). The region of the transcript that forms the stem-loop of the hairpin (upstream of −7) is the same region that interacts with the rudder and then passes under the βG flap. The flap is likely to be flexible (*6*), and formation of the bulky hairpin might cause it to open. In addition, invasion of the stem-loop into the upstream region of the RNA/DNA hybrid might disrupt the nucleic acid interactions with the rudder, triggering collapse of the transcription bubble from the upstream edge (*3*), either by partial melting of the hybrid or by forward translocation of the enzyme (*28*), resulting in disintegration of the TEC.

Zhang *et al.* (*6*) proposed that the secondary channel (Figs. 3C and 4A) provides access to the active center for the nucleoside triphosphate substrates, which cannot enter through the main channel, which is blocked by the nucleic acid framework. The structural basis for this hypothesis is illustrated in Fig. 3C, showing a clear path for the incoming nucleotide (green) through the secondary channel.

Because the secondary channel can accommodate only one diffusing nucleotide at a time, there is only a 25% chance that an incoming nucleotide would be correct. Thus, on average, 3 nt would be required to diffuse in and out for each correct incorporation. This may explain the extremely slow elongation rate of the cellular RNAPs. The elongation rate for the prokaryotic RNAPs is 50 to 100 nt/s, while the eukaryotic enzymes are even slower, at 5 to 20 nt/s. This is contrasted with 300 to 1000 nt/s elongation rates for the single-subunit nucleotide polymerases (*29*).

Another role for the secondary channel may be to accommodate the 3′-proximal single-stranded RNA fragment generated by RNAP backtracking (*4*, *14*, *15*). Cross-linking data (*30*) support this notion (Fig. 2). In a backtracked complex, up to 9 nt of the 3′-proximal RNA fragment are protected from ribonuclease digestion (*15*), consistent with the ~45 Å length of the secondary channel. The backtracked TEC is an obligate intermediate in GreA/GreB-induced transcript cleavage (*15*), suggesting that the Gre factors interact with the extruded 3′-segment. Consistent with this, the approximate location of the GreB binding site on core RNAP (*31*) is next to the entrance of the secondary channel (Fig. 3C).

A major constraint of the modeling was that conformational changes of the RNAP were not allowed. Nevertheless, the model suggests conformational changes of the protein that may occur between the crystal structure and the TEC. One of the strongest distance constraints applied during modeling was that the 3′-OH of the *i*-site nucleotide

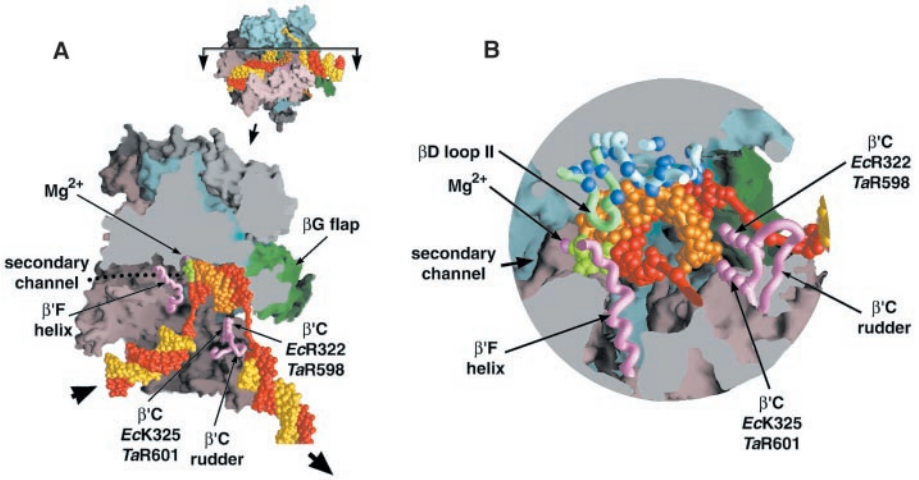

Fig. 4. Additional views of the TEC model. The *Taq* core RNAP crystal structure (*6*) is shown as molecular surfaces, color-coded as in Fig. 3. The nucleic acids are shown as in Fig. 3 (CPK for base-paired regions, backbone worm with CPK phosphates for single-strand regions). (**A**) Cross section of the TEC model illustrating the bending angle of the entering downstream duplex DNA and exiting upstream duplex DNA. The small diagram (upper right) illustrates how the cross section was generated. The view is generated from the view of Fig. 3A by slicing the molecule horizontally in half, removing the top half, then viewing the cross section from the top. Surfaces of the protein structure exposed by slicing through are colored gray. Two structural components, the β′C rudder and the β′F helix, are shown as fuchsia backbone worms without the corresponding surfaces. The two highly conserved basic residues on the rudder that face the RNA transcript are shown as fuchsia CPK atoms and labeled. The path of the secondary channel from the left of the molecule to the *i* + 1 site harboring the incoming nucleotide substrate (colored light green) is indicated by the black dotted line. (**B**) Magnified view of the region around the RNA/DNA hybrid, with obscuring parts of the protein and DNA sliced away. Protein surfaces exposed by slicing are colored gray. Parts of the β subunit harboring rifampicin resistance (Rif^r) substitutions are shown as cyan backbone worms (except βD loop II, which is colored green), which extend over the top of the RNA. The α-carbons of Rif^r substitutions are shown as blue spheres.

(position −1) should be chelated (within about 2 Å) with the catalytic Mg^{2+} ion, by analogy with crystal structures of single-subunit polymerases captured with bound substrates (*17*). However, satisfaction of this constraint results in steric clashes of the template strand DNA with the β′F helix. Because major rearrangements of the template DNA in this region are not likely, owing to base pairing with RNA, we envisage tilting or rolling of the β′F helix away from the template DNA, analagous to the conformational change in the "O" helix of the single-subunit polymerases needed to achieve catalytic activity (*17*).

While cross-links from specific positions of the downstream duplex DNA mapped to both β and β′ (Fig. 2), the β fragments had large exposed surfaces and were not useful as distance constraints in the modeling calculations (*18*). Application of the β′ distance constraints resulted in a close fit between the β′ trough and the DNA (Figs. 3 and 4) but pulled the DNA away from the β subunit, opening a gap between the roof of the main channel (the β subunit) and the nucleic acids. In the real TEC, this gap is likely to be closed owing to flexibility of the RNAP "jaws." Indeed, a hingelike rotation of two closely positioned segments of β [structural domain 2 (*6*) and the linker connecting structural domains 3 and 4, comprising mainly of βD] about 20° toward the downstream DNA results in a close, complementary fit between the DNA, the RNA/DNA hybrid, and the roof of the main RNAP channel (the "before" and "after" positions of these domains are indicated in Fig. 3D). Flexing of the RNAP jaws is also indicated from comparisons of the x-ray crystal structure of *Taq* core RNAP (*6*) with the structure of *E. coli* core RNAP determined by cryoelectron microscopy and image processing of helical crystals (*32*). As a result of jaw closing, two loops extending from β structural domain 2, centered at residues *Ec*163 (*Ta*154) and *Ec*191 (*Ta*179), fit into the major groove of the downstream duplex DNA at around +6/+7 (Fig. 3D). Around βD, a loop centered at residue *Ec*543 (*Ta*423) fits into a pocket formed at the downstream edge of the transcription bubble (between +1 and +3) and may play a role in its maintenance ["βD loop I" in Fig. 3 and Web figure 9C (*7*)]. A second loop, centered at residue *Ec*568 (*Ta*448), closes over the base pair between the incoming nucleotide substrate and the DNA template base at position +1 ("βD loop II" in Fig. 4B). Loop II is very highly conserved and includes a site cleaved by hydroxyl radicals generated at the active-site metal (*33*). Loop II is positioned immediately next to the β′F helix (Fig. 4B), suggesting that the two proposed conformational changes, movement of the β′F helix and closure of the β subunit segments down onto the nucleic acids, may be concerted.

In the open conformation of the jaws, some downstream DNA that is protected from hydroxyl-radical cleavage (Fig. 2) appears exposed in the model. With the proposed conformational change, the model corresponds more closely to the protection data (Fig. 3D). Although much of the downstream duplex DNA and RNA/DNA hybrid are enclosed inside the main RNAP channel, there still appears to be some exposure of protected regions. There are two likely explanations. First, there may be other conformational changes that result in further closure of the RNAP jaws to completely enclose the DNA/RNA hybrid and much of the downstream duplex DNA. Second, protection of nucleic acids was measured with *E. coli* RNAP, which harbors two large insertions of nonconserved domains in β compared with *Taq* (*34*). Within β structural domain 2, a 27-residue nonconserved segment of *Taq* (*Ta*208-234) is replaced by a 142-residue segment in *E. coli* (*Ec*221-362) (Fig. 3D). The position of this extra domain with respect to the DNA (Fig. 3D) may explain why *E. coli* RNAP protects the downstream DNA to about +13 to +15, whereas the *Taq* enzyme appears to cover the top face of the DNA to only about +8 or +9. The difference in the extent of the downstream footprints would reveal itself mainly on the nontemplate DNA strand.

Finally, a deletion mutant of *E. coli* RNAP [β′*Ec*Δ(215-220) (KKLTKR) or *Ta*Δ(490-495) (AKARKR)] forms unstable promoter complexes (*35*). These residues form part of a hairpinlike structure that inserts into the major groove of the downstream duplex DNA at about +12/+13 (Fig. 3D). The deletion would likely disrupt this structural feature, explaining the loss of complex stability and downstream footprint.

In conclusion, the model of the TEC presented here provides a structural basis for a wide range of genetic, biochemical, and biophysical experiments investigating the structure-function relation of RNAP. Yet the model is limited to core RNAP, which requires additional protein factors for initiation. The next challenge in the transcription field is to determine the structure of a promoter initiation complex and to understand its transition to the highly stable and processive TEC.

References and Notes

1. D. Sweetser, M. Nonet, R. A. Young, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1192 (1987); R. S. Jokerst, J. R. Weeks, W. A. Zehring, A. L. Greenleaf, *Mol. Gen. Genet.* **215**, 266 (1989); J. Archambault and J. D. Friesen, *Microbiol. Rev.* **57**, 703 (1993).
2. T. D. Yager and P. von Hippel, in Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology*, F. C. Neidhardt, Ed. (American Society for Microbiology, Washington, DC, 1987), p. 1241; M. J. Chamberlin, *Harvey Lect.* **88**, 1 (1993); J. Gelles and R. Landick, *Cell* **93**, 13 (1998); R. A. Mooney, I. Artsimovitch, R. Landick, *J. Bacteriol.* **180**, 3265 (1998).
3. N. Korzheva, A. Mustaev, E. Nudler, V. Nikiforov, A. Goldfarb, *Cold Spring Harbor Symp. Quant. Biol.* **63**, 337 (1998).
4. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, *Cell* **89**, 33 (1997).
5. E. Nudler, E. Avetissova, V. Markovtsov, A. Goldfarb, *Science* **273**, 211 (1996).
6. G. Zhang *et al.*, *Cell* **98**, 811 (1999).
7. See supplementary material at *Science* Online at www.sciencemag.org/feature/data/1051897.shl..
8. S. S. Daube and P. H. von Hippel, *Science* **258**, 1320 (1992).
9. M. A. Grachev *et al.*, *Eur. J. Biochem.* **180**, 577 (1989).
10. U. Siebenlist, R. B. Simpson, W. Gilbert, *Cell* **20**, 269 (1980); K. Kirkegaard, H. Buc, A. Spassky, J. C. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2544 (1983); J. Mecsas, D. W. Cowing, C. A. Gross, *J. Mol. Biol.* **220**, 585 (1991);
11. E. Zaychikov, L. Denissova, H. Heumann, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1739 (1995).
12. D. Wang and R. Landick, *J. Biol. Chem.* **272**, 5989 (1997).
13. J. A. Monforte, J. D. Kahn, J. E. Hearst, *Biochemistry* **29**, 7882 (1990); N. Komissarova and M. Kashlev, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14699 (1998).
14. T. C. Reeder and D. K. Hawley, *Cell* **87**, 767 (1996); N. Komissarova and M. Kashlev, *J. Biol. Chem.* **272**, 15329 (1997).
15. N. Komissarova and M. Kashlev, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1755 (1997).
16. PDB ID 1FIX [N. C. Horton and B. C. Finzel, *J. Mol. Biol.* **264**, 521 (1996)].
17. S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **391**, 251 (1998); J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **391**, 304 (1998); G. M. T. Cheetham, D. Jeruzalmi, T. A. Steitz, *Nature* **399**, 80 (1999).
18. To generate the TEC model, a 15-bp double-stranded, B-DNA fragment (corresponding to the downstream duplex DNA from +3 to +17) and a 9-bp RNA/DNA hybrid (*16*) were positioned manually into the Taq core RNAP crystal structure (*6*) [Protein Data Bank (PDB) ID 1DDQ] in a manner consistent with the cross-linking data (Fig. 2) by using the program O [T. A. Jones, J.-Y. Zou, S. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991)]. Selected cross-links were specified as nuclear Overhauser enhancement–type distance constraints (with a 7 Å distance between the modified backbone phosphate and the appropriate protein atoms for the nonspecific cross-linking reagent, and a 3 Å length for the Lys-specific reagent), and a rigid-body energy minimization was performed to minimize steric clashes and maximize the correspondence with the distance constraints by using CNS [P. D. Adams, N. S. Pannu, R. J. Read, A. T. Brunger, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5018 (1997)]. Only cross-links mapped to single amino acids or to protein fragments with relatively small exposed surfaces were used for the distance constraints. After this initial refinement, the additional segments of the nucleic acid framework were modeled manually. A second energy minimization was then performed in which the downstream duplex DNA and RNA/DNA hybrid were fixed in order to regularize the bond geometry and minimize steric clashes. Some additional distance constraints were introduced for the upstream single-strand RNA at this stage, in accordance with the cross-linking data.
19. E. Nudler, I. Gusarov, E. Avetissova, M. Kozlov, A. Goldfarb, *Science* **281**, 424 (1998).
20. The label *Ec* preceding amino acid numbers indicates *E. coli* numbering, *Ta* indicates *Taq* numbering. If no indication is given, *E. coli* numbering is assumed.
21. The evolutionarily conserved segments of the large β′ and β subunits (regions A to H and A to I, respectively) have been defined previously (*1*, *6*).
22. D. J. Jin and C. A. Gross, *J. Mol. Biol.* **202**, 45 (1988); K. Severinov, M. Soushko, A. Goldfarb, V. Nikiforov, *Mol. Gen. Genet.* **244**, 120 (1994).
23. W. R. McClure and C. L. Cech, *J. Biol. Chem.* **253**, 8949 (1978).
24. J. P. Richardson, *J. Mol. Biol.* **98**, 565 (1975).
25. W. A. Rees, R. W. Keller, J. P. Vesenka, G. Yang, C. Bustamante, *Science* **260**, 1646 (1993).
26. Y. d'Aubenton Carafa, I. Brody, C. Thermes, *J. Mol. Biol.* **216**, 835 (1990).
27. S. Lynn, L. Kasper, J. Gardner, *J. Biol. Chem.* **263**, 472 (1988); S.-W. Cheng *et al.*, *Science* **254**, 1205 (1992); K. Wilson and P. von Hippel, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8793 (1995).

28. I. Gusarov and E. Nudler, *Mol. Cell* **3**, 495 (1999); W. S. Yarnell and J. W. Roberts, *Science* **284**, 611 (1999).
29. S. M. Uptain, C. M. Kane, M. J. Chamberlin, *Annu. Rev. Biochem.* **66**, 117 (1997).
30. S. Borukhov, J. Lee, A. Goldfarb, *J. Biol. Chem.* **266**, 23932 (1991); V. Markovtsov, A. Mustaev, A. Goldfarb, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3221 (1996); V. Epshtein *et al*., in preparation.
31. A. Polyakov *et al*., *J. Mol. Biol.* **281**, 262 (1998).
32. S. A. Darst, A. Polyakov, C. Richter, G. Zhang, *Cold Spring Harbor Symp. Quant. Biol.* **63**, 269 (1998); S. A. Darst, N. Opalka, A. Polyakov, C. Richter, G. Zhang, in preparation.
33. E. Zaychikov *et al.*, *Science* **273**, 107 (1996); A. Mustaev *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6641 (1997).
34. S. Borukhov *et al*., *J. Biol. Chem.* **266**, 23921 (1991); K. Severinov *et al*., *J. Biol. Chem.* **269**, 14254 (1994).
35. M. S. Bartlett, T. Gaal, W. Ross, R. L. Gourse, *J. Mol. Biol.* **279**, 331 (1998).
36. J. F. W. Keana and S. X. Cai, *J. Org. Chem.* **55**, 3640 (1990).
37. K. Severinov *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4591 (1995); K. Severinov *et al*., *J. Biol. Chem.* **271**, 27969 (1996).
38. Figures 3 and 4 were made with the program GRASP [A. Nicholls, K. A. Sharp, B. Honig, *Prot. Struct. Funct. Genet.* **11**, 281 (1991)]. We thank K. Severinov for the strain overproducing the Xho 19 β subunit mutant and M. Chlenov for participation in development of the DNA-protein cross-linking methods. Supported in part by NIH grants GM30717 and GM49242 (to A.G.) and NIH grant GM53759 (to S.A.D.).

3 May 2000; accepted 5 June 2000

Soluble Adenylyl Cyclase as an Evolutionarily Conserved Bicarbonate Sensor

Yanqiu Chen,* Martin J. Cann,*† Tatiana N. Litvin, Vadim Iourgenko, Meeghan L. Sinclair, Lonny R. Levin, Jochen Buck

Spermatozoa undergo a poorly understood activation process induced by bicarbonate and mediated by cyclic adenosine 3′,5′-monophosphate (cAMP). It has been assumed that bicarbonate mediates its effects through changes in intracellular pH or membrane potential; however, we demonstrate here that bicarbonate directly stimulates mammalian soluble adenylyl cyclase (sAC) activity in vivo and in vitro in a pH-independent manner. sAC is most similar to adenylyl cyclases from cyanobacteria, and bicarbonate regulation of cyclase activity is conserved in these early forms of life. sAC is also expressed in other bicarbonate-responsive tissues, which suggests that bicarbonate regulation of cAMP signaling plays a fundamental role in many biological systems.

Ejaculated spermatozoa are not competent to fertilize an egg. They must first undergo a number of bicarbonate-induced processes, including the induction of hyperactivated motility, capacitation, and the acrosome reaction (*1*–*3*). Each of these processes is known to be cAMP-dependent, but they do not appear to involve the widely studied, hormone-responsive, transmembrane adenylyl cyclases (tmACs). We recently described the purification and cloning of a form of mammalian adenylyl cyclase, the sAC, that is structurally, molecularly, and biochemically distinct from the heterotrimeric guanosine triphosphate–binding protein (G protein)–regulated tmACs (*4*). The sAC cDNA encodes a 187-kD protein that is proteolytically processed to the mature 48-kD isoform purified from rat testes (*4*). Regulators of sAC activity have not yet been identified; tmAC modulators such as G proteins and forskolin do not affect sAC activity (*4*–*6*). Because sAC message is most abundantly expressed in male germ cells (*4*, *7*) and its activity is distinct from that of tmACs, we tested whether sAC mediated the bicarbonate-induced cAMP increase in sperm.

Department of Pharmacology, Joan and Sanford I. Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed.

Western blotting confirmed sAC's presence in bicarbonate-responsive spermatozoa. Anti-sAC antisera (*8*) detected the 48-kD mature form (*4*) and also detected higher molecular weight precursors in both rat testis and mouse epididymal sperm (Fig. 1B). Native sAC was also detected in other tissues known to regulate bicarbonate concentrations and reported to contain bicarbonate-stimulated AC activity (*9*), such as the kidney and the choroid plexus (Fig. 1B). These antisera specifically immunoprecipitated a bicarbonate-stimulated AC activity from the cytosol of rat testis (Fig. 1C) (*10*). The activity was not forskolin-responsive and is therefore unlikely to be caused by cross-reacting tmACs in the immunoprecipitate. Contrary to previous reports in which its in vitro activity required nonphysiologically relevant concentrations of Mn^{2+}–adenosine triphosphate (ATP) (*4*, *11*), sAC activity in these immunoprecipitates was measured in the presence of the more physiologically relevant substrate Mg^{2+}-ATP. These data suggest that sAC is responsible for bicarbonate-stimulated cAMP accumulation in the testis and sperm and that bicarbonate may be acting directly on sAC enzymatic activity.

The effect of bicarbonate on sAC activity was tested in a stable HEK293 cell line expressing the full-length (sAC_{fl}) cDNA (HEK293/sAC_{fl}). Addition of $NaHCO_3$ to the extracellular medium stimulated cAMP accumulation in HEK293/sAC_{fl} cells but not in vector-transfected HEK293 cells (HEK293/V) (Fig. 2A) (*12*). The bicarbonate concentrations used in this experiment mimic the increase observed in sperm (from ≤5 mM in caudal epididymal sperm before ejaculation to ≥25 mM after ejaculation) (*13*). Bicarbonate increased cAMP production within minutes of its addition (*14*), concomitant with an observed elevation of intracellular pH indicating that bicarbonate had entered the cell (*15*). These data demonstrate that sAC can be activated by bicarbonate in a cellular context in the absence of any additional testis- or sperm-specific factors.

To delineate the regions of sAC that mediate bicarbonate activation, we constructed an additional stable cell line (HEK293/sAC_t) expressing a catalytically active NH_2-terminal truncation consisting almost exclusively of the two catalytic domains (sAC_t), which approximates the native 48-kD species (*4*). Bicarbonate also stimulated cAMP accumulation in HEK293/sAC_t cells (*14*), revealing that bicarbonate stimulation of sAC activity does not require the large COOH-terminal domain. Bicarbonate also activated heterologously expressed sAC in vitro. AC activity was stimulated in cellular lysates from HEK293/sAC_{fl} and HEK293/sAC_t cells (Fig. 2B), which is consistent with the idea that the enzyme is directly modulated by bicarbonate ions.

To demonstrate that bicarbonate acts directly on sAC and to exclude the possibility of accessory factors mediating activation in preparations from testis (Fig. 1C) and stable cell lines (Fig. 2), we purified recombinant sAC_t protein (*16*). Purified enzyme was stimulated more than sevenfold (Fig. 3A) with a median effective concentration (EC_{50}) (25.4 ± 7.6 mM) within the physiologically relevant bicarbonate concentration in mammalian serum (22 to 26 mM) (*17*, *18*). Presumably, this direct activation of sAC accounts for the observed intracellular increase in cAMP generation in sAC-expressing cell lines (Fig. 2A).

Bicarbonate stimulation was not due to altered pH, because both Mg^{2+}-ATP alone and bicarbonate-stimulated sAC activities were completely insensitive to pH changes