PERSPECTIVES: STRUCTURAL BIOLOGY

A Marvellous Machine for Making Messages

Aaron Klug

The multisubunit enzyme RNA polymerase II (also called RNA polymerase b, Rpb, or Pol II) is the central enzyme of gene expression in eukaryotes. It reads the sequence of one strand of the DNA double helix (the template) and in so doing synthesizes messenger RNA (mRNA), which is then translated into protein. Two papers by Roger Kornberg's group (*1*, *2*) on pages 1863 and 1876 of this issue go a long way toward helping us to understand the structural basis of transcription at the atomic level.

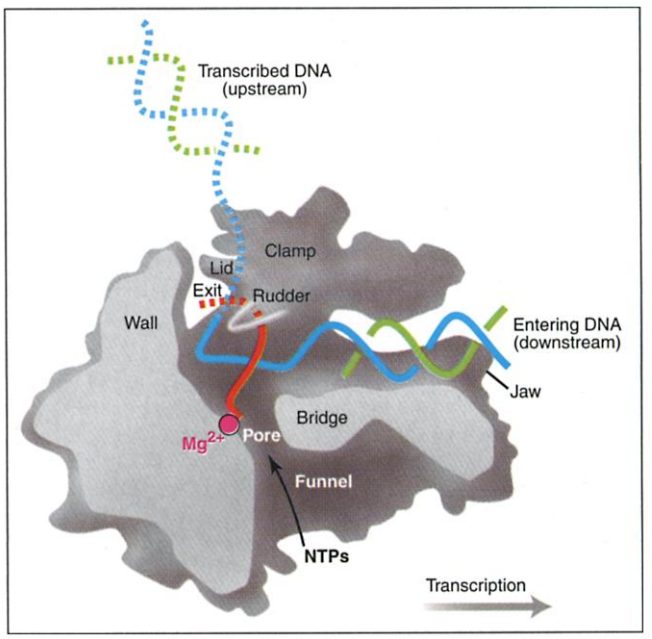
The first paper (*1*) describes two crystal forms of the 0.4-megadalton, 10-subunit RNA polymerase II of yeast at high resolution (3.1 and 2.8 angstroms, respectively). The crystal structures reveal the enzyme in two states: an open form and a partly closed form. These forms differ mainly in the position of a massive 50-kilodalton region of the enzyme called the clamp, which is thought to close over the DNA as it enters the enzyme. A set of protein loops at the base of the clamp appear to act as pivots for DNA movement.

The second paper (*2*) reports the structure of an actively transcribing complex of the enzyme with DNA at 3.3 angstrom resolution (determined by molecular replacement from the 2.8 angstrom enzyme structure). The electron density map shows, with conspicuous clarity, the synthesized RNA, the 9–base pair DNA-RNA hybrid in the transcription bubble, and even the three bases of the single-stranded DNA template known from biochemical studies to be unwound before it enters the hybrid duplex (see the figure, this page). The active site where the ester bond is broken in the substrate nucleoside triphosphates (NTPs) is marked by a metal ion at the base of the hybrid. Somewhat less clear, because it is partially disordered, is the DNA double helix situated in the cleft formed between the two largest enzyme subunits, Rpb1 and Rpb2.

As anticipated from electron microscopy (*3*) and cross-linking experiments, the DNA-RNA hybrid is bound in the cleft between the two large subunits, but makes a right-angle bend at the active center. The moving hybrid, as it were, comes to a "wall" (emanating from Rbp2) and then runs up against it, so that the nucleic acids exit the enzyme from its top and back. The synthesized RNA strand spun off by the enzyme runs through a groove and exits under the "lid."

Most remarkable of all, and unexpected at this stage of the analysis, the authors are able to propose a convincing model for the translocation step that must follow the addition of the last nucleotide to the elongating RNA chain (see the figure, next page). They propose that translocation is accomplished with the help of a protein helix (the "bridge helix") that spans the cleft between Rpb1 and Rpb2. This helix is also present in the structure of the bacterial RNA polymerase (which is homologous to the eukaryotic polymerase II), whose structure has been determined at somewhat lower resolution by Darst and colleagues (*4*). Amino acid side chains from the bridge helix (threonine and alanine) make hydrophobic contacts with the base of the coding nucleotide in the template strand at the active site (see the figure, next page). This region is straight in the yeast polymerase II structure, but bent in the bacterial version by about 3 angstroms along the direction of the template strand. Kornberg and colleagues (*2*) therefore suggest that the bridge helix acts as a ratchet, allowing the release of the DNA and RNA strands for translocation but maintaining its grip on the growing end of the hybrid, thus enabling the next step in the elongation cycle to take place.

The two Kornberg papers are thus a far-reaching stride toward comprehending the enzymatic mechanism underlying transcription. The first paper (*1*) sets out, along the amino acid sequences of Rpb1 and Rpb2, structural elements to which the authors have assigned names that explain their functions: wall, clamp, rudder, zipper. These structural elements do not directly correspond



Side (cutaway) view of the RNA polymerase II transcribing complex, showing the paths of the nucleic acids and the locations of some functional elements of the enzyme. Cut surfaces of the protein, in the front, are lightly shaded and the remainder, at the back, are darkly shaded. By convention, the polymerase is moving on the DNA from left to right (direction of large arrow, bottom). This view exposes the DNA duplex entering the enzyme: The template strand coding for RNA is in blue, the nontemplate strand is in green, and the RNA in the DNA-RNA hybrid in the active center region is in red. DNA entering the enzyme is gripped by protein "jaws" (the upper jaw is not seen in the cutaway view). The 3′ (growing) end of the RNA is located adjacent to an active site Mg^{2+} ion (pink sphere). A "wall" of protein blocks the straight passage of nucleic acids through the enzyme, as a result of which the axis of the DNA-RNA makes almost a right angle with the axis of the entering DNA. The bend exposes the end of the DNA-RNA hybrid for addition of substrate nucleoside triphosphates (NTPs). The NTPs may enter through a funnel-shaped opening on the underside of the enzyme and gain access to the active center through a pore. The 5′ end of the RNA abuts a loop of protein (the rudder), which prevents extension of the DNA-RNA hybrid beyond 9 base pairs, separating DNA from RNA. The exit path of the RNA passes beneath the rudder and beneath another loop of protein (the lid). Exiting RNA, DNA, and also the nontemplate strand of the DNA, are not seen in the electron density map (dashed red, blue, and green lines). The rudder and lid emanate from a massive clamp that swings over the active center region (from back to front in the view shown), restraining nucleic acids and likely contributing to the high processivity of transcription. [Adapted from (*5*)]

CREDIT: ROGER KORNBERG

The author is at the MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK.

to protein domains, such as one often finds in proteins involved in simpler biochemical processes, because some of these elements are small loops only about 20 amino acids long and may not fold independently. However, the naming has heuristic value not only in preparing researchers to confront the intricacies of transcription, which has no counterpart in the everyday world from which the terms derive, but, above all, in suggesting informative mutagenesis sites.

In view of its far-reaching impact on our understanding of transcription, it is instructive to review the advances over the last 10 years that have led to this outstanding achievement. Last year saw publication of a complete 3 angstrom backbone model of the enzyme (*5*), revealing many features suggestive of the enzyme's mode of action. The year before that heralded the announcement of a 5 angstrom model (*6*) and also, through three-dimensional (3D) electron microscope image reconstruction of 2D crystals, a transcribing complex (*3*) in which the nucleic acids were shown to occupy the large 25 angstrom cleft in the molecule. Indeed, electron microscopy has been crucial for the eventual determination of the x-ray crystal structure of RNA polymerase II. The variable stoichiometry of two small subunits (Rpb4 and Rpb7) proved to be an impediment to crystallization. This was discovered by correlating the size and degree of order of the 2D crystals obtained by careful monitoring with electron microscopy. These 2D crystals were formed on lipid bilayers in a process developed previously by Kornberg (*7*). Epitaxial growth of further layers on these crystals demonstrated a way of growing 3D crystals not only of polymerase II itself, but also of the related bacterial polymerase, later solved by Darst *et al.* (*4*).

Nevertheless, obtaining reproducible 3D crystals of this fragile enzyme complex has been no easy matter. In contrast to the bacterial RNA polymerase and, say, ribosomes (the molecular machines that make proteins), the RNA polymerase II complex is present in low amounts in yeast so that, even with the heterogeneity eliminated, determination of the first 5 angstrom crystal structure (*6*) required 10,000 liters of cultured yeast cells. The crystals themselves showed much variability in unit-cell dimensions, and only after shrinking the crystals in other buffers, were reproducible crystals formed of sufficiently good quality to diffract x-rays to 3.1 angstrom, the form solved last year (*5*). The second (open) form described by Kornberg and co-workers (*1*) was produced by further shrinkage of the crystals.

Growing crystals of a transcribing polymerase (*2*) presented even more difficulties because even highly purified preparations contained many inactive molecules, which could be identified individually under the electron microscope, but interfered with homogeneous 3D crystallization. The problem was solved by removing inactive enzyme molecules with columns of heparin (the heparin is able to bind to the empty DNA cleft in the enzyme) (*8*).

Although the goal of solving the enzymatic mechanism of transcription is an end in itself, it is part of a wider quest to understand the gene-specific regulation of transcription. For Kornberg, this quest began nearly 20 years ago, when he started to work on yeast at a time when most researchers were studying mammalian cells. It was a strategic decision based on the powerful potential of yeast genetics, especially when combined with biochemistry, and has now paid handsome dividends. At the time, however, it was not at all clear that it would, as no one, including Kornberg, had been able to make an extract from yeast that supported promoter-dependent polymerase II transcription (which had been achieved with mammalian cells by Roeder and colleagues) (*9*). Moreover, Roeder (*10*) had also prepared fractions from HeLa cells that supported accurate initiation at a promoter and that were later shown to contain the five general transcription factors (GTFs): TFIIB, D, E, F, and H.

The difficulties with yeast were finally overcome (*11*) and, despite the low amounts of RNA polymerase II in yeast, the Kornberg group finally purified the yeast GTFs (*12*). These turned out to be highly homologous to mammalian GTFs. An added bonus associated with purifying the yeast GTFs was a hint of the existence (later leading to the purification) of part of the transcription machinery responsible for gene-specific regulation—a 20-subunit protein called Mediator (*13*, *14*). Initially, it was thought that gene-specific transcription factors (or activators) worked by directly contacting the GTFs. But now we know that it is the Mediator complex that transduces the regulatory information from gene-specific factors binding to upstream activating sequences (or enhancers in mammalian cells) to the RNA polymerase itself. The huge Mediator complex appears to form a crescent shape that could embrace the RNA polymerase (*15*).

Where do the GTFs come into the picture? Again, a start was made by Kornberg's group, who analyzed 2D cocrystals of polymerase II together with TFIIB or TFIIE by electron microscopy (*16*). TFIIB brings the TATA box complex to the polymerase II surface at a point where the DNA can follow a straight path and the transcription start site is juxtaposed to the active center [see figure 10 in the Cramer *et al.* research article (*1*)]. A glimpse of what might be happening comes from the known properties and interactions of the GTFs. Once bound to polymerase II, the promoter DNA must be opened up by the helicase activity of TFIIH to enable the initiation of transcription. One might also imagine that this huge 500-kilodalton complex

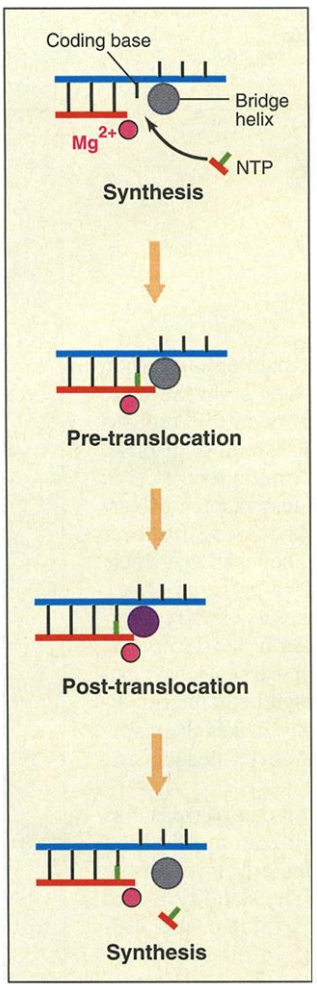


Diagram of a step in the transcription elongation cycle. The coding strand of the DNA is in blue and the RNA is in red. Bases are in black, except for the one in the substrate nucleoside triphosphate (NTP) entering the RNA polymerase, which is in green. The pre-translocation state shown is that actually observed in the transcribing complex structure, whereas the synthesis and post-translocation states are inferred. In the pre-translocation state, an active center Mg^{2+} ion (pink dot) is located between the nucleotide just added by synthesis and the next nucleotide in the RNA chain. The bridge helix of the protein (see figure, previous page) may exist in two conformations: The first abuts the coding base in the synthesis and pre-translocation states (gray sphere), and the second abuts this base in the post-translocation state (purple sphere) accompanying movement of the DNA and RNA along the protein surface. The bridge may therefore help the enzyme retain a grip on the growing end of the DNA-RNA hybrid during movement. [Adapted from (*2*)]

CREDIT: ROGER KORNBERG

(*17*)—which can straddle the whole polymerase enzyme—opens up the DNA further downstream and so disengages the rest of the DNA such that it does not begin to drag on the actual region of DNA being transcribed.

Whatever the case, we shall probably soon find out how GTFs cooperate in the initiation of transcription. The two Kornberg papers (*1*, *2*) mention work in progress on 3D cocrystals of RNA polymerase together with various GTFs. The determination of these structures should fill the gaps in our picture of promoter-dependent transcription. This is structural biology on a grand scale. So great is its power when combined with biochemistry and genetics that we should eventually even be able to see Mediator regulating the enzyme complex. We look forward to further chapters in the saga of this voyage of discovery.

References

1. P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1863 (2001).
2. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1876 (2001).
3. C. L. Poglitsch *et al.*, *Cell* **98**, 791 (1999).
4. G. Zhang *et al.*, *Cell* **98**, 811 (1999).
5. P. Cramer *et al.*, *Science* **288**, 640 (2000).
6. J. Fu *et al.*, *Cell* **98**, 799 (1999).
7. E. E. Uzgiris *et al.*, *Nature* **301**, 125 (1983).
8. A. L. Gnatt *et al.*, *J. Biol. Chem.* **272**, 30799 (1997).
9. P. A. Weil *et al.*, *Cell* **18**, 469 (1979).
10. T. Matsui *et al.*, *J. Biol. Chem.* **255**, 11992 (1980).
11. N. F. Lue and R. D. Kornberg, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8839 (1987).
12. M. H. Sayre *et al.*, *J. Biol. Chem.* **267**, 23376 (1992).
13. Y. J. Kim *et al.*, *Cell* **77**, 599 (1994).
14. L. C. Myers *et al.*, *Genes Dev.* **12**, 45 (1998).
15. F. J. Asturias, Y. W. Jiang, L. C. Myers, C. M. Gustafsson, R. D. Kornberg, *Science* **283**, 985 (1999).
16. K. K. Leuther *et al.*, *Cell* **85**, 773 (1996).
17. J. Q. Svejstrup *et al.*, *J. Biol. Chem.* **269**, 28044 (1994).

PERSPECTIVES: CHEMISTRY

Fleeting Molecules Extend Their Stay

Curt Wentrup

In its stable saturated compounds, such as methane (CH_4), carbon is bound to four other atoms. Molecules with carbon valencies other than 4—such as free radicals, carbocations, carbanions, and carbenes—are usually reactive intermediates with short lifetimes. These fleeting species can sometimes be stabilized. Stable free radicals were first prepared in 1900, but carbenes long remained elusive. On page 1901 of this issue, Solé *et al.* (*1*) report an important synthetic advance that enables the preparation of unusually stable yet highly reactive carbenes. Related species may find future application as efficient catalysts.

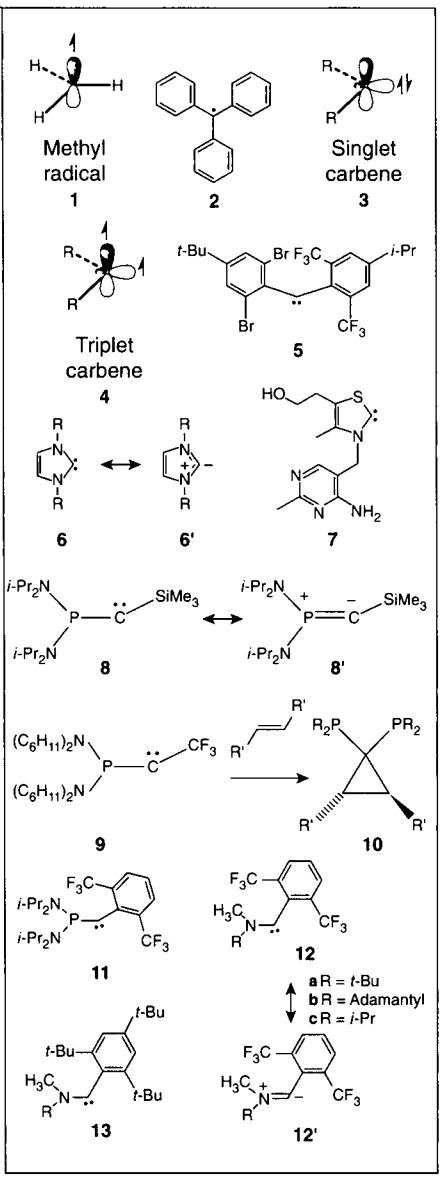
In free radicals, the carbon atom is bound to three atoms, with one nonbonded electron remaining on the carbon. A simple example is the methyl radical, CH_3 (**1**) (see the figure). Many free radicals are transient reactive intermediates, but numerous radicals have been made that are stable at room temperature, starting with the pioneering preparation of triphenylmethyl (**2**) by Gomberg in 1900 (*2*). This molecule is stabilized through delocalization of the free electron into the aromatic rings and remains stable because the bulky groups prevent access to the radical center ("steric hindrance").

It took much longer to prepare stable carbenes. The carbon atom in a carbene is surrounded by only six electrons, rather than the usual eight that confer electronic stability. With just two bound atoms, two nonbonded electrons remain on the carbon atom. These electrons may be in the same orbital (singlet state, **3**) or in different orbitals (triplet state, **4**). Attempts to make the simplest carbene, methylene (CH_2), were made before the tetravalency of carbon became generally known (*3*). Just over 100 years ago, Nef still believed that stable simple carbenes could be prepared (*4*, *5*). The existence of carbenes as reactive intermediates was established in the 1950s and 1960s (*6*, *7*). Since then, carbenes have become well-established synthetic intermediates. Their high reactivity makes them versatile targets for preparative, mechanistic, and theoretical work.

In recent years, much progress has been made in the stabilization of triplet di(aryl)carbenes, particularly through the work of Tomioka (*8*). The stability record is currently held by triplet carbene **5**, which has a half-life of 16 minutes in solution at room temperature and is stable indefinitely at –40°C (*9*, *10*). The stabilization of this and related carbenes largely results from steric hindrance of the carbene center by large groups or atoms.

In 1991, the first stable singlet di(amino)carbene (**6**) was prepared by Arduengo *et al.* (*11*). Di(amino)carbenes are stabilized in the singlet state by electron donation from the two nitrogen lone pairs into the vacant carbon p orbital on the carbon. This stabilization is also termed "push-push" stabilization. Many cyclic and acyclic di(amino)carbenes have been prepared, but they do not have typical carbene reactivity because of the electronic stabilization (best described by the resonance structure **6′**) (*12*, *13*). A related species (**7**) is involved in the action of the coenzyme thiamin (*14*, *15*).

The author is in the Department of Chemistry, School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia. E-mail: wentrup@chemistry.uq.edu.au



How to make stable radicals and carbenes. Free radicals (**1**) were stabilized many years ago (**2**). Carbenes (**3** and **4**) have proved more difficult to stabilize. Recent synthetic advances have led to the successful preparation of stable carbenes (**5** to **13**), some of which retain their reactivity. See text for detailed descriptions of the depicted molecules.