How Does III × II Make U6?

JAMES E. DAHLBERG AND ELSEBET LUND

YET ANOTHER SIMPLIFYING GENERALIZATION IN MOLECULAR biology seems to be on its way out—or at least destined for modification. Until recently, it was assumed that each of the three classes of nuclear RNA polymerases in eukaryotic cells (RNAPs I, II, and III) (*1, 2*) had its own set of associated factors. It is now becoming clear that not all transcription by RNAP III obeys this rule. In particular, synthesis of the U6 small nuclear RNA (snRNA) by RNAP III in vertebrate cells requires an RNAP II factor, TFIID (for transcription factor D of RNAP II); this factor recognizes a DNA sequence, the so-called TATA box, that is otherwise a hallmark of mRNA genes transcribed by RNAP II. The genes for several other small RNAs transcribed by RNAP III, such as 7SK, H1, MRP, and *c-myc* RNAs have structures similar to that of U6 RNA (*3*), and they may have similar requirements. In this Perspective we describe studies on the transcription of vertebrate and yeast U6 RNA genes (Fig. 1).

Four of the five snRNAs known to participate in processing of pre-mRNA (*4*) (called U1, U2, U4, and U5 RNA) are synthesized by RNAP II, whereas the fifth snRNA (U6) is made by RNAP III (*5*). However, snRNA genes are not typical of other RNAP II or RNAP III transcription units. For example, the vertebrate snRNA genes transcribed by RNAP II lack the TATA box sequence present in the promoters of most other RNAP II genes and contain a sequence that couples 3′ end formation to promoter function (*6*). Moreover, the nucleotide sequences required for transcription of vertebrate U6 genes by RNAP III complex lie entirely upstream of the coding region (*7*); this organization contrasts with that of most other RNAP III templates, like tRNA or 5*S* ribosomal RNA (rRNA) genes, in which sequences within the coding region are required for initiation (*2*).

Another intriguing feature of the snRNA genes of higher eukaryotes is that, regardless of whether they are transcribed by RNAP II or RNAP III, the promoters closely resemble each other in sequence. Thus, all vertebrate snRNA genes contain an essential, conserved region called the PSE (for proximal sequence element) located at precise distances upstream of the points where RNAP II and RNAP III initiate snRNA transcription (*5*) (centered at positions −55 and −60, respectively); in the case of RNAP II snRNA genes, this element fixes the position of the 5′ end of the RNA (*5*) and also is needed for 3′ end formation (*6*). About 175 bp upstream of the PSE is the DSE (distal sequence element), which functions as an snRNA promoter-specific enhancer of transcription (*5*).

In addition to these two common promoter elements, a paradoxical feature of vertebrate U6 genes is the presence of another essential promoter element, at about position −30 (*7*); this sequence element conforms to the consensus sequence of the TATA box. In fact, the TATA boxes of mRNA and U6 gene promoters can be interchanged without loss of function (*8*). Moreover, insertion of a TATA box into the promoter of an RNAP II–type snRNA gene converts it into a template for RNAP III rather than for RNAP II (*9*). Thus, the same sequence specifies initiation of transcription of mRNA genes by RNAP II and of snRNA genes by RNAP III.

Several laboratories have fractionated cell extracts and purified transcription factors to learn what trans-acting factors are needed for vertebrate U6 RNA synthesis (*8, 10–12*). Work by Reddy (*10*) first demonstrated that synthesis of U6 RNA requires both general RNAP III transcription factors and snRNA-specific factors. At least three factors are required in addition to RNAP III. They include TFIIIB, previously shown to be needed for initiation of transcription by RNAP III, plus two factors that bind specifically to the essential U6 gene promoter sequences described above, the PSE and the TATA box (*8, 11, 12*). Seifart and co-workers (*11*) demonstrated that a fraction highly enriched in TFIIIB is required for mammalian U6 RNA synthesis. Moreover, they isolated a protein (PSE binding protein, PBP) that binds to the PSE element and is necessary for U6 gene transcription in vitro. Concurrently, fractionation studies in the laboratories of Hernandez (*8*) and of Mattaj (*12*) showed that recognition of the TATA box in U6 genes is mediated by a heat labile factor that can be replaced by the product of the cloned gene encoding TFIID. Although these studies show that TFIID can promote transcription of the U6 gene by RNAP III, the in vitro system used is not as strictly dependent on the PSE as is U6 synthesis in vivo (*8*).

To understand how this assemblage of transcription factors might function during U6 gene transcription, it is informative to discuss the role of TFIIIB in RNAP III transcription. TFIIIB forms a stable complex with DNA sequences in tRNA and 5*S* rRNA genes just upstream of the point of transcription initiation and this complex is responsible for directing RNAP III to the correct start site. Formation of the TFIIIB-DNA complex, in turn, depends on the prior binding of another factor, TFIIIC, to sequences (the A- and B-blocks) located within the coding region of tRNA genes (*2, 13*). In 5*S* rRNA genes, a third factor, TFIIIA, that binds to an internal control region (ICR) within the coding region (*2*) also participates in this process (*13*).

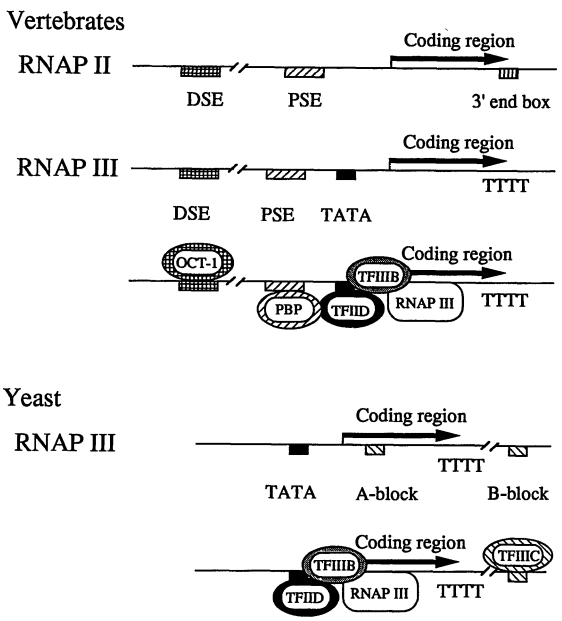

Fig. 1. Schematic representations of snRNA genes of vertebrate and yeast cells. In vertebrates both types of snRNA genes use upstream promoter elements (DSE and PSE), but the precise location of the PSE (relative to the transcription start site) differs between RNAP II and RNAP III genes. The yeast U6 gene utilizes promoter elements inside and downstream of the coding region (A- and B-blocks) and contains an upstream TATA box. Some of the factors needed for transcription of these genes are PBP, TFIIIC, TFIID, and octamer binding protein 1 (Oct-1). Structures are not drawn to scale.

Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI 53706.

Vertebrate U6 genes lack 5S gene–ICR, A-block, and B-block sequences and neither TFIIIA nor TFIIIC appears to be necessary for their transcription in vitro (*10, 11*). Thus, formation of a complex between TFIIIB and the promoter must occur by a mechanism different from that of tRNA or 5S rRNA genes. It is likely that PBP and TFIID, interacting with each other while bound to their respective cis-acting sequences in the 5′ flanking region of U6 genes, could facilitate formation of a TFIIIB-promoter complex, as has been proposed and diagrammed by several groups (*3, 8, 11, 12*). No evidence has been presented that PBP or TFIID remain on the template after TFIIIB is bound or that these three factors are the only ones needed.

The simultaneous interaction of PBP and TFIID with each other, their target sequences, and TFIIIB would be expected to place severe constraints on the spacing between the PSE and TATA box sequences in the DNA template. Indeed, this distance is invariant in all vertebrate U6 genes (*3, 5*), and experimental alterations increasing this distance severely impair RNAP III transcription (*8, 14*). The essential nature of this precise spacing is further illustrated by the snRNA genes of the plant *Arabadopsis*, in which the promoters utilized by RNAP II or RNAP III possess identical PSE-like and TATA sequences and only differ in the number of helical repeats that separate these elements (three versus two turns, respectively) (*15*).

Relatively little is known about the components of the system responsible for transcription of vertebrate U1 to U5 snRNA genes by RNAP II, although progress has been made in defining factors that interact with the DSE (*5, 8*). As with many other constitutively expressed RNAP II housekeeping genes, the U1 to U5 genes lack a TATA box to which TFIID might bind. By analogy with other TATA-less genes, this factor could still participate in the initiation of transcription of these genes through protein-protein interaction (*16*); paradoxically, introduction of a TATA sequence into the promoter of an RNAP II snRNA gene, at a location inappropriate for RNAP III, inhibits transcription by RNAP II (*8, 14*). Other factors that do not bind directly to the template may be required for transcription of snRNA genes by RNAP II; perhaps one of these is responsible for recognition of the snRNA-specific 3′ end signal that the RNAP II transcription complex encounters at the far end of the gene (*17*). The specificity of factors for RNAP II transcription of snRNA versus mRNA genes is indicated by the fact that DNA-dependent in vitro transcription systems for the U1 to U5 genes were developed only recently (*18*).

In contrast to the multicopy snRNA genes in higher eukaryotes, the single-copy snRNA genes of yeast appear to have promoters that more closely resemble those of other nuclear genes. For example, yeast equivalents of the RNAP II snRNA genes contain TATA boxes, and their promoters can be substituted with mRNA promoters (*5*). The U6 gene of *Saccharomyces cerevisiae* contains sequences that resemble the A- and B-blocks (located within and downstream of the coding sequence, respectively) as well as an upstream TATA box (*19*). The B-block is essential for in vivo function of the U6 RNA gene and in vitro transcription in crude extracts (*19*). Binding of TFIIIC to the B-block would presumably result in the formation of a stable complex between TFIIIB and the gene, as described above for tRNA genes (*13*). In contrast, a yeast U6 gene that lacks the downstream B-block element can be transcribed in vitro upon addition of sufficient TFIIIB (*20*); as expected, in vitro transcription of this truncated template does not require TFIIIC (*21*). Like vertebrate U6 genes, this transcription is dependent on added TFIID, suggesting that TFIID may bind to the TATA box in the yeast U6 gene; in the absence of TFIIIC, TFIID added in vitro might stabilize a complex between TFIIIB and the mutant template. It remains to be determined if this type of stabilization occurs during in vivo transcription of the wild-type gene.

Why, then, do higher eukaryotes use the PSE and TATA box to establish the presumptive interaction between TFIIIB and the U6 gene, whereas yeast do it with the B-block, similar to the way tRNA transcription is initiated? Perhaps the PSE is a cis-acting control element that is used by higher eukaryotes as a way of coordinating transcription of all snRNA genes. In yeast, where snRNAs are encoded by single copy genes and where the concentrations of snRNA are much lower, it may not be necessary to maintain such fine control of snRNA synthesis.

How might the PSE control two different classes of vertebrate snRNA genes? To answer these questions it is necessary to know if a single type of PBP or TFIID complex is able to effect transcription initiation on both RNAP II and RNAP III snRNA genes. We propose that, for both RNAP II and RNAP III transcription, PBP and TFIID interact with each other while PBP is bound to the PSE. When TFIID binds to an appropriately positioned TATA box, it would promote the subsequent binding of TFIIIB (*8*). In the absence of a TATA box, a PBP-TFIID complex would promote the binding of RNAP II to its transcription start site, probably utilizing snRNA-specific factors as well as others that are normally involved in basal transcription by RNAP II.

Thus the association of PBP with the common PSE element apparently is responsible for the surprising role reversal of TFIID. The mechanisms by which this occurs and by which TFIID activates transcription of snRNA genes, however, remain elusive.

REFERENCES AND NOTES

1. M. Sawadogo and A. Sentenac, *Annu. Rev. Biochem.* **59**, 711 (1990).
2. E. P. Geiduschek and G. P. Tocchini-Valentini, *ibid.* **57**, 873 (1988).
3. G. Kunkel, *Biochem. Biophys. Acta* **1088**, 1 (1991); N. Hernandez, in *Transcriptional Regulation*, S. L. McKnight and K. R. Yamamoto, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, in press).
4. J. A. Steitz *et al.*, in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*, M. L. Birnstiel Ed. (Springer-Verlag, Berlin, 1988), p. 115; C. Guthrie and B. Patterson, *Annu. Rev. Genet.* **22**, 387 (1988).
5. J. E. Dahlberg and E. Lund, in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*, M. L. Birnstiel, Ed. (Springer-Verlag, Berlin, 1988), p. 38.
6. N. Hernandez and R. Lucito, *EMBO J.* **7**, 3125 (1988); H. E. Neuman de Vegvar and J. E. Dahlberg, *Nucleic Acids Res.* **17**, 9305 (1989).
7. P. Carbon *et al.*, *Cell* **51**, 71 (1987); G. Das, D. Henning, D. Wright, R. Reddy, *EMBO J.* **7**, 503 (1988); G. Kunkel and T. Pederson, *Genes Dev.* **2**, 196 (1988).
8. S. M. Lobo, J. Lister, M. L. Sullivan, N. Hernandez, *Genes Dev.* **5**, 1477 (1991).
9. I. W. Mattaj, N. A. Dathan, H. W. Parry, P. Carbon, A. Krol, *Cell* **55**, 435 (1988); S. M. Lobo and N. Hernandez, *ibid.* **58**, 55 (1989).
10. R. Reddy, *J. Biol. Chem.* **263**, 15980 (1988).
11. R. Waldschmidt, I. Wanandi, K. H. Seifart, *EMBO J.* **10**, 2595 (1991).
12. K. A. Simmen *et al.*, *ibid.*, p. 1853.
13. G. A. Kassavetis, B. R. Braun, L. H. Nguyen, E. P. Geiduschek, *Cell* **60**, 235 (1990).
14. A. Lescure, P. Carbon, A. Krol, *Nucleic Acids Res.* **19**, 435 (1991).
15. G. Waibel and W. Filipowicz, *ibid.* **18**, 3451 (1990); *Nature* **346**, 199 (1990).
16. B. F. Pugh and R. J. Tijan, *Cell* **61**, 1187 (1990); M. Ptashne and A. A. F. Gann, *Nature* **346**, 329 (1990).
17. H. E. Neuman de Vegvar, E. Lund, J. E. Dahlberg, *Cell* **47**, 259 (1986); N. Hernandez and A. M. Weiner, *ibid.*, p. 249.
18. E. Lund and J. E. Dahlberg, *EMBO J.* **8**, 287 (1989); S. I. Gunderson, M. W. Knuth, R. R. Burgess, *Genes Dev.* **4**, 2048 (1990).
19. D. Brow and C. Guthrie, *Genes Dev.* **4**, 1345 (1990).
20. F. Margottin *et al.*, *Science* **251**, 424 (1991).
21. A. Moenne *et al.*, *EMBO J.* **9**, 271 (1990).

22. We thank D. A. Brow and M. P. Terns for thoughtful and critical comments on this manuscript. Supported by NIH grant GM30220.