RNA SPLICING

Structure of a transcribing RNA polymerase II–U1 snRNP complex

Suyang Zhang[1], Shintaro Aibara[1], Seychelle M. Vos[1]*, Dmitry E. Agafonov[2]†, Reinhard Lührmann[2], Patrick Cramer[1]‡

To initiate cotranscriptional splicing, RNA polymerase II (Pol II) recruits the U1 small nuclear ribonucleoprotein particle (U1 snRNP) to nascent precursor messenger RNA (pre-mRNA). Here, we report the cryo–electron microscopy structure of a mammalian transcribing Pol II–U1 snRNP complex. The structure reveals that Pol II and U1 snRNP interact directly. This interaction positions the pre-mRNA 5′ splice site near the RNA exit site of Pol II. Extension of pre-mRNA retains the 5′ splice site, leading to the formation of a "growing intron loop." Loop formation may facilitate scanning of nascent pre-mRNA for the 3′ splice site, functional pairing of distant intron ends, and prespliceosome assembly. Our results provide a starting point for a mechanistic analysis of cotranscriptional spliceosome assembly and the biogenesis of mRNA isoforms by alternative splicing.

The production of mRNA in eukaryotic cells involves precursor mRNA (pre-mRNA) synthesis and processing—in particular, 5′ end capping, splicing, and 3′ end cleavage and polyadenylation. During splicing, noncoding introns are predominantly removed from the pre-mRNA in a cotranscriptional manner as the nascent RNA emerges from RNA polymerase II (Pol II) (*1*–*6*). Cotranscriptional splicing enhances the efficiency and accuracy of pre-mRNA processing and explains why splicing is at least 10 times faster in vivo than in vitro (*7*). In metazoan cells, introns are often several thousand nucleotides long, which raises the question of how the ends of an intron are functionally paired for splicing. Cotranscriptional splicing has been suggested to facilitate juxtaposition of the 5′ splice site (5′SS) and the 3′SS (*8*). Consistent with this idea, the rate of Pol II elongation can affect selection of splice sites in pre-mRNA (*9*), leading to alternative splicing and different mRNA isoforms (*10*). Pol II can recruit splicing factors via its flexible C-terminal domain (CTD) [reviewed in (*6*)]; however, this is insufficient to stimulate splicing (*11*). Despite these advances, the mechanisms underlying cotranscriptional splicing remain unknown.

As a first step to investigate the mechanisms of cotranscriptional splicing, we studied the interaction between transcribing Pol II and the U1 small nuclear ribonucleoprotein particle (U1 snRNP) biochemically and structurally. U1 snRNP is the first building block of the spliceosome to engage with nascent pre-mRNA (*12*). Human U1 snRNP consists of U1 snRNA, seven Sm proteins, and three U1-specific proteins (U1-70k, U1-A, and U1-C) (*12*). U1 snRNA recognizes the 5′SS through base pairing (*13*). When the branch point sequence emerges on nascent pre-mRNA, U2 snRNP joins, forming the prespliceosome or A complex. The A complex later associates with the U4/U6.U5 tri-snRNP to form the pre-B complex (*14*), which is subsequently converted to the B complex and activated for splicing.

To assemble a transcribing Pol II–U1 snRNP complex, we used Pol II from the domestic pig, *Sus scrofa domesticus*, which has 99.9% sequence identity to human Pol II. We further used human U1 snRNP and a DNA-RNA scaffold that contains a DNA mismatch bubble and a modified MINX pre-mRNA with a 5′ cap. The scaffold enables formation of a 9–base pair DNA-RNA hybrid duplex inside the bubble and contains a 145-nucleotide (nt) RNA that comprises a 5′ exon and a truncated intron of 29 nts (Fig. 1A and fig. S1, A and B). We incubated purified Pol II with the DNA-RNA scaffold and phosphorylated the resulting complex with the kinase-positive transcription elongation factor b, which stabilizes an active Pol II elongation complex (*15*). The phosphorylated Pol II–DNA-RNA complex was purified by size exclusion chromatography and incubated with purified U1 snRNP (fig. S1C).

The obtained sample was subjected to single-particle cryo–electron microscopy (cryo-EM) analysis (see materials and methods). A tilt angle of 40° was applied during data collection to resolve orientation bias (fig. S1, D to F, and table S1). Processing of the cryo-EM data showed that 17.5% of Pol II particles contained densities corresponding to U1 snRNP (fig. S2, A to F). Three-dimensional (3D) classification followed by 3D refinement of the class with the best density for U1 snRNP yielded a final reconstruction of the Pol II–U1 snRNP complex at an overall resolution of 3.6 Å, and focused 3D refinement improved densities for Pol II, U1 snRNP, and the Pol II–U1 snRNP interface region (figs. S2, G to I, S3, and S4 and movie S1). Fitting and adjustment of the structures of Pol II (*15*) and U1 snRNP (*16*–*18*) resulted in a model of the Pol II–U1 snRNP complex with good stereochemistry (Fig. 1, B and C, table S1, and movie S2).

The structure shows that U1 snRNP binds directly to the surface of Pol II near upstream DNA and the site of RNA exit (Fig. 1 and movie S2). U1 snRNP engages Pol II via its conserved and functionally essential (*17*) subunit U1-70k, which contacts the protrusion domain in Pol II subunit RPB2 and the zinc finger domain in subunit RPB12 (Figs. 1, B and C, and 2A). U1-70k interacts with Pol II through its RNA recognition motif (RRM) domain that is resolved at ~3.5-Å resolution, revealing bulky side chains and enabling an unambiguous fit (figs. S3D and S4, B and C). We did not observe any densities indicative of the Pol II CTD.

Our structure provides details of the Pol II–U1 snRNP interface. The U1-70k RRM interacts with Pol II mainly through its two conserved α helices (human residues 116 to 126 and 154 to 164) (Fig. 2A). The Pol II–U1 snRNP interacting surfaces have reverse charges (fig. S5, A and B). The side chain of the positively charged U1-70k RRM residue Arg^{121} inserts into a negatively charged pocket formed by RPB2 and RPB12 (Fig. 2B). Three negatively charged residues (Glu^{124}, Glu^{152}, and Asp^{156}) of U1-70k RRM contact a positively charged surface of the RPB2 protrusion domain (Fig. 2C). In addition, Val^{125} of U1-70k RRM binds to a hydrophobic patch of RPB2 (Fig. 2D).

Residues in U1-70k that contact Pol II are highly conserved among metazoa (fig. S5C). RPB2 residues that interact with U1-70k are also well conserved among multicellular eukaryotes but differ from those in the counterpart subunits of Pol I and Pol III (fig. S5D). This indicates that the observed U1 snRNP interaction with Pol II is conserved among metazoa and specific to the Pol II transcription system. By contrast, the interacting residues in both U1-70k and RPB2 are only partially conserved in yeasts (fig. S5, C and D), implying that the mechanism of cotranscriptional splicing may differ in some unicellular eukaryotes, which contain fewer and shorter introns.

Superposition of our Pol II–U1 snRNP structure with the activated Pol II elongation complex EC* (*15*) revealed that binding of U1 snRNP, as observed in our structure, is compatible with the presence of the transcription elongation factors DSIF, SPT6, and PAF1 complex (PAF) on the Pol II surface (fig. S6). Consistent with this compatibility, we found that a stable complex containing EC* and U1 snRNP can be formed (Fig. 3, A to C, and fig. S7, A to F). In

[1]Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. [2]Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany.
*Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. †Present address: Department of Cellular Logistics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany.
‡Corresponding author. Email: patrick.cramer@mpibpc.mpg.de

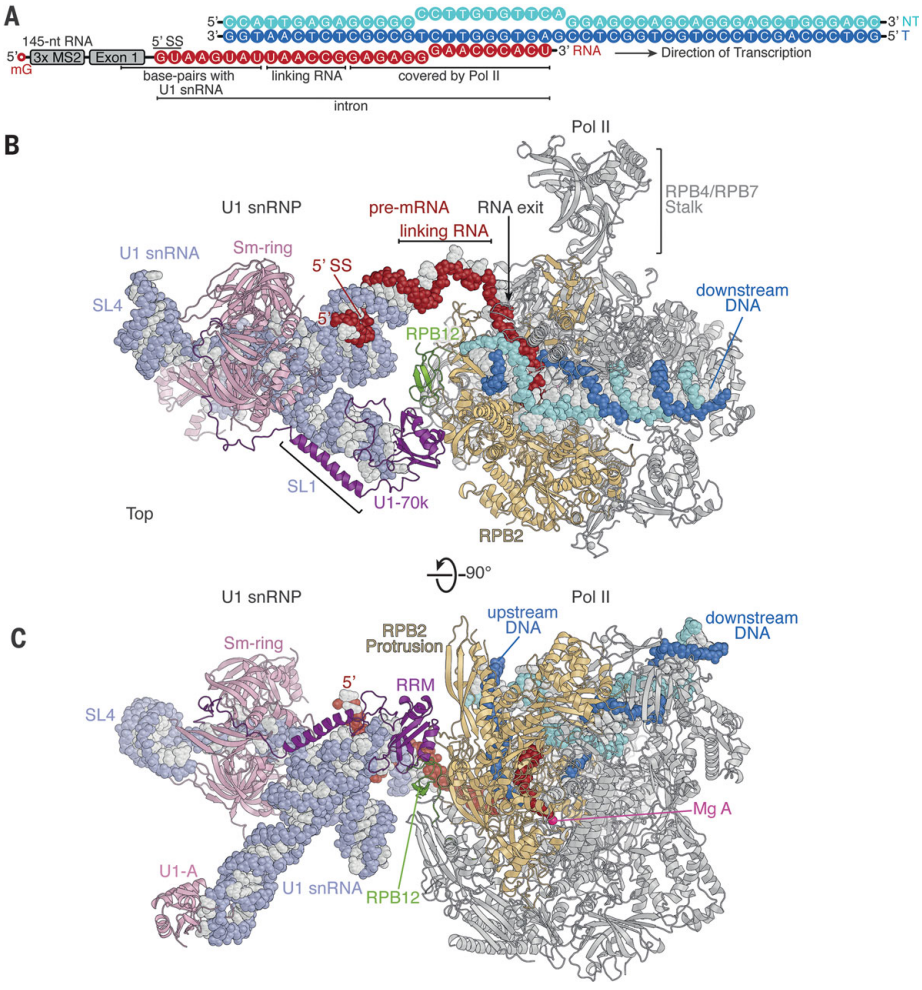

Fig. 1. Structure of the transcribing Pol II–U1 snRNP complex. (**A**) Nucleic acid scaffold with template DNA (T) in dark blue, nontemplate DNA (NT) in cyan, and RNA in red. mG, 5′ cap. (**B** and **C**) Two views of the structure. Nucleic acids are shown in spheres. The backbone of U1 snRNA is in pale slate and U1 snRNP proteins are in pink, except for U1-70k, which is in purple. Pol II subunits (ribbons) are in gray, except for RPB2 in gold and RPB12 in green. During transcription, Pol II moves to the right and RNA exits to the left. A magenta sphere depicts the Pol II active site. SL, stem loop.

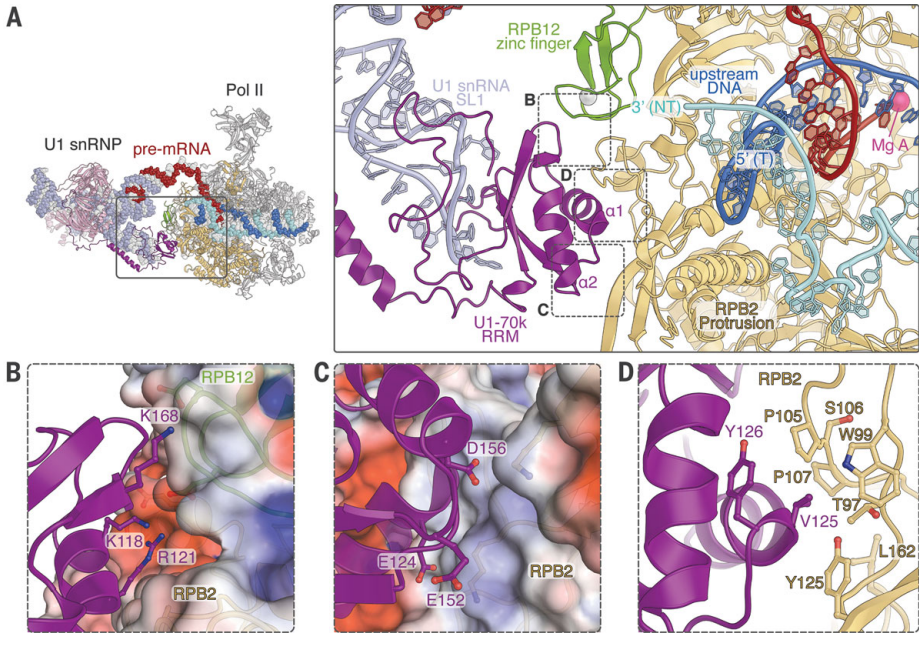

Fig. 2. Direct Pol II–U1 snRNP interaction. (**A**) Close-up view of the Pol II–U1 snRNP interface showing interactions between the RRM domain of U1-70k (purple) and Pol II subunits RPB2 (gold) and RPB12 (green). The Pol II active site is depicted with a magenta sphere. (**B** to **D**) Detailed views of interactions between residues of U1-70k RRM and Pol II. The electrostatic surface potential displays a range of ±5 k_BT/e (where k_B is the Boltzmann constant, T is temperature, and e is electron charge), where blue and red represent positively and negatively charged areas, respectively. Amino acid abbreviations: D, aspartic acid; E, glutamic acid; K, lysine; L, leucine; P, proline; R, arginine; S, serine; T, threonine; W, tryptophan; Y, tyrosine.

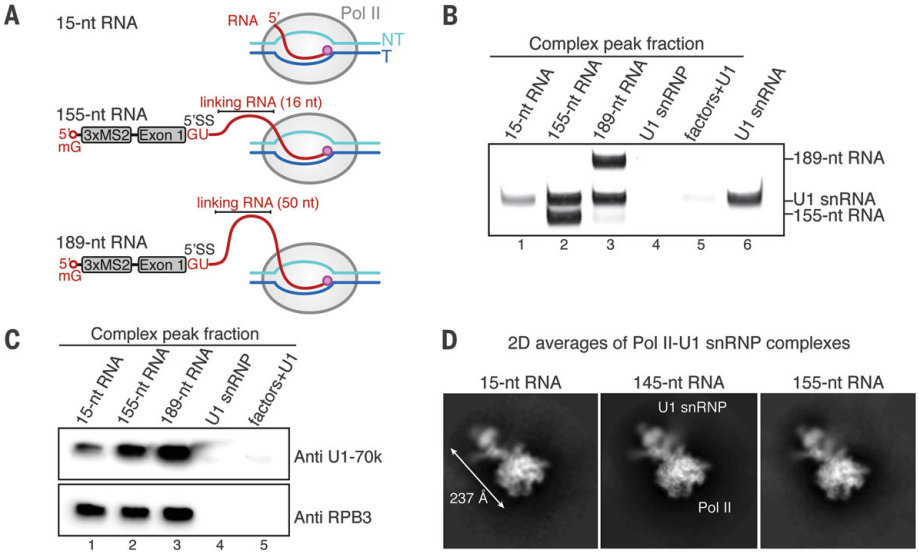

Fig. 3. RNA-independent Pol II–U1 snRNP interaction. (**A**) Schematic of Pol II complexes with various RNAs. The active site of Pol II is depicted with a magenta sphere. (**B** and **C**) EC* associates with U1 snRNP in the presence of RNA lacking the 5′SS (15-nt RNA) or RNAs with an extended linking region (155-nt and 189-nt RNAs). Shown is the analysis of the complex peak fraction (fraction 2 of fig. S7A) obtained by size exclusion chromatography using denaturing gel electrophoresis (B) or Western blotting (C). Lanes 4 and 5 are control runs with U1 snRNP alone or U1 snRNP with phosphorylated elongation factors. (**D**) Cryo-EM analysis of Pol II–U1 snRNP complexes containing 15-nt RNA or 155-nt RNA showed 2D averages resembling those of the refined structure with 145-nt RNA.

summary, these results reveal an unexpected direct Pol II–U1 snRNP interaction that is compatible with the presence of general elongation factors in the transcribing Pol II complex EC*.

Our structure additionally shows one turn of an RNA duplex formed by base pairing of the 5′SS with U1 snRNA (Fig. 1 and fig. S4, A and D), as observed in previous U1 snRNP structures (*16*–*18*). By RNA duplex formation, the pre-mRNA tethers Pol II to U1 snRNP with a 6-nt linking RNA region (Fig. 1A and fig. S4A). To investigate the role of RNA tethering, we performed analytical size exclusion assays using the activated elongation complex EC* (see materials and methods). We prepared EC* containing a minimal scaffold with a 15-nt RNA that lacks the 5′SS and is completely sequestered within the Pol II core (Fig. 3A and fig. S1, A and B), excluding any RNA-based tethering of Pol II and U1 snRNP. The resulting minimal EC* still bound U1 snRNP, although at apparently substoichiometric levels (Fig. 3, B and C, and fig. S7, A and B).

To confirm that the same direct Pol II–U1 snRNP interaction occurs in the absence of a tethering RNA, we subjected a Pol II–U1 snRNP complex containing the nucleic acid scaffold with the 15-nt RNA to cryo-EM analysis. We obtained the same two-dimensional (2D) particle averages as for our original structure (Fig. 3D). The percentage of Pol II–U1 snRNP particles was lower, hindering 3D reconstruction. Nevertheless, these results show that RNA tethering is not required for direct Pol II–U1 snRNP interaction in vitro and may explain how U1 snRNP can be recruited to transcription units independent of splicing in vivo (*19*).

Our observations suggested that transcription elongation may generate an RNA loop between Pol II and U1 snRNP while the direct Pol II–U1 snRNP interaction is maintained. To investigate this, we tested whether insertion of additional nucleotides into the linking RNA region allows for U1 snRNP binding to Pol II or EC*. We designed scaffolds that contained either a 155-nt RNA, which has a 10-nt extension in the linking RNA compared with the 145-nt RNA in the cryo-EM structure, or a 189-nt RNA, which contains a linking RNA of 50 nts and would be sufficient for formation of an active spliceosomal B complex (*20*) (Fig. 3A and fig. S1, A and B). We found that both scaffolds with extended RNAs allowed formation of an EC*-U1 snRNP complex (Fig. 3, B and C, and fig. S7, A to D). These results support the model in which, during elongation of the pre-mRNA, the 5′SS is retained near the Pol II surface whereas the intron loops out.

To confirm that the same direct Pol II–U1 snRNP interaction is maintained when the linking RNA is extended, we determined the cryo-EM structure of the Pol II–U1 snRNP complex with the scaffold containing the 155-nt RNA at a nominal resolution of 3.9 Å (fig. S8). The structure was essentially unaltered compared with our first structure, revealing the same direct Pol II–U1 snRNP interaction (Fig. 3D and fig. S8). In addition, in vitro RNA elongation activities of Pol II or EC* were not affected by the presence of U1 snRNP (fig. S7, G and H). These results indicate that U1 snRNP can remain bound to Pol II as pre-mRNA is elongated during transcription.

To investigate whether assembly of spliceosomal complexes may be possible on Pol II, we compared our structure with the yeast A complex (*21*) (fig. S9, A and B). This suggested that U2 snRNP can be accommodated and the A complex can form on the Pol II surface without moving U1 snRNP. Human A complex formation requires binding of stem loop 4 (SL4) of U1 snRNA to the U2 snRNP subunit SF3A1 (*22*). Consistent with this, SL4 faces away from Pol II in our structure and is available for SF3A1 interaction (Fig. 1, B and C). Superposition of our structure onto the human pre-B complex (*23*, *24*) resulted in clashes between Pol II and the U4/U6.U5 tri-snRNP (fig. S9C). However, the tri-snRNP is flexibly attached to the remainder of the pre-B complex (*14*), and structural adjustments may therefore allow formation of the pre-B complex on Pol II.

Our findings suggest a topological model for cotranscriptional spliceosome assembly that we refer to as the growing intron loop model (Fig. 4). In this model, a direct Pol II–U1 snRNP interaction retains the 5′SS near the RNA exit site of Pol II during transcription. This results in formation of a growing loop in the nascent pre-mRNA that facilitates scanning for the downstream branch point and 3′SS. Recruitment of U2 snRNP then leads to formation of the A complex on the Pol II surface. Formation of the A complex may be further stimulated by the downstream 5′SS (*25*) and binding of the downstream exon to regulatory factors that facilitate U2 snRNP recruitment (*26*). After A complex formation, the U4/U6.U5 tri-snRNP may join, leading to the pre-B complex. The subsequent transition to the B complex displaces U1 snRNP (*12*), liberating the precatalytic spliceosome from Pol II. Therefore, the growing intron loop model implies that the assembly of the spliceosome occurs cotranscriptionally, whereas the ensuing catalytic steps of splicing may be functionally independent of Pol II, although they occur predominantly on the nascent pre-mRNA.

Finally, U1 snRNP inhibits 3′ end RNA processing (*27*), suppresses premature transcription termination (*28*), and impairs premature termination in the sense direction of bidirectional promoters (*29*, *30*). To explain these findings, we speculate that U1 snRNP binding to the Pol II surface interferes with

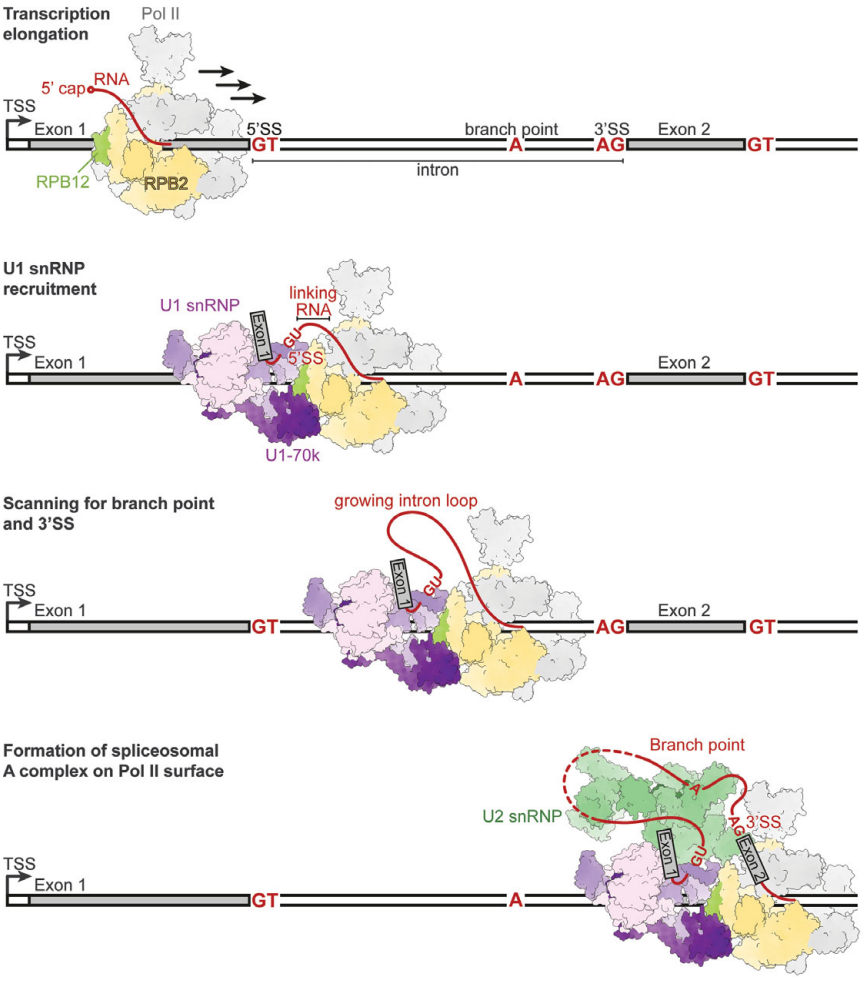

Fig. 4. Growing intron loop model of cotranscriptional spliceosome assembly. The intron is defined by the 5′SS, branch point, and 3′SS (conserved nucleotides in red). When the 5′SS emerges in nascent pre-mRNA, U1 snRNP (purple) is recruited and directly binds Pol II subunits RPB2 (gold) and RPB12 (green). U1 snRNP and the 5′SS are retained near the RNA exit site during transcription elongation, resulting in the formation of an RNA loop. Loop formation facilitates scanning for the downstream branch point and 3′SS and assembly of the A complex.

productive binding of the 3′ end processing machinery. Thereby, U1 snRNP may suppress premature termination and keep Pol II in an actively transcribing state until a genuine 3′ end processing signal is detected in the nascent RNA.

REFERENCES AND NOTES

1. K. M. Kotovic, D. Lockshon, L. Boric, K. M. Neugebauer, *Mol. Cell. Biol.* **23**, 5768–5779 (2003).
2. S. A. Lacadie, M. Rosbash, *Mol. Cell* **19**, 65–75 (2005).
3. I. Listerman, A. K. Sapra, K. M. Neugebauer, *Nat. Struct. Mol. Biol.* **13**, 815–822 (2006).
4. E. W. J. Wallace, J. D. Beggs, *RNA* **23**, 601–610 (2017).
5. G. Bird, D. A. Zorio, D. L. Bentley, *Mol. Cell. Biol.* **24**, 8963–8969 (2004).
6. M. Tellier, I. Maudlin, S. Murphy, *WIREs RNA* **11**, e1593 (2020).
7. P. J. Curtis, N. Mantei, C. Weissmann, *Cold Spring Harb. Symp. Quant. Biol.* **42**, 971–984 (1978).
8. M. J. Dye, N. Gromak, N. J. Proudfoot, *Mol. Cell* **21**, 849–859 (2006).
9. L. E. Giono, A. R. Kornblihtt, *Biochem. J.* **477**, 3091–3104 (2020).
10. M. de la Mata *et al.*, *Mol. Cell* **12**, 525–532 (2003).
11. B. J. Natalizio, N. D. Robson-Dixon, M. A. Garcia-Blanco, *J. Biol. Chem.* **284**, 8692–8702 (2009).
12. C. L. Will, R. Lührmann, *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
13. S. M. Mount, I. Pettersson, M. Hinterberger, A. Karmas, J. A. Steitz, *Cell* **33**, 509–518 (1983).
14. C. Boesler *et al.*, *Nat. Commun.* **7**, 11997 (2016).
15. S. M. Vos *et al.*, *Nature* **560**, 607–612 (2018).
16. G. Weber, S. Trowitzsch, B. Kastner, R. Lührmann, M. C. Wahl, *EMBO J.* **29**, 4172–4184 (2010).
17. Y. Kondo, C. Oubridge, A. M. van Roon, K. Nagai, *eLife* **4**, e04986 (2015).
18. D. A. Pomeranz Krummel, C. Oubridge, A. K. Leung, J. Li, K. Nagai, *Nature* **458**, 475–480 (2009).
19. B. Spiluttini *et al.*, *J. Cell Sci.* **123**, 2085–2093 (2010).
20. K. Bertram *et al.*, *Cell* **170**, 701–713.e11 (2017).
21. C. Plaschka, P. C. Lin, C. Charenton, K. Nagai, *Nature* **559**, 419–422 (2018).
22. S. Sharma, S. P. Wongpalee, A. Vashisht, J. A. Wohlschlegel, D. L. Black, *Genes Dev.* **28**, 2518–2531 (2014).
23. C. Charenton, M. E. Wilkinson, K. Nagai, *Science* **364**, 362–367 (2019).
24. X. Zhan, C. Yan, X. Zhang, J. Lei, Y. Shi, *Cell Res.* **28**, 1129–1140 (2018).
25. B. L. Robberson, G. J. Cote, S. M. Berget, *Mol. Cell. Biol.* **10**, 84–94 (1990).
26. J. Ule, B. J. Blencowe, *Mol. Cell* **76**, 329–345 (2019).
27. S. Vagner, U. Rüegsegger, S. I. Gunderson, W. Keller, I. W. Mattaj, *RNA* **6**, 178–188 (2000).
28. D. Kaida *et al.*, *Nature* **468**, 664–668 (2010).
29. A. E. Almada, X. Wu, A. J. Kriz, C. B. Burge, P. A. Sharp, *Nature* **499**, 360–363 (2013).
30. P. K. Andersen, S. Lykke-Andersen, T. H. Jensen, *Genes Dev.* **26**, 2169–2179 (2012).

ACKNOWLEDGMENTS

We thank C. Dienemann and U. Steuerwald for support at the microscope and maintaining the EM facility, T. Schulz for maintaining cells and thymus tissue, and P. Rus and U. Neef for running the insect cell facility. **Funding:** S.Z. was supported by an EMBO long-term fellowship (ALTF 830-2018). S.A. was supported by H2020 Marie Curie Individual Fellowship (894862). P.C. was supported by the Deutsche Forschungsgemeinschaft (EXC 2067/1-390729940) and the European Research Council Advanced Investigator Grant CHROMATRANS (grant agreement 882357). **Author contributions:** S.Z. designed and performed all experiments, collected and analyzed cryo-EM data, determined the structure, and carried out biochemical analysis. S.A. assisted with cryo-EM data processing. R.L., S.A., and S.M.V. provided advice. S.M.V. and D.E.A. performed related preliminary experiments. D.E.A. and R.L. provided U1 snRNP. P.C. supervised research. S.Z. and P.C. wrote the manuscript, with input from R.L. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The cryo-EM maps and atomic coordinates were deposited in the Electron Microscopy Data Bank under accession codes EMD-11972 (overall map), 11973 (focused refined map at interface), 11974 (focused refined map of U1 snRNP), and 11975 (map with 155-nt RNA) and in the Protein Data Bank under ID 7B0Y. Materials are available from P.C. upon request under a material transfer agreement with the Max Planck Society.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6526/305/suppl/DC1
Materials and Methods
Figs. S1 to S9
Table S1
References (*31*–*47*)
MDAR Reproducibility Checklist
Movies S1 and S2

View/request a protocol for this paper from *Bio-protocol*.

9 October 2020; accepted 7 December 2020
10.1126/science.abf1870

Science

Structure of a transcribing RNA polymerase II–U1 snRNP complex

Suyang Zhang, Shintaro Aibara, Seychelle M. Vos, Dmitry E. Agafonov, Reinhard Lührmann and Patrick Cramer

Science **371** (6526), 305-309.
DOI: 10.1126/science.abf1870

A tight couple makes messenger RNAs

Gene expression in eukaryotes first requires transcription of DNA to an RNA copy and then splicing to form the final, processed messenger RNA (mRNA). Zhang *et al.* investigated how gene transcription and RNA splicing are physically coupled. Using cryo–electron microscopy, they resolved the molecular structure of a complex of the transcription enzyme RNA polymerase II with part of the splicing machinery, the U1 small nuclear ribonucleoprotein particle. The results provide important details for our understanding of coupled mRNA production.

Science, this issue p. 305