

complexes is uncoupled from small-RNA processing. *Drosophila* is certainly not the only organism with functional differences between its Ago proteins, so it seems very likely that these conclusions will have broad significance. Nonetheless, it is curious that initial reports indicate that at least one species of particular interest—*Homo sapiens*—does not appear to sort siRNAs and miRNAs into distinct subsets of Ago proteins<sup>17,20</sup>. Our understanding of the functional distinctions between the human Ago proteins and their respective small RNA-loading factors await more detailed studies.

## COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

- Zamore, P.D. & Haley, B. *Science* **309**, 1519–1524 (2005).
- Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y. & Zamore, P.D. *Cell* (in the press).
- Tomari, Y., Du, T. & Zamore, P.D. *Cell* (in the press).
- Tomari, Y. & Zamore, P.D. *Genes Dev.* **19**, 517–529 (2005).
- Lee, Y.S. *et al.* *Cell* **117**, 69–81 (2004).
- Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M.C. *Genes Dev.* **18**, 1655–1666 (2004).
- Caudy, A.A., Myers, M., Hannon, G.J. & Hammond, S.M. *Genes Dev.* **16**, 2491–2496 (2002).
- Han, J. *et al.* *Cell* **125**, 887–901 (2006).
- Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W. & Sontheimer, E.J. *Cell* **117**, 83–94 (2004).
- Liu, Q. *et al.* *Science* **301**, 1921–1925 (2003).

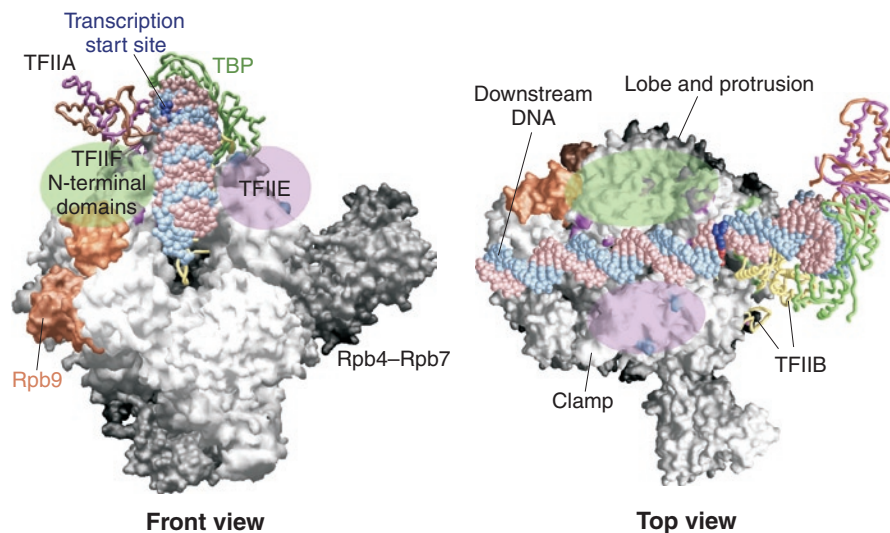
- Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P.D. *Science* **306**, 1377–1380 (2004).
- Liu, X., Jiang, F., Kalidas, S., Smith, D. & Liu, Q. *RNA* **12**, 1514–1520 (2006).
- Pham, J.W. & Sontheimer, E.J. *J. Biol. Chem.* **280**, 39278–39283 (2005).
- Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. & Cohen, S.M. *Cell* **113**, 25–36 (2003).
- Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M.C. *Genes Dev.* **19**, 2837–2848 (2005).
- Rand, T.A., Ginalski, K., Grishin, N.V. & Wang, X. *Proc. Natl. Acad. Sci. USA* **101**, 14385–14389 (2004).
- Liu, J. *et al.* *Science* **305**, 1437–1441 (2004).
- Song, J.J., Smith, S.K., Hannon, G.J. & Joshua-Tor, L. *Science* **305**, 1434–1437 (2004).
- Li, X. & Carthew, R.W. *Cell* **123**, 1267–1277 (2005).
- Meister, G. *et al.* *Mol. Cell* **15**, 185–197 (2004).

## Finding the right spot to start transcription

Patrick Cramer

**How does RNA polymerase II cooperate with initiation factors to locate transcription start sites throughout the genome? A new cross-linking approach reveals previously unknown initiation factor-binding sites on the polymerase surface. The resulting model of the transcription initiation complex suggests that initiation factors cooperate above and inside the polymerase active center cleft to open DNA and find the transcription start site.**

In all living cells, the expression of genetic programs is governed by the dynamic transcriptome, which is established mainly by regulated transcription initiation at gene promoters. Thus, if we want to understand gene regulation, we must know the mechanism of transcription initiation. Initiation by RNA polymerase II (Pol II), which catalyzes eukaryotic messenger RNA transcription, requires the general transcription factors TFIIB, TFIID (containing the TATA box-binding protein, TBP), TFIIE, TFIIIF and TFIIH, which assemble with the polymerase into a preinitiation complex (PIC) at the promoter<sup>1–4</sup>. The structure of the large and complex PIC remains under debate even at a topological level<sup>5</sup>. On page 696 of this issue, new work from Chen *et al.* of the Hahn laboratory<sup>6</sup> provides insights into the PIC architecture and demonstrates an approach for elucidating the topology of a transient multicomponent complex even when only partial structural information is available.



**Figure 1** Topological model of the transcription preinitiation complex according to Chen *et al.*<sup>6</sup>. Figure courtesy of Hung-Ta Chen and Steven Hahn.

Patrick Cramer is at the Gene Center Munich and the Center for Integrated Protein Science (CiPS<sup>M</sup>), Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany.  
e-mail: cramer@lmb.uni-muenchen.de

Pol II can form a stable complex on promoter DNA with only TBP, TFIIB and TFIIIF. In the presence of nucleoside triphosphates, the DNA is melted with the help of TFIIE and TFIIH, mRNA synthesis begins and Pol II escapes from the promoter to form an elongating complex. The initial structure of the Pol II core enzyme revealed an active center cleft that is blocked on one end by a protein wall

and flanked by a mobile clamp on one side and the so-called protrusion and lobe domains on the other<sup>7,8</sup> (Fig. 1). The complete initiation-competent Pol II also contains the Rpb4–Rpb7 subcomplex but has a closed clamp position that renders the cleft too narrow for passage of a DNA duplex<sup>9–12</sup>. This suggests that the clamp is closed during initiation, that the promoter duplex is initially bound above the

cleft and that only the template strand of the DNA, made available by duplex melting, slips inside the cleft and binds in the Pol II active center to initiate RNA synthesis.

Insights into initiation came from structural and biochemical studies of Pol II in complex with initiation factors. The location of initiation factors along promoter DNA was revealed by site-specific DNA-protein cross-linking<sup>13–16</sup>. With the use of photo-cross-linking and radical probing, the N-terminal domain of TFIIB was located at the polymerase dock domain<sup>17</sup> and its C-terminal domain was placed over the polymerase cleft and wall<sup>18</sup>. Crystallographic analysis of Pol II in complex with TFIIB confirmed the location of the TFIIB N terminus at the Pol II dock domain and further showed that a TFIIB 'finger', a region between the N- and C-terminal domains of TFIIB, extends into the active center<sup>19</sup>.

TFIIF consists of two subunits that heterodimerize via their N-terminal regions. EM has been used to elucidate the structure of the Pol II–TFIIF complex. The results suggested that the second-largest TFIIF subunit extends along the cleft, whereas density at the Rpb4–Rpb7 subcomplex was attributed to TFIIF's largest subunit<sup>20</sup>. It is likely that TFIIF, like TFIIB, extends a loop into the active center<sup>18,21</sup>. TFIIF, a large, ten-subunit complex that contains helicase activities involved in DNA melting, is located near downstream DNA<sup>14,15</sup>.

To further elucidate the PIC architecture, all initiation factor-binding sites would ideally be mapped on the polymerase surface. However, this requires the introduction of site-specific cross-linkers in the multiprotein polymerase, which is not easily achieved, as the polymerase is not available in recombinant form. Chen *et al.*<sup>6</sup> have now overcome this difficulty by adopting a genetic system initially developed in the Schultz laboratory<sup>22</sup>, which allowed them to incorporate a non-natural photoreactive amino acid at specific sites on the Pol II surface<sup>6</sup>.

The new cross-linking approach enabled the researchers to ask which factors bind at particular surfaces of the polymerase. Photoactivated cross-linking probes introduced onto the surfaces of the polymerase wall and protrusion domains bound to the TFIIB C-terminal domain, confirming the location of this domain above the cleft and validating the new experimental approach. Probes in the lobe and protrusion domains gave rise to extensive cross-linking to the TFIIF large subunit and one cross-link to the TFIIF small subunit. In

addition, two probes at the top of the clamp cross-linked to the two subunits of TFIIE. A more complete model of the PIC thus emerged. In the PIC model, upstream promoter DNA is anchored by the TBP–TFIIB complex above the polymerase wall, and the downstream DNA runs along the cleft. TFIIF and TFIIE flank the downstream DNA on both sides of the cleft.

The derived locations of the initiation factors elucidate their function in the determination of the transcription start site. In particular, the new data help explain previously described mutational defects in the PIC. Mutations in Pol II that affect selection of transcription start sites mapped to the protrusion and lobe<sup>23,24</sup>. Chen *et al.*<sup>6</sup> found that these mutations alter the affinity of the polymerase for TFIIF, suggesting that they produce a shift in the transcription start site by impairing normal TFIIF function<sup>6</sup>. Mutations leading to similar defects were found in the N-terminal region of the two TFIIF subunits<sup>21,25</sup>, which confer heterodimerization<sup>26</sup>. With the use of radical-generating probes at specific sites in TFIIF, the authors were able to place the N-terminal dimerization module of TFIIF near the Pol II lobe and protrusion<sup>6</sup>.

Comparison of the cross-linking results<sup>6,13</sup> with results from cryo-EM<sup>20</sup> and crystallography<sup>19</sup> reveals obvious differences in the proposed locations of TFIIF subunits and the TFIIB C-terminal domain, respectively. However, all techniques have their limitations. Assignment of low-resolution densities to particular regions in a highly modular initiation factor is difficult. Furthermore, the structure of an initiation factor in a binary complex with Pol II may be incomplete and may differ from the structure adopted in the highly cooperative PIC. Also, cross-linkers may disturb a natural complex. We should therefore expect the PIC model to be modified and extended as more data become available. However, the convergence of data from genetics, cross-linking, functional analysis and protein interaction assays suggests that the new PIC model<sup>6</sup> will withstand the test of time.

In the future, the new cross-linking approach should allow mapping of additional initiation factor-binding sites on the Pol II surface, and it may also identify other proteins that interact with certain Pol II regions. Alternative promoter sequences and TFIIF may be included in these studies to address a puzzling open question, namely how the same (or a similar) PIC can form on highly divergent promoter sequences. On the scale of

a genome, there are no universally conserved core promoter elements<sup>27</sup>, suggesting that the cooperative multiprotein network within the PIC accommodates different sets of initiation factor–promoter interactions or that there are promoter-specific PIC isoforms. During the initiation–elongation transition, the PIC probably undergoes large structural changes, which offer many possibilities for regulation. Elucidating the structural basis for this dynamic transition will remain a long-term challenge that must be tackled with a combination of structural, biochemical and genetic approaches, along the lines followed in the new study.

#### COMPETING INTERESTS STATEMENT

The author declares no competing financial interests.

- Kornberg, R.D. *Trends Cell Biol.* **9**, M46–M49 (1999).
- Lee, T.I. & Young, R.A. *Annu. Rev. Genet.* **34**, 77–137 (2000).
- Orphanides, G. & Reinberg, D. *Cell* **108**, 439–451 (2002).
- Roeder, R.G. *Trends Biochem. Sci.* **21**, 327–335 (1996).
- Asturias, F.J. *Nat. Struct. Mol. Biol.* **11**, 1031–1033 (2004).
- Chen, H.-T., Warfield, L. & Hahn, S. *Nat. Struct. Mol. Biol.* **14**, 696–703 (2007).
- Cramer, P. *et al. Science* **288**, 640–649 (2000).
- Cramer, P., Bushnell, D.A. & Kornberg, R.D. *Science* **292**, 1863–1876 (2001).
- Craighead, J.L., Chang, W.H. & Asturias, F.J. *Structure* **10**, 1117–1125 (2002).
- Armache, K.-J., Kettenberger, H. & Cramer, P. *Proc. Natl. Acad. Sci. USA* **100**, 6964–6968 (2003).
- Armache, K.-J., Mitterweger, S., Meinhart, A. & Cramer, P. *J. Biol. Chem.* **280**, 7131–7134 (2005).
- Bushnell, D.A. & Kornberg, R.D. *Proc. Natl. Acad. Sci. USA* **100**, 6969–6972 (2003).
- Chen, B.S., Mandal, S.S. & Hampsey, M. *Biochemistry* **43**, 12741–12749 (2004).
- Miller, G. & Hahn, S. *Nat. Struct. Mol. Biol.* **13**, 603–610 (2006).
- Kim, T.K., Ebricht, R.H. & Reinberg, D. *Science* **288**, 1418–1422 (2000).
- Bartlett, M.S., Thomm, M. & Geiduschek, E.P. *Nat. Struct. Mol. Biol.* **7**, 782–785 (2000).
- Chen, H.T. & Hahn, S. *Mol. Cell* **12**, 437–447 (2003).
- Chen, H.T. & Hahn, S. *Cell* **119**, 169–180 (2004).
- Bushnell, D.A., Westover, K.D., Davis, R.E. & Kornberg, R.D. *Science* **303**, 983–988 (2004).
- Chung, W.H. *et al. Mol. Cell* **12**, 1003–1013 (2003).
- Freire-Picos, M.A., Krishnamurthy, S., Sun, Z.W. & Hampsey, M. *Nucleic Acids Res.* **33**, 5045–5052 (2005).
- Chin, J.W. *et al. Science* **301**, 964–967 (2003).
- Chen, B.S. & Hampsey, M. *Mol. Cell Biol.* **24**, 3983–3991 (2004).
- Hekmatpanah, D.S. & Young, R.A. *Mol. Cell Biol.* **11**, 5781–5791 (1991).
- Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M. & Ponticelli, A.S. *Mol. Cell Biol.* **24**, 10975–10985 (2004).
- Gaiser, F., Tan, S. & Richmond, T.J. *J. Mol. Biol.* **302**, 1119–1127 (2000).
- Muller, F., Demeny, M.A. & Tora, L. *J. Biol. Chem.* **282**, 14685–14689 (2007).