Yeast Pol II start-site selection: the long and the short of it

The formation of a pre-initiation complex (PIC) is the first step that takes place during transcription by RNA polymerase II (Pol II). The TATA-binding protein binds to the TATA box, thereby positioning the polymerase in the PIC in such a way that it can unwind downstream DNA and lock the template strand in the active site (Bushnell *et al*, 2004). In metazoans, initiation takes place 25–30 base pairs (bp) downstream of the TATA box (Smale & Kadonaga, 2003), whereas in *Saccharomyces cerevisiae*, start sites are spread across 40–120 bp downstream of the TATA box (Hampsey, 1998). The rules that control selection of the Pol II start point are not well understood. In metazoans, start sites often occur at a conserved initiator sequence (Smale & Kadonaga, 2003), whereas yeast initiators are more divergent (Zhang & Dietrich, 2005) and the polymerase has been proposed to ‘scan’ the downstream sequences for optimal start sites (Giardina & Lis, 1993; Kuehner & Brow, 2008). Several articles recently published by the Brow, Lacroute, Libri and Reines laboratories show that start-site selection has an important role in regulating the expression of genes of the *S. cerevisiae* nucleotide-biosynthetic pathways (Jenks *et al*, 2008; Kuehner & Brow, 2008; Kwapisz *et al*, 2008; Thiebaut *et al*, 2008).

IMD2 and *URA2* encode the initial enzymes in the pathways that lead to the synthesis of GTP and UTP, respectively. The transcription of these genes is repressed by the end product NTP (Escobar-Henriques & Daignan-Fornier, 2001; Losson & Lacroute, 1981); however, despite years of searching, no repressors or activators have been identified. Studies on *IMD2* identified a guanine-responsive TATA box and a repressive element spanning the messenger RNA (mRNA) start site, yet how these *cis*-elements regulate *IMD2* expression was unclear (Escobar-Henriques *et al*, 2003; Shaw *et al*, 2001). Transcription profiling of strains disabled for nuclear exosome function revealed the existence of transcripts that originate upstream of

the *IMD2* and *URA2* mRNA start point (Davis & Ares, 2006; Thiebaut *et al*, 2008), and additional work has shown that these transcripts terminate before reaching the coding region (Fig 1; Jenks *et al*, 2008; Kuehner & Brow, 2008). Termination is directed by components of the non-poly(A) termination pathway, and include the RNA-binding proteins Nrd1 and Nab3, the helicase Sen1 and components of the nuclear exosome (Conrad *et al*, 2000; Steinmetz *et al*, 2001; Vasiljeva & Buratowski, 2006). This termination complex recognizes specific sequences in the nascent transcript, and directs its termination and degradation. Deletions or point mutations in the *IMD2* or *URA2* terminators—located upstream of the coding region—result in gene expression even under repressing conditions such as high NTP concentration, showing that the upstream non-coding transcripts are essential for complete repression (Kuehner & Brow, 2008; Kwapisz *et al*, 2008; Thiebaut *et al*, 2008).

What factors determine whether Pol II initiates at the upstream or downstream sites? Presumably, as Pol II scans the sequences downstream of the TATA box, potential start sites pass through the active site (Fig 2). Under GTP-replete or UTP-replete conditions, upstream start sites predominate, whereas depletion of the specific NTP leads to initiation at the downstream site. The upstream *IMD2* transcripts begin with the dinucleotide sequence GG, which led Steinmetz and colleagues to propose that the concentration of initiating nucleotide is an important determinant of start-site selection (Steinmetz *et al*, 2006). Recent results from the Brow laboratory show that mutations from G to A in position +1 result in the production of upstream transcripts in low-guanine conditions (Kuehner & Brow, 2008), which supports this hypothesis. When there is insufficient GTP to form the first phosphodiester bond, as would be necessary in the case of *IMD2*, the polymerase continues to scan downstream, eventually finding the mRNA start site. In the case of *URA2*, the situation is more complex: the upstream transcripts start with an A, which would argue against its regulation by the initiating NTP (Kwapisz *et al*, 2008; Thiebaut *et al*, 2008). However, the first few nucleotides of these transcripts are U-rich, which might trigger pausing and allow time for short abortive transcripts to be released from the PIC during scanning.

Regulatory non-coding RNAs have been described previously in yeast. For example, the non-coding *SRG1* transcript negatively regulates expression from the downstream *SER3* gene by destabilizing the assembly of a PIC on the downstream *SER3* promoter (Martens *et al*, 2004). This is not the case for *URA2*, however, as expression levels of both upstream and downstream transcripts are reduced by the mutation of an upstream TATA box (Kwapisz *et al*, 2008). In addition, Pol II is present at the *URA2* promoter even under repressing conditions (Kwapisz *et al*, 2008). A single TATA box has been identified for *IMD2*; however, mutations of this element reveal the presence of downstream cryptic TATA-like sequences (Kuehner & Brow, 2008). The increase in downstream start sites seen when the regulatory NTP is depleted does not seem to come at the expense of upstream start sites, indicating that either Pol II is always scanning downstream and initiates there only when NTP levels drop or

Fig 1 | Map of the upstream and downstream transcripts that regulate the expression of *IMD2* and *URA2*. mRNA, messenger RNA; ORF, open reading frame.

that there is increased recruitment to the promoter. In the first case, an unknown signal might stabilize the PIC at the downstream start site. Evidence for the second possibility was provided by experiments that showed a fivefold increase in *IMD2* mRNA even when the upstream start sites and terminator were deleted. Therefore, additional regulatory steps that involve the recruitment and/or stabilization of the PIC might be required for the regulation of *IMD2* and *URA2*.

The way in which Pol II scans for an initiation site is unknown. One model proposes that the transcription factor IIH (TFIIH) helicase activity pumps DNA downstream of the TATA box into the active site (Hampsey, 2006; Miller & Hahn, 2006; Roberts, 2006). The central cleft of the polymerase cannot accommodate a transcription bubble of more than 20 bases, and therefore it is thought that the single-stranded DNA loops out of the central cleft at the upstream end of the transcription bubble (Fig 2). This model is consistent with the idea that Pol II maintains contact with the general transcription factors (GTFs) during the scanning process, and with permanganate-footprinting results that show the melting of promoter DNA extending from 20 bp downstream of the TATA box to more than 100 bases downstream (Giardina & Lis, 1993). Although considerable energy is needed to unwind such an extended transcription bubble and to scan for distal start sites, some could be stored in the stressed DNA loops that extend from the polymerase active-site cleft.

In addition to the potential initiator sequences, both Pol II and the GTFs TFIIIF and TFIIIB are important for start-site selection (Hampsey, 2006). Several *rpb1* mutations near the active site reduce upstream start sites and increase downstream start sites, resulting in constitutive expression of *URA2* and *IMD2* (Kuehner & Brow, 2008; Kwapisz *et al*, 2008; Thiebaut *et al*, 2008). Similar results were obtained with mutations in the TFIIIB B finger (Kuehner & Brow, 2008; Thiebaut *et al*, 2008). By contrast, deletion of the non-essential *RPB9* subunit gene results in a shift to upstream start sites, thereby preventing the induction of *IMD2* (Jenks *et al*, 2008). Taken together, these results argue that the PIC acts as a sensor that monitors NTP concentration and uses this information during scanning to find the appropriate start site. Kwapisz and colleagues propose that the Switch 1 loop—located at the downstream end of the transcription bubble—acts as the sensor (Kwapisz *et al*, 2008). This loop is positioned in the DNA-binding channel rather than in the funnel through which NTPs are thought to enter the active site (Cramer *et al*, 2001), although NTPs have also been proposed to enter the active site through the DNA channel (Gong *et al*, 2005), in which case the Switch 1 loop would be well situated to act as an NTP sensor.

The regulation of nucleotide synthesis by start-site selection has previously been observed in bacteria (Sorensen & Neuhaard, 1991; Wilson *et al*, 1992). Transcription from the *PyrC* promoter initiates at a C residue when pyrimidine levels are high; however, these transcripts are not translated owing to a hairpin that forms at the 5' end and blocks translation. When pyrimidine levels are low, transcription initiates at a site several nucleotides downstream, and the shorter mRNA, which is unable to form a stable hairpin, is translated. In principle, this is the same as the mechanism used by *IMD2* and *URA2*; however, in yeast the alternative start sites are much further apart. In addition, the alternative RNA sequences included in the upstream transcript have different downstream targets: the ribosome in the bacterial case and the non-poly(A)-termination machinery for the yeast genes.

Fig 2 | The pre-initiation complex scans the downstream promoter DNA looking for an initiation site (indicated by a question mark). bp, base pairs; mRNA, messenger RNA; PIC, pre-initiation complex; Ter, non-poly(A) terminator.

Given the diversity in yeast start sites, it is not surprising that several of the four articles highlighted here offer preliminary evidence indicating that other yeast genes are regulated by alternative start sites. *URA8* and *ADE12* are crucial enzymes in the synthesis of CTP and ATP, respectively, and both genes have upstream non-coding transcripts (Thiebaut *et al*, 2008). This indicates that the synthesis of all four NTPs is regulated by a similar feedback mechanism. From the genomic analysis of known GTP starts, Kuehner and Brow propose that genes in the yeast ribosome-biosynthetic pathway are also regulated by the availability of GTP. Although there is no evidence of short non-coding RNAs to support this model, the authors note that the regulation by availability of the initiating NTP is similar to the regulation of ribosomal RNA synthesis in *Escherichia coli* (Gaal *et al*, 1997). Further analysis of non-coding RNAs in yeast and other organisms is bound to lead to more examples of regulation by start-site selection.

REFERENCES

- Bushnell DA, Westover KD, Davis RE, Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II–TFIIIB cocrystal at 4.5 Å. *Science* **303**: 983–988
- Conrad NK, Wilson SM, Steinmetz EJ, Patturajan M, Brow DA, Swanson MS, Corden JL (2000) A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **154**: 557–571
- Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**: 1863–1876
- Davis CA, Ares M Jr (2006) Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **103**: 3262–3267
- Escobar-Henriques M, Daignan-Fornier B (2001) Transcriptional regulation of the yeast gmp synthesis pathway by its end products. *J Biol Chem* **276**: 1523–1530
- Escobar-Henriques M, Daignan-Fornier B, Collart MA (2003) The critical cis-acting element required for *IMD2* feedback regulation by GDP is a TATA box located 202 nucleotides upstream of the transcription start site. *Mol Cell Biol* **23**: 6267–6278
- Gaal T, Bartlett MS, Ross W, Turnbough CL Jr, Gourse RL (1997) Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **278**: 2092–2097
- Giardina C, Lis JT (1993) DNA melting on yeast RNA polymerase II promoters. *Science* **261**: 759–762

- Gong XQ, Zhang C, Feig M, Burton ZF (2005) Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol Cell* **18**: 461–470
- Hampsey M (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* **62**: 465–503
- Hampsey M (2006) The Pol II initiation complex: finding a place to start. *Nat Struct Mol Biol* **13**: 564–566
- Jenks MH, O'Rourke TW, Reines D (2008) Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. *Mol Cell Biol* **28**: 3883–3893
- Kuehner JN, Brow DA (2008) Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. *Mol Cell* **31**: 201–211
- Kwapisz M, Wery M, Despres D, Ghavi-Helm Y, Soutourina J, Thuriaux P, Lacroute F (2008) Mutations of RNA polymerase II activate key genes of the nucleoside triphosphate biosynthetic pathways. *EMBO J* **27**: 2411–2421
- Losson R, Lacroute F (1981) Cloning of a eukaryotic regulatory gene. *Mol Gen Genet* **184**: 394–399
- Martens JA, Laprade L, Winston F (2004) Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature* **429**: 571–574
- Miller G, Hahn S (2006) A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nat Struct Mol Biol* **13**: 603–610
- Roberts JW (2006) Biochemistry. RNA polymerase, a scrunching machine. *Science* **314**: 1097–1098
- Shaw RJ, Wilson JL, Smith KT, Reines D (2001) Regulation of an IMP dehydrogenase gene and its overexpression in drug-sensitive transcription elongation mutants of yeast. *J Biol Chem* **276**: 32905–32916
- Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449–479
- Sorensen KI, Neuhaud J (1991) Dual transcriptional initiation sites from the *pyrC* promoter control expression of the gene in *Salmonella typhimurium*. *Mol Gen Genet* **225**: 249–256
- Steinmetz EJ, Conrad NK, Brow DA, Corden JL (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413**: 327–331
- Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell* **24**: 735–746
- Thiebaut M, Colin J, Neil H, Jacquier A, Seraphin B, Lacroute F, Libri D (2008) Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in *S. cerevisiae*. *Mol Cell* **31**: 671–682
- Vasiljeva L, Buratowski S (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell* **21**: 239–248
- Wilson HR, Archer CD, Liu JK, Turnbough CL Jr (1992) Translational control of *pyrC* expression mediated by nucleotide-sensitive selection of transcriptional start sites in *Escherichia coli*. *J Bacteriol* **174**: 514–524
- Zhang Z, Dietrich FS (2005) Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res* **33**: 2838–2851

Jeffry L. Corden is at the *Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.*
Tel: +1 410 955 4719; **Fax:** +1 410 502 6718;
E-mail: jcorden@jhmi.edu

Keywords: initiation; nucleotide biosynthesis; RNA polymerase; termination

Submitted 15 September 2008; accepted 18 September 2008; published online 10 October 2008

EMBO reports (2008) **9**, 1084–1086. doi:10.1038/embo.2008.192