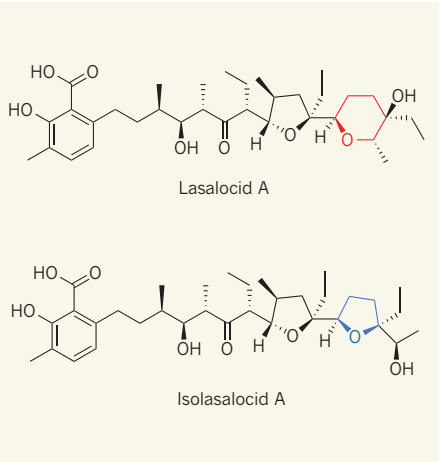

Figure 1 | Polyether antibiotics. The naturally occurring antibiotic lasalocid A is biosynthesized as a mixture with its isomer, isolasalocid A. The structures differ in that the terminal cyclic ether (red) in lasalocid A consists of six atoms, whereas the equivalent ether in isolasalocid A (blue) has five. Hotta *et al.*[3] report the mechanism of formation of the six-membered ether in lasalocid A.

protein corresponding to Lsd19A and a synthetic product analogue carrying two ether rings is bound to the active site of the Lsd19B domain. The structure clearly shows that the active-site cavities of each half of the protein have very similar overall architectures and general catalytic features, but different depths. The consequent differences in active-site shape and volume can reasonably account for the distinct role of each domain in catalysing different stages of the two-step, ether-forming reaction cascade. The high-resolution structure also reveals candidates for the catalytically important acidic and basic residues in the active sites of each protein subunit.

The authors then probed the enzyme in even more depth by carrying out theoretical studies to model the reactions catalysed by Lsd19B, using quantum-mechanical calculations that incorporated the experimentally determined positions of the catalytic residues in the active site of the Lsd19B domain. In particular, they wanted to learn how the enzyme preferentially catalyses the formation of a six-membered cyclic ether, given that a five-membered ring would normally form more easily in the non-enzyme-catalysed reaction (according to well-established rules[10] for predicting the outcomes of ring-formation reactions in solution).

Hotta *et al.* began by modelling ether-ring formation in the presence of either a simple acid or a base, but in the absence of an enzyme. As expected, the theoretical model predicted that the five-membered ring forms more quickly than the six-membered ring, although the preference is smaller in the case of the base-catalysed reaction. Despite this difference in formation rate, the authors concluded that the product containing the six-membered ring is thermodynamically more stable than that containing the five-membered ring, even though it is more difficult to obtain (slower to form).

By contrast, when they modelled ether-ring formation in the presence of the enzyme, the calculations showed that formation of the six-membered ether was favoured. This is a chemical process in which an aspartate amino acid acts as a catalytic base, and two other amino-acid residues act as acids to stabilize the transition state of the reaction. The results show that when enzymes have the choice of catalysing two competing chemical reactions, they can evolve to preferentially accelerate the intrinsically slower transformation, until the otherwise less-preferred product becomes dominant. In other words, enzyme catalysts can channel chemical 'dominoes' to fall in a different direction from that favoured in the absence of the enzyme.

This impressive study solidifies and extends earlier hard-won advances in the study of polyether biosynthesis, and has implications for the formation of many other polyethers. Still unanswered is how the latest findings might apply to the biosynthesis of polyethers that have more than two ether rings, and to the intriguing ladder-like polyethers found in toxins made by marine organisms. A complete picture of polyether biosynthesis will also require a better understanding of how the oxidative enzymes involved in earlier stages of polyether biosynthesis control the three-dimensional structure of the polyepoxides that serve as substrates for the intricate cascade of ether-forming reactions. ■

David E. Cane *is in the Department of Chemistry, Brown University, Providence, Rhode Island 02912-9108, USA.*
e-mail: david_cane@brown.edu

1. Callaway, T. R. *et al. Curr. Issues Intest. Microbiol.* **4,** 43–51 (2003).
2. Liu, T., Cane, D. E. & Deng, Z. *Meth. Enzymol.* **459,** 187–214 (2009).
3. Hotta, K. *et al. Nature* **483,** 355–358 (2012).
4. Cane, D. E., Celmer, W. D. & Westley, J. W. *J. Am. Chem. Soc.* **105,** 3594–3600 (1983).
5. Oliynyk, M. *et al. Mol. Microbiol.* **49,** 1179–1190 (2003).
6. Sun, Y. *et al. Chem. Biol.* **10,** 431–441 (2003).
7. Smith, L., Hong, H., Spencer, J. B. & Leadlay, P. F. *ChemBioChem* **9,** 2967–2975 (2008).
8. Migita, A. *et al. Biosci. Biotechnol. Biochem.* **73,** 169–176 (2009).
9. Minami, A. *et al. Org. Lett.* **13,** 1638–1641 (2011).
10. Baldwin, J. E. *J. Chem. Soc. Chem. Commun.* 734–736 (1976).

GENE EXPRESSION

Transcription initiation unwrapped

A genome-wide, high-resolution study of DNA-binding sites for proteins that transcribe DNA into RNA reveals details about how this process occurs *in vivo*.
SEE ARTICLE P.295

STEPHEN BURATOWSKI

Decades of research using purified molecules *in vitro* have produced a basic understanding of the enzymes and mechanisms that contribute to gene expression in eukaryotes (organisms such as animals, plants and fungi). But these results must be confirmed in living cells, and a productive approach has been to use chromatin immunoprecipitation. This technique reveals the genomic locations of DNA-binding proteins such as those forming nucleosomes — DNA segments wrapped around a histone protein core — and transcription factors. Although these studies have produced a wealth of data, they have not always provided mechanistic insight. However, a paper by Rhee and Pugh[1] on page 295 of this issue sheds light on several questions concerning the initiation of DNA's transcription into RNA.

In chromatin immunoprecipitation (ChIP), cells are first treated with a chemical that crosslinks proteins and DNA. The cells are then disrupted so that their DNA is fragmented. By using specific antibodies, a protein of interest is isolated together with any bound DNA pieces, and these can then be analysed by high-throughput DNA sequencing or other techniques. This allows a quick identification of all the binding sites for a protein, such as those that form protein complexes for transcription initiation, across an entire genome.

Rhee and Pugh have previously reported[2] a modification of ChIP, called ChIP-exo, in which an enzyme removes all DNA except that closest to the protein–DNA crosslink, markedly improving the technique's resolution to a few DNA nucleotides. The results can be quite striking, as illustrated by the excellent correlation that the authors observe in their present study[1] between the ChIP-exo crosslink sites for the transcription factors TBP (TATA-binding protein) and TFIIB, and the protein–DNA

contacts seen in their crystal structures.

The researchers analyse promoters — sequences that specify where to begin the transcription of DNA into RNA — in cells of the yeast *Saccharomyces cerevisiae*, a model eukaryotic organism. Most notable are their findings regarding how the enzyme that synthesizes messenger RNA, called RNA polymerase (Pol) II, is targeted to promoters. TBP is known to recognize the 'TATA box' — a specific DNA sequence found in many promoters — and to position Pol II and its associated factors at the transcription start site (TSS). However, only some promoters, typically those that alternate between repressed and highly active states, contain an obvious TATA box sequence, which represents the optimal TBP-binding site[3].

Surprisingly, Rhee and Pugh's analysis[1] of TBP-binding sites in 'TATA-less' promoters — more prevalent among 'housekeeping' genes that are expressed ubiquitously — reveals this to be a misnomer. These promoters do contain TATA boxes, but their sequences stray from the standard sequence by two or more DNA bases and so their binding to TBP is weaker. This finding echoes those from classic studies[4] on the yeast *HIS3* promoter, which contains two TATA boxes: a weak one for constitutive basal transcription and another, clearly recognizable, for maximal, regulated expression.

Although all Pol II promoters seem to share a common mode of binding to TBP and basal factors, Rhee and Pugh's data[1] help to explain a functional distinction that has been observed[5] between the two classes of promoter. The authors report[1] that, at promoters with obvious TATA sequences, the TSS — and even the initiation complex itself — often overlaps with the first nucleosome in the transcribed region. The expression of genes containing such promoters tends to depend on the presence of the SAGA protein complex[5], which facilitates nucleosome movement, and thus DNA unwrapping, by adding acetyl groups to the nucleosome's histones. Therefore, the first nucleosome probably represses the genes' transcription by blocking their TSS, and gene activation occurs when the histones are removed by targeted acetylation. Once the TSS-containing DNA is unwrapped, efficient binding of TBP, and Pol II and its associated factors, allows the promoter to be expressed at very high levels.

However, most genes have less-obvious TATA boxes, and their expression depends on other proteins known as TBP-associated factors (TAFs), which together with TBP constitute TFIID. *In vitro* studies[6] have shown that TAFs interact with DNA sequences downstream of the TATA box, including sequences around the TSS. These additional contacts may help to compensate for the weaker TBP binding to the DNA, but they probably have other functions. Rhee and Pugh[1] find that, at those promoters to which TFIID preferentially binds, TSSs are located near the upstream boundary of the first nucleosome. Therefore, TAFs may be positioned in such a way that they contact the first nucleosome, preventing it from encroaching on the promoter and thereby allowing basal gene expression. Indeed, some TAFs form a structure resembling the nucleosome histone core[7], suggesting that they might slot into position within an array of nucleosomes.

In addition to invalidating the concept of TATA-less promoters, Rhee and Pugh raise questions about two other recent hypotheses. The first proposes that Pol II promoters are intrinsically bidirectional, that is, a single TATA box can drive transcription in opposite directions. This idea seems plausible because TATA sequences are roughly palindromic, and transcript sequencing studies[8,9] have shown that the TSSs of many mRNAs are close to a non-coding RNA that is transcribed in the opposite direction. However, the authors' ChIP-exo data[1] show that the nucleosome-depleted regions between these divergent TSSs harbour two initiation complexes. In other words, bidirectional transcription is the result of two overlapping but divergent promoters driving transcription in opposite directions, rather than a single promoter that can fire in both directions.

The second hypothesis[10] is that 'gene looping' — the formation of a physical linkage between the beginning and end of active genes — is mediated, in part, by TFIIB. This model is based on observed interactions between mutations in genes that encode TFIIB and 3′-end processing factors (which modify the end of mRNA precursors), as well as ChIP localization of TFIIB (but not the rest of the initiation complex) at transcription-termination regions of selected genes in yeast[10]. However, the present study[1] and another genome-wide ChIP analysis[11] failed to localize TFIIB to 3′ ends, except in the context of initiation complexes at an adjacent promoter. Therefore, the general role of TFIIB in gene looping needs further scrutiny.

It is worth noting that two transcription factors occupying the same genomic location in ChIP experiments may not actually be there at the same time in the same cell, as this technique captures a snapshot of events in a cell population. *In vitro* experiments are therefore needed to probe the kinetics and intermediates of gene expression. Rhee and Pugh[1] use many biochemical and structural studies to inform the interpretation of their ChIP-exo data; the ChIP-exo data, in turn, provide an essential *in vivo* test for *in-vitro*-derived molecular models. This synergism underscores the necessity of applying both approaches to important questions in gene expression. ■

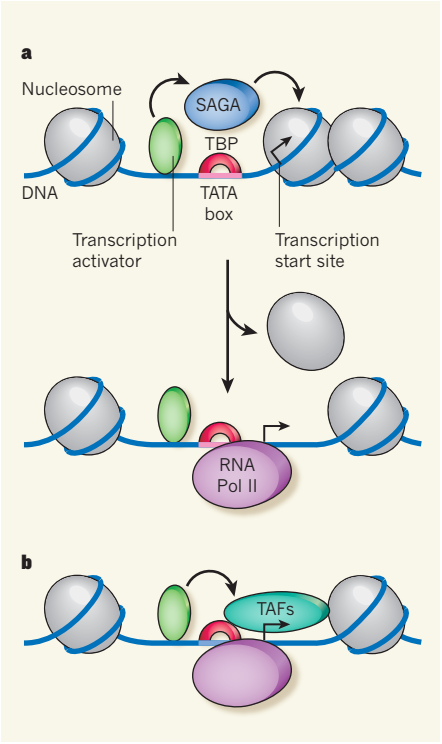

Figure 1 | Two ways of starting the synthesis of messenger RNA. Promoters are sequences within genes that specify where DNA's transcription into RNA starts. **a**, Some promoters have a clearly recognizable sequence, or TATA box, to which transcription factor TBP can bind to recruit the enzyme RNA polymerase (Pol) II, which synthesizes messenger RNA. These promoters can alternate between repressed and active states. In the repressed state (top), a nucleosome (protein–DNA complex) blocks the transcription start site. Transcription activator proteins can then recruit additional proteins, such as the SAGA complex, to trigger nucleosome removal, allowing access to Pol II and therefore activating transcription (bottom). **b**, Most promoters lack a clearly recognizable TATA box, and their expression depends on the presence of transcription factor TFIID, a complex formed by TAF proteins and TBP. Transcription activators could recruit TAFs, which, in turn, might interact with both TBP and a nucleosome to keep the transcription start site accessible.

Stephen Buratowski *is in the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.*
e-mail: steveb@hms.harvard.edu

1. Rhee, H. S. & Pugh, B. F. *Nature* **483,** 295–301 (2012).
2. Rhee, H. S. & Pugh, B. F. *Cell* **147,** 1408–1419 (2011).
3. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. *Cell* **116,** 699–709 (2004).
4. Struhl, K. *Mol. Cell. Biol.* **6,** 3847–3853 (1986).
5. Huisinga, K. L. & Pugh, B. F. *Mol. Cell* **13,** 573–585 (2004).
6. Verrijzer, C. P., Chen, J. L., Yokomori, K. & Tjian, R. *Cell* **81,** 1115–1125 (1995).
7. Selleck, W. *et al. Nature Struct. Mol. Biol.* **8,** 695–700 (2001).
8. Neil, H. *et al. Nature* **457,** 1038–1042 (2009).
9. Xu, Z. *et al. Nature* **457,** 1033–1037 (2009).
10. Singh, B. N. & Hampsey, M. *Mol. Cell* **27,** 806–816 (2007).
11. Mayer, A. *et al. Nature Struct. Mol. Biol.* **17,** 1272–1278 (2010).