MILESTONE 11

Getting specific

Although our understanding of gene expression is much more complex these days, the 'central dogma' — namely that DNA encodes RNA encodes protein — remains a valid principle. Back in the 1970s, proteins were known to be expressed in a highly regulated fashion in response to specific signals. How this regulation is achieved and how the signals are relayed has kept molecular biologists busy ever since. Regulated gene transcription by *trans*-acting sequence-specific transcription factors that bind *cis*-regulatory DNA sequences to regulate transcription by RNA polymerase (pol) I, II or III (see Milestone 7) was the first level of regulation to emerge. Today, this stands as a near universal mode of gene regulation.

The first detailed mapping of DNA sequences bound by a transcription factor was published in 1978. Tjian used an adenovirus–simian-virus-40 (SV40) hybrid, which supported higher expression of a protein functionally similar to SV40 T antigen in a system more amenable to purification. He showed that the purified protein bound in a sequential manner to tandem recognition sequences at the SV40 replication origin, which turned out to overlap with the promoter. Tjian drew comparisons to phage lambda-repressor DNA binding, characterized a few years earlier by Ptashne and colleagues (see Milestone 2). He noted that the binding region contained palindromic stretches now emblematic of many other transcription-factor-binding sites, and that the protein probably bound as a multimer. Notably, large T-antigen binding to this region had been shown previously by electron microscopy.

Around the same time, Roeder observed that although RNA pol III alone did not allow 5S RNA gene transcription, *Xenopus laevis* oocyte extracts contained an activity that allowed accurate transcription. Roeder and colleagues purified the activity — TFIIIA — in 1980, showed specificity for 5S genes and DNA binding independent of pol III, and mapped the promoter binding site by footprinting.

The following year, Yamamoto and colleagues reported specific *in vitro* binding of the glucocorticoid receptor to a 4.5-kb fragment of the mammary tumour virus, which they had shown previously to mediate hormone-responsive transcription. One year later, McKnight and Kingsbury were the first to apply a 'linker-scanning' mutagenesis approach to produce a detailed map of the promoter region of the viral thymidine kinase gene. In 1983, Dynan and Tjian isolated the human transcriptional activator Sp1 using an *in vitro* transcription assay, and located its binding sites upstream of the transcription-initiation site.

The next breakthrough came from experiments by Brent and Ptashne establishing the modular nature of transcription factors. Using a hybrid fragment that contained a DNA-binding region of the prokaryotic transcriptional repressor LexA fused to a DNA-binding-deficient fragment of the yeast transcriptional activator Gal4, they showed that this chimaera supported transcription only from yeast promoters containing the *lexA* operator. Remarkably, this operator could drive expression even when located downstream of the promoter.

MILESTONE 12

The right tools for the job

Nowadays, a chef does not pass muster unless he or she can wield a state-of-the-art, hand-held gas torch. Equally, scientists need the right tools to further their field of research. One such tool in the field of transcription was the development, by the Roeder laboratory in 1979, and the Sharp and Gefter laboratories in 1980, of an *in vitro* assay for promoter-specific transcription by RNA polymerase II (pol II).

At the time, it was not clear that eukaryotic pol II could accurately initiate transcription *in vitro*. However, using a well-characterized adenovirus promoter, the groups were able to show selective and accurate transcription initiation in the presence of either crude cell extracts with purified pol II or whole-cell extracts alone. This assay opened up the field for the biochemical characterization of the eukaryotic transcription machinery — by adding or subtracting specific fractions, the factors required to reconstitute transcription could be identified.

Indeed, soon thereafter, using this assay, it was discovered that numerous activities were required for transcription initiation, of which at least one could recognize the TATA-box region of promoters before pol II binding. These activities were linked to a set of 'general transcription factors', which turned out to be essential for transcription initiation, and assisted pol II to recognize, bind to and initiate transcription from the core regions of many promoters. Biochemical studies of transcription systems derived from HeLa cells, rat liver, fruitflies and yeast were carried out by the Chambon, Egly, Roeder, Reinberg, Conaway, Parker, Kornberg, Greenleaf and Tjian laboratories, among others, during the late 1980s and early 1990s. It is often forgotten that proof of the existence of general transcription factors depended on their, sometimes painstaking, isolation and the reconstitution of pol II transcription with purified proteins.

"This advance paved the way for all subsequent biochemical studies of eukaryotic transcription."
Michael Green

In 1991, Joan and Ronald Conaway and colleagues were the first to accomplish this feat using the rat liver system, followed by Roger Kornberg and co-workers in the yeast system in 1992, and subsequently by others in the fruitfly and human systems. The Conaway team fractionated rat liver and identified five distinct enzyme fractions that were essential for specific transcription: α, βγ, δ, ε and τ. Four of these could be replaced by purified proteins from rat liver: TFIIB (α), TFIIF (βγ), TFIIE (ε) and TFIIH (δ). With the purification of the final factor, TFIIE, and having in hand purified pol II, TFIIB, TFIIF and TFIIH, as well as recombinant TATA-binding protein (TBP), it was possible to show that promoter-specific transcription could be reconstituted *in vitro* with purified factors. Crucially, this allowed the definition of the minimal transcriptional machinery required for promoter-specific transcription by pol II.

Incidentally, TFIID in higher eukaryotes was notoriously difficult to purify, but recombinant yeast TBP — which had been

Arguably, sequence-specific transcription factors constitute the most important and diverse gene-regulatory mechanism. The combinatorial diversity afforded by transcription-factor binding is an effective means of coordinating the regulation of complex sets of genes. Moreover, signalling cascades regulate gene expression predominantly by modulating transcription-factor activity.

With transcription factors, the 'central dogma' had come full circle: protein regulates gene, hence message and protein.

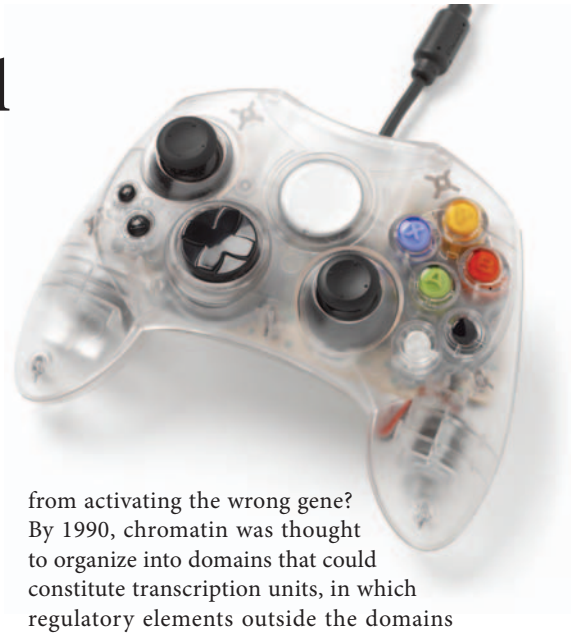
Bernd Pulverer, Editor, Nature Cell Biology

References and links

ORIGINAL RESEARCH PAPERS Tjian, R. The binding site on SV40 DNA for a T antigen-related protein. *Cell* **13**, 165–179 (1978) | Engelke, D. R., Ng, S.-Y., Shastry, B. S. & Roeder, R. G. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* **19**, 717–728 (1980) | Payvar, F. *et al*. Purified glucocorticoid receptors bind selectively *in vitro* to a cloned DNA fragment whose transcription is regulated by glucocorticoids *in vivo*. *Proc. Natl Acad. Sci. USA* **78**, 6628–6632 (1981) | McKnight, S. L. & Kingsbury, R. Transcriptional control signals of a eukaryotic promoter-encoding gene. *Science* **217**, 316–324 (1982) | Dynan, W. S. & Tjian, R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**, 79–87 (1983) | Brent, R. & Ptashne, M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729–736 (1985)

MILESTONE 13

Remote control



Although studies in prokaryotes have identified promoter (see Milestone 5) and operator DNA elements involved in transcription regulation, eukaryotes have markedly different mechanisms for this process. In 1980, it was known that an AT-rich region (the TATA box) ~30 base pairs (bp) upstream of a transcription-initiation site could function as a promoter *in vitro*. However, removal of the TATA box in the simian virus 40 (SV40) early genes did not abolish protein expression *in vivo*. Instead, transcription of these genes required ~200 bp of DNA upstream of the TATA box, including three tandem GC-rich 21-bp repeats and two 72-bp repeats. Benoist and Chambon showed that the presence of at least one of the two 72-bp repeats is necessary for expression of SV40 early genes, indicating that gene expression in eukaryotes can be influenced by remote DNA elements.

Is regulation by remote elements unique to the SV40 early genes? Studies during the early 1980s showed that other sequences upstream of the TATA box of certain genes were essential for expression. In 1981, Schaffner and colleagues provided further evidence that regulation by remote elements might be a general phenomenon. They showed that the SV40 72-bp repeats, which they called 'enhancers', could drive the expression of the heterologous rabbit haemoglobin β1 gene in HeLa cells. In addition, these enhancers could exert their effect even when placed thousands of base pairs upstream or downstream of the transcription-initiation site, independent of the orientation of the enhancer.

Using the native SV40 system, Fromm and Berg confirmed the long-range effect of the enhancers and showed that DNase I hypersensitivity, which occurs within the SV40 enhancer and is an indicator of open chromatin structure, was introduced into the sites where the enhancers were moved. These studies solidified the idea of long-range transcription regulation in eukaryotes, extended the reach of the remote regulatory elements and implicated open chromatin structures in the activity of the enhancers.

The surprising observation that remote enhancers could affect transcription on either side of a gene, irrespective of their orientation, was not readily accepted in the early 1980s. It took a decade of further analysis for this idea to take hold, which then raised another question: what prevents an enhancer from activating the wrong gene? By 1990, chromatin was thought to organize into domains that could constitute transcription units, in which regulatory elements outside the domains have no effect on the gene activity within them. To test this idea, Kellum and Schedl developed an assay using the *Drosophila melanogaster* heat-shock-gene boundary elements, scs and scs′. They showed that heterologous gene constructs enclosed within scs and scs′ are insulated from positive and negative regulatory effects of surrounding elements. Notably, scs and scs′ themselves do not have either positive or negative regulatory activity. This study established a functional definition for 'insulators', and provided a link between chromatin domains and transcription regulation.

The composite proximal and distal *cis*-regulatory elements are essential for combinatorial transcriptional regulation. This type of regulation is particularly important for complex organisms, in which diverse gene-expression patterns that define the many different cell types are driven by the mixing and matching of transcription factors. Much is now known about the transcriptional complexes that bind to some of these regulatory elements. Nevertheless, we still do not have a complete picture of how these elements communicate over long distances to affect transcription.

Hwa-ping Feng, Senior Editor,
Nature Structural & Molecular Biology

References and links

ORIGINAL RESEARCH PAPERS
Benoist, C. & Chambon, P. *In vivo* sequence requirements of the SV40 early promoter region. *Nature* **290**, 304–310 (1981) | Banerji, J., Rusconi, S. & Schaffner, W. Expression of β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981) | Fromm, M. & Berg, P. Simian virus 40 early- and late-region promoter functions are enhanced by the 71-base-pair repeat inserted at distant locations and inverted orientations. *Mol. Cell. Biol.* **3**, 991–999 (1983) | Kellum, R. & Schedl, P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941–950 (1991)

isolated and cloned by the Sharp and Guarente groups, and the Chambon group, in 1988 — could substitute for TFIID in certain circumstances.

Defining the minimal set of transcription factors required for pol-II-mediated transcription paved the way for subsequent studies of basal and activated transcription — experiments that would not have been possible without the right tools.

Arianne Heinrichs, Chief Editor,
Nature Reviews Molecular Cell Biology

References and links

ORIGINAL RESEARCH PAPERS Weil, P. A., Luse, D. S., Segall, J. S. & Roeder, R. G. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. *Cell* **18**, 469–484 (1979) | Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. *Proc. Natl Acad. Sci. USA* **77**, 3855–3859 (1980) | Conaway, J. W., Hanley, J. P., Pfeil Garrett, K. & Conaway, R. C. Transcription initiated by RNA polymerase II and transcription factors from liver: structure and action of transcription factors ε and τ. *J. Biol. Chem*. **266**, 7804–7811 (1991) | Conaway, R. C., Pfeil Garrett, K., Hanley, J. P. & Conaway, J. W. Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors α and βγ promote entry of polymerase into the preinitiation complex. *Proc. Natl Acad. Sci. USA* **88**, 6205–6209 (1991)

FURTHER READING Buratowski, S., Hahn, S., Sharp, P. A. & Guarente, L. Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* **334**, 37–42 (1988) | Cavallini, B. *et al*. A yeast activity can substitute for a HeLa cell TATA box factor. *Nature* **334**, 77–80 (1988) | Sayre, M. H., Tschochner, H. & Kornberg, R. D. Reconstitution of transcription with five purified initiation factors and RNA polymerase II from *Saccharomyces cerevisiae*. *J. Biol. Chem*. **267**, 23376–23382 (1992)