Gene expression

Protein contacts for promoter location in eukaryotes

from Andrew Travers

WHAT is the mechanism by which RNA polymerase locates a promoter site? In eubacteria, location requires the direct interaction of RNA polymerase with conserved sequences adjacent to the transcription start point. However for eukaryotes the potential for such a strategy of promoter identification must be severely constrained by the size of the genome. In a mammalian nucleus the concentration of a unique DNA-binding site is $\sim 10^2$ times per genome lower than in *Escherichia coli* while the potential number of random DNA-binding sites is 10^3–10^4 times greater[1]. These factors imply first that in a eukaryote, a DNA-binding protein must have 10^2–10^3-fold higher affinity for its binding site than would be necessary in *E. coli* — assuming all DNA to be equally accessible; and second, that to achieve efficient occupancy the sequence specificity of the interaction relative to random binding must be 10^3 times greater than for *E. coli*. The dilemma posed is made apparent by the i^{X86} mutation of the *E. coli lac* repressor which increases the affinity of the repressor for DNA, at both random and specific sites, by $\sim 10^2$. This repressor binds so tightly to random DNA that its efficiency in locating the operator and thereby repressing *lac* expression is severely impaired[2].

In eubacteria the ability of RNA polymerase to recognize promoter sequences is conferred by the core polymerase binding to the σ polypeptide independently of DNA binding. Thus in *Bacillus subtilis* replacement of the major vegetative σ factor by analogous polypeptides during endospore formation allows the polymerase to bind to different classes of promoter with different conserved sequences[3]. For eukaryotes one solution to the dilemma posed by an excess of irrelevant DNA-binding sites would be for a protein that can direct promoter-specific transcription to bind to the promoter site independently of the polymerase. The transcribing enzyme could then itself bind directly to the promoter-recognition protein. This strategy of promoter location could be common to cell types with widely varying amounts of accessible DNA, for example rapidly growing and highly differentiated cells, and would substitute a primary protein–protein recognition for the protein–DNA recognition used by eubacteria. The promoter-recognition proteins in eukaryotes would, like σ factors, be able to interact with both RNA polymerase and with specific sequences in the promoter site[4,5].

There is already substantial evidence that the DNA sequences defined *in vivo* as necessary for the activity and regulation of RNA polymerase II (or B) and RNA polymerase III promoters do not simply direct polymerase binding. Purified eukaryotic nuclear RNA polymerase behaves like the bacterial core polymerase in initiating on any single-stranded region of DNA without apparent sequence preference[6]. Moreover, in the case of polymerase II promoters, Davison *et al.*[7] have recently shown that the core promoter region containing the TATA box common to most polymerase II promoters binds in the absence of polymerase itself at least one of the several factors necessary for accurate initiation *in vitro*. This binding depends on the integrity of the TATA box and results in the formation of a stable preinitiation complex. Similarly, activity of promoters used by polymerase III for the initiation of 5S RNA, but not of tRNA or other viral polymerase III transcripts, depends on the binding of a protein, TFIIIA, to an internal region of the gene determining transcriptional activity and regulation[8,9].

Although the experiments of Davison *et al.* suggest that a specific DNA-binding protein is necessary for promoter identification by polymerase II they leave open the question of how the ability to distinguish between different promoters is achieved. The first strong evidence that selectivity can be obtained *in vitro* has recently been provided by Dynan and Tjian[10] who have partially purified a transcription factor that is required for transcription from the simian virus 40 (SV40) early and late promoters but not from other polymerase II promoters tested. It is not yet known whether the activity of this promoter-selecting factor depends on the upstream control sequences specific to the SV40 early promoter but it seems highly probable that the selectivity observed *in vitro* reflects sequence differences between the promoters tested.

How might such a selectivity factor act? One possibility is that a specific DNA-binding protein recognizes the upstream promoter element and interacts by protein–protein contact with either or both the polymerase and the promoter-recognition protein. Efficient binding of the polymerase at the promoter site would require either sequential or simultaneous interaction with both these protein elements. In this way the binding specificity of the promoter-recognition protein could be bootstrapped to provide high transcriptional selectivity using protein–protein contacts for the primary interaction.

A major consequence of invoking protein–protein contact as the primary mechanism for promoter location by RNA polymerase is that regulation of those interactions could be achieved by altering the protein–protein contact, for example by covalent modification. Such modification would allow, for example, polymerase II to interact with one set of selectivity factors and exclude interaction with a second set. In the case of a gross environmental perturbation, modification might permit binding of polymerase to only one class of gene. Such mechanisms could explain the switching off of transcription of other genes during the heat-shock response in *Drosophila*[11] or the constitutive expression of the major human heat-shock gene in the presence of the adenovirus E1a protein[12]. There are also hints that modulation of RNA polymerase II interactions may alter the developmental phenotype of particular tissues. For example a class of RNA polymerase II mutants isolated in rat myoblasts fail to undergo terminal differentiation[13].

This mechanism also has evolutionary implications. For any organism which utilises high-specificity DNA–protein interactions as the primary mode of regulating transcription, substantial increases in DNA content which are not accompanied by compensating increases in the content of specific DNA-binding proteins could be viewed as selectively disadvantageous by decreasing the efficiency of the response to environmental pertubations. For such organisms there would be a selective premium in maintaining a compact efficiently organized genome. By contrast, in eukaryotes, a reliance on protein–protein interactions as the major determinant of transcriptional control would free an organism from the deleterious effects of rapid expansion of the genome and thereby minimize the DNA load caused by the increased amount of competing DNA. Thus the potential for evolutionary flexibility would be increased. Nevertheless, a regulatory mode depending on the modification of multiple protein–protein interactions might respond less quickly to environmental perturbation than one depending on direct alteration of protein–DNA interactions, and consequently would be at a selective disadvantage in niches favoured by many eubacteria. □

Andrew Travers is at the Laboratory of Molecular Biology, MRC Centre, Hills Road, Cambridge CB2 2QH.

1. Linn, S. & Riggs, A.D. *Cell* **4**, 107 (1975).
2. Pfahl, M. *J. molec. Biol.* **106**, 857 (1976).
3. Johnson, W.C., Moran, C.P. & Sosick, R. *Nature* **302**, 800 (1983).
4. Simpson, R.B. *Cell* **18**, 277 (1979b).
5. Chenchick, A., Beabealashvilli, R. & Mirzabekov, A. *FEBS Lett.* **128**, 46 (1981).
6. Lewis, M.K. & Burgess, R.R. *J. biol. Chem.* **255**, 4928 (1980).
7. Davison, B.L. *et al. Nature* **301**, 680 (1983).
8. Engelke, D.R. *et al. Cell* **19**, 717 (1980).
9. Pelham, H.R.B. & Brown, D.D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4170 (1980).
10. Dynan, W.S. & Tjian, R. *Cell* **32**, 669 (1983).
11. Ashburner, M. & Bonner, J.J. *Cell* **17**, 241 (1979).
12. Nevins, J. *Cell* **29**, 913 (1982).
13. Pearson, M.L. & Crerar, M.M. in *Molecular and Cellular Control of Muscle Development* (Cold Spring Harbor, New York, in the press).

0028-0836/83/260755-02$01.00