Molecular Biology I: Transcriptional Regulation

Dr Nick Plant

Centre for Toxicology, Faculty of Health and Medical Sciences, University of Surrey, Guildford,
Surrey, GU2 7XH, UK

Email: N.Plant@Surrey.ac.uk

Nick Plant is a Senior Lecturer in Molecular Toxicology within the Centre for Toxicology at the University of Surrey, UK. His research interest include the use of systems modelling to understand the role of nuclear receptors in co-ordinating body responses to therapeutic drugs

Words (Total): 2103

Words (Abstract): 164

Figures: 1

Tables: 0

Keywords

Transcription; PCR; TaqMan; microarray; Quantitation; nuclear run on; EMSA

Conflicts of Interest

None declared

Abstract

Whereas the DNA of the cell may be envisaged as the blueprints for a human, it is the processes of transcription and translation that act as the production line to convert these blueprints into the active protein required to produce the complex biological interactions that allow a cell to function. At its very simplest level the process of transcription is concerned with the recruitment of RNA polymerase II (Pol II) to the transcription start site of the target gene: The faster this process occurs then the more transcripts can be produced in a given amount of time and hence the more mRNA is produced to be converted to protein.

In the current article we will examine the techniques available for examining the DNA:protein interactions within the regulatory regions of target genes, which result in the increased recruitment of RNA polymerase II. In addition, we will examine those techniques designed to measure the next stage, the transcription of gene coding regions to make mRNA transcripts.

General Introduction

The rate of Pol II recruitment to the transcription start site of a target gene is determined by the action of transcription factors; proteins that interact with specific sequences within the regulatory regions of a gene and act as a molecular bridge to grab Pol II and recruit it to the transcription start site (Figure 1). These transcription factors may be generic (e.g. Sp1, TBP), with their mere presence being enough to encourage recruitment, or they may respond to the presence of an activating chemical (e.g. PXR, PPARs) before they encourage recruitment. These two systems represent the *generic* and *responsive* modes of gene expression, with the former being important for basal expression of genes, whereas the latter is central to the body's response to altered chemical levels, such as is seen during pathophysiology or therapeutic treatment (Gibson and Skett 2001; Plant 2003). To be able to understand gene expression it is therefore necessary to be able to measure the protein:DNA interactions that occur within gene regulatory regions, as well as the resultant increase in mRNA transcript production rate.

DNA:Protein Interactions

The prerequisite for DNA transcription is the interaction of proteins with the regulatory regions of the target gene. As described above this can be as simple as the recruitment of Pol II to the transcription start site, to the complex interaction of multiple proteins within a regulatory region that ultimately results in the efficient recruitment of Pol II. A simple assay for such DNA:Protein interaction is the Electromobility Shift Assay (EMSA), alternatively referred to as the gel retardation assay. There is an obvious difference in both mass and bulk for protein-free DNA versus protein-bound DNA. Hence, when separated by electrophoresis they will migrate at different rates through the gel, and this is the principle behind EMSA. Initially, a short region of DNA (20-50bp) that you think your protein may bind to is selected and labelled for easy detection; common labels include radioactivity, fluorophores or biotinylated colourimetric markers. This DNA is then mixed with the protein to allow interaction and following incubation for 30 minutes, loaded onto a polyacrylamide

gel for separation by electrophoresis. After electrophoresis, the DNA is detected and its movement relative to protein-free DNA observed; protein:DNA interactions will cause the DNA to be 'retarded' and move less distance through the gel. However, it is important to demonstrate that any observed interaction is actually specific and not the result of a 'sticky-protein', which would bind to any piece of DNA regardless of the biological relevance. To check for specificity a competition assay is carried out, whereby excess unlabelled probe DNA is added to the labelled DNA/protein mix during the incubation step. If the interaction is specific then binding to both the labelled and unlabelled probe DNA will be equally likely; as the unlabelled DNA is in excess this will effectively compete out the labelled DNA and the retarded band will disappear. In comparison, if the interaction is non-specific then the band will not be competed out as efficiently by the unlabelled probe DNA and the retarded DNA will remain (Aouabdi et al. 2006). A final check for specificity that is often carried out is the so-called super-shift assay. Here an antibody to the target protein is also added to the labelled DNA and protein in the incubation mix. Obviously, the interaction of all three will result in an even larger complex that will be even further retarded in the gel, and as the antibody is specific for the target protein this super-shifted band will only appear if the interacting protein is indeed the protein of interest.

Transcriptional Activation

Whereas EMSA will show you that protein is interacting with DNA it will not tell you if that interaction results in an increase in transcription. To achieve this we need to use other techniques such as reporter gene assays and chromatin immunoprecipitation (ChIP).

For a reporter gene assay it is first necessary to attach the regulatory region of interest (up to several kilobases of DNA) upstream of an easily measurable reporter gene, such as fluorescent proteins or enzymes which convert chemicals for subsequent colourimetric/luminescent detection. This construct is contained within a plasmid, a small circular piece of self-replicating DNA, which can be transfected into a mammalian cell line of choice for your assay. The fact that this assay occurs

within a mammalian cell line is a major advantage over the EMSA as it puts the experiment into an environment that is much closer to the *in vivo* situation and hence increases confidence in the answer produced. In the basic assay, your reporter construct is transfected into the cell line of interest and then transcription stimulated through the addition of the protein of interest, a stimulating chemical, or a combination of both. Following stimulation for 12 to 72 hours the protein coded for by the reporter gene is measured, with an increase in production being indicative of transcriptional activation (El-Sankary et al. 2001).

The activation of transcription within a reporter gene assay is clearly indicative of the potential for transcription to occur within a cell. However, within a cell DNA exists as chromatin, where it is bound around nucleosomes, a octuplet of histone proteins that act to protect the DNA from degradation. In addition to protecting the DNA, histones also regulate transcription; simply put, when DNA is wound tightly around histones there is not space for transcription factors to bind to the regulatory regions of genes and initiate DNA. This level of control is extremely important in setting the overall gene expression profile within a cell and hence should be taken into account when studying transcription. Reporter gene assays use naked DNA in a plasmid that does not have any histones bound, and hence do not provide the full story. By comparison, ChIP studies transcription in the genomic DNA, with its full complement of histone binding. As such ChIP offers many advantages over reporter gene assays, although this does come at the cost of an increase in technical difficulty. In ChIP, interacting proteins are first fixed to genomic DNA using formaldehyde and then the genomic DNA broken into small (<500bp) fragments, usually by sonication. Using an antibody to the protein of interest any DNA bound to this protein is isolated through immunoprecipitation, followed by reversal of the formaldehyde cross links to free the DNA. Each fragment of isolated DNA represents a piece of DNA to which the target protein was bound and through sequencing or PCR these regions can be identified, telling you where in the genome your protein bound to the DNA. Carrying out this analysis for several different proteins, including histones, generic and responsive

transcription factors, as well as Pol II itself, lets you build up a picture of the interactions occurring in regulatory regions of genes *in vivo*.

mRNA Transcript Measurements

The previous techniques have examined the protein:DNA interactions required to initiate transcription. We shall now consider those technologies that look purely at the products of transcription and measure the level of mRNA transcripts. There are three major current technologies for measuring the level of mRNA transcript production within the cell: (Q)RT-PCR, Nuclear run on and DNA microarray. As we will see the first two measure the levels of a single mRNA transcript, whereas the latter allows measurement of thousands of different mRNA transcripts in parallel, allowing information on the transcription rates within a whole cell to be derived (the transcriptome). Polymerase chain reaction (PCR) is a simple technique using primers specific to a target region of DNA to achieve its exponential amplification. If PCR is coupled with an initial conversion of RNA into DNA, through the use of a reverse transcriptase enzyme, then it is capable of measuring the levels of RNA within a cell (RT-PCR). Finally, if this measurement is undertaken in a robust manner and compared to a standard curve of known amounts of target RNA/DNA then this will produce a fully quantitative measurement of transcript levels within the cell: QRT-PCR, often shortened to Q-PCR. For Q-PCR there exists two commonly used detection systems, Sybr-green and TaqMan. Sybr-green is a fluorescent chemical that intercalates into the major groove of DNA in a stoichiometric fashion, meaning that the level of fluorescence is directly related to the level of DNA in the solution. The benefit of this approach is its relative low cost, but as Sybr-green binds to DNA non-specifically then extra care must be taken to ensure that you are only measuring the level of your target transcript and not of anything else. In comparison the 5' nuclease assay (TaqMan) uses a specific probe against the target transcript that contains a fluorophore and quencher. Each amplification round degrades the probe, freeing the fluorophore in a manner directly proportional to the level of target transcript. Through the use of a specific probe, TaqMan has the benefit of guaranteed high specificity towards

your target transcript, but this increased specificity comes at an increased cost per reaction (Plant 2003).

One potential issue with Q-PCR is that it measures the size of the mRNA transcript pool. As the size of the pool is determined by both the rate of mRNA production and degradation it is not a direct measurement of transcriptional activation. It is possible, for example, for a stimulus to decrease the rate of mRNA degradation, which would increase the size of the mRNA transcript pool without altering transcription rates (El-Sankary et al. 2001). The nuclear run-on assay is designed to study just the rate of mRNA transcript production and as such gets around this potential problem. However, it is technically quite challenging and hence many people just use Q-PCR and accept the limitations. In nuclear run-on assays nuclei are first extracted and then incubated with a labelled nucleotide (usually ^{32}P -UTP) to produce labelled *de novo* mRNA. This mRNA is then hybridised against a macroarray (<100 spots) of probes specific to target mRNA transcripts, washed to remove non-specific binding and quantified by densitometry. Each band represents the level of *de novo* mRNA transcript production for the target transcript, and is reflective of transcription rates alone.

Whereas both TaqMan and nuclear run-on assays look at the level of a single mRNA transcripts DNA microarrays are capable of measuring the level of *every* mRNA transcript within the cell simultaneously. Such technology has the power to examine whole transcriptome differences between normal and tumour tissue (Ramaswamy et al. 2001), normal and chemically stimulated tissue (Pogribny et al. 2007) or for mapping mRNA transcript levels across a large range of samples/tissues (Plant et al. 2006). Each microarray is comprised of thousands of spots of DNA, each of which is specific for a target mRNA transcript. RNA is extracted from the cell/tissue of interest and then labelled with a fluorescent dye for later detection. This mixture of labelled mRNAs is then hybridized to the microarray, and brighter fluorescence for a given spot reflects more of that mRNA transcript being present in the original RNA mixture. Such techniques can be performed as one- or two-colour arrays: In a one-colour array RNA from both normal and stimulated cells are labelled with

the same fluorescent marker and then hybridised to different microarrays. Each is quantified separately and the results compared to see differential expression. In a two-colour system, normal and stimulated cells are mixed with different fluorescent markers, usually red and green, and then hybridised to the same microarray. If an mRNA transcript is present at the same level in both normal and stimulated cells then the resultant spot will be orange; if more mRNA is present in the normal sample then the green fluorophore will be predominant and the spot will be green; more mRNA present in the stimulated sample will produce a red spot. In essence, both the one- and two-colour systems provide the same answer, the relative expression of a mRNA transcripts in one sample compared to another, but use subtly different procedures to achieve this.

Conclusions

For the body to function, and for it to be able for it to respond to the alterations away from homeostasis caused by pathogenesis or chemical exposure, it is vital that there is a flexible, responsive system to allow the expression of proteins as and when required by the cell. Control of transcription is the first stage of this response network, whereby information from the body's blueprints (DNA) is turned into the mRNA transcripts required to produce the basic machine parts for body functioning, the proteins. The next stage in this response system is the conversion of these individual proteins into the complex, multi-protein machines required to make the body run, which will be examined in the next article.

References

- Aouabdi, S., G. G. Gibson and N. Plant. Transcriptional regulation of PXR: Identification of a PPRE within the proximal promoter responsible for fibrate-mediated transcriptional activation of PXR. *Drug Met. Disp.* 2006; **34**(1): 138-144.
- El-Sankary, W., G. G. Gibson, A. Ayrton, et al. Use of a reporter gene assay to predict and rank the potency and efficacy of CYP3A4 inducers. *Drug Met. Disp.* 2001; **29**(11): 1499-1504.
- Gibson, G. G. and P. Skett. *Introduction to Drug Metabolism*. Cheltenham: Nelson Thornes 2001.
- Plant, K. E., D. M. Everett, G. Gordon Gibson, et al. Transcriptomic and phylogenetic analysis of Kpna genes: a family of nuclear import factors modulated in xenobiotic-mediated liver growth. *Pharmacogenet. Genomics* 2006; **16**(9): 647-658.
- Plant, N. *Molecular Toxicology*. London: BIOS 2003.
- Pogribny, I. P., T. V. Bagnyukova, V. P. Tryndyak, et al. Gene expression profiling reveals underlying molecular mechanisms of the early stages of tamoxifen-induced rat hepatocarcinogenesis. *Toxicol. Appl. Pharm.* 2007; **225**(1): 61-69. .
- Ramaswamy, S., P. Tamayo, R. Rifkin, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. USA* 2001; **98**(26): 15149-15154.

Figure 1: The Basic Transcriptional Process: Ligand-responsive and generic transcription factors bind to response elements within the regulatory regions of a gene. This increases the recruitment of RNA polymerase II to the transcription start site, allowing Pol II to move along the protein coding region of the gene producing mRNA transcripts. These transcripts are then translated by the ribosome to produce the final protein.