Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Microbiology

Direct methods for studying transcription regulatory proteins and RNA polymerase in bacteria

David C Grainger[1], David J Lee[2] and Stephen JW Busby[2]

Transcription factors and sigma factors play a major role in bacterial gene regulation by guiding the distribution of RNA polymerase between the promoters of different transcription units in response to changes in the environment. For 40 years *Escherichia coli* K-12 has been the paradigm for investigating this regulation and most studies have focused on small numbers of promoters studied by a combination of genetics and biochemistry. Since the first complete sequence for a bacterial genome was reported, the emphasis has switched to studying transcription on a global scale, with transcriptomics and bioinformatics becoming the methods of choice. Here we discuss two complementary direct experimental methods for studying transcription factors and sigma factors and we outline their potential use in rapidly establishing the regulatory networks in newly sequenced bacteria.

Addresses
[1] Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK
[2] School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK

Corresponding author: Busby, Stephen JW (s.j.w.busby@bham.ac.uk)

Current Opinion in Microbiology 2009, **12**:531–535

This review comes from a themed issue on
Genomics
Edited by Carmen Buchrieser and Stewart Cole

Available online 15th September 2009

1369-5274/$ – see front matter

DOI 10.1016/j.mib.2009.08.006

Introduction

Gene expression in all bacteria is tightly controlled, with transcription initiation being the principal point of regulation for many genes. Since the beginning of molecular biology, *Escherichia coli* K-12 has been the organism of choice for the study of transcriptional regulation and it is clear that this regulation is due to a complex network of transcription factors and sigma factors that control the expression of ~1800 transcription units in response to changes in the environment [1,2]. The *E. coli* genome encodes over 250 gene regulatory proteins that range from highly specific transcription factors such as the lactose operon repressor (Lac repressor), which controls a single transcription unit, through to global regulatory proteins, such as the cyclic AMP receptor protein, which controls scores of transcription units. In addition, the nucleoid-associated proteins, which are needed for maintaining chromosome folding and compaction, play important roles in transcriptional regulation. Many of these are present in large quantities that vary according to growth conditions, and they play key roles in upregulating or downregulating specific promoters [3,4].

Over 50 years, the network of *E. coli* gene regulatory proteins has been established by integrating information from studies on individual promoters and transcription factors. Before the arrival of whole genome sequences, most investigators would begin with their 'favourite' promoter or factor and exploit a toolbox of genetic tricks to select and characterise mutants in which the activity or regulation of the promoter or factor was altered. The advent of cloning gave access to a battery of biochemical approaches for studying protein binding at specific promoters, and bioinformatic approaches, based on establishing and exploiting consensus sequences, were widely applied. Up to the arrival of large-scale shotgun sequencing, only a small number of gene regulatory regions were studied in depth. However, whole genome sequences, in combination with transcriptomics and bioinformatics, opened the way to pan-genome viewing of transcriptional regulation. It was quickly established that some transcription factors regulated scores or more of transcription units. This led to the ideas of transcriptional regulatory networks and transcription factor hierarchies, and it is from these studies, taken together with the data from years of laborious effort, that we have now such a comprehensive view of transcription in *E. coli* K-12 [1,3]. This, of course, begs the question of whether we can now find more rapid easier routes to establish transcription networks in other bacteria. We want to argue that the rate limiting step in the post-genomic era has been the dearth of direct methods to detect what is happening at any gene regulatory region. Thus, any transcriptomics or proteomics experiment depends on measuring the consequences of the actions of gene regulatory proteins (that is, production of RNA or protein) rather than their interactions directly, and disentangling direct from indirect effects is not trivial. Furthermore, following up such experiments, or following up bioinformatic predictions, using genetic or biochemical methods is time consuming and may not be possible for some bacteria. Here we present chromatin immunoprecipitation (ChIP) as a method of choice for the rapid analysis of binding targets for gene regulatory proteins in any bacterium for which the genome

sequence is known. We also discuss a newly developed complementary method, DNA sampling, which permits an 'audit' of the regulatory proteins interacting at any locus.

Chromatin immunoprecipitation

ChIP represents a powerful tool, since it identifies protein–DNA interactions *in vivo* directly, independent of the biological consequences of binding (different technical aspects are reviewed in [5–7]). Briefly, bacterial cells are exposed to formaldehyde, thereby instantly crosslinking DNA binding proteins to the chromosome. After cell lysis and shearing of chromosomal DNA by sonication, the protein of interest is immunoprecipitated with specific antibodies, together with crosslinked DNA fragments. After reversal of the crosslinks and purification, the immunoprecipitated DNA is analysed in order to detect enrichment of the sequences bound by the protein of interest. Using ChIP in conjunction with DNA microarray analysis (ChIP-on-chip) permits DNA binding to be measured on a chromosome-wide scale (Figure 1).

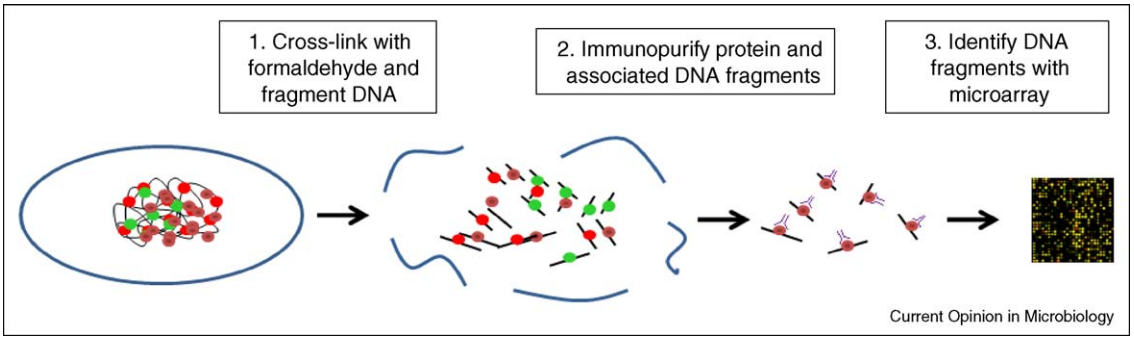
ChIP and ChIP-on-chip were first developed for eukaryotic cells and have since found applications in organisms as diverse as bacteriophage, yeast and mammals, where the binding of transcription factors, as well as other DNA-associated proteins, was investigated. For bacteria, ChIP-on-chip was first used with *Caulobacter crescentus* [8] but has been most applied to *E. coli* [9•]. In principle, it can be used in any bacterium, such as *Bacillus subtilis* [10,11], *Salmonella enterica* [12], *Helicobacter pylori* [13], *Mycobacterium bovis* BCG [14] and *Mycobacterium tuberculosis* [15•]. A crucial point is that protein binding locations can be identified without the need for a lot of prior knowledge of the bacterium under study. Indeed, because it is now possible to identify immunoprecipated DNA by sequencing (ChIP-seq), in principle, it is possible to derive information from bacteria whose chromosome sequence has not been determined.

Applications of ChIP to study transcription factors

The most straightforward use of ChIP in bacterial systems is in the location of transcription factors. Following the global analysis of the *C. crescentus* CtrA regulon [8] and the *B. subtillis* Spo0A and CodY regulators [10,11], ChIP-on-chip has now been applied to many other bacterial systems, including several pathogens. For instance, the *H. pylori* Fur protein has been studied and found to bind at about 200 genomic loci in an iron-dependent manner, supporting the idea that this protein acts as a pleiotropic regulator [13]. The previously uncharacterised BlaI transcription factor from *M. tuberculosis* was recently shown to regulate five DNA loci including the *blaI* gene itself, and others involved in resistance to β-lactam antibiotics [15•]. Unexpectedly, BlaI was found to bind upstream of the operon encoding ATP synthase, suggesting links between cell wall damage and ATP production. In the case of the *Salmonella enterica* PhoP transcription regulator, ChIP experiments were used to understand the hierarchy with which different promoters are 'served' as concentrations of the trigger ligand, magnesium ions, change [12].

Studies performed in *E. coli* K-12 represent the paradigm for ChIP-on-chip studies of transcription factors. The distribution of ArcA [16], CRP [17], FNR [18], LexA [19], Lrp [20], MelR [21], NsrR [22•,23] and RutR [24] has been determined. These studies show that some factors recognise single binding sites (as in the case of MelR) whilst others have more complex distributions (for example, CRP and LexA). There have also been some surprises. Thus, CRP binds throughout the *E. coli* chromosome to nearly 1000 sites and, at many targets, appears to have no effect on transcription [17]. With RutR, most of its 20 binding sites mapped within coding regions, suggesting that it may play some other, as yet undiscovered, role [24]. It is plausible that the binding of transcription factors to specific sites without any function is a by-product of evolution, perhaps due to inadequate 'purging' after horizontal gene transfer.

Figure 1



Schematic outline of ChIP-on-chip analysis. The coloured dots represent different proteins bound to DNA in a bacterium.