Genome-wide expression profile of RNA polymerase II subunit mutant of yeast using microarray technology

Beena Pillai[†], Samir K. Brahmachari[#] and Parag P. Sadhale[†,*]

[†]Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India
[#]Functional Genomics Unit, Centre for Biochemical Technology, New Delhi 110 007, India

Rpb4, the non-essential core subunit of RNA polymerase II has been assigned a function of regulating stress response in *S. cerevisiae* based mainly on phenotypes associated with its deletion. The actual mechanism has been elusive, although various hypotheses have been put forth. We have shown previously that it plays a significant role in activation of a subset of genes, rather than causing generalized defect in transcription. We used the microarray technology to look at the effect of this RNA polymerase subunit on the expression pattern of the entire *S. cerevisiae* genome. Many surprises emerged when we compared the genome-wide expression patterns of wild type and a mutant lacking the RPB4 gene (*rpb4Δ*) subjected to heat shock. The initial analysis of genes down-regulated in the mutant showed that the co-regulation of genes is not position-dependent, although the locus carrying the deletion had unexpectedly a large cluster of down-regulated genes. We also found that among the known down-regulated genes, a majority is involved in hexose uptake and utilization. We speculate that this could potentially contribute to the slow growth rate of the mutant. Compared to the other components of the transcription machinery, the Rpb4 subunit affects a unique set of genes.

THE most highly regulated step in gene expression is transcription. Many viruses transcribe their genes using highly efficient single polypeptide RNA polymerases. Prokaryotes employ a core RNA polymerase composed of five polypeptides ($2a$, b, b', w). The specificity of the polymerase is achieved through the sixth subunit designated as sigma. The task of eukaryotic transcription is shared by three different RNA polymerases I, II and III (also called A, B and C). The core RNA polymerase II (RNA pol II), which is composed of 12 subunits, transcribes all messenger RNAs in the cell. All these polymerases interact with many other proteins and together they achieve a highly orchestrated gene expression which is essential for a cell to rapidly adapt to changes in its immediate environment. The number and complexity of these additional factors also increase markedly from viruses to eukaryotes. The very large number of multisubunit complexes which together constitute eukaryotic transcription machinery has been collectively called the transcriptosome (not to be confused with the complete set of expressed RNAs of a cell at a given time–transcriptome)[1,2].

The eukaryotic RNA pol II is highly conserved from yeast to humans. The yeast RNA pol II is assembled from 12 subunits, Rpb1–12. Rpb1–3 makes up the enzymatic core of the polymerase. Some of the smaller subunits provide the structural integrity to the polymerase, while others may act as targets for regulation[3]. We have been studying Rpb4, a 25 kDa peripheral, non-essential protein of the polymerase over the past few years. The deletion of this subunit leaves the cell viable but compromised in survival at extreme temperatures[4]. Molecular genetic experiments have shown that *rpb4Δ* is defective in activation of many regulated promoters *in vitro* and *in vivo*[5–7]. *In vivo*, the stoichiometry of this subunit within the polymerase changes with the growth phase[8,9]. These features taken together suggest that Rpb4 could functionally be the eukaryotic counterpart of the prokaryotic sigma subunit[10]. There are conflicting reports in the literature cited above regarding the role of Rpb4 in transcription. The proposed roles include an effect on basal transcription of almost all genes, a more pronounced effect on activated transcription and a stress-related but gene-specific role, all of which may result in the phenotypes shown by *rpb4Δ* cells. This dilemma can be ultimately solved only by studying the whole genome expression pattern in *rpb4Δ* compared to the wild type. We report here the results of our attempt to study the role of this subunit in the backdrop of transcriptosome – the effect of the deletion of *RPB4* on the genome-wide expression pattern of yeast.

Ever since yeast became the first eukaryote whose genome was completely sequenced, newly available tools have opened up the opportunity to ask many new questions which could not previously be asked. The availability of the whole yeast genome microarray, an ordered assembly of DNA from yeast genes immobilized on a 5 cm × 2 cm area of a glass slide, has added yet another powerful tool to the geneticists' toolbox. Using microarray analyses, Holstege *et al.*[11] have identified the transcriptome of cells in which key components of the

*For correspondence. (e-mail: pps@mcbl.iisc.ernet.in)

transcriptional apparatus (one protein in each subcomplex) have been deleted or mutated, including Rpb1, the largest and essential subunit of RNA pol II, general transcription factors, histone acetylation factors, suppressor of RNA polymerase B (SRB) and TATA-box-binding-protein-associated factor (TAF) proteins (see Table 1). Amongst the twelve core RNA pol II subunits, the influence of only Rpb1 and Rpb9 on whole genome transcription pattern has been studied. Rpb1 is required for expression of practically all messenger RNAs and Rpb9 is required for the expression of only a few metabolic genes. Therefore, the widely acknowledged need for a subunit-specific, genome-wide expression map of the core RNA pol II remains largely unfulfilled.

The microarrays used in this study were procured from Ontario Cancer Institute, Canada. Each slide carries all the 6215 annotated yeast ORFs spotted in duplicate and arranged in 32 individual arrays of 400 spots each. *rpb4Δ* cells are incapable of survival at temperatures above 37°C. Therefore, *rpb4Δ* and *RPB4* yeast cells were grown to mid-log phase at 25°C and then shifted to 39°C for an hour before isolating RNA by conventional methods. Initially, we had done direct labelling of 10 μg of total RNA with Cy3-dUTP and Cy5-dUTP using Superscript RT II (Invitrogen Life Science Technologies), as recommended by the array manufacturer. The intensity of the signal was very low with this amount of total RNA. In our experience, the tyramide signal amplification (Perkin Elmer Life Science)-mediated system is more convenient, primarily due to the low amount of RNA required per experiment. The RNA was used without further purification to generate biotin-dCTP and fluorescein-dCTP incorporated cDNA using reverse transcriptase. Reciprocal labelling was also carried out simultaneously, so that both RNA samples were separately labelled with both biotin and fluorescein. The cDNA was purified and the labelling efficiency was tested by spotting the cDNAs on membranes and detecting them using streptavidin/anti-fluorescein antibody conjugated to HRP. The mutant and wild type cDNAs labelled with biotin and fluorescein were mixed and hybridized to the array overnight at 65°C in a hybridization volume of 50 μl, under a coverslip. Corning hybridization chambers were used to incubate the slides submerged in a water-bath at 65°C. All further steps were carried out according to the manufacturer's recommendations (http://www.nen.com/pdf/penen264-mmaxaminated_card. pdf), except for the final washes following addition of Cy5-tyramide (Perkin Elmer Life Science). We found that at least three vigorous washes (and sometimes more) for 15 min each were necessary to reduce Cy5 background on the slide surface, as against the 3×5 min wash recommended by the manufacturer. The slides were scanned immediately and after four days using Scanarray-lite (GSI Lumonics). It was seen that slides can be scanned days after the experiments without any significant loss of signal as long as they are stored in dust-free, air-tight, dark containers. The images were analysed using Quantarray software (GSI Lumonics). The software generates a composite image wherein a spot is given yellow colour if there is equal contribution from both fluorophores; a higher contribution of red or green in a spot implies differential expression in the samples tested. Figure 1 shows a representative part of the array covering 400 spots–red spots correspond to Cy3 signal and green spots correspond to Cy5. The results presented in Figure 1 are from a typical experiment out of three such experiments carried out. Blank spots and plant DNA spots in each array have been used to rule out any non-specific hybridization. The Cy5 signal was plotted against the Cy3 signal on a scatter plot. *ACT1*, the gene that codes for actin was used as a control, since it is widely accepted that actin is expressed constitutively and at high levels. The ratio of Cy5 to Cy3 was normalized to that of *ACT1*, which was arbitrarily fixed at 1. Only data points that showed consistent results in

Table 1. Positive and negative effects of the yeast transcription machinery on a genome-wide scale

Complex	Subunit[a]		Description		Affected fraction (%)	No. of genes		Total scored
						Up	Down	
RNA pol II	Rpb1		Largest subunit, essential for all mRNAs		100	14	4842	5590
	Rpb9		Transcription start site selection	(Rpb4 and Rpb9 are the only two dispensable subunits)	0.6	11	28	~ 6400
	Rpb4		Proposed role in stress response		11	524	190	~ 6400
SRB/mediator (core)	Srb4		Target of GAL4 activator		93	29	4766	5440
	Srb5		Unknown function		16	59	675	4876
	Med6		Activates some genes		10	14	478	5695
SRB/CDK	Srb10		CTD kinase		3	168	9	5626
SWI/SNF	Swi2		Chromatin remodelling		6	200	116	5695
GTFs	TFIID	TAF145	Large TBP-associated factor, histone acetylase		16	54	1425	5441
		TAF17	Component of TFIID and SAGA		67	169	3488	5349
	TFIIE		Promoter opening		54	52	4243	6082
	TFIIH		CTD kinase		87	37	4659	3236
SAGA	GCN5		Histone acetylase		5	78	179	4912
	TAF17		Component of TFIID and SAGA		67	169	3488	5349

[a]After refs 11 and 12.

multiple experiments, reciprocal labelling and between duplicate spots (on the same slide) were used for further analyses.

RPB4 has been cloned for over a decade now. Using the phenotypes as a clue, many workers have tried to identify the genes whose expression is affected by this subunit: the search has been largely disappointing. Studying expression of individual candidate genes only served to confirm that certain heat shock proteins are not induced effectively[6–8]. Suppressor analyses yielded extragenic suppressors which were shown to function via indirect means, for instance, by non-specifically increasing RNA stability[6]. Using the phenotypes as a direction, other workers have tried to identify the transcriptional regulators which can suppress the phenotypes associated with Rpb4; this attempt has also not been successful. We had appreciated the need for understanding the genome-wide pattern even before microarray experiments were feasible in India. We have earlier reported our findings from RNA Arbitrarily Primed PCR[13] of RNA samples from *rpb4Δ* and *RPB4*, but the method suffered from limitations of low repro-

ducibility. It also required cloning and sequencing to identify the differentially expressed clones. Since Rpb4 is a component of the core RNA polymerase, it has been assumed that its deletion will result in under-expression of many genes. Figure 1 summarizes the pattern of gene expression in *rpb4Δ* mutant compared to the wild type. The genome-wide analyses (both from RAP-PCR and from the microarray experiments reported here) have led to a surprising result. Deletion of *RPB4*, in fact resulted in overexpression rather than under-expression of many more genes. Five hundred and twenty-four genes, which account for 8% of the genome are overexpressed (henceforth referred to as 'up' genes), whereas 190 genes which account for 3% of the genome were under-expressed ('down' genes) in the mutant. Table 1 summarizes the effect of various other subunits of the transcription machinery in yeast and compares our observations with these reports. Mutations in most of the proteins in the transcription machinery result in down-regulation of a large set of genes. The set of up-regulated genes is larger than the set of down-regulated genes for only three
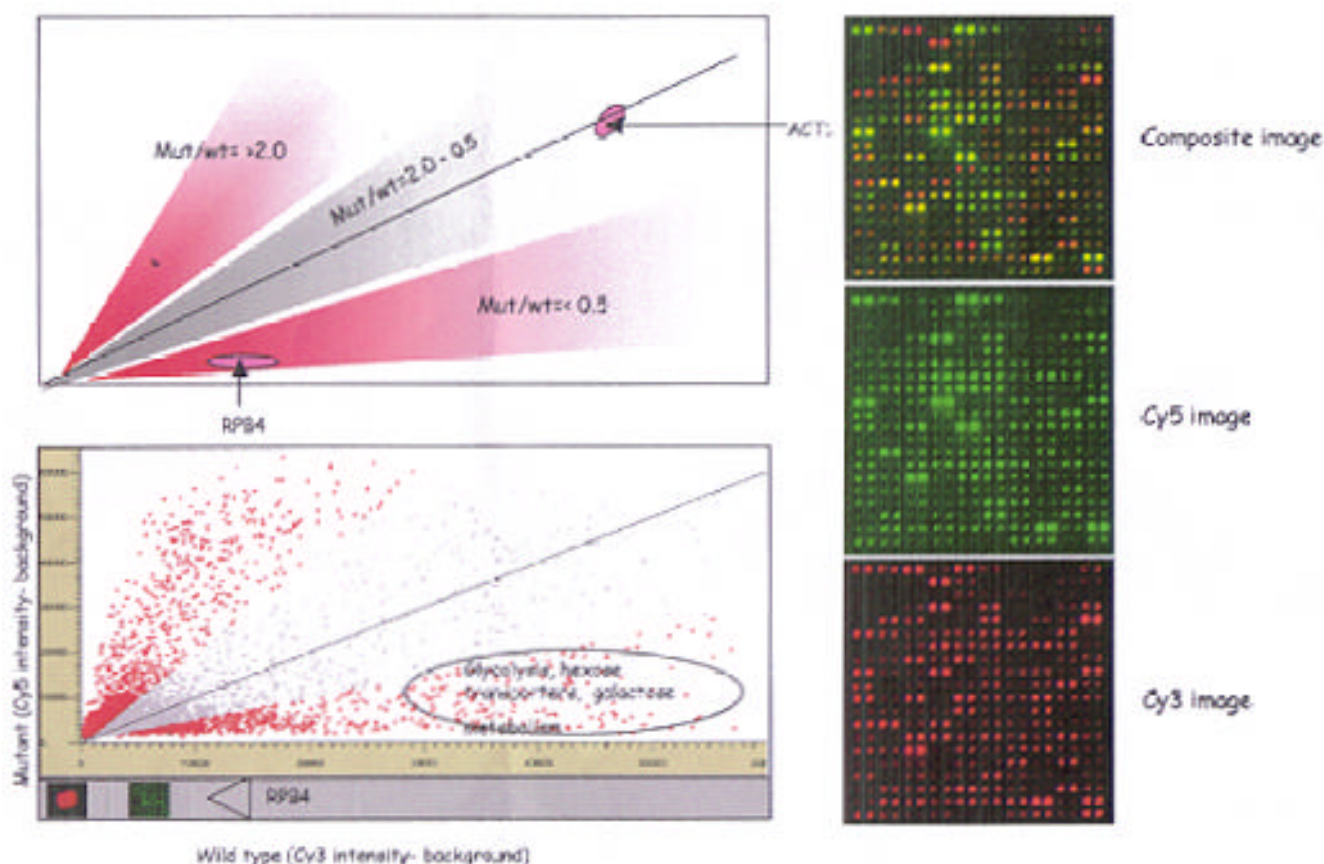

Figure 1. Subarray consisting of 400 spots is shown here. Cy5 (green) image corresponds to the expression pattern of *rpb4Δ* mutant and Cy3 (red) image corresponds to wild type. Composite image shows yellow spots wherever expression of the mutant and wild type cells were similar. Red spots in the composite image imply low level of expression in mutant and green spots imply that the corresponding gene was highly expressed in mutant. The background intensity of a fixed area around each spot was reduced from the intensity of each spot. The background corrected Cy5 signals were plotted against Cy3 signals in a scatter plot. The ratios of Cy5 : Cy3 signals that fall between 2 and 0.5 are in grey. Spots of interest corresponding to genes overexpressed in mutant (above the grey area) and under-expressed (below the grey area) are in red. In the schematic diagram, the areas marked in pink show the location of *ACT1* gene and *RPB4*. Spots corresponding to *RPB4* are shown enlarged below the scatter plot.
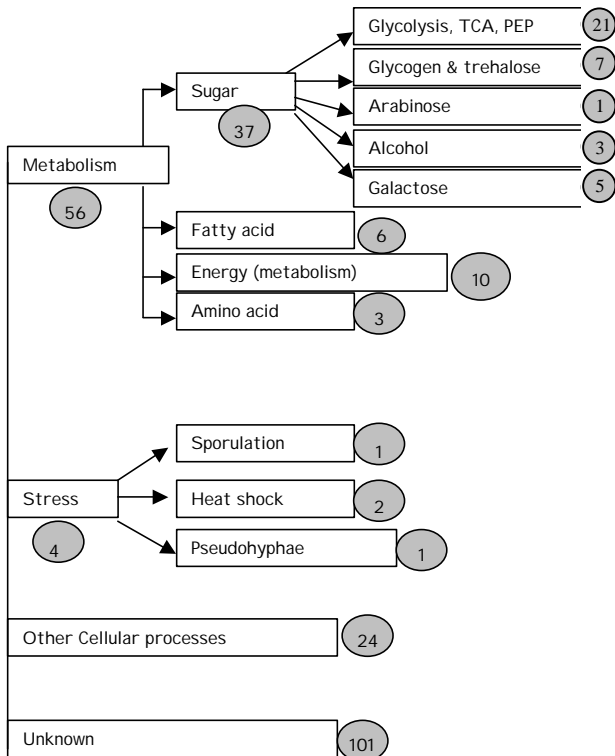
Figure 2. Functional clustering of genes under-expressed in *rpb4Δ*. ORF names of the under-expressed genes were used to retrieve the corresponding annotation from Saccharomyces Genome Database. Genes were classified into various categories depending on their function. Numbers in circles to the right side of the group refer to the total number of genes that fall into that category. Numbers do not add up to 190 because ambigous annotations have not been included.

subunits, Srb10, Swi2 and Rpb4. Srb10 represses stress-related genes under non-stress conditions and Swi2 is involved in silencing and chromatin remodelling[11]. There is no previous report of Rpb4 acting as a repressor of any gene.lb Detailed analysis of the up-regulated genes, to further understand the common features of these genes and the reason for their co-regulation is being carried out. Here we report the preliminary analyses of the genes whose expression drops in response to deletion of *RPB4*.

We have categorized the down genes into groups based on their functions, according to the annotation provided by Saccharomyces Genome Database (genome-www. stanford.edu/Saccharomyces; Figure 2). A very large proportion (50% of down genes) are of unknown function. The level of expression and the promoter sequences of these genes can now be used to classify them with the known genes in the cluster. These strategies may provide valuable clues to assign functions to these unknown genes. Amongst the known genes, 21 are related to glycolysis; further, many of them are hexose transporters. The inability to utilize sugars effectively could be responsible for the slow growth defect seen in *rpb4Δ* mutants.

We generated a chromosomal location map (Figure 3) for the genes under-expressed in *rpb4Δ*, to rule out the possibility that they show positional bias. Glycolysis genes are amongst the most highly expressed yeast genes. Therefore, there is a possibility that the effect of Rpb4 on these genes is not due to any direct correlation to glycolysis, but due to a non-specific effect on regions of high expression. We do not see any significant overlap with
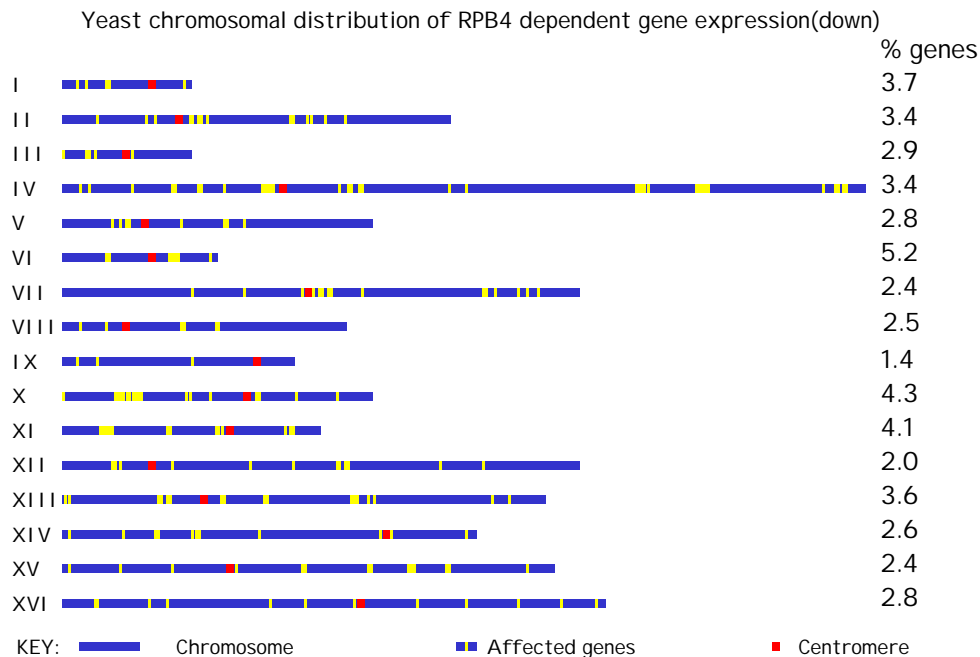

Figure 3. Chromosomal position of genes down-regulated in *rpb4Δ*. Blue bars represent chromosomes and a single yellow line represents one gene. Regions which have many consecutive down-regulated genes have proportionately longer regions in yellow. The lengths of the genes have not been considered. Similar maps for gene density, function and transcriptional activity are available. These maps were overlapped, but have not been reproduced here.
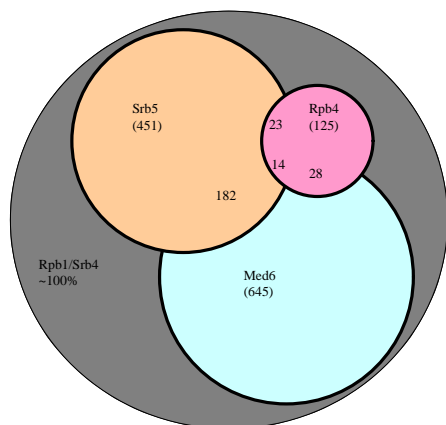
Figure 4. Genome-wide dependence on *RPB4* compared to that of other key components of the transcription machinery. Circles are labelled with gene names and the area of the circle (except the largest) is proportional to the total number of genes down-regulated by the mutation. Numbers in brackets refer to the number of genes unique to each set. The number of genes affected by both mutations is shown in the respective overlapping regions. Thus 14 genes are affected by all the three genes shown.

maps of regions with high or low gene expression activity[1], silenced regions, etc. Coincidentally, the largest cluster of 9 consecutive affected genes is on the left arm of chromosome X at the *RPB4* locus.

These observations are likely to be more meaningful when studied in the backdrop of the transcription machinery. We have integrated our results with those of similar experiments done by Holstege *et al.*[11] and Hemming *et al.*[12] using other components of the transcription machinery. As seen in Table 1 per cent of genes affected by Rpb4 is in the range affected by some general transcription factors. The venn diagram (Figure 4) shows that the set of genes affected by Rpb4 overlaps partially with corresponding sets of Srb5 and Med6. Fourteen genes are affected by all these proteins. We have also included Rpb9 in the analyses, though the data are not represented in the venn diagram. Rpb9, the only other non-essential subunit of RNA pol II besides Rpb4, affects a much smaller number (28) of genes. Four of the 14 genes mentioned above are also affected by Rpb9. Rpb4 affects the expression of 125 genes which are not affected by the other subunits. This implies that the transient association of Rpb4 with the polymerase can be a step in recruiting the polymerase to these promoters.

Technically, the experiments can be carried out easily in a standard molecular biology laboratory. Centralized scanning facilities can therefore support users who wish to do the experiments in their own lab and send the arrays for scanning. The major hurdle that keeps this powerful technique inaccessible to most laboratories is the cost of the arrays. The entry of more non-commercial centres for printing arrays (e.g. the yeast arrays used in this work) and the diminishing prices of arrays from commercial organizations promise that the costs will not continue to be prohibitive for very long.

In summary, using a whole genome-based approach to understanding the role of a subunit of RNA polymerase, we report that this subunit has a hitherto unsuspected repressive effect on a very large number of genes. Genes related to sugar metabolism are severely down-regulated in its absence. The set of genes affected by Rpb4 is not a subset of the ones affected by any other key component of the transcription machinery. Thus there is a unique set of 125 genes whose expression depends on the presence of *RPB4*. We intend to pursue these leads using bioinformatics, to study the common structural and functional features of the promoters affected by Rpb4 and conventional molecular genetics and biochemistry, to check if indeed Rpb4 is a regulator of polymerase specificity at these promoters.

1. Velculescu, V. E. *et al.*, *Cell*, 1997, **88**, 243–251.
2. Greenblatt, J., *Curr. Opin. Cell Biol.*, 1997, **9**, 310–319.
3. Hampsey, M., *Microbiol. Mol. Biol. Rev.*, 1998, **62**, 465–503.
4. Woychik, N. and Young, R. A., *Mol. Cell. Biol.*, 1989, **9**, 2854–2859.
5. Edwards, A. M., Kane, C. M., Young, R. A. and Kornberg, R. D., *J. Biol. Chem.*, 1991, **266**, 71–75.
6. Tan, Q., Li, X., Sadhale, P. P., Miyao, T. and Woychik, N. A., *Mol. Cell. Biol.*, 2000, **20**, 8124–8133.
7. Pillai, B., Sampath, V., Sharma, N. and Sadhale, P., *J. Biol. Chem.*, 2001, **276**, 30641–30647.
8. Choder, M. and Young, R. A., *Mol. Cell. Biol.*, 1993, **13**, 6984–6991.
9. Choder, M., *J. Bacteriol.*, 1993, **175**, 6358–6363.
10. Jaehning, J. A., *Science*, 1991, **253**, 859.
11. Holstege, F. C. *et al.*, *Cell*, 1998, **95**, 717–728.
12. Hemming, S. A., Jansma, D. B., Macgregor, P. F., Goryachev, A., Friesen, J. D. and Edwards, A. M., *J. Biol. Chem.*, 2000, **45**, 35506–35511.
13. Sharma, N. and Sadhale, P. P., *J. Genet.*, 1999, **78**, 149–156.