Nutrient-regulated gene expression in eukaryotes

Richard J. Reece¹, Laila Beynon, Stacey Holden, Amanda D. Hughes, Karine Rébora and Christopher A. Sellick

Faculty of Life Sciences, University of Manchester, The Michael Smith Building, Oxford Road, Manchester M13 9PT, U.K.

Abstract

The recognition of changes in environmental conditions, and the ability to adapt to these changes, is essential for the viability of cells. There are numerous well characterized systems by which the presence or absence of an individual metabolite may be recognized by a cell. However, the recognition of a metabolite is just one step in a process that often results in changes in the expression of whole sets of genes required to respond to that metabolite. In higher eukaryotes, the signalling pathway between metabolite recognition and transcriptional control can be complex. Recent evidence from the relatively simple eukaryote yeast suggests that complex signalling pathways may be circumvented through the direct interaction between individual metabolites and regulators of RNA polymerase II-mediated transcription. Biochemical and structural analyses are beginning to unravel these elegant genetic control elements.

Introduction

The genome of an organism contains within its DNA sequence all of the information required to define not only cell type, but also the ability to respond to a variety of external conditions and signals. For example, the genome of the simple eukaryotic yeast Saccharomyces cerevisiae contains 12 million base pairs of DNA, split into 16 separate chromosomes and comprising approximately 6000 different genes. The protein products encoded by these genes rarely act individually, and whole pathways are often regulated in a co-ordinated fashion in response to a particular signal. The complex processes of the cell depend on the differential expression of sets of genes in particular cell types, at a particular time, and/or

¹To whom correspondence should be addressed (email Richard.Reece@manchester.ac.uk).
under certain environmental conditions. Efficient genetic switches within the cell permit the co-ordinated transcription of sets of genes in response to a signal.

It is becoming increasingly apparent that what were traditionally thought of as metabolic enzymes can play diverse roles in a wide range of cellular processes [1]. Similarly, metabolites appear not just to serve as substrates for enzymes. Indeed, the link between metabolic substrates and their ability to influence transcription directly is long established in prokaryotes. For example, in *Escherichia coli*, the interaction between lactose and the *lac* repressor [2], or galactose and GalR [3], in modulating the DNA-binding properties of each respective protein is well understood. Here we will discuss two eukaryotic genetic switches and the ways in which they are regulated by the action of small-molecule metabolites. Insights into the processes of metabolically regulated gene expression in eukaryotes have been gained from the study of the relatively simple genetic switches of yeast – although the mechanistic details have turned out to be somewhat complex.

**Metabolic signalling through protein complex formation – GAL induction**

The expression of the *GAL* genes of the yeast *Saccharomyces cerevisiae* is a paradigm for eukaryotic transcriptional regulation. The *GAL* genetic switch provides the transcriptional control of genes whose products are required for the metabolism of the sugar galactose – the enzymes of the Leloir pathway [4]. When yeast cells are grown in a medium without galactose, the *GAL* genes are essentially transcriptionally inert. However, if galactose is the only source of carbon available to the cell, then the *GAL* genes will become transcriptionally active – both rapidly and to a high level – and galactose is converted into glucose 1-phosphate [5]. Galactose is a comparatively poor sugar source for the cell, and yeast will preferentially metabolize other sources of carbon (e.g. glucose) to galactose even if a mixture of the two sugars is available [6]. Glucose triggers catabolite repression of the *GAL* genes, and the expression of the activator upon which they depend, Gal4p, is severely reduced when cells are grown on glucose [7]. In the presence of other carbon sources, e.g. raffinose or glycerol, Gal4p is produced in the cell, and can be found tethered upstream of the *GAL* genes, but does not bring about transcriptional activation. The transcriptional activity of DNA-bound Gal4p is inhibited through its interaction with another protein, Gal80p [8]. Activation of *GAL* genes by Gal4p is thought to involve the ordered recruitment of various protein complexes. The transcriptionally active form of DNA-bound Gal4p interacts directly with the Tra1p subunit of the SAGA (Spt/Ada/Gcn5 acetyltransferase) co-activator/histone modifying complex [9–11]. Other subunits of SAGA (Spt3p and Spt8p) interact directly with specific subunits of the RNA polymerase II mediator complex (Med2p and Pgd1p respectively) [12]. Gal4p may also interact with components of the mediator directly [13]. Finally, the mediator recruits RNA polymerase II holoenzyme to the promoter, and the initiation of transcription can occur.
Activation of the transcriptional properties of Gal4p

Although the presence of galactose within the cell triggers the activation of Gal4p, neither Gal4p nor Gal80p acts as the galactose sensor. A transcriptional inducer, Gal3p, interacts with the transcriptional repressor, Gal80p, in a galactose- and ATP-dependent manner [14]. Gal3p appears to require galactose and ATP so that it can adopt a conformation to allow it to interact with Gal80p [14]. The net result of this interaction is that Gal4p becomes active, and transcription of the GAL genes proceeds.

The molecular mechanism by which the activation of the GAL genes occurs has been the focus of a much debate. Two somewhat conflicting models for its molecular mode of action have been proposed (Figure 1). It has been suggested that the induction of the GAL genes may occur through the association of a tripartite complex formed between Gal4p, Gal80p and Gal3p [8]. In favour of this model (Figure 1A) are the observations that (1) Gal4p purified from yeast grown in either the presence or the absence of galactose is associated with Gal80p [15], (2) artificially constructed Gal80p molecules that contain an activation domain are able to regulate transcription in the presence and absence of galactose [16], (3) the three proteins assemble in vitro in a gel-shift assay [8], and (4) Gal4p and Gal80p do not dissociate from each other in the presence or absence of galactose, as demonstrated recently using FRET (fluorescence...
resonance energy transfer) [11]. Other evidence, however, suggests that this model is incorrect, and that Gal80p dissociates from Gal4p and interacts with Gal3p in the cytoplasm of yeast cells [17]. This would result in the freeing of Gal4p from the inhibitory effects of Gal80p, allowing transcriptional activation to occur [18]. The dissociation model (Figure 1B) is supported by data indicating that Gal3p is located predominately in the cytoplasm of yeast cells [17] and that the expression of a myristoylated version of the protein (which targets it to the plasma membrane) does not unduly impair the induction of the GAL genes [18]. In addition, ChIP (chromatin immunoprecipitation) experiments [18] and pull-down assays [19] suggest that the Gal4p–Gal80p complex is somewhat weakened (although perhaps not completely dissociated) when cells are grown in the presence of galactose. In either of the above models, the ability of Gal3p to interact with Gal80p in a galactose-dependent fashion is essential for the transcriptional induction of the GAL genes. Understanding this interaction is therefore key to understanding the transcriptional activation in response to galactose.

Gal3p is highly similar to galactokinases

The analysis of Gal3p function has been greatly assisted by its similarity to Gal1p – yeast galactokinase. Gal1p and Gal3p are ∼70% identical and ∼90% similar at the amino acid level [20,21]. Both proteins require galactose and ATP to function: Gal3p requires these to interact with Gal80p [22], while Gal1p produces galactose 1-phosphate from galactose at the expense of ATP. Despite the high level of similarity between the two proteins, Gal3p does not possess a galactokinase activity [23]. The removal of Gal1p from yeast cells results in an inability to grow on galactose, whereas Gal3p-depleted cells can grow on galactose, but show a severe long-term adaptation phenotype. Rather than inducing the GAL genes within 30 min of addition of the sugar to cultures, cells lacking Gal3p only fully induce the GAL genes some 3–4 days later [24]. The effects of the loss of Gal3p can be overcome if Gal1p is produced at higher than wild-type levels [25]. However, yeast cells missing both GAL1 and GAL3 function are unable to induce the GAL genes at all [24]. Recent modelling data have suggested that the GAL genes become induced in the presence of Gal1p, but the absence of Gal3p due to stochastic changes in the cellular concentration of Gal80p [26,27]. Other galactokinases, e.g. those from E. coli [24], are able to substitute for the loss of Gal1p in yeast to rescue the galactokinase deficiency, but are unable to overcome the transcriptional defect in the GAL genes resulting from the loss of Gal3p. Combined, these data suggest that although Gal1p is primarily a galactokinase enzyme, it can also function as a weak transcriptional inducer. This conclusion has also been supported through a number of in vitro experiments [8,23]. Other yeast species, e.g. the milk yeast Kluyveromyces lactis, contain a single galactokinase-like molecule that functions both as an efficient transcriptional inducer (Gal3p-like function) and as a galactokinase enzyme (Gal1p-like function) [28].

Gal1p and Gal3p appear to have evolved from each other. Gal3p is a defective galactokinase because it contains a two-amino-acid deletion within its galactokinase signature motif [23]. The re-introduction of these residues (a serine
Nutrient-regulated gene expression

© 2006 The Biochemical Society

and an alanine residue starting at position 164) imparts the resulting protein with a measurable galactokinase activity in vitro and in vivo [23]. The structures of the galactokinase enzymes from Lactococcus lactis [29], Pyrococcus furiosus [30] and human [31] have recently been solved. The equivalent residue to serine-164 in both the P. furiosus and human galactokinase structures interacts with the magnesium ion and the β-phosphate of the nucleotide [30,31]. Despite the relative

Figure 2 Conserved ligand-binding site in different galactokinases.
The galactose-binding sites from the L. lactis, P. furiosus, human and S. cerevisiae enzymes are shown, together with the relative position of either a phosphate ion, ADP or ADPNP, present in the crystal structure as indicated. A magnesium ion (green sphere) is also present in some structures. The dashed lines indicate distances of between 2.3 and 3.3 Å between the protein and the ligands, and may indicate potential hydrogen bond interactions. Amino acids with the capacity to form hydrogen bonds with the sugar are labelled using the single letter amino acid code.
lack of overall sequence similarity between the enzymes from diverse organisms, the galactose- and nucleotide-binding sites of each crystallized protein show a high level of conservation (Figure 2). A model for the bifunctional Gal1p from \textit{K. lactis} has been proposed based on its similarity to other sugar kinases whose structures have been solved [32]. The overall shape of the molecule within the region of the protein involved in the interaction with the ligands may indeed be similar to the galactokinases, but the likely Gal80p interaction site is not complete in these models. It therefore seems probable that Gal3p also shares similar galactose- and nucleotide-binding pockets and is likely to interact with its small molecules in a manner similar to the galactokinases. Crystals of \textit{S. cerevisiae} Gal1p that diffract to 2.1 Å have recently been grown in the presence of galactose and the non-hydrolysable ATP analogue ADPNP (adenosine 5'-[β,γ-imido]triphosphate), and the structure has been solved [32a]. The overall structure of the protein is, as expected, highly similar to those of the previously solved galactokinases. In addition, the galactose- and ATP-binding sites are also highly conserved (Figure 2). The structure of Gal1p, and consequently the implied structure of Gal3p, is generating superb insights into the differences between the purely enzymatic proteins and those capable of transcriptional induction. The structure also allows the precise mapping of mutations in \textit{GAL3} that give rise to a constitutive phenotype (activation of transcription in the absence of galactose). In a linear amino acid map, these mutations are spread throughout the length of the protein [33,34]; however, in the structure, they cluster at the interface between the N- and C-terminal domains of the protein. It is tempting to speculate that this represents the site of Gal80p binding.

Mutations in Gal80p that result in an impaired interaction with Gal3p are also spread throughout the length of the protein [35,36]. Gal80p has been suggested to interact with Gal4p as a dimeric protein [37], but its interaction with Gal3p occurs as a monomer only in the presence of galactose and ATP [38]. The influence of Gal3p and Gal4p upon the monomer/dimer interaction of Gal80p is not fully characterized, and the structure of Gal80p is not currently known. Modelling the primary amino sequence of Gal80p on to the previously solved structure of the sequence-related, but functionally distinct, glucose–fructose oxidoreductase from \textit{Zymomonas mobilis} [39] suggested that the mutations in Gal80p that affect its interaction with Gal3p may cluster on one surface of the protein, while mutations that affect the interaction with Gal4p reside on another surface [36]. Caution should be exercised in interpreting a homology model based on two distantly related proteins (15% amino acid identity; 30% similarity), but the similarity between Gal80p and an oxidoreductase may suggest that Gal80p is also a dinucleotide-binding protein. This could provide yet another layer of complexity to the control of expression of the \textit{GAL} genes.

An additional impediment to the biochemical analysis of Gal3p is that the interaction between galactose and Gal3p, or Gal1p, appears to be somewhat weak in nature. The \textit{GAL} genetic switch is activated maximally \textit{in vitro} using galactose at a concentration 1.5 mM [8], and the \textit{Km} for galactose of Gal1p is 1.2 mM [21]. The weakness of this interaction may be a controlling influence on the overall activation of the \textit{GAL} genes \textit{in vivo}. The weak interaction between Gal3p and galactose may mean that a relatively high concentration of the
metabolite must be reached inside the cell before the GAL genes are switched on fully. The GAL system thus offers an intriguing, and complex, interplay between a transcriptional activator, regulator proteins and small molecules. The recent structural analysis of related proteins is providing a mechanism to dissect the molecular details of this elegant switch.

Direct modulation of transcription factor function – Put3p

*S. cerevisiae* can utilize a variety of compounds as potential sources of nitrogen. When preferred nitrogen sources, such as ammonia or glutamate, are available, the transcription of genes involved in the utilization of poorer nitrogen sources is repressed [40]. Proline, a comparatively poor nitrogen source, is transported into the cell by the general amino acid permease Gap1p and by the proline-specific transporter Put4p (Figure 3). Both of these are located in the plasma membrane, and the activity of each is regulated in response to the quality of the nitrogen available to the cell [41,42]. For its utilization as a nitrogen source, intracellular proline is converted into glutamate in the mitochondria by the products of the *PUT1* and *PUT2* genes, encoding proline oxidase and Δ¹-pyrroline-5-carboxylate dehydrogenase respectively [43]. The presence of proline, but the absence of preferred nitrogen sources, results in the high-level transcriptional induction of both *PUT1* and *PUT2* [44]. This induction of transcription is regulated on several levels. In response to the overall quality of nitrogen available to the cell, Nil1p and Ure2p – two GATA factor proteins – regulate the expression of both *PUT1* and *PUT2* [45,46]. However, neither Nil1p nor Ure2p is essential for the growth of yeast on proline as the sole source

![Figure 3 Proline utilization by yeast cells.](image)

Proline enters the cell via one of two transporters – Gap1p or Put4p. Once inside the cell, proline is enzymatically converted into glutamate in the mitochondria. Additionally, proline probably enters the nucleus to convert Put3p into a transcriptionally active form. The presence of proline, but only as the sole source of nitrogen, results in induction of Put3p, which activates transcription of the proline utilization genes, *PUT1* and *PUT2*. The enzymes encoded by these genes, Put1p and Put2p, are transported into the mitochondria, where they convert proline into glutamate.
of nitrogen, whereas the specific regulator of the proline utilization pathway, Put3p, is absolutely required for growth on proline [45].

**Put3p architecture**

Put3p is a transcriptional activator that has an overall architecture typical of the Zn(II)$_2$Cys$_6$ binuclear cluster family of proteins [47]. It possesses an N-terminal DNA-binding and dimerization domain [48] and an acidic activation domain at its C-terminus [49]. The protein is bound to its target DNA sites upstream of the *PUT1* and *PUT2* genes irrespective of whether those genes are transcriptionally active [50]. Put3p is a phosphoprotein, and the extent of phosphorylation has been correlated with transcriptional activity [51]. However, phosphorylation does not appear to be dependent on the presence of proline. Rather, Put3p is differentially phosphorylated in response to the quality of the nitrogen source available to cells. This was further confirmed by the observation that treatment of yeast cells with rapamycin, which mimics nitrogen-limiting conditions, resulted in the hyperphosphorylation of Put3p [46].

**Conformational changes in Put3p**

The activation of transcription by Put3p in response to proline appears to be mediated by a conformational change within the protein. The addition of proline to immunoprecipitated Put3p resulted in increased resistance to digestion by endoproteinase Glu-C [52]. Additionally, GST (glutathione S-transferase)–Put3p fusion protein immunoprecipitated from proline-grown yeast cells had increased resistance to thrombin-mediated cleavage of the GST tag compared with the same protein prepared from ammonia-grown cells [52]. Combined, these data suggest that Put3p undergoes a conformational change in the presence of proline. It is conceivable that this conformational change results in unmasking of the activation domain, thereby allowing the recruitment of the transcriptional machinery.

**Transcriptional activation induced by proline**

*In vivo* footprinting clearly demonstrated that Put3p binds to its DNA recognition sites in a proline-independent manner [53]. Therefore transcriptional activation by Put3p cannot be regulated through the control of its DNA-binding properties, but instead changes in the transcriptional activity of DNA-tethered Put3p must occur in response to proline. *In vitro* transcription experiments demonstrated that the activity of Put3p is induced in the presence of proline itself [54].

**Direct interaction between Put3p and proline**

The ability of proline to modulate the activity of Put3p *in vitro* suggests that the metabolite is directly affecting the activity of the transcriptional regulator. Proline affinity chromatography demonstrated that the recognition of proline by Put3p is indeed via a direct physical interaction. When Put3p was applied to a column bearing immobilized proline attached through its carboxy group, the protein was able to bind the matrix and could be eluted with proline [54]. Put3p was eluted from the column at a proline concentration of approx. 25 mM, which suggests that the interaction between Put3p and proline is weak [54].
However, this is a level of the metabolite comparable with that required for the activation of PUT1 and PUT2 in vivo [51], and to elicit a response using in vitro transcription assays [54]. If a similar column was produced in which the proline was attached through its imino-group nitrogen atom, then no Put3p binding could be detected. This result, and other data, suggested that that the pyrrolidine ring of proline is vital for recognition by Put3p. Pyrrolidine itself was found to be able to activate Put3p in an in vitro transcription assay [54], while the addition of hydroxy groups to the pyrrolidine ring ablated its ability to modulate the activity of Put3p. Therefore Put3p appears to recognize and interact with a simple pyrrolidine ring, and the carboxy group found in proline may be removed or modified without unduly affecting the interaction with the protein.

The interaction between Put3p and proline is sufficient to bring about a transcriptional response. However, the position and structure of the proline-binding site within Put3p are not known. The activation domain of Put3p is not proline-responsive and, when fused to a heterologous DNA-binding domain, it activates transcription in a constitutive manner [49]. This suggests that regulation of the transcriptional activity of Put3p occurs elsewhere in the protein. We are currently striving to identify the proline-binding site within Put3p.

Model for transcriptional activation by Put3p

At present, the model for transcriptional activation by Put3p is that a dimeric, DNA-bound form of Put3p ‘senses’ the presence of proline within a cell (Figure 4). The presence of proline, but only as the sole source of nitrogen, results in the activation of transcription of PUT1 and PUT2. The conversion of Put3p from a transcriptionally inert form into a transcriptionally active form is thought to occur via a conformational change in the protein. It is possible that this conformational change results in unmasking of the activation domain in the presence of proline, thus allowing the recruitment of the transcriptional machinery.

Phosphorylation may also play a role in regulating the overall transcriptional activity of Put3p. It has been proposed that phosphorylation of Put3p is required for transcriptional activation and that underphosphorylated forms of the protein are unable to activate transcription [51]. However, analysis of the phosphorylation profile of recombinant purified Put3p that activated transcription in vitro revealed that the protein was unphosphorylated [54]. This suggests that proline is able to act on unphosphorylated Put3p to convert it into a transcriptional activator in vitro, and raises the question as to the role of phosphorylation in regulating transcriptional activation by Put3p. It is possible that phosphorylated forms of Put3p are able to respond to lower levels of proline; alternatively, phosphorylation of Put3p may be involved in the global regulation of the proline utilization pathway by the GATA factors. The mechanism by which the overall quality of the nitrogen source available to a yeast cell is detected is still largely unknown. In addition, the molecular mechanism by which this information is relayed to the PUT genes is unclear. A complex interplay between nitrogen transporters, GATA factors and Put3p is likely to be involved.
Other metabolically regulated transcriptional switches

The systems described above represent just two well-characterized transcriptional programmes that respond directly to changes in the intracellular levels of metabolites. There are, however, others where activation is controlled by metabolites. For example, (1) the expression of genes required for pyrimidine biosynthesis is regulated by orotic acid (a biosynthetic intermediate of the pyrimidine pathway), which controls the activity of the transcriptional regulator Ppr1p [55]; (2) the formation of an active transcriptional complex between Bas1p and Bas2p to control de novo AMP biosynthesis is dependent on 5′-phosphoribosyl-5-aminomidazole-4-N-succinocarboxamide [56]; (3) leucine biosynthesis is regulated by the action of α-isopropyl malate on Leu3p [57]; (4) lysine biosynthesis is regulated by a presumed interaction between α-aminoadipate semialdehyde and Lys14p [58]; (5) arginine biosynthesis is regulated by the interaction between arginine and Arg81p, which promotes the DNA-binding properties of the Arg81p–Mcm1p complex [59]; and (6) in response to weak acid stress, War1p activates certain stress response genes through the likely interaction between the protein and weak acids [60,61]. In each of the above cases, the small molecule is presumed to modulate the transcriptional properties of the activator protein by some mechanism yet to be determined.

![Figure 4 Proline-specific transcriptional activation by Put3p](image)

**Figure 4 Proline-specific transcriptional activation by Put3p.** ADNA-tethered Put3p dimer acts as a proline sensor. When the proline concentration is sufficiently high, proline interacts directly with Put3p. The proline-binding domain of Put3p is likely to be within the central domain of the protein. The direct interaction results in a conformational change in Put3p that unmasksthe activation domain and allows the protein to activate transcription.
**Conclusion**

The direct modulation of transcription factor function by small-molecule metabolites is well established in prokaryotes. As we have shown here, this mechanism of transcriptional regulation is also common in single-celled eukaryotes. Yeasts contain many of the signalling pathways found in higher eukaryotes, e.g. MAPK (mitogen-activated protein kinase) cascades [62–64], but lack the metabolic homoeostasis systems present in higher, multicellular, eukaryotes. In higher eukaryotes, the interactions between, say, hormones and transcription factors, are understood in detail [65], but it would appear that direct modulation of transcription factor function by metabolites is largely lacking. Single-celled organisms have generally less control over their metabolic content, and this appears to have been exploited to provide a series of elegant and precise genetic controls. Understanding the diverse ways in which metabolites can directly influence and regulate transcription is vital if transcription and metabolism are to be understood in a fully integrated fashion.

We are grateful to Hazel Holden and David Timson for discussions and their comments on, and suggestions for, this chapter. Work in the authors’ laboratory is funded by The Wellcome Trust and the BBSRC.

**References**