

Gene regulation in *Aspergillus*: From genetics to genomics

MARK X. CADDICK, CHRIS DOBSON, IGOR Y. MOROZOV & MERIEL G. JONES

School of Biological Sciences, University of Liverpool, Liverpool, UK

A fundamental aspect of any organism's success is the ability to monitor and respond effectively to its environment, a process which is largely achieved through the appropriate regulation of gene expression. There are few better examples than fungi, which inhabit diverse and often hostile environments, ranging from leaf litter to the human body. Regulation can occur at many levels, and as we investigate specific genes in detail, the paradigm is one of increasing complexity. We will briefly review the different levels at which regulation is known to occur in *Aspergillus* and the insights gained from the available genome sequences.

Keywords *Aspergillus*, gene regulation, promoter elements, MEME, bioinformatics

Introduction

Aspergillus nidulans has played a significant role in developing our understanding of the processes and mechanisms underlying gene regulation. Fundamental concepts relating to promoters, regulatory transcription factors and the coordinated response of different genes to overlapping signals having been elegantly dissected by classical genetics since the 1970s [1]. The legacy is a detailed understanding of various processes and networks such as fungal development and the regulation of primary metabolism. In many instances this has been translated to an understanding of the equivalent systems in organisms of medical and biotechnological importance. The era of molecular biology allowed this fertile field to be further developed but the availability of the genome sequences and the associated technologies should lead to a revolution in our understanding, providing the tools and technologies for global analysis.

Levels of regulation

Generally the initiation of transcription is seen as the most important point at which gene expression is regulated. Significant effort has been focused on characterizing the regulatory proteins that interact at specific promoters to either activate or repress tran-

scription. These proteins can be separated into two classes, the first being those with a wide domain of function, such as AreA which mediates nitrogen metabolite repression [2] or PacC, which mediates the response to ambient pH [3]. Other genes act at very specific groups of genes, for example NirA which mediates nitrate/nitrite induction, only regulate a handful of genes [2]. A variety of these proteins have been characterized and shown to have sequence-specific DNA binding, usually mediated by well-conserved domains found in other species. Simplistically, the binding of these proteins leads to the activation or repression of transcription. However, there are various additional levels of complexity including: (i) how the activity of the protein is modulated; (ii) which other regulatory proteins, co-activators or repressors interact; (iii) how this varies with different conditions and at different promoters; (iv) the specific molecular targets of their activation/repression function; and (v) how their function is affected by the location and specific sequence of the binding element.

The regulation of the genes involved in nitrate utilization is a relatively simple and well-characterized example. Expression of the four known genes involved in nitrate metabolism requires two signals – (i) the limitation of the available nitrogen within the cell, which leads to activation of AreA and (ii) the presence of nitrate, which activates the pathway specific transcription factor NirA. Individually neither protein is able to activate transcription. It is only in the presence of nitrate and general limitation of nitrogen within the cell that the two transcription factors act synergistically and activate transcription. At each promoter a number

Correspondence: Mark X. Caddick, School of Biological Sciences, The Biosciences Building, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK. Tel: 0151 795 4566; Fax: 0151 795 4410; E-mail: caddick@liv.ac.uk

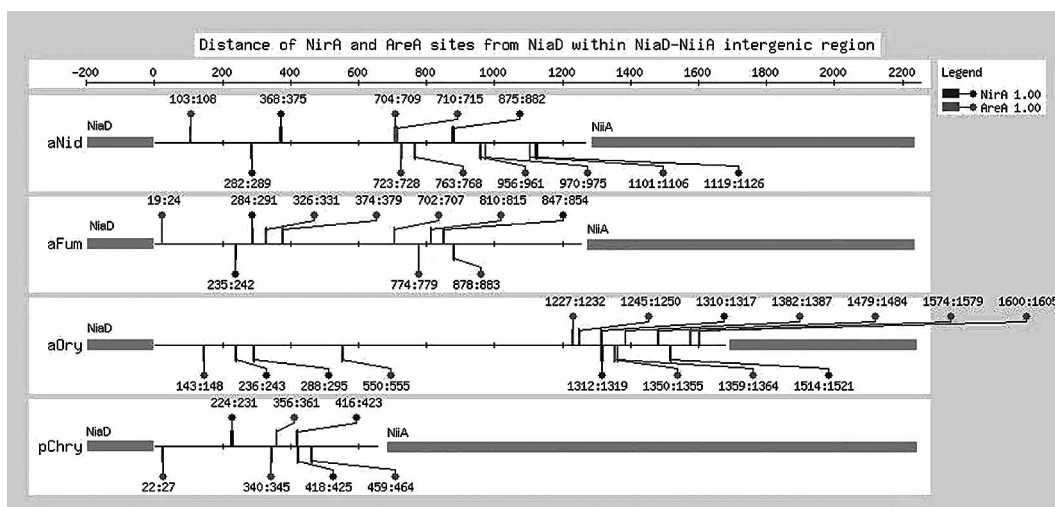


Fig. 1 Location of NirA and AreA binding sites within the *niaD-niaA* intragenic region. The relative position of potential binding sites is indicated. As can be seen, the number, orientation and location is poorly conserved. However, in all three cases the two structural genes are divergently transcribed and multiple binding sites for both factors are present.

of NirA and AreA binding motifs are present (Fig. 1). However, not all these appear to influence transcription directly [4]. Additionally, the DNA is not naked but is associated with nucleosomes, which are re-positioned in order for activation to take place [5]. How NirA is modulated is still unknown, although classical genetics has identified specific regions of the gene involved [6]. NirA is shuttled between the nucleus and cytoplasm and this, as with many transcription factors, is subject to regulation [7]. Modulation of AreA is complex, involving different signalling processes acting at different levels [2]. For example, the intracellular level of Gln determines the level of transcript stability and consequently affects the level of AreA synthesis [8] and also the AreA protein interacts with a repressor protein NmrA [9]. AreA shuttles between the nucleus and cytoplasm and nuclear retention is stimulated by nitrogen starvation [10]. There is also growing evidence that phosphorylation and protein degradation may also play a role in modulation. Finally regulatory proteins such as AreA rarely work in isolation but function as part of a complex. It is known that the co-activator TamA acts with AreA at some genes [11], there is evidence that the repressor protein NmrA, may have a minor role in activation [MX. Caddick, unpublished data] and NirA may interact directly with AreA [12]. In addition to the regulation of transcription the mRNA of all four structural genes is subject to regulation at the level of stability which relates directly to the nitrogen state of the cell [Morozov, Jones, Caddick, unpublished] and there is good evidence that in the case

of NiaD, protein stability is also regulated [13] as is the case in plants.

Another regulatory protein that may be involved is AreB, which shares significant homology to AreA at the DNA binding domain. Consequently it appears to compete with AreA for specific binding sites where it acts as a repressor [14]. Expression of *areB* is itself regulated by AreA and one of its transcripts is regulated at the level of stability in a similar way to *areA* [Morozov, Conlon, Caddick, unpublished data]. *areB* has two other interesting regulatory features. Firstly, it is subject to differential splicing and has multiple promoters. Both are common features in animals and plants but have not up to this point been commonly reported in fungi. However, a significant level of differential splicing has been observed in *Cryptococcus neoformans* [15]. Another feature of *areB* is the apparent use of non-canonical start codons GUG and CUG in place of AUG. The frequency of this across the genome is unknown and the biological consequences poorly understood but probably relates to regulation at the level of translation. From the genome sequence it has been predicted that 21% of genes are regulated by the presence of short upstream open reading frames [16]. There is also likely to be global regulation of translation and post-translational modification is prevalent.

What does the genome sequence tell us?

By definition the genome sequence has all the information required for appropriate gene regulation, but the

problem for the biologist is identifying and interpreting the key features. The first problem is to predict gene position; the transcription and translation start sites, location of splice sites and the 3' end of the transcript being fundamental. Early annotations of *A. nidulans*, *A. fumigatus* and *A. oryzae* included a significant number of errors as the gene finding programmes were not sophisticated enough to interpret the sequence fully. The majority of genes have been identified but include errors, usually in defining the start and intron location, and we would predict that as with the human genome the predicted number of genes should evolve significantly over the coming months and years. The annotation is improving with the careful analysis of predicted genes, searches for genes not initially identified and most importantly cDNA sequences. It is therefore important as a community that the information is shared and the genome annotation continuously revised. However, assuring the resources for this is not trivial and has not yet been achieved.

Despite these shortfalls it is possible to conduct informative analyses. Comparing the sequences upstream and downstream from the predicted open-reading frames in all three species led to the identification of conserved sequence elements adjacent to orthologous genes [16]. In certain cases these short elements are related to motifs recognized by known DNA or RNA binding proteins and/ or the associated proteins had related functions or cellular locations. For example, the CpcA/Gcn4p element was identified as a conserved sequence upstream of a number of genes involved in amino acid transport and metabolism, consistent with the known function of CpcA [17]. Elements enriched in the 3' UTR included the Puf-binding site, and as would be expected from work in *Saccharomyces cerevisiae* these are probably involved in post-transcriptional regulation of genes involved in mitochondrial function. The various novel motifs identified may relate to as yet unidentified binding proteins.

This type of approach does however have limitations. For example an element related to the GATA sequence, which is known to be bound by the AreA transcription factor, was found associated with a number of genes with no apparent common functional role. What we knew prior to this work is that there are multiple GATA factors in the *Aspergilli*, and although those tested bind very similar sequences *in vitro*, they have radically different functions *in vivo*. These include the regulation of nitrogen metabolism, iron sequestration, sexual development, light and circadian responses [2]. It is therefore important to be aware that the function of motifs is context dependent and regulatory proteins can differentiate between them *in vivo* if not *in*

vitro. The most highly conserved sequence within untranslated regions of the *Aspergillus* genomes was the *areA* element which mediates regulated transcript stability [18] suggesting this regulatory process is conserved across all three species. However, recently we have found a number of genes that share this regulation but do not have a related element [Caddick, Jones, Morozov, unpublished data], perhaps suggesting a more complex story is yet to be uncovered.

To investigate promoter elements within the *A. nidulans*-, *A. fumigatus*- and *A. oryzae*-genomes we applied three different approaches. Taking the promoter sequences from orthologous genes we asked: what sequence elements can be found in all three promoters and how well is their positioning conserved? Doing this with the *niiA-niaD* intragenic region, the motif search program MEME [<http://meme.sdsc.edu/meme/intro.html>] (accessed on 10/7/2006) identified conserved elements which relate to the known pathway specific transcription factor, NirA, responsible for signalling the presence of NO_3^- . However, the relative location and orientation within the promoters was not conserved (Fig. 1). Using MEME we were unable to identify the GATA elements bound by AreA. The reason for this disparity relates to the complexity of the two sequence elements involved which relates to their function. NirA is probably only involved in regulating a handful of genes, induced by the presence of nitrate, while AreA regulates hundreds [Caddick, Jones, Morozov, unpublished data] as it has a role in coordinating nitrogen metabolism as a whole.

Taking orthologues for 14 different genes known to be regulated by AreA we were able to predict the AreA binding site [Dobson, Caddick, unpublished data]. Similar analysis also identified the SreA binding site, which is involved in the regulating iron assimilation. It is therefore reasonable to assume that if we take a group of genes that are co-regulated, we will be able to use computational analysis of the promoters to identify the regulatory elements responsible. For example microarray analysis could be used to define the transcriptional response to important factors associated with pathogenicity such as the presence of activated myocytes or oxidative stress. Subsequent computational analysis of the associated intergenic regions could potentially lead to the identification of key regulatory elements responsible for coordinating the cellular response. If such elements can be identified we will then have a handle on identifying key regulators involved, for example by using affinity chromatography to purify the protein which can then be identified by mass-spectrometry.

The third approach we employed was to look for promoter elements that are known functional binding

sites for transcription factors. If a specific motif occurred in the promoters for orthologous genes in all three species we could then determine if this can be used to predict the gene regulation. For example, using the NirA binding site we identified 13 genes. This included all four genes known to be regulated by *nirA*, and these were the only ones which consistently had multiple elements. Interestingly, one of the other genes encodes a putative nitrate reductase. The expression of five of these genes was tested. The novel nitrate reductase failed to show expression suggesting that there is some other level of regulation, for example developmental which restricted the analysis. Two of the remaining four genes were co-regulated by NirA. Similarly, based on sequence analysis we have been able to predict regulation of a gene by PrnA, which induces the proline metabolism cluster [Dobson, Caddick, unpublished observations].

Conclusion

The available genome sequences have facilitated the identification of promoter elements and other important features involved in gene regulation for the three *Aspergillus* spp. However, with respect to promoters the level of conservation is limited with no clear conservation of organisation. In the case of yeast species, the most informative comparisons relate to very close relatives [19], and it will be very interesting whether the inclusion of other sequenced *Aspergillus* genomes will provide the data required in order to confidently identify key elements and predict expression profiles. Undoubtedly, this type of analysis will help us uncover important elements, and develop a more sophisticated appreciation of transcriptional regulation.

References

- Martinelli SD, Kinghorn JR. *Aspergillus*: 50 years on. Amsterdam: Elsevier, 1994.
- Caddick MX. Nitrogen Regulation in Mycelial fungi. In: Brambl R, Marzluf GA (eds). *The Mycota III: Biochemistry and Molecular Biology*, 2nd edn. Berlin: Springer-Verlag, 2004: 349–368.
- Arst HN, Penalva MA. pH regulation in *Aspergillus* and parallels with higher eukaryotic regulatory systems. *Trends Genet* 2003; **19**: 224–231.
- Punt PJ, Strauss J, Smit R, *et al.* The intergenic region between the divergently transcribed *niiA* and *niaD* genes of *Aspergillus nidulans* contains multiple NirA binding-sites which act bidirectionally. *Mol Cell Biol* 1995; **15**: 5688–5699.
- Muro-Pastor MI, Gonzalez R, Strauss J, *et al.* The GATA factor AreA is essential for chromatin remodelling in a eukaryotic bidirectional promoter. *EMBO J* 1999; **18**: 1584–1597.
- Cove DJ. Genetic studies of nitrate assimilation in *Aspergillus nidulans*. *Biol Rev* 1979; **54**: 291–327.
- Berger H, Pachlinger R, Morozov I, *et al.* The GATA factor AreA regulates localization and *in vivo* binding site occupancy of the nitrate activator NirA. *Mol Microbiol* 2006; **59**: 433–446 JAN.
- Morozov IY, Galbis-Martinez M, Jones MG, *et al.* Characterization of nitrogen metabolite signalling in *Aspergillus* via the regulated degradation of areA mRNA. *Mol Microbiol* 2001; **42**: 269–277.
- Andrianopoulos A, Kouirambas S, Sharp JA, *et al.* Characterization of the *Aspergillus nidulans* *nmrA* gene involved in nitrogen metabolite repression. *J Bacteriol* 1998; **180**: 1973–1977.
- Todd RB, Fraser JA, Wong KH, *et al.* Nuclear accumulation of the GATA factor AreA in response to complete nitrogen starvation by regulation of nuclear export. *Euk Cell* 2005; **4**: 1646–1653.
- Small AJ, Todd RB, Zanker MC, *et al.* Functional analysis of TamA, a coactivator of nitrogen-regulated gene expression in *Aspergillus nidulans*. *Mol Gen Genom* 2001; **265**: 636–646.
- Narendja F, Goller SP, Wolschek M, *et al.* Nitrate and the GATA factor AreA are necessary for *in vivo* binding of NirA, the pathway-specific transcriptional activator of *Aspergillus nidulans*. *Mol Micro* 2002; **44**: 573–583.
- Hynes MJ. The effect of lack of carbon source on nitrate reductase activity in *Aspergillus nidulans*. *J Gen Microbiol* 1973; **79**: 155–157.
- Conlon H, Zadra I, Haas H, *et al.* The *Aspergillus nidulans* GATA transcription factor gene *areB* encodes at least three proteins and features three classes of mutation. *Mol Micro* 2001; **40**: 361–375.
- Loftus BJ, Fung E, Roncaglia P, *et al.* The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 2005; **307**: 1321–1324.
- Galagan JE, Calvo SE, Cuomo C, *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 2005; **438**: 1105–1115.
- Hoffmann B, Valerius O, Andermann M, *et al.* Transcriptional autoregulation and inhibition of mRNA translation of amino acid regulator gene *cpcA* of filamentous fungus *Aspergillus nidulans*. *Mol Biol Cell* 2001; **12**: 2846–2857.
- Platt A, Langdon T, Arst HN Jr., *et al.* Nitrogen metabolite signalling involves the C-terminus and the GATA domain of the *Aspergillus* transcription factor AREA and the 3' untranslated region of its mRNA. *EMBO J* 1996; **11**: 2791–2801.
- Kellis M, Patterson N, Endrizzi M, *et al.* Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 2003; **423**: 241–254.