The EMBO Journal Vol.16 No.13 pp.3974–3986, 1997

Subtle hydrophobic interactions between the seventh residue of the zinc finger loop and the first base of an HGATAR sequence determine promoter-specific recognition by the *Aspergillus nidulans* GATA factor AreA

Adriana Ravagnani[1,2], Lisette Gorfinkiel[3,4], Timothy Langdon[1,5], George Diallinas[3,6], Elisabeth Adjadj[7], Stéphane Demais[3], Diana Gorton[8,9], Herbert N.Arst,Jr[1] and Claudio Scazzocchio[3,8,10]

[1]Department of Infectious Diseases and Bacteriology, Royal Postgraduate Medical School, Ducane Road, London W12 0NN, UK, [3]Institut de Génétique et Microbiologie, Unité de Recherche Associé au CNRS, 1354, Bâtiment 409, Université Paris-Sud, Centre d'Orsay and [7]U350 INSERM, Institut Curie, Centre Universitaire, Bâtiment 112, 91405 Orsay Cedex, France and [8]Department of Biology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, Essex, UK

[2]Present address: Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA, UK
[4]Present address: Seccion Bioquímica, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay
[5]Present address: Institute of Grassland and Environmental Research, Plas Gogerddan, Aberystwyth, Dyfed SY23 3EB, UK
[6]Present address: Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, PO Box 1527, Heraklion, 71 110, Crete, Greece
[9]Present address: Enfield and Haringey Health Authority, Alexandra Place, Lower Park Road, London N11 1ST, UK

[10]Corresponding author at the Orsay address

L.Gorfinkiel and T.Langdon contributed equally to this work

A change of a universally conserved leucine to valine in the DNA-binding domain of the GATA factor AreA results in inability to activate some AreA-dependent promoters, including that of the *uapA* gene encoding a specific urate–xanthine permease. Some other AreA-dependent promoters become able to function more efficiently than in the wild-type context. A methionine in the same position results in a less extreme, but opposite effect. Suppressors of the AreA(Val) mutation mapping in the *uapA* promoter show that the nature of the base in the first position of an HGATAR (where H stands for A, T or C) sequence determines the relative affinity of the promoter for the wild-type and mutant forms of AreA. *In vitro* binding studies of wild-type and mutant AreA proteins are completely consistent with the phenotypes *in vivo*. Molecular models of the wild-type and mutant AreA–DNA complexes derived from the atomic coordinates of the GATA-1–AGATAA complex account both for the phenotypes observed *in vivo* and the binding differences observed *in vitro*. Our work extends the consensus of physiologically relevant binding sites from WGATAR to HGATAR, and provides a rationale for the almost universal evolutionary conservation of leucine at the seventh position of the Zn finger of GATA factors. This work shows *inter alia* that the sequence CGATAGagA-GATAA, comprising two almost adjacent AreA-binding sites, is sufficient to ensure activation of transcription of the *uapA* gene.

Keywords: *Aspergillus nidulans*/GATA transcription factors/molecular modelling/promoter recognition

Introduction

The *areA* gene encodes a factor necessary for the efficient transcription of >100 genes involved in the utilization of nitrogen sources in *Aspergillus nidulans* (Arst and Cove, 1973). An isofunctional, homologous gene, *nit-2*, has been described in *Neurospora crassa* (Fu and Marzluf, 1990), while the *GLN3*, *DAL80*, *NIL1* and *NIL2* genes regulate in a less straightforward fashion the utilization of some nitrogen sources in *Saccharomyces cerevisiæ* (Cunningham and Cooper, 1991; Minehart and Magasanik, 1991; Coffman *et al.*, 1995; Stanbrough *et al.*, 1995; D.Rowen, N.Esiobu and B.Magasanik, personal communication). The single DNA-binding domains of these fungal proteins show striking similarities to the zinc fingers and adjacent downstream basic regions of the metazoan GATA factors (Kudla *et al.*, 1990), more particularly to the carboxy-terminus of the two DNA-binding domains. The AreA protein mediates nitrogen metabolite repression by responding to the glutamine concentration in the cell. Glutamine is thought to destabilize the *areA* mRNA through an unknown mechanism, and to prevent, via its interaction with another protein, the binding of AreA to its cognate DNA sites (Platt *et al.*, 1996a). Even within the same pathway, different promoters show varying degrees of sensitivity to nitrogen metabolite repression. Thus, in the purine degradation pathway, the promoters of the permeases coded by the *uapA* and *uapC* genes are exquisitively sensitive to nitrogen metabolite repression, while the expression of the *hxA* and *uaZ* genes is never repressed completely (Oestreicher and Scazzocchio, 1995).

Two mutations that alter differentially the expression of a number of genes controlled by AreA result in substitutions of a universally conserved leucine in the seventh position of the Zn finger loop (Leu683) by valine and methionine respectively (Kudla *et al.*, 1990). Strains carrying the valine substitution (*areA102*) show, *inter alia*, a drastically increased expression of the *amdS* gene, encoding acetamidase, and almost null expression of the genes coding for the specific uric acid–xanthine permease and for the broad specificity purine permease (*uapA* and *uapC* respectively, Arst and Scazzocchio, 1975; Hynes, 1975; Scazzocchio and Arst, 1978; Gorfinkiel *et al.*, 1993; Diallinas *et al.*, 1995). Strains carrying the methionine substitution (*areA30*, *areA31*) have a phenotype which is qualitatively a less extreme mirror image of that of strains carrying the valine substitution (Arst and Scazzocchio, 1975; Gorton, 1983). Strains carrying the methionine

substitution have been isolated as revertants of *areA102* on uric acid as nitrogen source (see below, Arst and Scazzocchio, 1975) A strain with an identical phenotype, isolated in another laboratory (*areA200*, Hynes, 1975), was shown to carry an identical substitution (T.Langdon and H.N.Arst,Jr, unpublished data).

As a consequence of the total loss of uric acid uptake activity, *areA102* strains cannot utilize uric acid as nitrogen source. We have shown that it is possible to obtain revertants which grow on uric acid. The *cis*-acting mutations responsible for the suppression of *areA102* specifically for growth on uric acid are adjacent to either the *uapA* or the *uapC* genes (Arst and Scazzocchio, 1975; Gorfinkiel *et al.*, 1993; Diallinas *et al.*, 1995). We speculated that such revertants may have acquired AreA-binding sites similar to those present in genes which are expressed normally or are overexpressed in the *areA102* strains. One such suppressor, *uapA100* (previously called *uap100*), was shown to affect drastically the expression of the *uapA* gene. In an *areA*$^+$ background, *uapA100* results in constitutivity and overexpression of the *uapA* gene. However, this mutation does not suppress a total loss-of-function *areA* allele, nor a null mutation in *uaY*, the gene coding for the transcription factor mediating specific induction by uric acid. *uapA100* is a 164 bp duplication. This duplication includes the single specific UaY-binding site (Suárez *et al.*, 1995) and a couple of putative, almost adjacent, AreA-binding sites, the sequence being CGATAGagAGATAA. Data from a number of laboratories have shown that while AreA and other GATA factors are able to bind to single WGATAR sites (Merika and Orkin, 1993; Ko and Engels, 1993; D.Gómez, B.Cubero and C.Scazzocchio, unpublished data; A.Ravagnani and H.N.Arst,Jr, unpublished data), physiologically relevant sites occur often in direct or inverted repeats separated by a few base pairs (Martin *et al.*, 1989; Strauss *et al.*, 1992; Feng *et al.*, 1993; Langdon *et al.*, 1995; M.I.Muro, J.Strauss, and C.Scazzocchio, unpublished data). It was shown *in vitro* that low affinity or even WGATTR sites can be bound if nearby there is a strong binding site. This cooperativity is evident even in fusion proteins containing little more than the DNA-binding domain (Feng *et al.*, 1993; D.Gómez, B.Cubero and C.Scazzocchio, unpublished data). Although the consensus for GATA factors is usually written WGATAR, CGATAR sites are recovered at low frequency in random selection experiments (Ko and Engel, 1993; Merika and Orkin, 1993). A physiological function for such CGATAR sites has not been described to date. We speculated that the non-canonical cytosine present in the CGATAGagAGATAA sequence of the *uapA* promoter is indeed the base which results in the inability of the *uapA* promoter to respond to an AreA102 mutated protein. We thus analysed a number of mutations which were obtained as specific revertants on uric acid but that showed a rather different phenotype from *uapA100*. These mutations accommodate the AreA102 (AreAVal) protein without affecting the non-induced levels of *uapA* expression. The NMR solution structure of the GATA-1–DNA complex showed that the conserved leucine in position 7 of the loop of the Zn finger makes a number of hydrophobic contacts, including one with the first base of the AGATAA sequence used in NMR studies (Omichinski *et al.*, 1993). We had speculated that the gene-specific effects of the mutations at position 683 might be due to the loss or strengthening of these hydrophobic interactions (Scazzocchio, 1990).

Results

Selection and genetic characterization of specific suppressors of areA102

We previously have described promoter mutations which suppress the *areA102* mutation specifically for growth on uric acid and xanthine. A number of these are adjacent to the *uapC* locus (Diallinas *et al.*, 1995), but only one, a 164 bp duplication, is a *cis*-acting mutation upstream of the coding region of the *uapA* gene (Arst and Scazzocchio, 1975; Gorfinkiel *et al.*, 1993). We sought to isolate additional suppressors mapping in the promoter of *uapA*. Spores from a *biA1 areA102* strain were plated on medium containing uric acid as sole nitrogen source. Strains able to grow on uric acid were selected either as spontaneous mutants or after UV mutagenesis (15% survival). Wild-type revertants or those able to grow on uric acid because the Val683 has been substituted by a methionine can be distinguished easily from suppressors specific for utilization of uric acid and presumably mapping in the *uapA* or *uapC* promoters by their effects on growth on a number of unrelated nitrogen sources (Arst and Scazzocchio, 1975; Kudla *et al.*, 1990). One spontaneous and seven UV-induced mutant strains carrying uric acid–xanthine specific suppressor mutations were found. Extragenic, specific suppression was shown for these eight strains by recovery, in every case, of 25% *areA102* non-suppressed progeny in crosses with *areA*$^+$ strains. These eight strains were then crossed to *uapA100 areA102* and *uapC201 areA102* strains carrying previously characterized *uapA* and *uapC* promoter mutations respectively. At least 300 progeny were analysed for each cross. Four strains each carry a suppressor mutation unlinked either to *uapA* or *uapC* and were not analysed further. Two strains carry suppressor mutations (designated *uapC305* and *uapC307*) in or very near the *uapC* promoter. These will be described in a separate publication.

Three strains carry suppressors of *areA102* which map in or very near the *uapA* promoter: they are designated *uapA302*, *uapA310* and *uapA349*. In particular, the phenotypically identical mutations *uapA302* and *uapA310* did not yield any *uapA*$^+$ recombinants in, respectively, 1110 and 2680 progeny in crosses in repulsion to *uapA100*. Mitotic haploidization analysis showed the *uapA349* strain to carry a I–VII chromosome translocation. *uapA* maps to chromosome I (Arst, 1988), and thus this translocation might be identical to the suppressor mutation. The extent of suppression afforded by *uapA310* is shown in Figure 1.

Molecular characterization of the uapA promoter suppressors of areA102

The phenotypically identical mutations *uapA302* and *uapA310* consist of a single base pair change in the *uapA* promoter: **C**GATAGagAGATAA is mutated to **T**GATAGagAGATAA (positions from nucleotide 653 to 666 in Gorfinkiel *et al.*, 1993, relevant bases in bold). For *uapA302*, we have sequenced 874 bp upstream of the

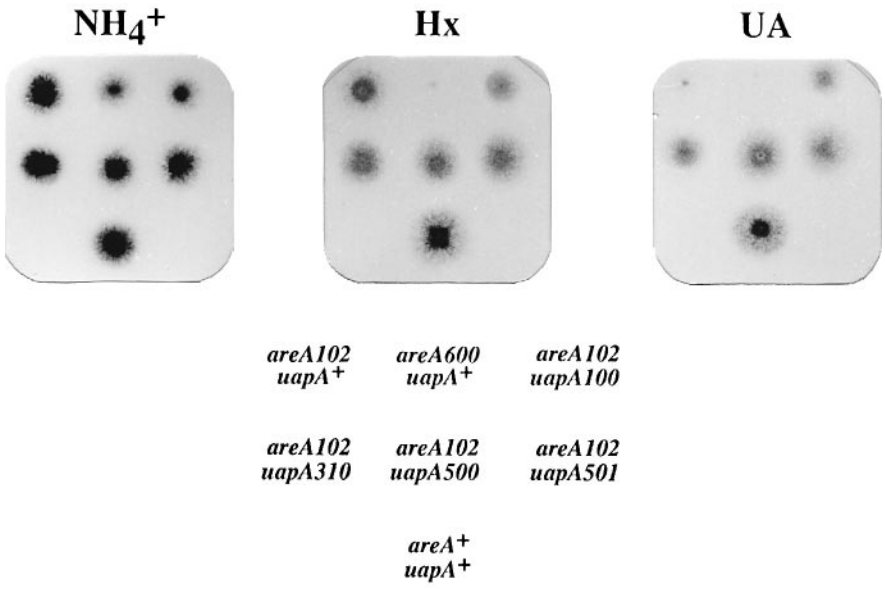

Fig. 1. The specific suppression of the *areA102* mutation by different mutations in the HGATAR sites in the *uapA* promoter is shown. The Petri dishes contain *A.nidulans* minimal medium with 5 mM ammonium (+) tartrate (NH_4^+), 700 μM hypoxanthine (Hx) or 700 μM uric acid (UA) as sole nitrogen sources. The genotype of each strain is indicated. *areA*$^+$ is the wild-type allele (AreALeu), *areA600* is a total loss of function, chain termination mutation (Kudla *et al.*, 1990) and is included as a control, *areA102* codes for the AreA(Val) protein, *uapA*$^+$ is the wild-type promoter (CGATAGagAGTATA). *uapA100* is a 164 bp duplication within the *uapA* promoter containing the CGATAGagAGTATA sequence described previously (Gorfinkiel *et al.*, 1993). *uapA310* carries the **T**GATAGagAGATAA pair of sites, *uapA500* the CGATAGag**T**GATAA pair and *uapA501* the **T**GATAAag**T**GATAA pair of sites (see text). Mutant bases are in bold. Mutation *uapA501* contains, in addition to the substitutions in the first base of the HGATAR sites, an inversion between nucleotides 452 and 802 of the *uapA* promoter (numbering as in Gorfinkiel *et al.*, 1993). This has presumably occurred because a very similar 9 bp sequence flanks these nucleotides. Despite several attempts, it was not possible to obtain a transformant which did not also carry this inversion. Note that this inversion inverts the relative orientation of AreA- and UaY-binding sites in relation to the *uapA* transcription startpoint.

uapA initiation codon**.** For *uapA310*, 160 bp comprising the AreA- and UaY-binding sites were sequenced. It was shown by transformation of a suitable *areA102* strain with the 160 bp mutated fragment that the mutation found was sufficient to suppress the *areA102* phenotype on uric acid and that the degree of suppression was identical to that found in the original *uapA302* and *uapA310* stains (Figure 1). No change was detected in the cognate region of *uapA349*. Southern blots hybridized with a number of probes upstream from the *uapA* gene showed a translocation breakpoint ~3.2 kb from the transcription start point (results not shown). This translocation is presumably associated with the weak suppression of *areA102* in strains carrying *uapA349*. As this work concerns specifically the GATA-binding sites, this translocation was not analysed further.

Levels of uapA transcripts in strains carrying point mutations in the uapA promoter

These are shown in Figure 2. The weak suppression of *areA102* by the two identical point mutations is clear, and is similar to the degree of suppression conferred by the *uapA100* duplication (Figure 5 of Gorfinkiel *et al.*, 1993). However, while the duplication results in constitutivity and considerably increased expression in an *areA*$^+$ background (Arst and Scazzocchio, 1975; Gorfinkiel *et al.*, 1993), *uapA310* does not result in constitutivity and, in fact, somewhat decreases *uapA*-induced expression in an *areA*$^+$ [AreA(Leu)] background.

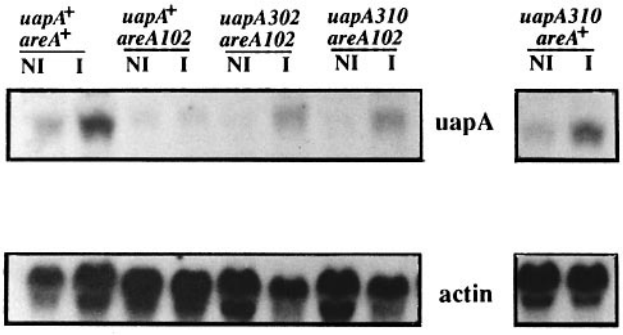

Fig. 2. Northern blots showing the steady-state levels of *uapA* mRNA in *areA*$^+$, *areA102* and suppressor mutations located in the *uapA* promoter as described in the text. In the right hand panel, the steady-state levels for a *uapA310* single mutant are shown. All procedures were carried out in parallel and the RNAs run in the same gel. NI, mycelia grown under non-inducing conditions, I mycelia grown under induced conditions as described in Materials and methods.

In vitro binding of wild-type and mutant AreA fusion proteins to the uapA promoter

The binding of a His fusion protein containing the AreA DNA-binding domain (residues 468–729) to the relevant region of the *uapA* promoter was investigated. Figure 3 (top panel) shows that the AreA(Leu) protein binds specifically to a 162 bp probe derived from the *uapA* promoter and containing the CGATAGagAGATAA sequence. A mutant probe in which both sites have been mutagenized to CG**GG**AGag**C**G**GG**AA (substituted bases in bold) does not bind to AreA(Leu), nor is it able to

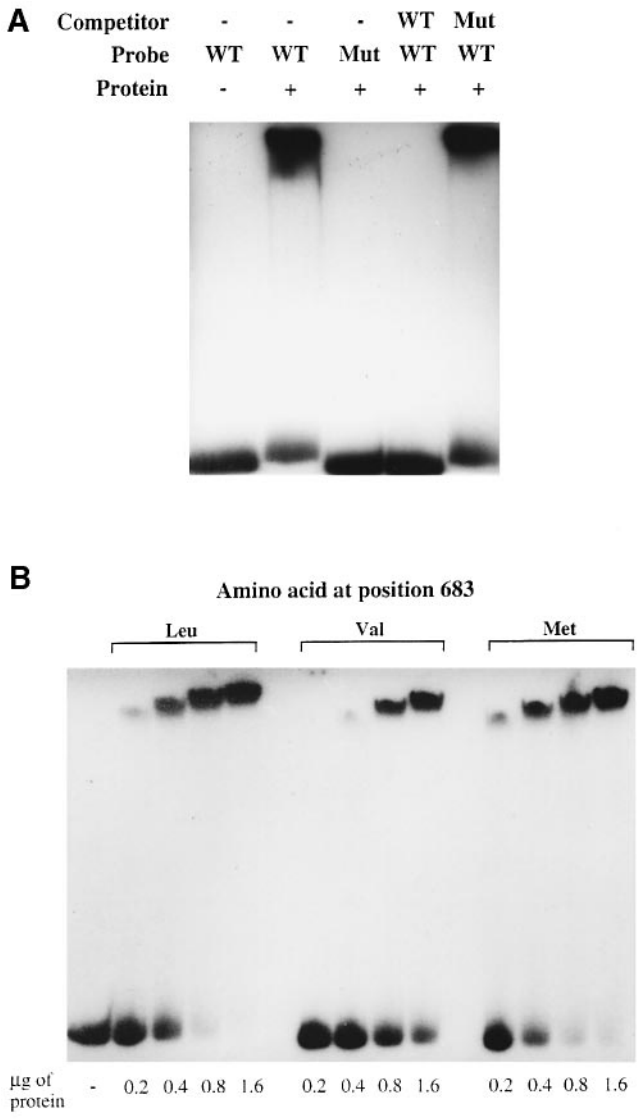

Fig. 3. (**A**) Specificity of binding of the AreA(Leu) protein. Four hundred ng of a His-AreA(Leu) protein were incubated with a 162 bp *uapA* promoter probe, prepared by PCR following the standard conditions indicated in Materials and methods and comprising the CGATAGagAGATAA sites, or with the same probe mutated in the cognate sites (mutated to CG**GG**AGag**C**G**GG**AA as described in Platt *et al.*, 1996b, see Materials and methods; mutagenized bases in bold). The presence and absence of the His-AreA protein is indicated as '–' and '+' respectively. Competition was with 100 times excess of cold probe. WT designates the wild-type and Mut the mutant probe. (**B**) Relative affinities of the AreA(Leu), AreA(Val) and AreA(Met) protein for the 162 bp *uapA* promoter region containing the 5′CGATAGagAGATAA sequence. '–' indicates probe incubated without protein, 'Leu', 'Val' and 'Met' the three different His-AreA fusion proteins used. Increasing quantities of proteins are indicated at the bottom of the figure.

compete with the binding of a wild-type probe. Figure 3 (bottom panel) shows that the wild-type probe described above has less affinity for the AreA(Val) protein than for the wild-type AreA(Leu) or AreA(Met) proteins.

Figure 4 (left panel) shows DNase I footprints of the non-template strands of the wild-type *uapA* promoter, a promoter carrying the *uapA310*/*uapA302* C→T substitution in the first base of the (upstream) CGATAG site and a promoter carrying an A→T mutation in the first base of the (downstream) AGATAA site. The striking result is the considerable loss of protection of the upstream site seen with AreA(Val) protein in the CGATAGagAGATAA promoter. The protection is recovered using both the **T**GATAGagAGATAA and the CGATAGag**T**GATAA probes (mutant bases in bold type). Note that improved binding of the AreA(Val) protein to the downstream site of the latter probe restores protection to the pattern observed with the AreA(Leu) protein. This is consistent with a cooperative effect between the AreA molecules occupying the two sites. Interestingly, both mutant probes show additional protected bases with the AreA(Val) mutant. The first bases of the wild-type probe (the G opposite the C in the first position of the upstream site and the T opposite the A in the first position of the downstream site) are not protected by the AreA(Leu), AreA(Val) or AreA(Met) proteins. The A opposite the first T of the upstream site in the **T**GATAGagAGATAA mutant probe and the A opposite the T in the downstream site of the CGATAGag**T**GATAA mutant probe (relevant bases in bold type) show clearly increased protection by the AreA(Val) protein. This implies that the AreA(Val) protein binds to TGATAR sites more tightly than the AreA(Leu) or AreA(Met) proteins and/or binds in a conformation resulting in extended protection. The footprint of the template strand (Figure 4, right panel) shows a slight but uniform decrease in protection of the CGATAGagAGATAA and **T**GATAGagAGATAA probes by the AreA(Val) protein when compared with that afforded by the AreA(Leu) and AreA(Met) proteins. For the latter probe, the T in the fourth position of the downstream site is less protected by the AreA(Val) protein than by AreA(Leu) and AreA(Met). For the CGATAGagTGATAA probe, the G in the second position of the downstream site is sensitive to DNase I in the presence of the AreA(Leu) and AreA(Met) proteins but is clearly protected by the AreA(Val) protein. These subtle changes are consistent with conformational differences between the different AreA–DNA complexes.

The effect of a T in the first position of an HGATAR site (where H stands for A, T or C) is shown clearly by the dUTP interference experiment of Figure 5. In this method, random substitution of thymines by uracil residues is used to investigate the role of a methyl group in position 5 of the pyrimidine ring in a given protein–DNA interaction (Pu and Struhl, 1992). A striking interference is seen for the TGATAGagAGATAA probe with the AreA(Val) protein. Thus, a 5-methyl group in the first base is only essential when a valine is the amino acid in position 7 of the Zn finger loop. Moreover, there is clear preferential binding with the uracil-substituted probe for the AreA(Leu) and AreA(Met) proteins (comparing the lanes corresponding to the bound and unbound probes). These results show clearly that a methyl group on carbon 5 facilitates the binding of AreA(Val) while it interferes with the binding of AreA(Leu) and perhaps, even more strongly, AreA(Met).

Interaction between the hydrophobic amino acid in position 7 and the first base pair of an HGATAR sequence

The results obtained by changing systematically the first base of the CGATAG sequence in two different contexts are shown in Figure 6. In one context, the downstream binding site is the AGATAA wild-type sequence, in the other this sequence was mutated to TGATAA. We have

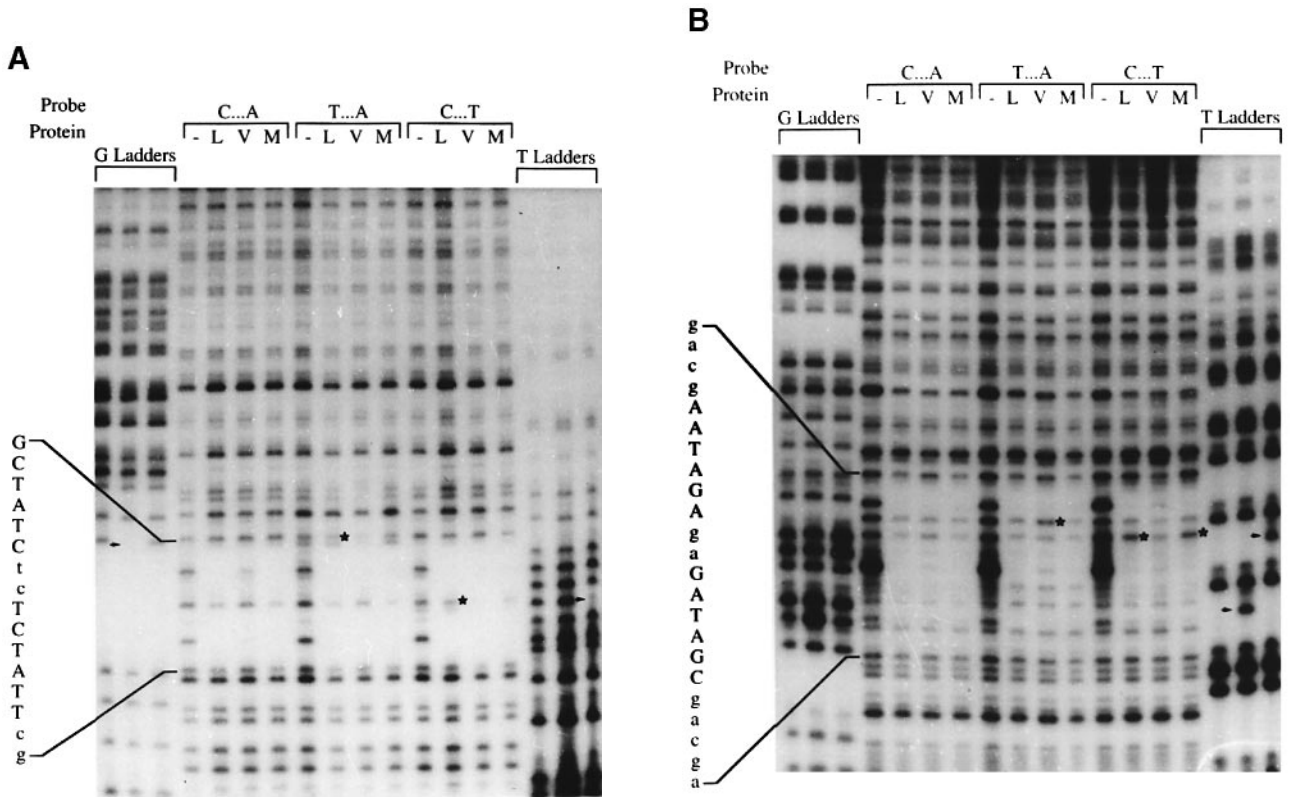

Fig. 4. DNase I protection of the 162 bp *uapA* promoter sequence containing two HGATAR sites. 'Probe' indicates whether the sites were **C**GATAGag**A**GATAA (wild-type promoter), **T**GATAGag**A**GATAA (*uapA302* and *uapA310* mutations) or **C**GATAGag**T**GATAA (*uapA500* mutation, see below). Only the first bases of the HGATAR sites, here indicated in bold, are shown. 'G ladders' and 'T ladders' show the cognate Maxam and Gilbert (1977) reaction for each of the probes respectively from left to right. DNase I digestions of each probe, in the absence of any protein and in the presence of AreA(Leu), AreA(Val) and AreA(Met), are indicated by –, L, V and M respectively. (**A**) Non-template strand. The sequence on the side is the protected sequence (complement of CGATAGagAGATAAgc). The arrows indicate the positions of the mutations, visible as the absence of a band, respectively in the G and T ladders. The stars at the left of weak or missing bands indicate the extended protection conferred by AreA(Val) to the mutant probes as discussed in the text. (**B**) Template strand. The protected sequence, including the CGATAGagAGATAA sequence, is shown on the side. The arrows indicate the mutations, visible in this strand as additional Ts. The stars to the right of bands indicate the specific loss of protection as described in the text.

investigated qualitatively, in conditions of excess probe, the binding of AreA(Leu), AreA(Val) and AreA(Met) to these different probes. To investigate the importance of a 5-methyl (pyrimidine) group in the first position of the upstream site, we have constructed, in addition to probes with thymine at this position, also probes with 5-methylcytosine.

Complexes of two different mobilities are seen. We shall assume in the discussion which follows that the high mobility complex contains one molecule of AreA protein attached to the probe and the low mobility complex contains two molecules attached to the probe. If this assumption is correct, a number of conclusions follow.

Firstly, there is no AreA binding to GGATAR sites, in other words, G in the first position is not acceptable. Independent evidence for the latter has been obtained through a detailed analysis of the *prnD–B* intergenic region (D.Gómez, B.Cubero and C.Scazzocchio, unpublished data). With a GGATAGagAGATAA probe, the AreA(Leu) and AreA(Met) proteins show binding to only one site while the AreA(Val) protein shows no binding whatsoever. This is consistent with very weak binding of AreA(Val) to AGATAR sites. An almost opposite situation is seen for the GGATAGagTGATAA probe. The well defined, high mobility complex seen with AreA(Val) implies that it binds only to the downstream site.

It is, however, striking that AreA(Val) gives only the low mobility complex with the wild-type CGATAGagAGATAA probe. The affinity is reduced drastically, but both sites are occupied. This is even true for the traces of complex formed with the AGATAGagAGATAA probe. Overexposure of the gels shown in Figure 6 revealed the low mobility complex clearly, but failed to show the high mobility complex for these two probes (not shown). These results are perfectly consistent with the DNase I protection experiments. Modification of the pattern on one strand, while the pattern of protection of the other strand is not modified, is consistent with a change of conformation of the complex, leading presumably to an increase in the dissociation constant, rather than to occupation of only one site. The gel retardation and DNase I footprinting results with the mutant TGATAGagAGATAA probe are both perfectly consistent with tight binding of the AreA(Val) protein to both HGATAR sites. There is, however, an apparent contradiction. The AreA(Val) protein shows only the high mobility complex with the CGATAGagTGATAA probe. When the longer version (162 bp) of the same probe was used, only the low mobility complex was seen (not shown) and protection of both sites was observed in DNase I footprinting (Figure 4). A footprint carried out with AreA(Leu) and AreA(Val) and the 52 bp CGATAGagTGATAA probe showed occupancy of both GATA

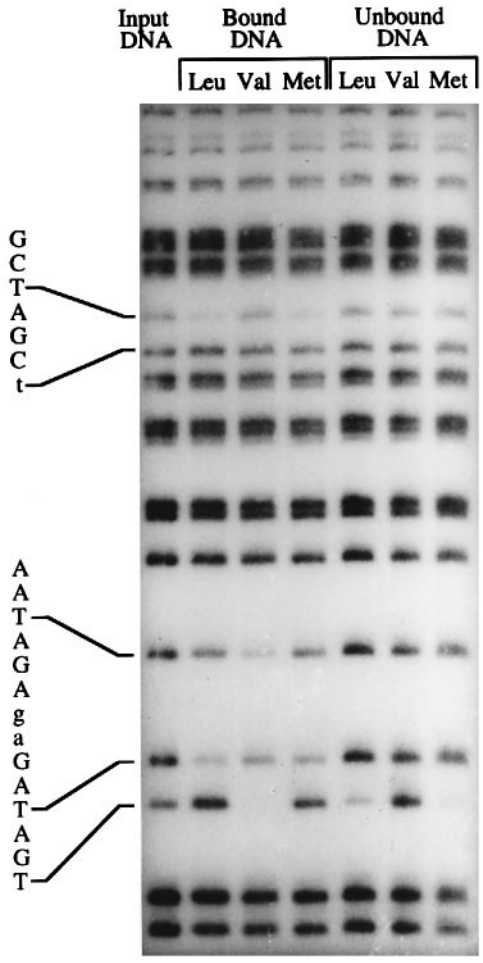

Fig. 5. dUTP interference of a *uapA* promoter probe containing a TGATAGagAGATAA mutant sequence. Leu, Val and Met designate respectively the amino acid present at the seventh position of the Zn finger loop. Differential interference is seen for the T in the first position of the upstream site (see text), while interference for the internal Ts is seen with all three proteins. Note that this gel also shows binding to a non-canonical CGATCG site 27 bp downstream from the two canonical sites. A clear, albeit partial interference is seen for all proteins, for the T in position 4 of the site. This binding and the interference of internal Ts of the canonical GATA sites constitute suitable internal controls for this experiment.

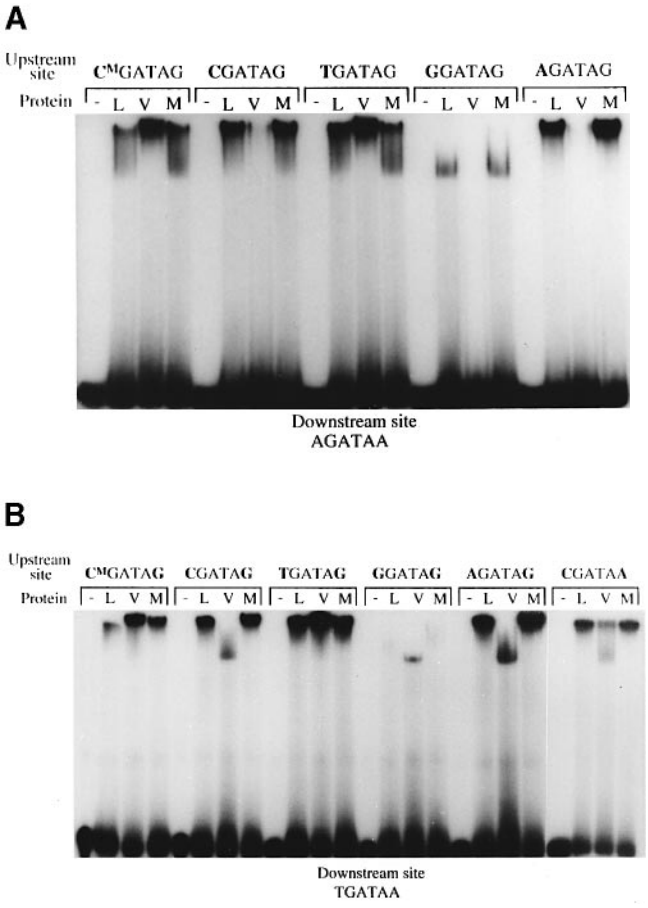

Fig. 6. Systematic variation of the first base of the upstream GATA site (5′CGATAG) of the *uapA* promoter. In all experiments, 52 bp probes, derived from the wild-type promoter as described in Materials and methods, were used. The probes were all prepared using a standard protocol and thus are in approximately equimolecular amounts and in excess for the binding reaction with all proteins. Mutations were carried out in (**A**) wild-type (AGATAA) downstream site context and (**B**) in a mutant (TGATAA) downstream site context. The bases mutated in one or more of the different probes are shown in bold. C^M indicates 5-methylcytosine Each group of four tracks corresponds to a given probe, '–' indicates absence of protein and L, V and M, fusion proteins with leucine, valine and methionine, respectively in the seventh position of the Zn finger loop. An excess of labelled probe was incubated with 400 ng of protein in each case.

sites for both proteins (results not shown). The DNase I treatment is carried out in solution, before the separation of the bound and unbound probes. These results strongly suggest that an unstable two-site complex is formed in solution but that the AreA molecule bound to one of the two sites (presumably the CGATAG site) dissociates during electrophoresis. The instability of the complex formed with the 52 bp probe accounts also for the upward smears observed in the gel retardation experiments. Thus the contradiction between the set of experiments in Figure 6 (52 bp probes) and Figure 4 (DNase I footprints, 162 bp probes) is only apparent and reflects a lower stability of the protein–DNA complexes obtained with the shorter probes.

The presence of a methyl group, whether in thymine or 5-methylcytosine, in position 1 of either or both sites drastically improves the binding of the AreA(Val) protein. 5-Methylcytosine in the first position of the site diminishes the binding of the AreA(Leu) and, to a lesser extent, AreA(Met) proteins. The Ara(Met) protein binds somewhat less than the AreA(Leu) protein to the TGATAGagAGATAA probe. Another subtle difference between the AreA(Leu) and AreA(Met) proteins is that the latter binds somewhat better to both the AGATAGagAGATAA and the AGATAGagTGATAA probes.

We have noted above that TGATAGagAGATAA is not equivalent to AGATAGagTGATAA. With the AreA(Val) protein, a low mobility complex is formed in the first case while a fast mobility complex (albeit with an upward smear) is seen in the second case. This might be due to the difference in the base in the sixth position between the upstream and downstream sites. That the base in the sixth position can influence the relative affinities of binding sites is supported by the difference observed between a CGATA**G**agTGATAA and a CGATA**A**agTGATAA probe (relevant bases in bold). With the former, we see only the high mobility complex, while with the latter we see mainly the low mobility complex.

Phenotypes of strains carrying CGATAGagTGATAA and TGATAGagTGATAA mutations in the uapA promoter

The binding experiments described above establish that both CGATAR and AGATAR sequences bind poorly to the AreA(Val) protein and that a mutation from CGATAGag**A**GATAA to CGATAGag**T**GATAA is sufficient to restore the binding pattern seen with the AreA(Leu) protein. We have therefore constructed, by *in vitro* mutagenesis followed by transformation, strains carrying a CGATAGag**T**GATAA (to be called *uapA500*) or a **T**GATAGag**T**GATAA mutation (to be called *uapA501*) in the *uapA* promoter (mutant bases in bold type). Figure 1 shows that these mutations are able to suppress the uric acid phenotype of an *areA102* mutation somewhat better than the **T**GATAGagAGATAA mutation previously described. It is clear that a thymine in the first position of either or both sites affords a clear suppression of the phenotype of *areA102*(Val) on uric acid and xanthine (only uric acid shown). This experiment and the phenotype of mutations *uapA302* and *uapA310* show that, while the sequence CGATAGagAGATAA is unable to promote efficient transcription in a AreA(Val) context, mutation in the first base of either (or both) HGATAR site to T results in levels of expression of *uapA* sufficient to allow growth on uric acid as sole nitrogen source.

Phenotypes of strains carrying other amino acid substitutions in position 7 of the Zn finger loop

The *in vitro* binding experiments showed that the AreA-(Val) protein binds poorly to both CGATAR and AGATAR sites. We have constructed, by *in vitro* mutagenesis and transformation, strains having alanine, threonine, isoleucine or asparagine, respectively, in position 7 of the AreA Zn finger loop. These strains were tested on a number of nitrogen sources (including ammonium, nitrate, nitrite, uric acid, L-proline, L-arginine, acetamide, acrylamide, L-histidine, L-citrulline and urea). The threonine and alanine substitutions behave like the valine substitution, as expected (see below). The isoleucine substitution is an interesting case. It results in slightly better growth than the wild-type (leucine) on nitrogen sources such as acrylamide and citrulline, which are favoured by the valine, threonine and alanine substitutions but, unlike them, hardly affects growth on uric acid and xanthine. The asparagine substitution strain grows somewhat less than the wild-type strain on all nitrogen sources tested except ammonium, but does not show striking differential effects. Some examples of the phenotypes are shown in Figure 7.

Modelling the interactions between the wild-type and mutant AreA Zn fingers and an HGATAA sequence

The high level of sequence identity of the DNA-binding regions between chicken GATA-1 and AreA permitted direct use of the atomic coordinates of the GATA-1–DNA complex (Omichinski *et al.*, 1993; pdb1gat.ent Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973) to model the wild-type AreA(Leu)–AGATAA complex (Figure 8).

Most of the important specific contacts in the GATA-1–AGATAA complex are of a hydrophobic nature. In particular, the conserved leucine in the seventh position

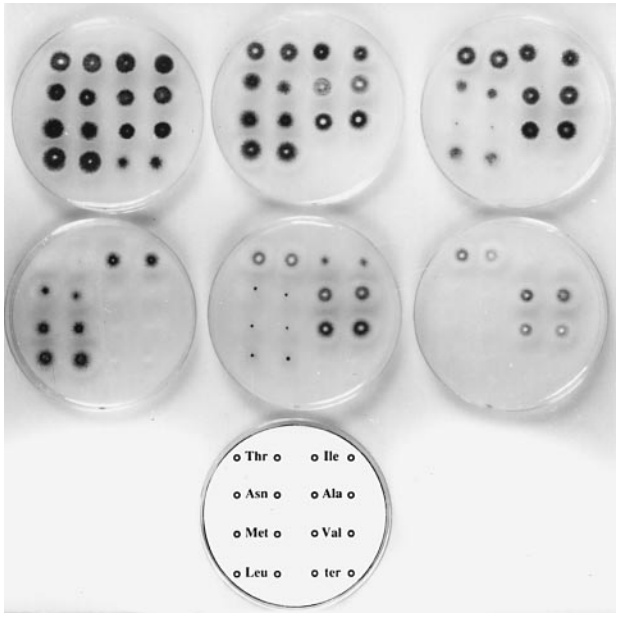

Fig. 7. Phenotypes of strain carrying substitutions in the seventh position of the Zn finger loop of AreA. 'Leu' indicates the wild-type *areA* gene, 'ter' indicates a strain carrying a total loss of function, chain termination mutation in the *areA* gene (*areA600*, Kudla *et al.*, 1990). Other substitutions are as indicated. The plates contain *A.nidulans* minimal medium suplemented with the following nitrogen sources: from left to right, top row ammonium [as the (+)-tartrate], nitrate (as the sodium salt), acetamide; bottom row, uric acid, citrulline, acrylamide. The different nitrogen sources are used at the standard concentrations (Arst and Cove, 1973).

of the zinc finger loop makes hydrophobic contacts with adenine-1, guanine-2 and the thymine opposite to adenine-3 (Omichinski *et al.*, 1993; Clore *et al.*, 1994). In order to provide a structural rationale for the differences in affinity and specificity observed both *in vivo* and *in vitro*, the binding of AreA(Leu), AreA(Val) and AreA(Met) with the AGATAA sequence was modelled. Substitutions of the first adenine by cytosine or thymine were studied in the same way. To remove steric hindrance effects resulting from these substitutions, the model corresponding to each complex was subjected to 100 conjugate gradient minimization steps. No water molecules were included in the modelling as these should be excluded from the protein–DNA interface (Clore *et al.*, 1994). The models obtained are shown in Figure 9. Each change results in obvious modification of the van der Waals distances between the methyl group(s) of the amino acid side chain and the base in position 1.

An examination of these models shows a very satisfying correlation of the specificity and affinity differences described in other sections of this article with the hydrophobic packing of the residue in the seventh position of the zinc finger loop and the first base of the HGATAA sequence. A striking result is the shortening of the van der Waals distances in the AreA(Val)–TGATAA complex, which accounts nicely for the experimental findings. The drastic effect of a valine substitution on the binding to an AGATAA sequence is neatly explained by this model. It is less straightforward to account for the *in vivo* and *in vitro* results obtained with CGATAR sites and the AreA(Val) protein as the lengthening of the van der Waals distances is not particularly drastic. The explanation of

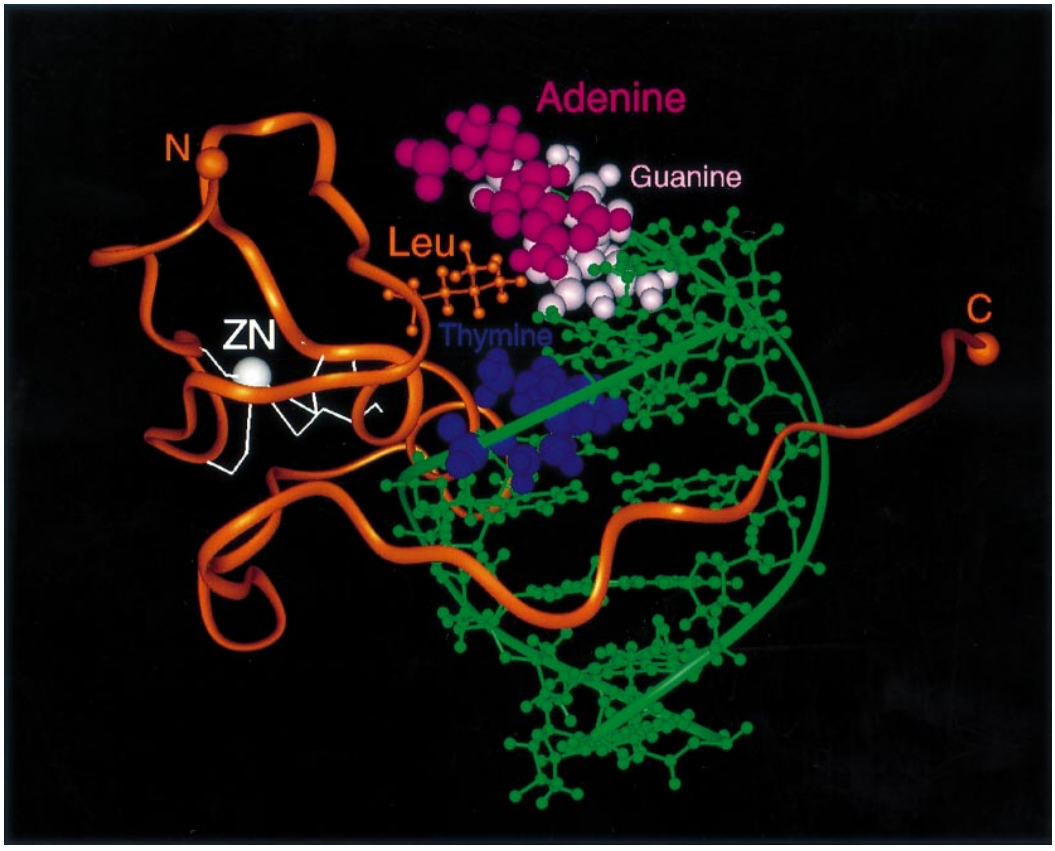

Fig. 8. Model of the wild-type AreA(Leu)–AGATAA complex as derived from the solution structure of the GATA-1–AGATAA complex (Omichinski *et al.*, 1993). The protein backbone is shown as a ribbon representation in orange. The DNA double helix is shown in green. The side chain of the leucine in the seventh position of the Zn finger loop is shown. The side chain of this leucine is within the van der Waals distance of three bases of the major groove of the DNA. These are shown in a space filling form in different colours as indicated. The adenine and guanine are respectively the first and second bases in the AGATAA sequence, the thymine is the base complementary to the adenine in the third position.

this apparent disagreement between the model and the experimental results probably lies in the observation of Clore *et al.* (1994) that water molecules are excluded from the interface between the CD1 methyl group of the leucine and the adenine in the first position in the GATA-1–AGATAA complex. The modest increase in distance between valine and cytosine (and, *a fortiori*, adenine) might well be sufficient to allow water molecules in the interface and therefore to prevent favourable hydrophobic interactions.

The strengthening of the hydrophobic interactions in the AreA(Met)–AGATAA and AreA(Met)–CGATAA complexes is also predicted. The spatial arrangement shown for the AreA(Met)–TGATAA complex is the only one permitted by the minimization procedure, other positions of the methyl group of the methionine being sterically impossible. Thus the possibility of hydrophobic interactions between the methyl groups of thymine and methionine is almost nil.

Leucine appears to be the only amino acid capable of forming a strong hydrophobic bond with all three bases found in the first position of physiological HGATAR sites, which perhaps accounts for its universal occurrence in the most abundant sub-class of GATA factors.

Discussion

We previously have suggested, that '*areA* binding sites might exist in discrete classes related to each other by changes involving only a few base pairs...Pairs of *areA* alleles with antisymmetrical properties and their respective specific initiator suppressors could well provide the means to identify and classify initiator sites in an eukaryote by purely genetic means.' (Arst and Scazzocchio, 1975). This work shows that the response of promoters to a set of mutant GATA factors depends on the first base of a HGATAR DNA-binding sequence.

This work provides qualitative evidence that different combinations of HGATAR sites and residues at the seventh position of the Zn finger loop result in differences in binding strength of the AreA–HGATAR complexes. The agreement between the results *in vivo*, the binding of AreA proteins *in vitro* and the modelling of the different AreA–HGATAR complexes is excellent. While we have not analysed in detail the kinetics of binding of the different proteins to different probes, or established which kinetic parameter(s) is affected by the various mutations, it is clear that valine only works well when a methylated base is present in the first position. It is satisfying that suppressor mutations obtained long before evidence that *areA102* results in a Leu→Val change (Gorton, 1983) turned out to be C→T transitions in the first position of an HGATAR site. The effects of valine in the seventh position on the *in vivo* phenotypes are drastic; those of methionine are qualitatively mirror images of those of valine but are more subtle [see Figure 7, note the poor growth on acetamide of the AreA(Met) strain]. Our

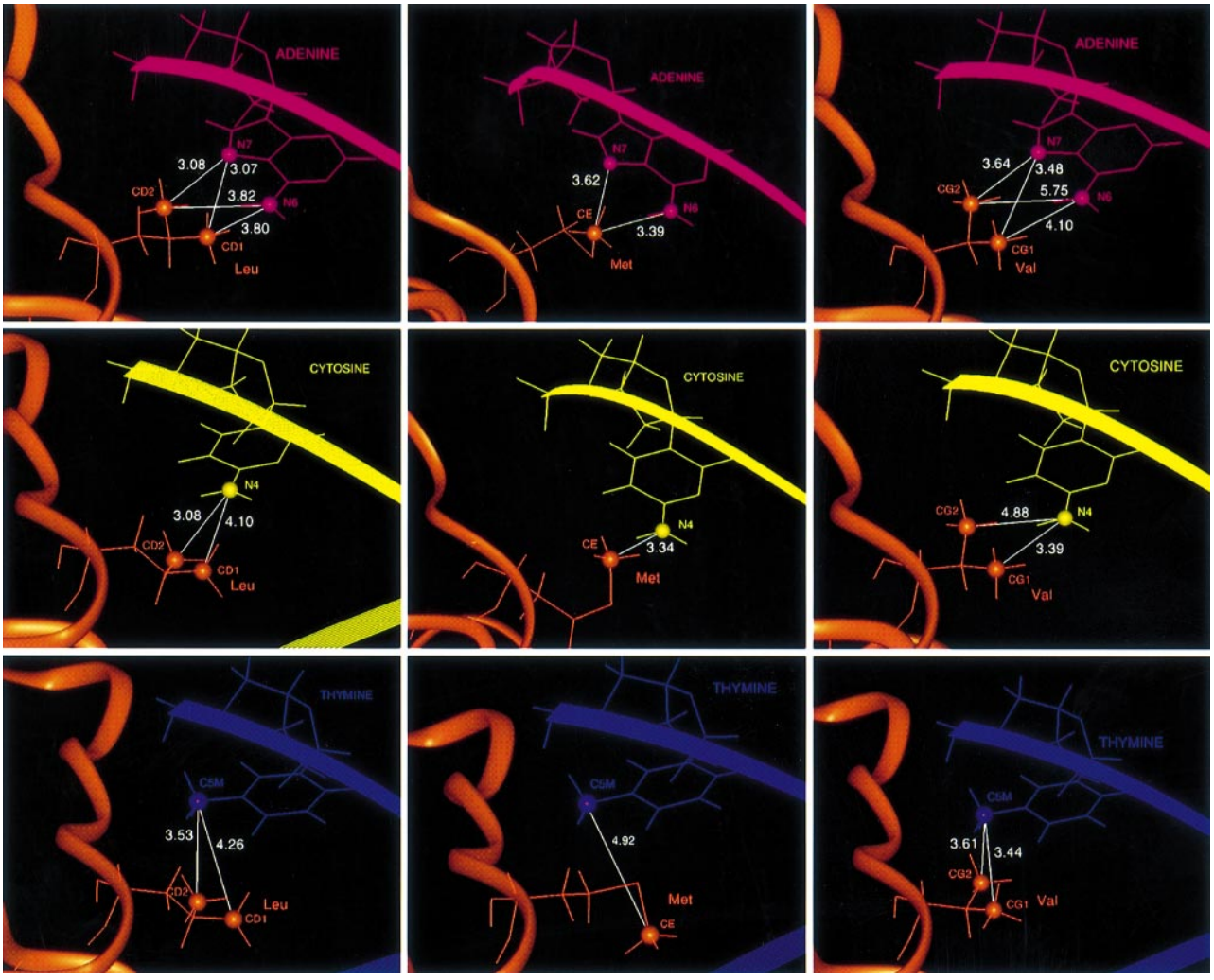

Fig. 9. Systematic variation of the van der Waals distances obtained from specific substitutions in the model shown in Figure 1. The relevant distances between the amino acid in position seven of the Zn finger and the first base of the HGATAA sequence are indicated**.** The three possible HGATAA sequences are shown in complexes with AreA-binding domains containing leucine, valine or methionine in the seventh position of the Zn finger loop.

qualitative binding studies show relatively little difference between the AreA(Leu) and AreA(Met) proteins (for example with the TGATAGagAGATAA and the C^{M}GATAGagTGAGAA probes in Figure 6). Subtle differences are apparent also in the dUTP interference experiment. The differences noted are consistent with the *in vivo* phenotypes.

The amino acid which is changed in *areA102* and *areA30* (and *31*) is strictly conserved among all characterized GATA factors having 17 amino acid Zn finger loops. This is true for both the amino- and carboxy-terminal fingers of those GATA factors of metazoans and fungi which have two Zn fingers. Omichinski *et al.* (1993) have shown that the leucine residue contributes substantially to specific DNA binding with three different hydrophobic interactions. Our work shows that the length of the hydrophobic chain of loop residue 7 is the parameter which determines whether a given AreA protein will prefer AGATAR, CGATAR or TGATAR sites. Model building based on published atomic coordinates for the GATA-1–AGATAA complex provides a sound structural rationale for the *in vivo* and *in vitro* experimental results.

We show for the first time that CGATAR sites are used *in vivo* in a GATA factor-responsive promoter, and that the wild-type AreA(Leu) is able to accommodate all HGATAR sequences. This might have an interesting implication. The C of a CGATAR sequence can potentially be methylated. 5-Methylcytosine at the first position clearly impairs the binding of an AreA(Leu) protein. This observation suggests a novel and unexpected possible mechanism to down-regulate GATA-factor-dependent promoters in a variety of organisms including metazoans.

It was unexpected that substitution of the hydrophobic leucine by hydrophilic asparagine would allow at least some activation of virtually all promoters subject to AreA control. Further work should disclose whether new hydrophilic interactions enable the binding of mutant GATA factors having a hydrophilic amino acid at position 7. A very restricted number of natural GATA factors have a hydrophilic amino acid in the seventh position of the Zn finger loop, but all of these have an 18 residue Zn finger loop, with an additional basic amino acid present at position 16. No details on their mode of binding have been reported (Ballario *et al.*, 1996).

The CGATAGagAGATAA sequence is 31 bp upstream from the *uapA* transcription startpoint. In the *uapA* promoter there is only one other canonical GATA site (AGATAG), and it is 391 bp upstream from the *uapA* transcription startpoint. It is not known whether this site has any physiological importance. This site is presumably inactive in an AreA(Val) context and thus the mutations described here demonstrate the crucial and sufficient role of the CGATAGagAGATAA pair of binding sites. Mutation of either (this work) or duplication of both of these sites (Gorfinkiel *et al.*, 1993) is sufficient to allow at least some *uapA* transcription in an AreA(Val) context. It should also be noted that these sites are 35 bp downstream from the only binding site for the UaY protein, which mediates pathway-specific induction (Suárez *et al.*, 1995, see below). We have characterized another gene which is also virtually not expressed in an AreA(Val) context. This is the *uapC* gene encoding a wide specificity purine permease. The structure of its promoter is very similar to that of *uapA*. There are three closely spaced MGATAA sites downstream (17 bp for the site nearest) from a unique canonical UaY-binding site. There is also a single TGATAG site 244 bp upstream of the UaY-binding site. Presumably, one or more of the MGATAA sites are relevant to *uapC* transcription while the upstream TGATAG site is not (see below; see Diallinas *et al.*, 1995 for the sequence of the *uapC* promoter).

areA102, the Leu→Val substitution, was selected for growth on acrylamide as sole nitrogen source (Hynes and Pateman, 1970). Acrylamide is not utilized by the wild-type but its utilization is enabled by overexpression of the *amdS* gene, coding for acetamidase. It is of interest to examine HGATAR sites in the *amdS* promoter, the prediction being that at least one relevant physiological site should be a TGATAR site. This examination reveals an interesting paradox. Hynes *et al.* (1988) have proposed that the AreA-binding site(s) of *amdS* are downstream of nucleotide –111 (numbering from the initiation codon). Transformants lacking sequences upstream of –111 were still dependent on AreA for utilization of acetamide. However, no HGATAR site lies downstream of nucleotide –111. There are only three such sites from position –111 to position –650, and all three belong to the TGATAR class. We can propose one resolution of this paradox: one or more of the TGATAR sites is the physiological site(s) for *amdS* expression. This is also consistent with the poor growth on acetamide of strains carrying the AreA(Met) substitution. The transformants of Hynes *et al.* (1988) were selected for growth on acetamide in a recipient deleted for the *amdS* gene and contained the transforming *amdS* sequences inserted at unknown places in the genome. The selection method necessitates that the gene is transcribed. That is possible if it is placed next to a constitutive promoter (which would have resulted in AreA independence) or suitably positioned HGATAR sites, which are not rare in the genome. Crosses of these transformants with strains carrying the *areA102* and *areA30* alleles would have revealed the nature (whether CGATAR or AGATAR as opposed to TGATAR) of these sites. Recent results (M.Hynes, J.A.Sharp and M.A.Davis, personal communication) are completely consistent with this interpretation, and demonstrate a physiological role for the most downstream TGATAR site of the *amdS* promoter. Moreover, the *Aspergillus oryzae amdS* gene has three AGATAA sequences and one CGATAA sequence in its promoter. This gene is well expressed when introduced into *A.nidulans* AreA(Leu) and AreA(Met) backgrounds but expressed very poorly when introduced into an AreA(Val) background (M.Hynes and M.A.Davis, personal communication).

Strains carrying *areA102*, in contrast to wild-type strains, also utilize as nitrogen sources certain natural metabolites such as histidine and citrulline (see Figure 7). This implies the presence of TGATAR sites suitably positioned in the promoters of the relevant structural genes. Almost all AreA-responsive genes are also subject to pathway-specific induction, mediated by positive-acting transcription factors. Mutations in a number of regulatory genes mediating induction (*nirA*, *uaY*, *amdR*) can eliminate nitrogen metabolite repression and/or by-pass the need for the AreA protein (Arst and Cove, 1973; Rand and Arst, 1978; Tollervey and Arst, 1981; Andrianopoulous and Hynes, 1990; Burger *et al.*, 1991; Oestreicher and Scazzocchio, 1995). Furthermore, *cis*-acting mutations, altering the response of the *amdS* gene to specific regulators, can be obtained as suppressors of *areA* null mutations (reviewed by Davis and Hynes, 1991). The explanation for the acquisition by *areA102* strains of the ability to utilize nitrogen sources not utilized by the wild-type resides in the increased affinity for the AreA(Val) transcriptional activator, thereby obviating the likely need for a pathway-specific regulator, which presumably responds to a specific co-inducer. The inability of the wild-type strain to utilize the nitrogen source would thus imply that the physiological inducer is a metabolite not directly involved in the pathway. For example, histidase synthesis is not induced by histidine (Polkinghorne and Hynes, 1975) but might nevertheless be induced by an unidentified and possibly unrelated metabolite. The regulation of the *amdS* gene provides a precedent. It is induced by acetamide, through its conversion to acetyl-CoA (Hynes, 1978), but also by an unrelated metabolite, β-alanine (Arst, 1976). The possibility of uncoupling induction from metabolism is a necessary corollary of the concept of gratuity, as elaborated by Monod (1970), and a potential source of evolutionary variation. It is interesting that a single amino acid change in a transcription factor can focus on this wider evolutionary process.

Materials and methods

Aspergillus nidulans strains and genetic techniques

Aspergillus nidulans strains carried auxotrophic or colour markers in standard use (Arst and Cove, 1973; Rand and Arst, 1977; Kudla *et al.*, 1990; Clutterbuck, 1993; Stankovich *et al.*, 1993), and genetic techniques followed Pontecorvo *et al.* (1953) with the modifications of Clutterbuck (1974). Transformation of *A.nidulans* was carried out according to Tilburn *et al.* (1983). Transformation with small DNA fragments was carried out according to Burger *et al.* (1991).

Sequencing of mutations

A 10 kb *Eco*RI fragment comprising the *uapA* gene region of a strain carrying the *uapA310* mutation was cloned in λEMBL4. M13 subclones derived from this phage were used to sequence the whole promoter region from nucleotide 1 to 906 (numbering as in Gorfinkiel *et al.*, 1993). The sequence change found was checked as indicated below for strains carrying mutations *uapA302* and *uapA349*.

For strains carrying the *uapA302* and *uapA349* mutations, the promoter region comprising the regulatory elements (AreA- and UaY-binding

Table I. Sequence and position of the oligonucleotides used for the construction of the *uapA* promoter probes

Oligonucleotide name	Sequence	Start position
MEC CONT	CGTCCAAGCCACTCTAGCAG	–172
MEC UAP	CGTCCAAGCCACTCTAGCAGCAGc^{M}GATA	–172
UAP33	TGgGaTcCAAGAGAAAGATAACCATG	25 R
UAP55	AGGACTATCCACCCTGTATGG	–465
UAP 102 SUP	GTCAAGCTAGCTGCTTATCWCTCTATCNCTGCTAGAG	–125 R
UAP CORE 3	AACTGTCAAGCTAGCTGCTT	–121 R
UAP ΔGATA	CCACTCTAGCAG(Δ18)CTAGCTTGACAG	–163 R
UAP GGATA	CGTCCAAGCCACTCTAGCAGgGATA	–172
UFOOT3	ATTAGCGTGGCGTCAGCCACAG	–68 R
UFOOT5	TCTCATCCTCCGCACCCGCTGTC	–229
UAP GATAA	CGTCCAAGCCACTCTAGCAGCGATAaAG	–172
U/S UAP INT15	CGTCCAAGCCACTCTAGCAGCGggAGAG	–172

Nucleotide changes are shown in lower case. c^{M} indicates the 5-methylcytosine replacement. Reverse oligonucleotides are indicated by an 'R' following the startpoint.

sites) was amplified (from nucleotide 540 to 704, numbering as in Gorfinkiel *et al.*, 1993) with suitable primers, and cloned in the appropriate Bluescript plasmid. Four independent clones were sequenced for each strain. Sequencing was carried out by the dideoxynucleotide termination method (Sanger *et al.*, 1977). The mutations generated *in vitro* in the AreA DNA-binding region and in the *uapA* promoter were sequenced directly from PCR products obtained after amplification of a DNA fragment from nucleotide 540 to 704.

Detection of transcription levels of uapA

Growth of mycelia, induction conditions, extraction of total RNA and Northern blots have been described previously (Suárez *et al.*, 1991; Oestreicher and Scazzocchio, 1995). The gratuitous inducer 2-thiouric acid was used thoughout to induce transcription of the *uapA* gene (Oestreicher and Scazzocchio, 1995).

Plasmid construction, expression and purification of the oligohistidine fusion proteins

The construct pARBD4 expressing the wild-type oligohistidine–AreA-(Leu683) fusion protein (residues 468–729) was described in Langdon *et al.* (1995). The constructs expressing the AreA(Val683) and the AreA(Met683) fusion proteins were made by replacing a 220 bp *Sph*I–*Dra*III fragment in pARBD4 with the corresponding fragments obtained from PCR products amplified from the *areA102* and *areA30* mutant strains described in Kudla *et al.* (1990). All constructs were checked by sequencing.

Expression, purification and quantitation of wild-type and mutant fusion proteins were described in Langdon *et al.* (1995) and Ravagnani and Arst (1995).

Wild-type and mutant probes

All probes were generated by PCR amplification using one ^{32}P-end-labelled and one cold oligonucleotide, followed by gel purification. The plasmid constructs p5UA3 (wild-type) and pUADS-B3 (in which both GATA sites were destroyed) which served as templates were described in Platt *et al.* (1996b). All the other templates were prepared as follows (see Table I for oligonucleotide sequence and position). Two PCR fragments were amplified using UAP55/UAP 102 SUP and UAPΔGATA/UAP33. The two products were combined and re-amplified using the external oligonucleotides UAP55 and UAP33. The final product was cut with *Nco*I and used to replace the equivalent fragment in the wild-type construct p5UA3. Of the eight possible combinations potentially generated by the UAP 102 SUP (...NGATAGagWGATAA...), we were unable to obtain the constructs containing the sequence ...**G**GATA-GagWGATAA.... The corresponding probes were obtained using a specific oligonucleotides (see Table I).

The 52 bp probes used in the experiment shown in Figure 6 were constructed using the primers shown above (Table II) in combination with oligonucloetide UAP CORE3. In Table II, we show for each probe the original sequence, the final GATA site sequence and the forward primers used.

The gel retardation assays shown in Figure 3, the DNase I footprintings and the dUTP interference were performed using 162 bp PCR probes generated from the appropriate templates using the oligonucleotides UFOOT5 and UFOOT3.

Table II. PCR amplification of the 52 bp *uapA* promoter probes

Template sequence	Forward oligonucleotide	Probe sequence
CGATAGAGAGATAA	MEC CONT	CGATAGAGAGATAA
CGATAGAGAGATAA	MEC UAP	c^{M}GATAGAGAGATAA
tGATAGAGAGATAA	MEC CONT	tGATAGAGAGATAA
aGATAGAGAGATAA	MEC CONT	aGATAGAGAGATAA
CGATAGAGAGATAA	UAP GGATA	gGATAGAGAGATAA
CGATAGAGtGATAA	MEC CONT	CGATAGAGtGATAA
CGATAGAGtGATAA	MEC UAP	c^{M}GATAGAGtGATAA
tGATAGAGtGATAA	MEC CONT	tGATAGAGtGATAA
aGATAGAGtGATAA	MEC CONT	aGATAGAGtGATAA
CGATAGAGtGATAA	UAP GGATA	gGATAGAGtGATAA
CGATAGagAGATAA	UAP GATAA	CGATAaAGAGATAA

All forward oligonucleotides have been used in combination with UAP CORE3 reverse oligonucleotide. Nucleotide changes are shown in lower case. c^{M} indicates the 5-methylcytosine replacement.

Gel retardation assays, DNase I footprinting and uracil interference.

Gel retardation assays and DNase I footprinting were performed as described in Langdon *et al.* (1995).

The dUTP interference was adapted from Pu and Struhl (1992). A PCR probe was prepared as described above, including 30 μM dUTP (final concentration) in the reaction. The binding reaction (containing 1.5 μg of fusion protein and 10^5 c.p.m. of probe) was subjected to gel retardation assay to separate bound and unbound probe. The DNAs were recovered and, following uracil-*N*-glycosylase and piperidine treatments, were subjected to electrophoresis on a 6% denaturing gel.

Directed mutagenesis of the AreA DNA-binding region

Overlap PCR was used to create the site-directed mutant DNA fragments used for transformation. The PCR products were transformed directly into *A.nidulans* and the mutants characterized as described previously (Langdon *et al.*, 1995; Tilburn *et al.*, 1995; Platt *et al.*, 1996a,b). The different substitutions at codon 683 (codon numbering according to Langdon *et al.*, 1995) were constructed by transforming an *areA5 inoB2 fwA1* strain, where *areA5* is a substitution of the first Cys of the Zn finger (residue 673) by a Tyr, resulting in complete loss of function (Platt *et al.*, 1996a). Transformants were selected for rescue of the *areA5* mutation, they were checked by crossing with an *areA*$^{+}$ strain, Southern hybridization and sequencing throughout the DNA-binding region. In each case, the new mutation segregated 1:1, the Southern blot showed a normal restriction pattern for the *areA* locus, no extraneous sequence changes were detected and sequencing revealed the desired changes

Construction of A.nidulans strains carrying mutations in the HGATAR sites

Plasmids used to transform an *A.nidulans areA102* strain were constructed by the following procedure. Plasmid p5UA3 contains a 976 bp *Sac*I–*Nar*I fragment comprising the entire *uapA* promoter region (Platt

et al., 1996b). An *Nco*I fragment of the above insert, containing the CGATAGagAGATAA sequence, was substituted by a cognate fragment carrying either a CGATAGag**T**GATAA or a **T**GATAGag**T**GATAA (mutant bases in bold) sequence, constructed as detailed above (see wild-type and mutant probes). The resulting plasmids were used to transform an *A.nidulans areA102* strain. Transformants able to grow on uric acid were selected and, from these, one transformant for each of the mutant plasmids was shown to carry only a simple replacement of the wild-type sequence. The introduction of the mutant sequences in the *uapA* promoter was checked by DNA sequencing after PCR amplification.

Modelling procedures

Models of AreA–AGATAA complexes were generated using INSIGHT/DISCOVER modules from the Biosym/MSI package. (Biosym/MSI, 9685 Scranton Road, San Diego, CA 92121-3752).

Acknowledgements

We thank Ramón González for helpful discussion and critical reading of the manuscript, Michael Hynes for allowing unpublished work to be cited and Alan Sheerins for technical assistance. This work was supported by the European Union (Contract SC1*CT92-0815), the CNRS and the BBSRC (grant GR/48510 to H.N.A.). A.R. was supported by a studentships of the SERC and of the 'Institute Pasteur–Fondazione Cenci Bolognetti', L.G. and G.D. by fellowships of the European Union and D.G. by the SRC.

References

Andrianopoulous,A. and Hynes,M.J. (1990) Sequence and functional analysis of the positively acting regulatory gene *amdR* from *Aspergillus nidulans*. *Mol. Cell. Biol.*, **10**, 3194–3203.

Arst,H.N.,Jr (1976) Integrator gene in *Aspergillus nidulans*. *Nature*, **262**, 231–234.

Arst,H.N.,Jr (1988) Localisation of several chromosome I genes of *Aspergillus nidulans*: implications for mitotic recombination. *Mol. Gen. Genet.*, **213**, 545–547.

Arst,H.N.,Jr and Cove,D.J. (1973) Nitrogen metabolite repression in *Aspergillus nidulans*. *Mol. Gen. Genet*., **126**, 111–141.

Arst,H.N.,Jr and Scazzocchio,C. (1975) Initiator constitutive mutation with an 'up-promoter' effect in *Aspergillus nidulans*. *Nature*, **254**, 31–34.

Ballario,P., Vittorioso,P., Magrelli,A., Talora,C., Cabibbo,A. and Macino,G. (1996) White collar-1, a central regulator of blue light responses in *Neurospora* is a zinc finger protein. *EMBO J.*, **15**, 1650–1657.

Burger,G., Strauss,J., Scazzocchio,C. and Lang,F.B. (1991) *nirA*, the pathway specific regulatory gene of nitrate assimilation in *Aspergillus nidulans*, encodes a putative GAL4-type zinc finger protein and contains four introns in highly conserved regions. *Mol. Cell. Biol.*, **11**, 5746–5755.

Clore,G., Bax,A., Omichinski,J.G. and Gronenborn,A. (1994) Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA. *Structure*, **2**, 89–94.

Clutterbuck,A.J. (1974) *Aspergillus nidulans*. In King,R.C. (ed.), *Handbook of Genetics*. Plenum Press, New York, Vol. 1, pp. 447–510.

Clutterbuck,A.J. (1993) *Aspergillus nidulans*. In O'Brien,S.J. (ed.), *Genetic Maps*, *Locus Maps of Complex Genomes*. 6th edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 3, pp. 3.71–3.84.

Coffman,J.A., Rai,R. and Cooper,T.G. (1995) Genetic evidence for Gln3p-independent, nitrogen catabolite repression-sensitive gene expression in *Saccharomyces cerevisiae*. *J. Bacteriol.*, **177**, 6910–6918.

Cunningham,T.S. and Cooper,T.G. (1991) Expression of the DAL80 gene whose product is homologous to the GATA factors and is a negative regulator of multiple catabolic genes in *Saccharomyces cerevisiæ*, is sensitive to nitrogen catabolite repression. *Mol. Cell. Biol.*, **11**, 6205–6215.

Davis,M.A. and Hynes,M.J. (1991) Regulatory circuits in *Aspergillus nidulans*. In Bennett,J.W. and Lasure,L.L. (eds), *More Gene Manipulations in Fungi*. Academic Press, San Diego, CA, pp. 151–189.

Diallinas,G., Gorfinkiel,L., Arst,H.N.,Jr, Cecchetto,G. and Scazzocchio,C. (1995) Genetic and molecular characterisation of a wide specificity purine permease gene of *Aspergillus nidulans* reveals a novel family of transporters conserved in prokaryotes and eukaryotes. *J. Biol. Chem.*, **270**, 8610–8622.

Feng,B., Xiao,X. and Marzluf,G.A. (1993) Recognition of specific nucleotide bases and cooperative binding by the trans-acting nitrogen regulatory protein NIT-2 of *Neurospora crassa*. *Nucleic Acids Res.*, **21**, 3989–3996.

Fu,Y.H. and Marzluf,G.A. (1990) *nit-2*, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain. *Mol. Cell. Biol*., **10**, 1056–1065.

Gorfinkiel,L., Diallinas,G. and Scazzocchio,C. (1993) Sequence and regulation of the *uapA* gene, encoding a uric acid–xanthine permease in the fungus *Aspergillus nidulans*. *J. Biol. Chem.*, **268**, 23376–23381.

Gorton,D.J. (1983) Genetical and biochemical studies of the uptake of purines and their degradation products in *Aspergillus nidulans*. Ph.D. Thesis, University of Essex, UK.

Hynes,M.J. (1975) Studies on the role of the *areA* gene in the regulation of nitrogen catabolism in *Aspergillus nidulans*. *Aust. J. Biol. Sci*., **28**, 301–313.

Hynes,M.J. (1978) Multiple independent control mechanisms affecting the acetamidase of *Aspergillus nidulans*. *Mol. Gen. Genet.*, **161**, 59–65.

Hynes,M.J. and Pateman,A.J. (1970) The genetic analysis of regulation of acetamidase synthesis in *Aspergillus nidulans*. I Mutants able to utilise acrylamide. *Mol. Gen. Genet.*, **108**, 97–106.

Hynes,M.J., Corrick,C.M., Kelly,J.M. and Littlejohn,T.G. (1988) Identification of the sites of action for regulatory genes controlling the *amdS* gene of *Aspergillus nidulans*. *Mol. Cell. Biol.*, **8**, 2589–2596.

Ko,L. and Engel,D.J. (1993) DNA binding specificities of the GATA transcription factor family. *Mol. Cell. Biol*., **13**, 4011–4022.

Kudla,B., Caddick,M.X., Langdon,T., Martinez-Rossi,N.M., Bennett, C.F., Sibley,S., Davis,R.W. and Arst,H.N.,Jr (1990) The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specific gene activation alter a loop residue of a putative zinc finger. *EMBO J*., **9**, 1355–1364.

Langdon,T., Sheerins,A., Ravagnani,A., Gielkens,M., Caddick,M.X. and Arst,H.N.,Jr (1995) Mutational analysis reveals dispensability of the N-terminal region of the *Aspergillus* transcription factor mediating nitrogen metabolite repression. *Mol. Microbiol*., **17**, 877–888.

Martin,D.I.K., Tsai,S.F. and Orkin,S.H. (1989) Increased γ-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature*, **338**, 435–438.

Maxam,A.M. and Gilbert,W. (1977) A new method for sequencing DNA. *Proc. Natl Acad. Sci. USA*, **74**, 560–564.

Merika,M. and Orkin,S.H. (1993) DNA-binding specificity of GATA family transcription factors. *Mol. Cell. Biol.*, **13**, 3999–4010.

Minehart,P.L. and Magasanik,B. (1991) Sequence and expression of GLN3, a positive nitrogen regulatory gene of *Saccharomyces cerevisiae* encoding a protein with a putative zinc finger DNA-binding domain. *Mol. Cell. Biol.*, **11**, 6216–6228.

Monod,J. (1970) *L'Hasard et la Nécessité*. Editions du Seuil, Paris.

Oestreicher,N. and Scazzocchio,C. (1995) A single amino acid change in a pathway-specific transcription factor results in differing degrees of constitutivity, hyperinducibility and derepression of several structural genes. *J. Mol. Biol.*, **249**, 693–699.

Omichinski,J.G., Clore,G.M., Schaad,O., Felsenfeld,G., Trainor,C., Appella,E., Stahl,S.J. and Gronenborn,A.M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA-binding domain of GATA-1. *Science*, **261**, 438–446.

Platt,A., Langdon,T., Arst,H.N.,Jr, Kirk,D., Tollervey,D., Mates Sanchez, J.M. and Caddick,M.X. (1996a) 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*., **15**, 2791–2801.

Platt,A., Ravagnani,A., Arst,H.N.,Jr, Kirk,D., Langdon,T. and Caddick, M.X. (1996b) Mutational analysis of the C-terminal region of AREA, the transcription factor mediating nitrogen metabolite repression in *Aspergillus nidulans*. *Mol. Gen. Genet.*, **250**, 106–114.

Polkinghorne,M. and Hynes,M.J. (1975) Mutants affecting histidine utilization in *Aspergillus nidulans*. *Genet. Res. Camb*., **25**, 119–135.

Pontecorvo,G., Roper,J.A., Hemmons,L.M., Macdonald,K.D. and Bufton,A.W.J. (1953) The genetics of *Aspergillus nidulans*. *Adv. Genet.*, **5**, 141–238

Pu,W.T. and Struhl,K. (1992) Uracil interference, a rapid and general method for defining protein–DNA interactions involving the 5-methyl group of thymines: the GCN4–DNA complex. *Nucleic Acids Res.*, **20**, 771–775.

Rand,K.N. and Arst,H.N.,Jr (1977) A mutation in *Aspergillus nidulans* which affects the regulation of nitrate reductase and is tightly linked to its structural gene. *Mol. Gen. Genet.*, **155**, 67–75.

Rand,K.N. and Arst,H.N.,Jr (1978) Mutations in *nirA* gene of *Aspergillus nidulans* and nitrogen metabolism. *Nature*, **272**, 732–734.

Ravagnani,A. and Arst,H.N.,Jr (1995) Quantitation of oligohistidine fusion proteins using $^{63}Ni^{2+}$ chelation. *Nucleic Acids Res.*, **23**, 3069–3070.

Sanger,F., Nickels,S. and Coulson,A.R. (1977) DNA sequencing with chain termination inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Scazzocchio,C. (1990) Of moulds and men, or two fingers are not better than one. *Trends Genet.*, **6**, 311–313.

Scazzocchio,C. and Arst,H.N.,Jr. (1978) The nature of an initiator constitutive mutation in *Aspergillus nidulans*. *Nature*, **274**, 177–179.

Stanbrough,M., Rowen,D.W. and Magasanik,B. (1995) Role of the GATA factors Gln3p and Nil1p of *Saccharomyces cerevisiae* in the expression of nitrogen regulated genes. *Proc. Natl Acad. Sci. USA*, **92**, 9450–9454.

Stankovich,M., Platt,A., Caddick,M.X., Langdon,T., Shaffer,P.M. and Arst,H.N.,Jr (1993) C-terminal truncation of the transcriptional activator encoded by *areA* in *Aspergillus nidulans* results in both loss-of-function and gain-of-function phenotypes. *Mol. Microbiol.*, **7**, 81–87.

Strauss,E.C., Andrews,N.C., Higgs,D.R. and Orkin,S.H. (1992) *In vivo* footprinting of the human α-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. *Mol. Cell. Biol.*, **12**, 2135–2142.

Suárez,T., Oestreicher,N., Kelly,J., Ong,G., Sankarsing,T. and Scazzocchio,C. (1991) The *uaY* positive control gene of *Aspergillus nidulans*: fine structure, isolation of constitutive mutants and reversion patterns. *Mol. Gen, Genet.*, **230**, 359–368.

Suárez,T., Vieira de Queiroz,M., Oestreicher,N. and Scazzocchio,C. (1995) The sequence and binding specificity of UaY, the specific regulator of the purine utilisation pathway in *Aspergillus nidulans*, suggest an evolutionary relationship with the PPR1 protein of *Saccharomyces cerevisiae*. *EMBO J.*, **14**, 1453–1467.

Tilburn,J., Scazzocchio,C., Taylor,G.G., Zabicky-Zissman,J.H., Lockington,R.A. and Davies,R.W. (1983) Transformation by integration in *Aspergillus nidulans*. *Gene*, **26**, 205–221.

Tilburn,J., Sarkar,S., Widdick,D.A., Espeso,E.A., Orejas,M., Mungroo,J., Peñalva,M.A. and Arst,H.N.,Jr (1995) The *Aspergillus* PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH. *EMBO J.*, **14**, 779–790.

Tollervey,D.W. and Arst,H.N.,Jr (1981) Mutations to constitutivity and derepression are separate and separable in a regulatory gene of *Aspergillus nidulans*. *Curr. Genet.*, **4**, 63–68.

Received on January 24, 1997; revised on March 14, 1997