Sequence and Regulation of the \textit{uapA} Gene Encoding a Uric Acid-xanthine Permease in the Fungus \textit{Aspergillus nidulans}*

(Received for publication, March 1, 1993, and in revised form, June 2, 1993)

Lisette Gorfinkel\textsuperscript{,}, George Diallinas\textsuperscript{,} and Claudio Scanzocchio\textsuperscript{,}

\textit{From the Institut de Génétique et Microbiologie, Unité Associée au Centre National de la Recherche Scientifique 1354, Université Paris-Sud, Bâtiment 409, Centre d’Orsay, F-91405 Orsay, France}

The nucleotide sequence of the \textit{uapA} gene, coding for the uric acid-xanthine permease of \textit{Aspergillus nidulans}, has been determined. The predicted \textit{uapA} gene product comprises 595 amino acids (\(M_r, 63,365\)); it is a highly hydrophobic protein with 12–14 putative transmembrane segments and shows no striking similarity to any other membrane protein of either prokaryotes or eukaryotes, except for a short highly hydrophobic amino acid sequence conserved in a number of different permeases. The presence of an acidic, amphipathic region overlapping with the last hydrophobic segment of \textit{UAPA} could also be of interest. The results presented suggest that the \textit{UAPA} permease represents a new type of membrane protein, not described previously. The transcription of \textit{uapA} is inducible by 2-thiouric acid, and it is highly repressible by ammonium. It is almost absolutely dependent on the presence of functional \textit{uAY} and \textit{areA} regulatory gene products. A specific mutation in the \textit{GATA} binding zinc finger of the \textit{AREA} protein nearly abolishes \textit{uapA} transcription. The \textit{uap100} cis-acting, ups-promoter, constitutive mutation is a duplication that comprises two \textit{GATA} sites and suppresses weakly the \textit{AREA} zinc finger mutation but does not alleviate the need for functional \textit{UAY} and \textit{AREA} proteins.

* This work was supported in part by the Centre National de la Recherche Scientifique and by scholarships from the Directorate General XII of the Commission of the European Communities (to L. G. and G. D.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank\textsuperscript{TM}/EMBL Data Bank with accession number(s) X71807.

† Supported by a European Community grant. Present address: Depto. de Bioquimica y Biologia, Facultad de Ciencias, Universidad de la Republica, Montevideo, Uruguay.

‡ Supported by a European Community grant. Present address: Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, P. O. Box 1527, Heraklion, Crete, Greece.

\textsuperscript{1} To whom correspondence and reprint requests should be addressed. Tel.: 33-1-6841-6356; Fax: 33-1-6841-7809.

\textsuperscript{*} L. Gorfinkel, H. N. Arst, Jr., D. J. Gorton, and C. Scanzocchio, unpublished data.

\textsuperscript{2} L. Gorfinkel, H. N. Arst, Jr., and C. Scanzocchio, unpublished data.
Uric acid-xanthine Permease of Aspergillus nidulans

(Artz and Sczzocchio, 1975; Sczzocchio and Artz, 1978; Diallinas and Sczzocchio, 1989), an up-promoter, constitutive mutation that suppresses the areA102 allele by uric acid as nitrogen source while still leaving upA gene products. A portion of the first strand product was annealed with the primer cUAP5 (S'GAAGCG-TTGACCGG CATAGGGGTGACGGC3') to establish the position of the 3' end. The primer used was a 35-nucleotide oligonucleotide with 15 dT residues and a restriction site, and reiterative polymerase chain reaction cycles using both primers amplified the double-stranded cDNA of 1,297 bp as described in the protocol of the GeneAmp DNA Amplification Reagent Kit 5 (Perkin-Elmer-Cetus Instruments). To obtain the 3' end we used a 35-nucleotide oligonucleotide with 15 dT residues and an adapter sequence containing the ClaI, SalI, and PstI recognition sites to create the first DNA strand. For the amplification a gene-specific primer near the 3' end (InC; 5'GGCAGGCTGTGCGCCACTG3') and the adapter primer were used. The appropriate fragment of 271 bp was obtained. Both fragments were cloned in the appropriate M13 vectors and sequenced. In the S1 nuclease protection experiments performed to determine the transcription start 209-bp Sau3AI and 643-bp Nhel fragments (see Fig. 1) and in the primer extension experiments a 5'GCTTTATCTCGAG3' oligonucleotide was used.

Computer Methods—The handling of sequences, their analysis, and comparisons were accomplished with the DNA Strider™ 1.0 and FASTA software. Protein homology searches were conducted with the European Molecular Biology laboratory protein data base SwissProt.

RESULTS

Sequence of the upA Gene—We have established previously the position of the upA gene within the BglII-Sall fragment of a plasmid described earlier, pAN503 (Diallinas and Sczzocchio, 1989). The complete 3.4-kilobase pair BglIII-Sall fragment was sequenced in both directions. Fig. 1 shows the strategy adapted to determine the sequence. The nucleotide sequence of two upA cDNA clones (Fig. 1) was also determined to establish the presence of three putative introns predicted from the 5', 3', and lariat consensus sequences of filamentous fungi (Gurr et al., 1987) and to map the 3' end of the upA transcript. Our results reveal a long, open reading frame of 1,785 bp interrupted by three short introns of 59, 64, and 58 bp, respectively (Fig. 2). The size of the open reading frame is consistent with the estimated size of the upA transcript (2.1 kilobases) from Northern blot analysis (Diallinas and Sczzocchio, 1989). The GC content of the upA sequence is high, a feature common in all filamentous fungi genes. The sequence around the start (ATG) codon shows excellent agreement with translation initiation sites of lower eukaryotes (Proudfoot, 1979; Kozak, 1984; Gurr et al., 1987); that is, nucleotide -3 is occupied by an A; -4 by a C; -14 by an A; and +6 by a T.

The 5'-noncoding region of the upA gene does not show typical TATA or CAAT consensus sequences present in eukaryotic genes (Proudfoot, 1979; Braithwaich and Chambon, 1981; Myeus et al., 1986). The sequences 5'GATA3' or its complement are indicated by a hatched bar below. The direction of transcription of the gene is from left to right (Diallinas and Sczzocchio, 1989). Bi, BglII; S, SalI; s, Sau3AI; A, AscI; H, HindIII; P, PstI; V, SacI; W, SacII; N, Nhel; X, Xbal.

EXPERIMENTAL PROCEDURES

Media and Growth Conditions—The standard media and growth conditions for A. nidulans were used (Cove, 1966; Sczzocchio et al., 1982). Supplements were added when appropriate. Noninducing, inducing, and repressing conditions have been described by Suizres et al. (1991a). The gratuitous inducer 2-thiouric acid, rather than the natural inducer uric acid, was used throughout. The identity of the response of the induction system to these chemicals has been shown previously (Sealy-Lewis et al., 1987; Sczzocchio et al., 1982).

Strains—The following A. nidulans strains were used: KA1; KA2, uapA100; pabaAl, uapY207; fuA1, pyna4, areAl02; pabaAl, uap100, uapY207; areAl02; pabaAl, uap100, areAl02; pabaAl, uap100, areAl02; pabaAl, uap100, areAl02. The double mutant strains were constructed using standard genetic techniques. uapA100 is a cis-acting regulatory mutation that affects the response of the upA gene to such mutations.

DNA and RNA Manipulations—Isolation of A. nidulans nucleic acids has been described previously (Lockington et al., 1985). DNA extraction from E. coli and DNA manipulations were described by Manistis et al. (1982). S1 nuclease experiments were described by Auger (1994). DNA sequencing using the dideoxyxynucleotide chain termination procedure (Sanger et al., 1987) was performed in M13mp18 and M13mp19 single-stranded templates. 32P-dATP-labeled DNA molecules were prepared by using random hexamucleotide primers following the suppliers' instructions (Amersham, Little Chalfont).

Transcript Mapping—To obtain a cDNA clone that would cover the gene (except the 5' and 3' ends) 20 μg of poly(A)2 RNAs of a wild type strain grown under inducing conditions were denatured with hydroxymethylmercuric according to Reesschaert et al. (1987) and reverse transcribed with 100 units of avian myeloblastosis virus reverse transcriptase to create a cDNA(-) strand. The primer used was a 35-nucleotide oligonucleotide with an adapter sequence that contained the HindIII restriction end (uCUP5; 5'ATACGGTACCGTTGACCGG CATAGGGGTGACGGC3'), a portion of the first strand product was annealed with the primer uCUP5 (5'GAGAGGCG-

The abbreviation used is: bp, base pair(s).

FIG. 1. Physical map and sequencing strategy for the upA gene. The solid black bar represents the approximate position of the upA open reading frame. Introns are shown as white bars within the upA open reading frame. The sequenced regions of the cDNA clones are indicated by a hatched bar below. The direction of transcription of the gene is from left to right (Diallinas and Sczzocchio, 1989). Bi, BglII; S, SalI; s, Sau3AI; A, AscI; H, HindIII; P, PstI; V, SacI; W, SacII; N, Nhel; X, Xbal.
![Fig. 2. Nucleotide sequence and deduced amino acid sequence of the uapA gene. Noncoding regions and introns are presented in lowercase letters. Introns 5', 3', and lariat consensus sequences, and the 3' putative cleavage sequence are underlined. Putative ARE binding sites are indicated with thin arrows. The CACC sequence is indicated with a thick arrow. The region duplicated in uapB100 mutation is underlined with a thick line. Triangles indicate start and stop of transcription. Protein sequences of particular interest are boxed (see "Discussion").](image)

complementary 5'TACTC3' are present four times upstream from the transcription start (see below). These are candidate sequences for AREA binding sites as deduced by work on the isofunctional Neurospora crassa NIT2 transcription factor (Fu and Marzluf, 1990a, 1990b) and footprints conducted with fusion proteins containing the A. nidulans AREA zinc finger. We have underlined the sequence duplicated in the uap100 mutation (Fig. 2). This sequence is 164 bp and not 139 bp as it was reported erroneously in our previous publication (Del accepts the ATG as the start codon, but the GAL7 gene transcript in Saccharomyces cerevisiae (Abe et al., 1990).

Transcript Mapping of the uapA Gene—A single mRNA of ~2.1 kilobases is seen in the wild type or in strains carrying the uap100 duplication in all growth conditions tested. The start of transcription of the uapA gene was mapped both by S1 nuclease protection and primer extension experiments in the wild type strain. The start of transcription of the uapA mRNA of a strain carrying the GAL7 gene transcript in Saccharomyces cerevisiae (Abe et al., 1990).
the uapA gene in an uap100 background. Moreover, reprobing of such a blot with an Nhel 288-bp probe, which overlaps the duplicated region and which would reveal any transcript initiating at the upstream predicted start, failed to show any hybridizing message (results not shown). As for other genes of A. nidulans, the start of uapA transcription is proximal to a pyrimidine-rich region. The 3' end of the uapA transcript was mapped by sequencing the appropriate 3' cDNA clone, 45 bp downstream from the translation stop codon and 11 bp upstream from the proposed signal for cleavage (see "Experimental Procedures"). The size of the uapA open reading frame and its transcript map predict an mRNA size of 1,940 ± 2 nucleotides (Fig. 1).

Analysis of the Predicted uapA Gene Product—The uapA open reading frame encodes a protein of 595 amino acids with a predicted molecular mass of 63,365 Da (Fig. 2). No codon bias is apparent. The UAPA protein is extremely hydrophobic (Fig. 3a), in agreement with the fact that uapA encodes a permease. The UAPA protein lacks a hydrophobic amino-terminal signal peptide. Both the amino-terminal region of 95 amino acids and the carboxyl-terminal region of 92 amino acids, are hydrophilic. 12 to 14 distinct putative transmembrane segments have an average hydropathy greater or close to +1.6 and can thus be considered to be membrane-spanning segments (Kyte and Doolittle, 1982). The limits of the 12–14 potential transmembrane segments are defined by proline residues potentially breaking α-helices (prolines 96, 129, 161, 185, 205, 207, 245, 277, 290, 312, 345, 438, 469, 496, 519, 527, 564) and by charged amino acids (see Fig. 3b). This is in line with what is found in other membrane-spanning proteins (Devillers-Thiéry et al., 1983; Hoffmann, 1985; von Heijne, 1988; Dalbey, 1990). The hydrophathy profile of the UAPA transporter, when compared with that of all other transporters of prokaryotes and lower eukaryotes, shows a particularly high hydrophilicity. In addition, UAPA has an unique region (residues 530–569), overlapping with the last hydrophobic segment (residues 540–574) which could form an acidic, amphipathic α-helix (residues 530–547) with a net charge of -5. A single putative glycosylation site (Struck et al., 1978), Asn10-Lys11-Ser12, exists in the amino-terminal segment.

We compared the primary amino acid sequence of the UAPA transporter with that of all known prokaryotic and eukaryotic transport proteins; the uapA transporter showed no or very little similarity with other membrane-spanning proteins. However, these studies revealed that the sequence X-Leu-X-Ile-Ile-Gly/Aln-X-X-Ile-X, where X is any hydrophobic amino acid, is present in a very hydrophobic middle segment of many different permeases (Fig. 4).

Regulation of uapA Expression—We have shown previously that uapA transcription is inducible by the gratuitous inducer 2-thiouric acid (and by implication by the natural inducer uric acid) and that the uap100 duplication results in a 5-fold increase of the uapA transcript in mycelia grown under non-inducing conditions and in an 8-fold increase in mycelia grown under inducing conditions (Diállinas and Scanzocchio, 1989). Fig. 5a establishes that uapA transcription is inducible by the gratuitous inducer 2-thiouric acid and that uap100 results in constitutivity and hyperinducibility. It also shows that uapA transcription is highly repressible by ammonium and that the uap100 mutation does not bypass repression of uapA transcription. Fig. 5b shows that (i) In the absence of a functional uapY product, the uapA gene has a low basal level transcription which may be somewhat increased in the presence of the uap100 mutation; (ii) in the absence of a functional areA product (areA600) or in the presence of the areA102 allele, uapA transcription is barely detectable; (iii) the presence of the uap100 mutation suppresses partially the phenotype of the areA102 allele but does not affect the phenotype of the null areA600 mutation.

DISCUSSION

The uapA gene is the first specific purine permease gene to be sequenced. A permease of S. cerevisiae published previously transports cytosine in addition to adenine, and it does not transport the oxidized purines xanthine and uric acid (FCY2; Weber et al., 1988, 1990). The gene coding for a permease specific for allantoin in S. cerevisiae has been cloned and sequenced (Rai et al., 1988; Yoo et al., 1992) and shows no obvious similarities with UAPA. Thus UAPA belongs to a
The uapA open reading frame reveals a structural gene that codes for a polytopic membrane protein containing 595 amino acids. The UAPA permease has an amino-terminal hydrophilic region of 85 amino acids. This is a charged region where an acidic stretch lies between two basic stretches. The hydrophobic core of the protein has a balanced number of basic and acidic residues which in their great majority fall together with α-helix-breaking proline residues, into regions of reduced hydrophobicity defining the borders of transmembrane segments (von Heijne, 1989; Dalbey, 1990). The “positive inside” rule predicts that positively charged segments of polytopic, membrane-spanning proteins face the cytoplasm (von Heijne, 1989). The UAPA permease could have 12–14 transmembrane helical segments, and thus its amino and carboxyl termini could be cytoplasmic. The UAPA protein shares with other membrane proteins a remarkable correlation between the distribution of charged amino acids and prolines and its predicted trans-membrane topology (Dalbey, 1990; Blobel, 1980; Haenpilte et al., 1989). The carboxyl-terminal part of the UAPA protein has a striking feature: a predicted α-helical, acidic, and amphipathic segment overlaps the last hydrophobic domain. This domain could participate in the interaction of UAPA with other protein(s) involved in its functioning or localization in the membrane.

Although the UAPA permease showed no extensive similarity with other protein sequences present in the EMBL protein data base Swiss-Prot, we observed the presence of the sequence X-Leu-X-Ile-Ile-Gly/Ala-X-X-Ile-X in a number of very different transport proteins of A. nidulans and S. cerevisiae. This motif is always found in one of the most hydrophobic segments of these proteins.

There is only one mRNA start in the uap100 mutant. Of the two possible starts only the downstream one is used in vivo. This implies that sequences downstream from the start are necessary for transcription. This has shown to be the case for a number of S. cerevisiae and mammalian promoters, in the latter case; and when no TATA box is present, specific transcription factors binding downstream from the mRNA start have been demonstrated (Maicas and Friesen, 1990; Smale et al., 1991; Harirahan et al., 1991).

The induction ratio of uapA is of the same order as other inducible genes involved in the utilization of nitrogen sources of A. nidulans, and its dependence on a functional uaY product is of the same order as other genes of the purine degradation pathway. On the other hand, the transcription of the uapA gene is more sensitive to ammonia repression and more dependent on the presence of a functional areA gene product than that of any other gene involved in purine utilization (Scanzocchio and Darlington, 1968; Arst and Cove, 1973; Oestreicher and Scanzocchio, 1993). The repressibility by ammonium of the uapA gene is higher than that of the genes of the proline utilization cluster, including the proline permease (Sophianopoulou et al., 1993), and it is of the same level than that of the nitrate permease (Unkles et al., 1991). Arst and co-workers have already proposed that uptake systems may be the preferred target for nitrogen metabolite repression (Arst et al., 1980).

The dependence of uapA expression vis à vis an AREA active protein does not explain the very low level of expression of uapA on an areA102 background. The areA102 mutation does not result in generalized partial loss of function (see the Introduction). Only a degeneracy of the DNA binding sites upstream the genes controlled by areA and a specificity of their response to the conservative change extant in the areA102 mutation are consistent with all of the data.

The uap100 mutation was isolated as a suppressor of areA102. Thus the double mutant necessarily has a level of uric acid uptake such as to satisfy partially the nitrogen requirements of the cell, and this was shown to be the case (Arst and Scanzocchio, 1975). The uap100 mutation does not alter the ammonium repressibility of the uapA gene, which also means that it does not alter the absolute requirement for an active areA gene product. The Northern blot of Fig. 5b shows that in the presence of a total loss of function areA mutation (areA600), uap100 does not affect the transcription of uapA, in spite of the overcharging and overexposing of lanes 7 and 8, the uapA transcript is barely visible. There is, however, a small but noticeable and repeatable difference between a strain carrying the areA102 and the strain carrying both the areA102 and the uap100 mutations. It is interesting that this modest increase in transcription results in a clear growth phenotype.

The duplicated region contains two directly repeated GATA...
sites (CGATAGAGAGATA) 34 bp upstream the start of transcription. Many of the NIT2 and AREA binding sites for both N. crassa and A. nidulans genes are arranged in direct or reverse repeats (Fu and Marzluf, 1990a). It is an intriguing possibility that more similar sites (CGATAGAGAGATAA) 34 bp upstream the start of two of the most closely related GATA sites present might mediate the transcriptional activation of the uapA promoter by the AREA protein, it is reasonable that a duplication of these sites has no effect when the AREA protein is completely absent or inactive (as in arcA600) but a measureable effect in the presence of a conservative replacement in the zinc domain (as in arcA102).

The uap100 duplication does not suppress the effect of a loss of function mutation in the uapA gene. In agreement with in vivo and uptake data (Arst and Scazzocchio, 1975; Scazzocchio and Arst, 1978), the constitutive expression of the uapA gene in the presence of an uap100 mutation is strictly dependent on an UAY functional protein, a promoter carrying the uap100 duplication maintaining a near absolute requirement for an active uay gene product. We do not know as yet if the duplicated sequence contains, as expected, UAY binding sites. The steady-state RNA levels of an uay207 and of an uay207 uap100 double mutant are, however, higher than the steady-state level of an arcA102 uap100 double mutant. There is no contradiction between this fact and the in vivo phenotypes of these strains. Strains carrying uay mutations do not grow on uric acid, not because of the impairment in uptake, but because they have very low levels of urate oxidase, which is also under uay control (Scazzocchio and Darlington, 1968; Scazzocchio et al., 1982; Oestreicher and Scazzocchio, 1993).

It is not clear why the uap100 duplication results in a striking (around eight times) increase in the steady-state amount of the uapA mRNA in an arcA uay background. Within the duplicated sequence a CAACCC sequence is found. It has been reported that this sequence is an enhancer necessary for the expression of genes controlled by GATA factors in the erythroid cell line (Tasie et al., 1991). We have observed that this sequence is present upstream of a number of genes controlled by AREA. It would be tantalizing if the conservation of the GATA factors and their cognate DNA binding sequences between fungi and metazoa would extend to the conservation of the associated enhancer elements and enhancer-binding proteins.

Acknowledgments—We are most grateful to Timothy Langdon who pointed out an important sequencing error in the noncoding upstream sequence of uapA and to Teresa Suarez who checked the correctness of this region independently and to H. N. Arst and co-workers for unpublished data.

A. Ravagnani, T. Langdon, and H. N. Arst, Jr., personal communication.