nit-2, the Major Nitrogen Regulatory Gene of Neurospora crassa, Encodes a Protein with a Putative Zinc Finger DNA-Binding Domain

YING-HUI FU AND GEORGE A. MARZLUF*

Department of Biochemistry, The Ohio State University, Columbus, Ohio 43210-1292

Received 30 August 1989/Accepted 14 November 1989

The nitrogen regulatory circuit of *Neurospora crassa* consists of a set of unlinked structural genes which specify various nitrogen catabolic enzymes plus control genes and metabolic effectors which regulate their expression. The positive-acting *nit-2* regulatory gene is required to turn on the expression of the nitrogen catabolic enzymes during conditions of nitrogen limitation. The complete nucleotide sequence of the *nit-2* gene was determined. The *nit-2* mRNA is 4.3 kilobases long and has a long nontranslated sequence at both its 5' and 3' ends. The *nit-2* gene nucleotide sequence can be translated to yield a protein containing 1,036 amino acid residues with a molecular weight of approximately 110,000. Deletion analyses demonstrated that approximately 21% of the NIT2 protein at its carboxy terminus can be removed without loss of function. The *nit-2* protein contains a single putative Cys_2/Cys_2 zinc finger domain which appears to function in DNA binding and which has striking homology to a mammalian *trans*-acting factor, GF-1.

In Neurospora crassa, limitation of preferred nitrogen sources, e.g., ammonia or glutamine, leads to the increased synthesis of various enzymes which are required for the use of a variety of secondary nitrogen sources, including nitrate, nitrite, purines, amino acids, and proteins (22). This response is governed by the nitrogen control circuit, one of several global metabolic regulatory circuits of N. crassa. The nitrogen regulatory circuit has been studied extensively and includes a number of unlinked structural genes which are controlled in parallel by both major and minor regulatory genes as well as by metabolic inducers and nitrogen catabolite repression (12, 17, 24). The nit-3 gene, which encodes nitrate reductase, is a well-characterized structural gene of the nitrogen circuit, and expression of nit-3 has been shown to be highly regulated by nitrogen repression, nitrate induction, and the activity of three separate control genes, nit-2, *nit-4*, and *nmr* (12).

Synthesis of enzymes of particular pathways within the nitrogen circuit requires specific induction of substrates or intermediates, mediated by minor control genes. For example, in the presence of the inducer nitrate, a pathway-specific control gene, *nit-4*, turns on the expression of the nitrate and nitrite reductase structural genes, *nit-3* and *nit-6*, respectively (11, 22).

The major positive-acting regulatory gene nit-2 plays a central role in the nitrogen circuit and turns on the expression of the structural genes during conditions of nitrogen limitation (13, 22). In contrast to nit-2, a distinct and unlinked regulatory gene, nmr (for nitrogen metabolic regulation), acts in a negative manner (10, 15, 27). In nmr mutants, various nitrogen-related enzymes are expressed constitutively, even in the presence of sufficient levels of preferred nitrogen sources to fully repress enzyme synthesis in nmr^+ strains. It appears that the nit-2 gene and the nmr gene are both directly involved in nitrogen catabolite repression in N. crassa, but the manner in which they may interact is not yet understood.

It was demonstrated that the nit-2 gene product encoded in

one nucleus is freely diffusible throughout the cell and can turn on the expression of structural genes not only in the same nucleus but also in other nuclei of a heterokaryon (5), suggesting that the nit-2 product is a protein. The isolation of a suppressible amber nonsense nit-2 mutant implies that the nit-2 gene encodes a protein (26). Several different hypotheses have been proposed concerning the expression of the nit-2 gene itself and the mechanism of operation of the nitrogen circuit (8, 9, 16). In each case, a functional nit-2 product was postulated to turn on nitrogen structural gene expression. These hypotheses include (i) constitutive expression of the *nit-2* gene, with the activity of its protein product being sensitive to glutamine (16); (ii) negative control of *nit-2* expression by the *nmr* gene, whose product was envisaged to be a repressor that becomes active upon binding glutamine (8); and (iii) the enzyme glutamine synthetase, postulated to also act as a repressor which prevents nit-2 expression (9). The recent molecular cloning of the nit-2 gene has allowed us to examine directly this nitrogen regulatory gene (12, 31). The nit-2 gene is transcribed to give a single 4.3-kilobase (kb) poly(A)⁺ RNA whose content was shown to increase three- to fourfold during nitrogen limitation, suggesting that expression of the *nit-2* control gene is itself subject to regulation (12). However, the results obtained did not support the possibility that nit-2 was controlled either by nmr or by glutamine synthetase. It appears obvious that an understanding of the operation of the nitrogen control circuit will be facilitated by a complete characterization of each of the regulatory genes. Here we report the complete nucleotide sequence of the nit-2 gene and show that it contains two small introns near the 5' end of the gene. The nit-2 gene appears to encode a protein composed of 1,036 amino acids with a molecular weight of approximately 110,000. The translated nit-2 protein appears to possess a single zinc finger domain that is essential for function and that probably plays a direct role in DNA binding. The single zinc finger of the NIT2 protein is remarkably similar to the two zinc finger domains found in a mammalian trans-acting factor, GF1.

^{*} Corresponding author.



FIG. 1. *nit-2* gene structure and sequencing strategy. The direction of transcription is indicated with the bold horizontal arrow which represents the 4.5-kb transcript. The shaded box shows the predicted *nit-2* protein composed of 1,036 amino acid residues. The horizontal arrows show the sequencing strategy that was achieved with deletion clones and with oligonucleotide primers. Restriction sites and length of the genomic segment sequenced (in kilobases) are shown.

MATERIALS AND METHODS

Strains. The *N. crassa* wild-type strain 74OR231A and *nit-2* mutant strains were obtained from the Fungal Genetics Stock Center (University of Kansas Medical Center). Cultures were grown in Vogels liquid medium supplemented as indicated for each experiment with shaking at 30° C as described previously (12, 14).

DNA sequencing and S1 nuclease mapping. DNA sequencing was accomplished by the dideoxy-chain termination method (30) with $\left[\alpha^{-32}P\right]dATP$ and a modified T7 bacteriophage DNA polymerase, Sequenase (United States Biochemical Corp., Cleveland, Ohio). dITP was successfully used in place of dGTP to sequence through compression regions; some regions were sequenced at 70°C with the heat-resistant DNA polymerase of Thermus aquaticus (Promega Biotec, Madison, Wis.). Plasmid DNA templates were prepared as minipreparations (3). Oligonucleotide primers for DNA sequencing and site-directed mutagenesis were synthesized on an Applied Biosystems model 380B DNA synthesizer by the Ohio State University Biochemical Instrument Center. S1 mapping experiments were conducted by the method of Berk and Sharp (2). Primer extension experiments were accomplished by mixing a 5'-end-labeled 17-mer oligonucleotide primer that hybridizes with $poly(A)^+$ RNA as a template for Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Inc., Gaithersburg, Md.).

Site-directed mutagenesis. The cloned nit-2 gene and the flanking region of 4.7 kb proved to be too large for efficient mutagenesis. Accordingly, a restriction fragment from the nit-2 gene of approximately 250 base pairs (bp) that included a potential zinc finger domain was subcloned into the Bluescript vector, and after mutagenesis and confirmation by sequencing, it was moved back into an otherwise wild-type nit-2 gene. Site-directed mutagenesis was performed as described by Kunkel (19).

Isolation of RNA and cDNA clones. N. crassa total RNA was prepared by the method of Reinert et al. (28). The poly(A)⁺ RNA fraction was isolated by oligo(dT)-cellulose chromatography (1) for use in S1 mapping and primer extension studies and for cDNA synthesis. A cDNA library was constructed in λ gt10 with *Eco*RI adaptors to eliminate the need for methylation of the double-stranded cDNA and *Eco*RI digestion before ligation into the vector arms, followed by packaging with the use of Packagene (Promega).

Four rounds of plaque hybridization with pNit2 as a probe were required to isolate several *nit-2* cDNA clones. Lambda DNA was purified and digested with *Eco*RI, and the insert cDNA was subcloned into the Bluescript plasmid vector and sequenced as described above. In some cases, sequencing was accomplished directly with the recombinant lambda bacteriophage DNA. Plasmid DNAs for use as probes were labeled with [^{32}P]dCTP by nick translation (29).

nit-2 deletions and transformation assays. Deletion clones of pNit2 were constructed by the use of DNA polymerase I with the Cyclone system (International Biotechnologies, Inc., New Haven, Conn.) as described previously (13). Mutated or partially deleted *nit-2* genes were assayed for function by transformation into *nit-2* mutant protoplasts, with appropriate positive and negative controls (12, 13). Functional *nit-2* genes gave a high transformation rate, similar to that obtained with the wild-type *nit-2* gene, whereas nonfunctional genes did not transform at all.

Computer methods. The handling of sequences, their analysis and translation, and hydropathy and codon bias analyses were accomplished with Pustell software (International Biotechnologies). Protein homology searches were conducted with GenBank, the Protein Identification Resource (National Biomedical Research Foundation), which contains 3,800 different protein sequences.

RESULTS

nit-2 nucleotide sequence. A restriction map of the *nit-2* gene and the strategy used to sequence it are shown in Fig. 1. The dideoxy sequencing method was used to sequence the entire 5.2-kb region which encompasses the *nit-2* gene and flanking DNA. Both DNA strands were sequenced except for part of the 3' noncoding region of the gene; overlapping clones and synthetic primers were employed to confirm the entire sequence. The nucleotide sequence of the *nit-2* gene and its flanking regions is presented in Fig. 2.

Intervening sequences in the *nit-2* gene. We determined the nucleotide sequence of two overlapping *nit-2* cDNA clones, which together compose a full-length cDNA copy of the *nit-2* transcript. The cDNA sequence agreed completely with the genomic sequence, except for the presence of introns, whose locations were determined by comparing the cDNA sequence with the genomic sequence (Fig. 2). The *nit-2* gene is interrupted by two introns of 98 and 78 bp which occur in the protein-coding region near the 5' end of the gene. Both

	40 -630 -63		20) -610			500	-59	0	-580		-570		-560 -5		i0	-54	0	-530			
CCCCGGG	T SCAGGI	66CCC	¥ ITCGAI	CAGCI	1 100071	AGTCGA	* ACCTAC	I ICCTCA		I BRCART	AATTGA	T CCCAGA	GAACO	1 Sgaaci	GACGGA	8 704440	1 240402					
	081001			JOEGG									JUUURUIL		UUUAUA	GARUU	JURRO		IVAAA	UTURU	-	
-520	-520 -510 -50		-50)	-49	0	-480		-410	0 -460		-450		-44	0	-430	0 -420			-410		
GCTGGC	TGCCCC	GTGCGC	TGAGCAG	Caagti	GGAAG	CACT	GCAGT	CACCAC	CCGAAC	Xetgeggi	AACTT	TTCCCTTCCACAACTAC			• TCTAAT	CCACI	FTTGT	GCCTCT	* CTACC	CTG		
-400	-400 -390 -		-380		-370)	-36	•	-350	-340		-330		-320		-310	-300			-290		
ACTEGAC	CGGAGG	AGTTCT	TECTETE	ITATG	TTC	TCTT	CGGCT	GCTGCA	ACTOCI	ICCTGGC	TGCCG	CTT	GGCTACCO	TAATCO	ACGTTC	CCTG	CTGCT	GAGCCO	TCCTG	GCAACT	GTC	
-280	-2	70	-260		-250		-240		-230	-2	20	-1	210	-200	-	190		-180	-	-170		
1		1	1		1		1		1		1		1	1		1		1		1		
CGTGCTGCTACACGCTACCACCACTAGTTCCAACGCCCCAAAAGTTTTGGACCTACGACGGCCGGGGACTACGTACCTCGCGGGGCCCATCGCCTCCATCTTGGAAGATCCAGGAAC															ICCT							
-160	-15	0	-140	-	130		-120	-	110	-10)	-9	90	-80	-	70		-60	-	50		
1		1	1		8.		t 1			1 1			1	1		1		1		1		
CCACTACGACGGAAAGTCAATTTCACACACGGCATCGGCCGCCCATCTACTTGGTTGTCCAAACTATAGCTGGAACAACTCCTGTGGTCGTCGTCGTCGCGTCGCGCGCGCGCGTCG															ACAG							
-40	-30		-20	•	10		1	10		20		30		40	5	50		60	1	0		
1	1		1	1			<u> </u>		*	1			1	1		:		1	1			
TCACGTTCCAATTCACGCTGTTCGALAGAGAGAGAGAGAGAGAA AGCGCAGTCGAGACTGACCGTCCGTGACCAGCTTCTTTGACCTCCCCCGCCTCTAGGTCCTTAGTACACTACACTAC																						
80	90		100	11	0	1	20	13	0	140		150		160	170)	18	10	190			
1	1		1		*		1		1	1		1		1	1	* *****		t CTCAGACTCTT		1 0000001010100		
ATAUCTE	ATACCTECCTACETACCTCTAGETAGECCACECCCTTCCCTETCACCTETCCETCCACCCCCACEACCCCEGAATCCCTTCCCCTCCACCTCTCAGACTCTTTCCCTCAGACACE																					
	210 220																	_				
200	210	2	20	230		24	0	250)	260	1	270	2	80	29	90		30	0	:	310	
200 1	210 1	2	20 1	230 1		24	0 1 400000	250) :	260 1		270 1	2	80 1	25	90 1 20 00	- LC	30	0 \$		310	
200 I Cagagte	210 1 AAGGTC	2 AACCGC	20 1 CAACGTC	230 E Ecato	CTCCI	24 LACCC	0 X ACCCC	250 1 CGGCAG	GTCTC	260 8 G <u>CTCCTC</u>	t CTTCC/	270 t AGTT	2 CCTCTGT	80 1 GCGACA	29 ATG GG Het A	90 I CG GC Ia Al	C AGC a Sei	30 C ACC r Thr	0 # ACA AC Thr Th	C CCA Ir Pro	310 ± ACC Thr	
200 X Cagagtc.	210 X AAGGTC 320	2 AACCGC	20 X CAACGTC 330	230 t ICATC	CTCCI	24 14CCC	0 X ACCCC	250 1 CGGCAG 350) IGTCTC)	260 8 G <u>CTCCTC</u> 3	: <u>CTTCC</u> / 60	270 T AGTT	2 CCTCTGT 370	80 t GCGACA	29 ATG GG Net Al 380	90 1 CG GC 1a Al	ic Age a Sei	30 C ACC r Thr 390	0 X ACA AC Thr Th	C CCA Ir Pro 40	310 8 ACC Thr	
200 8 Cagagte	210 T AAGGTC 320 T	2 AACCGC	20 2 CAACGTC 330 8	230 t tcatc	CTCCI	24 LACCC. 840 1	0 8 ACCCC	250 8 CGGCAG 350	GTCTC	260 1 G <u>CTCCTC</u> 3	CTTCCI	270 X AGTT	2 CCTCTGT 370 8	BO T GCGACA	25 ATG GG Het A 380 1	90 I CG GC In Al	C AGG a Sei	30 C ACC F Thr 390 T	0 t ACA AC Thr Th	C CCA r Pro 40	310 acc Thr 0 t	
200 E CAGAGTC	210 1 AAGGTC 320 1 ACA A	2 AACCGC Gg CCG	20 I CAACGTC 330 I ITT IT Pho Ph	230 E TCATC	CTCC/	24 MACCC 840 8 MAT	O XACCCCC	250 1 CGGCAG 350 1 CG ACA	GTCTC	260 t G <u>CTCCTC</u> 3 GAG CAC	CTTCC/ 50 8 GAC 1	270 ¥ AGTT TTT Pbo	2 CCTCTGT 370 8 CGC TTC Arg Phe	BO E GCGACA CCG CI Pro Al	25 ATG GU Net A 380 t 3C AGG	CG GC la Al CCA	C AGC a Sei GGT (G) - J	30 C ACC Thr 390 BAT TC	0 # ACA AC Thr Th C ATG r Net	C CCA ir Pro 401 GCC G	310 I ACC Thr 0 I GC 1v	
200 E CAGAGTC GCC ACG Ala Thr	210 T AAGGTC 320 T ACA A Thr A	2 AACCGC GG CCG rg Pro	20 8 330 8 TTT TT Phe Ph	230 t TCATC C ACC t Thr	CTCC/ S ATG Het	24 AACCC 340 3 AAT ASN	0 8 ACCCCC CCC A Pro T	250 3 CGGCAG 350 1 CG ACA hr Thr	GTCTC	260 x g <u>ctcctc</u> 3 gAG CAC g1u Bis	CTTCC/ SO S GAC 1 Asp 1	270 I AGTT TTT Phe	2 CCTCTGT 310 2 CGC TTC Arg Phe	80 T GCGACA CCG CI Pro An	29 ATG GU Net A 380 t SC AGG rg Arg	CG GC A CCA Pro	C AGG a Sei GGT (Gly /	30 C ACC r Thr 390 s GAT TC Amp Se	0 8 ACA AC Thr Th C ATG r Het	C CCA r Pro 40 GCC G Ala G	310 * ACC Thr 0 * GC ly	
200 E CAGAGTC GCC ACG Als Thr	210 3 AAGGTC 320 1 ACA A Thr A 410	2 AACCOC GG CCG rg Pro	20 1 CAACGTC 330 1 TTT TT Phe Ph 420 1	230 E TCATC C ACC E Thr	CTCC/ 3 ATG Het	24 AACCC 840 8 AAT Asn 130 8	0 1 ACCCCC CCC A Pro T	250 1 CGGCAG 350 1 CG ACA hr Thr 440	GTCTC	260 t G <u>CTCCTC</u> 3 GAG CAC Glu Bis 4	CTTCC/ SO E GAC 1 Asp 1 50 E	270 X AGTT TTT Phe	2 CCTCTGT 370 2 CGC TTC Arg Phe 460 1	80 E GCGACA CCG CI Pro An	29 ATG GU Net A 380 1 GC AGG 1 ATG 470 1	90 EG GC la Al CCA Pro	C AGC a Sei GGT (Gly /	30 C ACC Thr 390 8 GAT TC Asp Se 480 8	0 # ACA AC Thr Th C ATG r Het	C CCA ir Pro 40 GCC G Ala G 49	310 x ACC Thr 0 x GC 1y 0 x	
200 3 CAGAGTC GCC ACG Ala Thr ACC GGA	210 1 AAGGTC 320 1 ACA A Thr A 410 1 CTT G	2 AACCGC GG CCG rg Pro GG GGT	20 I CAACGTC 330 I TTT TT Phe Ph 420 I SCC GC	230 t TCATC C ACC t Thr C ATG	CTCCJ S ATG Het TCA	24 ACCC AAT AAT ASN I30 TCG	0 I ACCCCC CCC A Pro T TCC T	250 8 CGGCAG 350 1 CG ACA hr Thr 440 1 CA GCC	GTCTC	260 t g <u>ctcctc</u> 3 gag cac giu Eis 4 AAC AAC	CTTCCI SO EGAC Amp 50 E CAC	270 I AGTT TTT Phe	Z CCTCTGT 370 z CGC TTC Arg Phe 460 z CAG CAC	BO I GCGACA CCG CI Pro An	29 ATG GU Het A 380 1 380 1 380 1 380 1 380 1 470 1 500 ATG	CG GC In Al CCA Pro	GCA SCA	30 C ACC T Thr 390 EAT TC Asp Se 480 E TTC AA	0 # ACA AC Thr Th C ATG r Het C CAC	C CCA Ir Pro 40 GCC G Ala G 49 CAC C	310 x ACC Thr 0 x GC ly 0 x AC	
200 8 CAGAGTC GCC ACG Ala Thr ACC GGA Thr Gly	210 x AAGGTC 320 x ACA A Thr A 410 x CTT G Leu G	2 AACCGC GG CCG rg Pro GG GGT ly Gly	20 8 CAACGTC 330 8 TTT TT Phe Ph 420 8 GCC GC Ala Al	230 E TCATC C ACC e Thr C ATG a Het	CTCC/ ATG Het TCA Ser	24 AACCC 840 8 AAT 430 8 TCG Ser	0 8 ACCCCC CCC A Pro T TCC T Ser S	250 8 CGGCAG 350 1 CG ACA hr Thr 440 1 CA GCC er Ala	GTCTC	250 3 GCTCCTC 3 GAG CAC Glu Eis 4 AAC AAC Ash Ash	CTTCC/ SO E GAC SO E CAC His	270 TTT Phe AAC Asn	2 CCTCTGT 370 ± CGC TTC Arg Phe 460 ± CAG CAC Gln His	80 T GCGACA CCG C(Pro A) CAC C(His P)	29 ATG GU Net A 380 1 3C AGG 12 ATG 470 1 5C ATG 10 Net	90 1 CG GC la Al CCA Pro AGC Ser	C AGO a Ser GGT (Gly / GCA (Ala)	30 C ACC F Thr 390 EAT TC Asp Se 480 E TTC AA Phe As	O Thr Th C ATG r Het C CAC a His	C CCA r Pro 40 GCC G Ala G 49 CAC C His H	310 * ACC Thr 0 * GC 1y 0 * AC is	
200 8 CAGAGTC GCC ACG Ala Thr ACC GGA Thr Gly	210 x AAGGTC 320 x ACA A Thr A 410 x CTT G Leu G 500	2 AACCGC GG CCG rg Pro GG GGT ly Gly	20 8 CAACGTC 330 8 TTT TT Phe Ph 420 8 GCC GC Ala Al 510	230 E ICATC C ACC e Thr C ATG B Het	CTCCJ S ATG Het TCA Ser	24 AACCC 340 3 AAT 4 Sen 330 3 TCG Ser 520	0 8 ACCCC CCC A Pro T TCC T Ser S	250 3 CGGCAG 350 1 CG ACA hr Thr 440 1 CA GCC er Ala 530	GTCTC	260 8 G <u>CTCCTC</u> 3 GAG CAC Glu Eis 4 AAC AAC ASB ASB	CTTCC/ SO 8 GAC 1 Asp 1 50 8 CAC 4 His 4	270 8 AGTT TTT Phe AAC	2 CCTCTGT 370 2 CGC TTC Arg Phe 460 3 CAG CAC Gln His 550	SO T GCGACA CCG CI Pro An CAC Ci His Pi	29 ATG GG Het A 380 ± GC AGG tr Arg 470 ± 560	90 1 CG GC 1a Al CCA Pro AGC Ser	C AGC a Sei GGT (Gly / GCA 1 Ala 1	30 30 390 390 390 390 30 390 390 390 390	O RACA AO Thr Th C ATG r Het C CAC n His	C CCA r Pro 40 GCC G Ala G 49 CAC C His H	310 x ACC Thr 0 x GC 1y 0 x AC is	
200 8 CAGAGTC GCC ACG Ala Thr ACC GGA Thr Gly	210 320 320 4 ACA A Thr A 410 3 CTT G Lew G 500 3	2 AACCGC GG CCG rg Pro GG GGT ly Gly	20 8 CAACGTC 330 8 TTT TT Phe Ph 420 8 GCC GC Ala Al 510 8	230 E TCATC C ACC e Thr C ATG a Het	CTCCJ ATG Het TCA Ser	24 AACCCC 349 8 AAT Asn 130 8 TCG Ser 520 8	0 8 ACCCC CCC A Pro T TCC T Ser S	250 1 CCGCAQ 350 1 CCG ACJ 1 CCG ACJ 1 CCA GCC CA GCC CA GCC CA GCC 1 1 CA GCC 1 1 CA GCC 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	GTCTCC	260 8 G <u>CTCCTC</u> 3 GAG CAC Glu Bis 4 AAC AAC Asn Asn 5	CTTCC/ SO 3 GAC 1 SO 50 50 50 50 40 8	270 E AGTT TTT Phe AAC Asn	2 CCTCTGT 370 2 CGC TTC Arg Phe 460 3 CAG CAC Gln His 550 1	80 EGGACA CCG CI Pro An CAC CI His Pi	2: ATG G4 Het A 380 * 3C AGG rs Ars 470 * CC ATG ro Net 560 *	OO I CG GC I A Pro AGC Ser	C AGG a Ser GGT (Gly / GCA (Ala)	30 2 ACC r Thr 390 8 330 8 8 480 8 480 8 7TC AA 9Phe As 570 8	0 8 ACA AC Thr Th C ATG r Het C CAC a His	C CCA r Pro 40 GCC G Ala G 49 CAC C His H 58	310 x ACC Thr 0 x GC ly 0 x AC is 0 x	
200 8 CAGAGTC GCC ACG Ala Thr ACC GGA Thr Gly CAC CAC	210 320 320 5 ACA A Thr A 410 500 500 500 500 500 500	2 AACCGC GG CCC GG CCC rg Pro GG GGT 1y Gly CT GCT	20 2 330 2 TTT TT Phe Ph 420 5 GCC GC Ala Al 510 2 GGC AG	230 I TCATC C ACC e Thr C ATG B Het	CTCCJ S ATG Het TCA Ser	24 AACCCC 840 1 AAT 4sn 130 1 Ser 520 1 CGC 520	0 8 ACCCCC CCC A Pro T TCC T Ser S CGT G	250 8 350 8 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	GUTCTC	260 E GCTCCTC 3 GAG CAC Glu Bis 4 AAC AAC Ash Ash 5 GGT CGC Clu A	CTTCC/ 60 2 GAC 1 Asp 1 50 2 CAC 4 His 4 40 2 CCC 4	270 E AGTT TTT Phe AAC Asn	2 CCTCTGT 370 2 CGC TTC Arg Phe 460 2 CAG CAC Gla Bis 550 2 AGC AGC	BO T GCGACA Pro An CAC C(HIS P)	23 ATG G4 Het A 380 E C AGG E C AGG E C AGG E C AGG E C AGG E C AGG E C AGC E C AGC E C AGC E C AGC E C AGG E C AGG AGG E C AGG E C AGG AGG E C AGG E C AGG E C AGG AGG AGG AGG AGG AGG AGG AGG AGG AGG	BO I CG GC Ia Al CCA Pro AGC Ser ACC Ser	C AGC a Ser GGT (Gly / Ala 1 AGC /	30 2 ACC 7 Thr 390 8 390 8 8 480 8 1 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	O ACA AC Thr Th C ATG r Het C CAC n His C TTT Pho	C CCA r Pro 40 GCC G Ala G 49 CAC C His H 58 GTC G	310 x ACC Thr 0 x GC ly 0 x AC is is CCC	
200 8 CAGAGTC GCC ACG Ala Thr ACC GGA Thr Gly CAC CAC Bis Bis	210 x AAAGGTC 320 x ACA A Thr A 410 x CTT G Leu G 500 x AAT G AAT A	2 AAACCGC GG CCG rg Pro GG GGT ly Gly CT GCT la Ala	20 330 TTT TT Phe Ph 420 5 GCC GC Ala Al 510 5 GGC AG GIT Se	230 t TCATC C ACC e Thr C ATG a Het T GCC r Ala	CTCCI ATG Het TCA Ser : : : : : : : : : : : : : : : : : : :	24 AACCCC 340 3 AAT 430 3 TCG Ser 520 3 CGC Gly	0 3 ACCCCC Pro T TCC T Ser S CGT G Arg A	250 350 1 CGGCAG 1 CG ACA 1 Thr Thr 440 1 CA GCC CA GCC CA GCC 1 3 XT AGC 530 59 Set	GTCTC:	260 t G <u>CTCCTC</u> 3 GAG CAC Glu Bis 4 AAC AAC Asn Asn 5 GGT CGC Gly Arg	CTTCC. CTTCC. CGAC 1 Amp 1 50 1 CGAC 1 His 1 40 5 CCC 1 Pro 1	270 TH AGTT Phe AAC Asn AGC Ser	2 CCTCTGT 370 2 CGC TTC Arg Phe 460 2 CAG CAC Gln His 550 2 AGC AGC Ser Ser	80 1 GCGACA Pro A1 CAC CC Bis P1 Ash A1	23 ATG GI Net Al 380 2 C AGG 470 3 C AGG 470 5 C ATG 5 5 60 8 C AAC 8 AC AAC 8 AC ASB	OO X CG GC Ia Al Pro AGC Ser ACC Thr	C AGC a Ser GGT (Gly / GCA 1 Ala 1 Ser /	30 C ACC r Thr 390 t CASP Se 480 t TTC AA SPhe As 570 t AASP GI	O S ACA AC Thr Th C ATG r Het C CAC n His C TTT y Phe	C CCA r Pro 400 GCC G Ala G 49 CAC C His H 58 GTG G Val A	310 x ACC Thr 0 x GC 1y 0 x AC is CC 1a	
200 8 CAGAGTC GCC ACG Ala Thr ACC GGA Thr Gly CAC CAC Bis Bis	210 320 320 320 320 320 320 320 500 500 500 500 500 500 500 5	2 AAACCGC GG CCG rg Pro GG GGT ly Gly CT GCT la Ala	20 330 1 TTT TT Phe Ph 420 5 GCC GC Ala Al 510 1 GGC AG 6G7 Se 600	230 t TCATC C ACC e Thr C ACC e Thr C ATG a Het T GCC r Ala	CTCCJ ATG Het TCA Ser Ser Ser	244 AACCCC 340 3 AAT 330 3 TCG Ser 520 3 GGC Gly 510	0 3 ACCCCC CCC A Pro T TCC T Ser S CGT G Arg A	250 1 CGGCAG 350 1 CG ACA hr Thr 440 1 CA GCC CA GCC CA GCC 1 530 2 S 7 AT AGC 521 521 521	A ACA I A ACA I Thr I A ACA I A ASA I A ASA I A ASA I A ASA I A ASA I	260 t g <u>crtctrc</u> 3 GAG CAC Glu Tis 4 AAC AAC Asn Asn 5 GGT CGC GI Ard	CTTCCC GAC 1 GAC 1 Asp 1 550 1 CAC 1 Bis 4 40 1 CCC 1 Pro 1 30	270 TTT Phe AAC Asn AGC Ser	2 CCTCTGT 370 2 CGC TTC Arg Phe 460 2 CAG CAC Gla His 550 2 AGC AGC Ser Ser 640	80 t GCGACA Pro Ai Eis Pi i Aac Ai Ash Ai	2: ATG GU Net Al 380 8 C AGG 470 8 C ATG 70 Net 560 8 AC AAC 8 8 AC AAC	90 1 CG GC 1a A1 CCA Pro AGC Ser ACC Thr	GGT (GIT (GIT / GCA (Ala) AGC (Ser (30 30 7 ACC 7 Thr 390 8 390 8 480 8 7 7 7 7 7 7 7 8 8 8 7 7 7 8 8 8 8 8	0 s ACA AC Thr Th C ATG r Het C CAC a Bis C TTT y Phe	C CCA r Pro 400 GCC G Ala G 49 CAC C His H 58 GTG G Val A 67	310 x ACC Thr 0 x GC ly 0 x AC is is icc ia	
200 8 CAGAGTC GCC ACG Ala Thr ACC GGA Thr Gly CAC CAC Bis Bis	210 320 320 320 320 320 5 4 410 5 500 5 4 410 5 500 5 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5	2 AAACCGC GG CCG rg Pro GG GGT ly Gly CT GCT la Ala	20 t CAACGTC 3300 t TTT TT Phe Ph 420 t GCC GC AGC Ala Al 510 t GGC AGC Gly Se 600 t	230 E FCATC C ACC e Thr C ATG a Het T GCC r Ala	CTCCJ S ATG Het (TCA Ser ! : CGT Arg	24 AACCC 340 8 AAT 430 8 TCG 8er 520 1 CGC Gly 510 8	0 8 ACCCCC CCC A Pro T TCC T Ser S CGT G Arg A	250 1 CCGCCAQ 350 1 CCG ACA 1 CCG ACA 1 Thr Thr 440 1 CA GCC 1 CA GCC 1 S3 5 3 1 AT AGC 59 8 1 1 C	AACA I AACA I Thr I AACA I AAA I AACA I AAAAAA AAAAAA AAAAAA AAAAAAA AAAAAA AAAA	260 8 G <u>GTCCTCC</u> 3 GAG CACC AGC AACC AACC AACC AACC AACC AGC AACC AGC CGCC GGT CGCC GGT CGCC GGT CGCC GGT CGCC	CTTCC. GO CAC 1 GAC 1 CAC 1 His 1 CAC 1 His 2 CCC 1 Pro 1 30 2	270 TTT Phe AAC Asn AGC Ser	2 CCCCTCTGT 3700 z CCCC TPC 460 z CCG CAC CAC CCG CAC CAC SCO Ser Ser 640 z SCO Ser Ser	SO T GCGACA CCG CC Pro An CAC CC Bis P: Aan A Aan A	23 ATG GG Net Al 380 8 36 AGG AGG 8 470 8 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	90 1 CG GC 1a A1 CCA Pro AGC Ser ACC Thr	GGT (GGT (GIy / GCA (Ala) Ser /	30 30 30 2 ACC 2 Thr 390 3 3 390 3 3 390 3 3 3 3 3 3 3 3 3 3 3 3 3	O ACA AC Thr Th C ATG C ATG T Het C CAC a His C TTT y Phe	CC CCA r Pro 40 GCC G Ala G 49 CAC C His H 58 GTG G GTG G Val A 67	310 aCC Thr 0 aCC aCC 1y 0 aCC 1y 0 aCC aCC aCC aCC aCC aCC aCC	

FIG. 2. Nucleotide sequence of *nit-2* and its flanking regions. The adenine base which composes the single transcription start site is numbered +1. Vertical arrows indicate 5' and 3' termini of the *nit-2* transcript. The 5' terminus was determined by primer extension analysis, and the 3' end was identified as being adjacent to the poly(A) tract of a cDNA clone. Transcription initiation begins at the 3' end of a 21-bp repeated sequence composed entirely of purine bases in the coding strand (underlined and overlined). Pyrimidine-rich and thymidine-rich tracts within the 5' and 3' noncoding regions of the *nit-2* mRNA are underlined. The translated amino acid sequence which encodes a protein of 1,036 amino acids is shown beneath the DNA sequence. The putative Cys_2/Cys_2 zinc finger domain is underlined. Intron 1 (98 bp) and intron 2 (78 bp) both have the consensus 5' GT and 3' AG splice site sequence.

introns have good consensus 5' and 3' splice site sequences. Removal of the two introns from the primary transcript yields an mRNA of 4.3 kb which contains a protein-coding region of 3.1 kb and long 5' and 3' nontranslated segments.

Organization of *nit-2* **gene.** The initiation site for transcription of the *nit-2* gene was determined by primer extension analysis. The results indicated that transcription initiation occurs at a single site located 284 bp upstream of the first AUG codon which begins a long open reading frame (Fig. 2). No consensus TATA box sequence is found within 150 bp upstream of the initiation site. Initiation occurs with an adenine residue located at the 3' end of an unusual, highly symmetrical 21-bp sequence which consists entirely of purine bases on the coding strand (Fig. 2). The 5' nontranslated region of the transcript is very rich in pyrimidine bases and

contains four stretches (of 9, 11, 11, and 16 bp) composed entirely of C and T residues (Fig. 2). The first AUG codon is closely followed by four additional in-frame AUG codons. It is unknown which of these five in-frame AUG codons actually represents the initiation codon; however, the sequence surrounding the second AUG codon has the best match to the *N. crassa* consensus translational start sequence. Translating from the initial AUG codon to the UAA stop codon yields a protein of 1,036 amino acids, with a molecular weight of approximately 110,000. Comparison of the sequences of genomic and cDNA clones revealed that polyadenylation of the *nit-2* transcript occurs 933 bp downstream of the UAA stop codon. Therefore, the *nit-2* transcript has long nontranslated regions at both its 5' and 3' ends. The very long 3' nontranslated region of the *nit-2*

		680	0 690 700							710 720							730					740				750			
		1				1			1		~~~	1		~~~		1	CAC	223	1 CTC	607	660	I ATG	ACC	CAG	AGC	¥ CCG	GAC	GAG	ATG
GGC Cl #	GCC	GCA Ala	GCC	TAT Typ	GAT	Leu	Leu	Arg	Ser	Ser	Ala	Phe	Pro	Pro	Phe	Gla	Asp	Gly	Leu	Ala	Gly	llet	Thr	Gla	Ser	Pro	Asp	Glu	Het
•1,																		,				830			84	0		8	50
		770			11	1		1	130			1			•	1			1			1			•••	1			8
CAG	***	CAG	GAC	CCG	CTG	GCT	ACC	CAG	CTC	TGG		TAC	TTT N	GCT	AAG	ACC	AAG	CTG	GCC	CTG	CCC	AAC	CAG	GAG C1.	CGC	ATG Nat	GAG C1m	AAC	TTG Leu
Gla	Ly:	Gla	Asp	Pro	Leu	Ala	ħr	GIN	Val	Trp	Lys	TJP	rne	A18	LYE	Ter	ьув	Per	A18	Per.	rro		918	016		86.			
		860			81	70		1				890			91	0		1	910		92	0		930		94	0		950 1
107	900	1	140	ATC	900	1	CCG	CTG	T CAG	ACC	TAT	T CGC	ACC	CAA	ATG	T GAG	ACT	GAC	CG	GTAT		• Igaai	AGA	IAGA(TAT	TAAG	IAGAC	GTCG	TGTGT
Thr	Trp	Arg	llet	Ket	Ala	Lys	Pro	Les	Gla	ħr	tyr	Arg	Arg	Gla	Het	Glu	ħr	٨sp	Ar										
							•		446		180			1818			102	20		1	030		1	040			10	i0	,
	301	1		1			:		1		1	1		1				1			1			1				1	
TGT	CCCC	AGC	STCT	GGCC	AGCA	AGCT	MTT	1744	TTGC	ATGT	GTTG	CAACI		AG G	ACA	CAC	CGT	TTT Phe	TCG Ser	GAA Glu	TCA Ser	GCC	CCT Pro	CAA Gla	LVE	TCG Ser	ACC Thr	AGC Ser	GUC Gly
														•	181					•					-,-				•
1	060			1070			10	80		1	990			1100			11	10		1	120			1130			114	10	
ATC	1 6C6	CGA	TTG	T CGC	AAG	TCC	TCG	I GAG	CAG	ACC	cia	TCC	CAG	GGG	TCT	GAC	CTC	ATG		CTG	GAC	GAT	TTC	ATC	AAC	GGC	GAA	AAC	ATA
Ile	Ala	Arg	Leu	Arg	Lys	Ser	ler	Glu	Gla	ħr	Gla	Ser	Gla	Gly	Ser	٨ŋ	Leu	llet	Asn	Leu	Å\$P	Asb	Phe	Ile	Åsa	Gly	Glu	Åsn	Ile
1	150			1160			11	78		1	180			1190			12	00		1	210			1220			12	30	
	1			1			••	1		-	ŧ			1				1			1			:				1	
AGC	ACG	CCA	GCC	GGC	CTG	TCT	CTT	GCG	CCT	TCG	CCC	GAA	ACA	TCC Ser	TCC	- AAA - L.v.e	ATG Net	GCC	GAC	GAC	GGA	ACT The	GCT	CAC	Lis	TCG Ser	ACG Thr	GUT	TUA Ser
aer	180				16.6	961	DC C					••••				-,-													
1	240			1250)		12	60		1	270			1280			12	90 1		1	300 1			1319			13	20 1	
GCT	ATC	CCT	ATC	:	GCT	CGC		GAC	CAG		TCT	CAG	CAC	ATC	ATC	CCT	CAG	TCO	C GTC	: CC1	GCC	GCC	CTG	CAC	CAC	CCA	AGG	ATG	CAG
41a	Ile	Pro	Ile	e Lyn	A la	. Arg	Lys	. Asy	Glu	Gla	Ser	Gla	li	llet	Ile	Pro	Glm	Sei	r Val	Pro	Ala	Ala	Leu	lis	lis	Pro	Arg	Het	Gla
1	330			1340)		13	50		1	360			1370)		13	80		1	390		140	0	1	410		142	0
	1			1	t			1			1			1				1			1			1		1			1
ACC	GAA G1=	TTI Phe	GGG	C 141 • 141	t CTI - Lei	1 CC1 1 Pro	t CGA	CAC I II II	CTI 1 Lei	CGC Ari	: Ш : Цуя	ACC The	AGC Sei	: ATO : Ile	; GAC : Ast	GAG Glu	i ACG i Thr	i AGU Sei	С ААЛ С Цун	l CGG L Ari	; GTG I	AGTA	GUTT	m	UTUT	TATU	AGTU	TUBU	UTUTTU
	••••																		•									••	
14	30		144		14	150		1460) t		1470) t		14	10 1		14	90 1			1500			151	1		15	20 1	
ATC	TGGT	GTAI	TCT	GTC	IGAT/	TAA	TATO	XTT	IGTCI	N DAY	AC C	200 I Pro I	AC (GC I		GC (CC G	In la	GAT 1 Asp 1	rtc : Phe i	ICT (ler l	CC 0 'ro 1	AC G is 1	TC 1 al 1	ICC G ler A	CG G la V	TG A	CT C hr P	CT AGO TO Ser
1	30			1540			155)		1!	560		1	.570	-	-	1586)	-	1	i 90		1	600			1610)	
_	1			1			1	1			1			1			1				1			1			1		
TAC Tyl	; GT(; Va)	i ACC Thi	: AA : As	C GG n Gl;	C CTI V Lei	C GA(N Asj	C GC(P Al:	C GA' A Asi	T AC: p Thi	r GA! r As;	r CTI p Lei	C CAN I III	r GA' I As	r TAI p Ty:	r Sei	i UTI i Lei	r GAN 1 Asy	r CA 9 Hi	r aci 8 Thi	G AG' r Sei	r CAN r Hin	r GAC 1 Asy	; GG() Gly	; TT(Lei	i CCG 1 Pro	Pro	Gli	ACC 1 The	GCC Ala
1	.96			1694			164		-	1	55A			1664			1670			1		-	1				1764	,	
1	1			1030 1			104	1		1	1			1000			1010	1		1	1		1	1			1141	1	
CC	T TC	TCI	G GT	t cc	C TA	C GC	T CT	C GA	T AC	C GT	C GG	C CTI	C GA	T GC	T GAT			C AT	C AC	C TC	t GC	C GG1			C CAO			110	TCT
Pr	s Sei	t Sei	r Va	i Pr	0 T y	r Al	a Le	L ÅS	p Th	r Va	1 G1;	y Lei E14		ן אן ר		p Thi	r M (, 11 ,	e Th	r Je	r Ali	L GI	f Pro	o Pho	e Gli	Gli	A 81	h Pho	e Ser
												L10	J.	2	U OI	niin	uec	ι.											

mRNA also contains multiple pyrimidine-rich, especially thymidine-rich, stretches (Fig. 2).

Analysis of 3' deletion clones of nit-2. Deletion clones which truncated the 3' end of the nit-2 gene were constructed as described in Materials and Methods (Fig. 3). Each of these deletion clones was transformed into the nit-2 mutant strain to test its functional activity. The subclone which contains the entire 4.3-kb EcoRI fragment and the deletion clones 41 and 22 all complemented the nit-2 mutant, indicating that each was expressed to give a functional nit-2 gene product. However, deletion clones 76 and 19 failed to transform the *nit-2* mutant. These results suggest that amino acid residues in the region between the sites defined by deletions 22 and 76 are important for function of the nit-2 protein. However, the manner in which the nit-2 deletions were made resulted in fusion proteins in which each truncated nit-2 protein was fused to amino acids specified by plasmid sequences (deletions 19, 22, 41, and 76 have, respectively, 51, 11, 12, and 50 extra amino acids fused to the truncated nit-2 protein). It is

not clear whether these additional amino acids at the carboxy terminus affected the activity of the various proteins. However, it should be emphasized that the truncated *nit-2* protein encoded by deletion 22, which lacks the carboxyterminal 214 amino acids, was functional, thus demonstrating that these amino acids residues are dispensable for activity.

The *nit-2* protein deduced from the nucleotide sequence is composed of 1,036 amino acids and contains only four cysteine residues, all of which are located in the region near the endpoint of deletion 76. The cysteine residues form a consensus zinc finger motif CYS-Xaa₂-CYS-Xaa₁₇-Cys-Xaa₂-Cys, in which the cysteine residues are postulated to serve as zinc-binding ligands (Fig. 4). Deletion 76 occurs in the 17-amino-acid loop of the potential zinc finger domain; thus, the lack of function of the deletion 76 protein is consistent with the suggestion that the *nit-2* protein is a DNA-binding protein with an essential zinc finger domain.

Mutagenesis of potential nit-2 zinc finger motif. To investi-

17	10	1720				1730			1740				1750 #			1	760 1	1760 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		177	0		11	80 1		1	790 1		
TTC Phe	TCG Ser	CCG Pro	TCC Ser	ACG Thr	TCA Ser	CCG Pro	ATG Net	GTT Val	AGC Ser	CAC His	GAT Asp	CCA Pro	TTT Phe	ACC Thr	GCC Ala	ATG Het	TTT Phe	GGC Gly	CCC Pro	AAC Asn	AAC Asn	TCT Ser	TCT Ser	ATG Net	CAC Hib	AGC Ser	GGC Gly	CCT Pro	ATC Ile
180	0		1810			1820			1830				1840			1	850			186	0		11	370		1	880		
	I GGC	AAC	AAC	*	TAC	TCG	* CCC	CCT	GCC	TCC	X GCA	TTC	CAG	¥ TCG	ACT	GCC	TCT	ACG	CCT	CAT	I CCT	ATG	AAC	¥ GAG	GCC	GGT	I GAT	AAC	TTT
Åsn	Gly	Asn	Asn	Phe	Tyr	Ser	Pro	Pro	Ala	Ser	Ala	Phe	Glm	Ser	thr	Ala	Ser	Thr	Pro	H is	Pro	Net	Asn	Glu	Gly	Gly	Asp	Asn	Phe
18	90	1900					1910		1920				1930			1940 *				195	i0 1		19	960 1		1	970 1		
TAC	TTT	GGT	GTC	GAT	ATG	CGC	CGC	GCT	CGC	CAG	CAG	CCT	TAT	CAG	CCA	GGA	AAC		GGA Cl-	ATG	GGC	AAC	GCC	ATG	GCT	CAC	CAG	TTC	CCC
Tyr	rae	617	V81	ABP	aet	AFE	Arg	AIS	ALE		010	riv	. yr		riv	GIY ASE BIS G			913	****							010		
13	80 1		1	1 330			2000			2010 ¥			2020			1					1			1			1		
TAT Tyr	GCC Ala	GGT Gly	AAC Asb	GGT Gly	AAC Asb	ATG Net	ATG Het	TTC Phe	CCG Pro	GCG Ala	TCT Ser	TCG Ser	GCT Ala	GGC Gly	CAG Gla	GAT Asp	CCG Pro	ACG Thr	CCT Pro	TCC Ser	TTT Phe	GCG Ala	GUT Ala	CCG Pro	Ast Ast	TCG Ser	Phe	Ser	Gly
20	70		2	080			2090			210	00		2	110		;	2120			213	10		2	140		:	150		
CAC	X ATT	GAC	ccc	X ACC	CAA	GTG	TTC	CAT	AAC	GAA	X Caa	GCA	GTC	I CGA	TCA	CCA	# 66C	ATG	AGT	GTT	I TTG	CAA	GAC	I Agt	CTA	TTC	I ACA	TTC	GGA
lis	Ile	Asp	Pro	Thr	Gla	Val	Phe	lis	Asb	Glu	Gln	Ala	Val	Arg	Ser	Pro	Gly	llet	Ser	Val	Leu	Glm	Asp	Ser	Leu	Phe	Thr	Phe	Gly
21	60 1		2170			2180				2190			2200			2210 ¥			2220				2	230		3	2240		
GCA Ala	GAA Clu	TCT	GAT	GGA Cl w	GAC	GAG Glu	GAG Gla	GAT	GGT Glv	GGC Cly	GCA Als	TTT Phe	GCC Ala	GAT	CGC	AAT Asp	CTA Les	TCG Ser	ATT	TCT Ser	CAT Nie	GAC	TTT Phe	TCT	TCC	CAG Gla	GGC Glv	ATG Net	GAG Glu
•••	50		<i>P</i>	76N			· · · · ·		2280				2290			9260				9316				120		2330			••••
	1			1		12/0			1				1 1			2300 1					1		2320			2330 8			
GAA Glu	Pro	GCA Ala	TTT Phe	Asp	Ser	Pro	Ser	Het	Gly	Trp	6AT Åsp	Pro	Ser	Leu	Pro	GGA Gly	Ast	Phe	AGC Ser	ACG Thr	Gla	Ala	Ala	AGG	tat Tyr	Pro	GUT Gly	GUC Gly	CCC Pro
23	40		2	350		2360			2370				2380			2390				2400				2410			2420		
CCG	X CGA	***	CAG	I GTT	ACG	ATT	¥ GGC	GCT	ACA	ACA	I ACC	GAT	TAT	t GTT	GAT	AAC	I ACT	GGA	GAA	tgg	1 Gat	GGA	AGC	t Ggt	CTT	CCC	1 AGG	TCT	CAG
Pro	Arg	L ys	Gla	Val	Thr	Ile	Gly	Ala	Thr	Thr	Thr	Asp	fyr	Val	Asp	Asa	Thr	Gly	Glu	Trp	Asp	G1 y	Ser	Gly	Leu	Pro	Arg	Ser	Gln
24	30 1		2	440 1			2450 ¥			24	50 1		2470				2480 1			24	90 1		2	500 1		2510			
TCG Ser	CAG Glb	TCA Ser	TTC Phe		CAG Gla	AGC Ser	GAC	CTG Lea	CGT		GGC Gly	AAG Lys	ATG Het	TCT Ser	AGG Are	ACG Thr	GCA Ala	TCG Ser	ACC	CCA Pro	GGT Glv	CTG Leu	TCC Ser	GCT Ala	AGA Are	ATG Net	AAC	CCG	TTC
25	20		2	530			2540			25	50	-,-	2	560			2570			25	80		,	590			 •		INC
	1		-	1	848	1370			792A			000	-	1			1				1			1			1		
Glu	Arg	Leu	Ala	Gln	Ser	Ala	Ser	lis	Ser	Pro	Pro	Ala	GAC Asp	Val	Gly	Arg	Ser	Ser	Gly	Leu	Ser	TCG Ser	UTA Val	Pro	GCA	AGC Ser	CGG Arg	CCT Pro	TCA Ser
26	10		2	620		2630			2640				2650			2660				2670				680		2690			
TCC	I CCT	CCT	CCT	T GGA	GCA		1 CAA	GGT	TCG	ACA	X ACG	AAC	TTG	I CAG	GGC	GCA	t GCT	GGA	AAC	TCG	¥ ACA	GAC	ACA	t CCA	ACA	407	1 101	100	440
Ser	Pro	Pro	Pro	Gly	Ala	Ly	Gla	Gly	Ser	Thr	Thr	Asn	Leu	Gla	Gly	Ala	Ala	Gly	Asa	Ser	Thr	Asp	The	Pro	Thr	Thr	Cys	Thr	Asp
												FI	G.	2-	-Col	ntin	uec	1.											

gate whether integrity of the potential zinc finger domain was necessary for function of the *nit-2* protein, we changed two of the four cysteine residues simultaneously by site-directed mutagenesis; cysteine residues 743 and 746 were changed to serine and glycine, respectively. The mutated gene segment was sequenced to confirm that the desired changes had been obtained, and then a restriction fragment containing the altered sequence was inserted into the corresponding position of an otherwise wild-type *nit-2* gene. The mutant gene thus constructed was found to be nonfunctional, i.e., incapable of transforming the *nit-2* mutant strain, demonstrating that the cysteine residues are essential, consistent with their potential role in formation of a zinc finger DNA-binding motif.

DISCUSSION

The *nit-2* gene is the major positive-acting nitrogen control gene of N. crassa. During conditions of nitrogen limitation,

the nit-2 gene product is required to turn on the expression of an entire series of unlinked structural genes which specify nitrogen catabolic enzymes. Here we report the complete nucleotide sequence of the nit-2 gene and its flanking regions. The 4.3-kb nit-2 mRNA has a long nontranslated 5' leader (284 bp) and an unusually long 3' nontranslated sequence (933 bp), both of which contain stretches that are pyrimidine rich. Although the significance of these extensive nontranslated regions of the nit-2 mRNA is unknown, they represent potential areas for control of mRNA translation or stability. The presence of the five closely spaced in-frame AUG codons at the beginning of the protein-coding region also suggests a potential for translational control of nit-2, perhaps similar to that suggested for the N. crassa cpc-l and the yeast GCN-4 proteins (25, 32). No other potential start sites or open reading frames occur in the 284-bp 5' leader sequence which precedes the first of the five closely spaced AUG codons.

TEC TTT ACE CAA ACE ACE CCA TTE TEE CEC CET AAC CCA GAT GEA CAA CCC CTC TEC AAC ECT TET GEC TTE TTE AAG CTC CAT GET Cys Phe Thr Gla Thr Thr Pro Leu Trp Arg Arg Asa Pro Asp Gly Gla Pro Leu Cys Asa Ala Cys Gly Leu Phe Leu Lys Leu Bis Gly GTC GTG AGA CCG CTA AGT TTG AAG ACA GAT GTT ATC AAG AAA CGG AAC CGC GGT TCG GGA GCG AGC CTG CCT GTG GGC GGT ACG AGC ACG Val Val Arg Pro Leu Ser Leu Lys Thr Asp Val Ile Lys Lys Arg Asa Arg Gly Ser Gly Ala Ser Leu Pro Val Gly Gly Thr Ser Thr CEG TEC ANG ANG ANT GEA AGE ATE AGT GEA GES GET ESG ANG AND TEG ACT TTE TEC ATT ACE TEC AND GEC AND AND CAN CEA CEA CEC GEC Arg Ser Lys Lys Ass Als Ser Het Ser Als Als Als Arg Lys Ass Ser Thr Les Ser Ile Thr Ser Ass Als Ass Ass Gla Pro Pro Als . CAR GTE GEG ACA DEG DEG GET CAG CAA GTE CET GEC AGE AGT GTE CAAC GAG AGE GAA AGE CET GEG AGT GGE CEC GET TEG GGT GGE Gin Val Ala Thr Pro Pro Ala Gin Gin Gin Val Arg Ala Ser Ser Val Ann Giu Ser Giu Ser Pro Ala Ser Giy Pro Ala Ser Giy Giy Asm Thr Ala Gly Ser Thr Pro Thr Ser Tyr His Gly Ser Thr Gly Ser Thr Ser Gly Ala Val Gly Gly Lys Ser Val Ile Pro Ile Ala Ser Ala Pro Pro Lys Ser Ala Pro Gly Pro Gly Ala Gly Ser Het Ser Arg Arg Asp Thr Ile Ser Ser Lys Arg Gla Arg Arg His Ser ANG AGE GEE GGA AGE GAT CAG CET GTE AGT GEE GGA GET GTE AGE AGE AGE GGA ATE GAE GTT GAT AGT CEE GEE AAC TEG ACA GGA TET Lys Ser Ala Gly Ser Asp Gla Pro Val Ser Ala Gly Ala Val Ser Ser Ser Gly Het Asp Val Asp Ser Pro Ala Asa Ser Thr Gly Ser AAT GAA ACA ATE CCC ACC TTC AAC CCC GGC GGT GCC TTT TCT GGG CTT CCC CCA ACG ACT CAG AGT AGC CTT GGC TTC GGC AAT GGG TAT Ass Glu The Met Pro The Phe Ass Pro Gly Gly Ala Phe See Gly Les Pro Pro The The Gla See See Les Gly Phe Gly Ass Gly Tyr ATC AAC ACC CCT CGT CCC ATG GTT GGG CCT GGC GGT ATG ATG GGA ATG CCG AAC GGC CAA GCT GGT CAG ATG ATG GGT GCG AGC AGC AGC Ile Asa Thr Pro Arg Pro Met Val Gly Pro Gly Gly Het Met Gly Met Pro Asa Gly Gla Ala Gly Gla Met Het Gly Ala Ser Ser Ser . AGT GGG CCT GGT AGC GGT CCT TCT CGT ACT GGA GCC GAG TGG GAA TGG CTC ACA ATG AGT CTC TAATGTGGAGGTCACGACACAGCGCACATGCGTTTTAT Ser Gly Pro Gly Ser Gly Pro Ser Arg Thr Gly Ala Glu Trp Glu Trp Leu Thr Het Ser Leu BHD . GCAGCAATCGGGGGCTTTGGAATTCACACAAGGCAATGGGGGCATTATGGTGTGGCAATGGCATAGACGGTCTTGATTTTTGCTTTTTACTCTGGCCTTTTTAAACGGTACGGCTGTTGAAG CTGTCAAGTTTCGAAGGACAGCTTTTCTAAACATTTCTCCCCCGCACGTCTTATTTCCTGGTACACAAAGCGTTGATAGAAGGAGGCCGGGCTGATGAGGATGGGAAGAGGGGCCCCAGG 3970 3980 AGCGGGGAAAGATGGAGATGGGAAGCTGGGAACGCGCACATGCATTTCAACAAGACACCATGCCAGCGAATTGTCGGCCGGGAAGTCTAGTCTCTCGACCATGAAAGTGCTATGAAGGAC ACCCGGCTGTTGAACGAAGAAGAAGCGACGATGACAGCAGAAGAATAAAGGGCAAAGGCAAACGCAACTTCGGGTTGGAGCATACATCTGCTTTGGGTGTTCTTTGGTGGACGAAG ATGTGGATGGCTGCTGGGCTGAATTTTCAAGTCCCGCGCGTCTACCGTTTTTGTCCATATACGAGAAAGATGATGATGATACCCGGAATTTTTCTTTTCAAGCTTGCTCTTCCTGGCTG 4500 4510 ATTOTTGTTTACGATAGGCAGATACCATCAATGAGAATAGTAAGCATGCAAATCTTACGTTGCCTCCATGTCACCACTCTTTCACA

FIG. 2-Continued.



FIG. 3. *nit-2* gene structure. The *nit-2* gene has five in-frame potential AUG initiation codons which occur in the 5' end of the mRNA at positions corresponding to amino acid residues 1, 8, 12, 51, 62, and 68 (displayed as solid triangles). The *nit-2* sequence can be translated to yield a protein of 1,036 amino acids, ending with a UAA termination codon. Two short introns of 98 and 78 bp occur near the 5' end of the gene within the protein-coding region. A single putative zinc finger element is composed of four cysteine residues at positions 743, 746, 764, and 767. Two acidic regions of the translated *nit-2* protein begin at residue 362 (A1, net charge, -6) and at residue 559 (A2, net charge, -11). Basic regions begin at residues 330 (B1) and 912 (B2) and immediately downstream of the putative zinc finger element (not shown). Four different 3' deletions are shown. Deletions 22 and 41 result in a functional *nit-2* protein as assayed by transformation, whereas deletions 19 and 76 are inactive. WT, Wild type.

The translated nit-2 protein contains 11 hydroxyl amino acids within the first 23 amino-terminal residues, which could possibly be phosphorylation sites for posttranslational control of nit-2 protein activity. An amino acid sequence which begins at residue 48, S-S-S-A-N-N-N, is repeated two more times starting at residue 87 and again at residue 105 (except that the later two repeats lack the central alanine), suggesting that duplication events have helped shape the nit-2 gene and its protein product. The hydroxyl amino acids compose 22% of the residues of the nit-2 protein, with serine being the most abundant individual amino acid (14%). Proline represents nearly 10% (99 residues) of the protein, whereas all the remaining nonpolar amino acids compose only 27% of the total residues; in contrast, 63% of the nit-2 protein is composed of polar or charged amino acids. Surprisingly, a computer homology search revealed that the nit-2 protein more closely resembles collagen than any other protein in the data base; collagen is known to be deficient in alpha-helical or beta-sheet structures. A hydropathy plot for the deduced nit-2 protein showed that it is an extremely hydrophilic protein with only a few hydrophobic regions (data not shown). Moreover, the predicted secondary structure of the protein suggests only a small amount of alphahelical and beta-sheet structures. The nit-2 protein contains 40 more basic than acidic amino acids and has a predicted pI of 9.0. The translated nit-2 protein has two acidic regions and three basic regions whose significance, if any, is unknown. Acidic regions have been shown to be responsible for transcriptional activation by the yeast GAL4 and GCN4 proteins (18, 21). A highly basic region, with a net charge of +12, occurs in the 50 amino acids immediately downstream of the putative zinc finger (see below).

The *nit-2* protein contains a single putative Cys_2/Cys_2 -type zinc finger domain which has a loop composed of 17 amino acid residues. Preliminary results indicate that a domain of the *nit-2* protein containing this potential zinc finger functions in DNA binding (Y.-H. Fu and G. A. Marzluf, unpublished data). Almost no homology is apparent when the zinc finger domain of *nit-2* is compared with the single Cys_2/Cys_2 -type zinc finger structures of five yeast regulatory

proteins, GAL4, PPR1, ARGRII, and LEU3 of Saccharomyces cerevisiae (20, 23, 35) and LAC9 of Kluyveromyces lactis (34), each of which has 13 amino acids in the internal loop, except for LEU3, which has 16 residues (Fig. 4). A region of the NIT2 protein on the carboxy side of the zinc finger motif is highly basic (net charge, +12); this basic region may represent part of a DNA-binding domain (Fig. 4), perhaps acting as a specificity region, as has been found in the yeast LAC9, GAL4, and PPR1 proteins (6, 34). Each of the single zinc finger proteins has a basic region immediately downstream of the zinc finger structure. The qa-1F gene of N. crassa encodes an activator protein for quinic acid catabolic genes and appears to contain a single zinc finger that shows significant homology to the zinc finger region of GAL4 and other yeast regulatory proteins.

GF1, a trans-acting DNA-binding protein which appears to activate gene expression specifically in mammalian cells of the erythroid lineage, possesses two zinc finger domains, each with a central loop composed of 17 amino acids (33). The two zinc finger structures of GF1 and their immediate downstream basic regions are remarkably similar to the single zinc finger and basic region of NIT2 (Fig. 4). This segment of the NIT2 protein has 32 of 50 amino acids identical to those in the second GF1 zinc finger region, i.e., 64% homology in this region, and it also displays strong homology to the first GF1 zinc finger. In fact, the zinc finger of NIT2 is more homologous to each of the GF1 fingers than they are to each other. The GF1 protein is composed of 413 amino acid residues and is particularly rich in serine (12%), threonine (8%), and proline (9.4%), very similar to the NIT2 protein, which is also rich in serine (14%), threonine (8%), and proline (9.9%). However, despite this similarity in amino acid composition, only the zinc finger regions of GF1 and NIT2 display any homology in amino acid sequence. The GF1 protein recognizes a core consensus DNA sequence, TATCT, present in promoter and enhancer elements of alpha, beta, and gamma globin and related genes of chickens, mice, and humans (33). This sequence is also present at least once in the 5' promoter region of structural genes controlled by the nit-2 gene and thus may represent all or



FIG. 4. (A) Putative single zinc finger element of the *nit-2* protein. The four cysteine residues are shown coordinating a zinc atom with a loop of 17 amino acid residues. The negatively charged region immediately downstream of the zinc finger is shown (basic amino acids are circled). (B) Comparison of *nit-2* zinc finger and downstream basic region with yeast single zinc finger proteins GAL4, LAC9, PPR1, ARGRII, and LEU3 and with the mammalian factor GF1, which has two zinc fingers (GF1a and GF1b). YCS, Yeast consensus residues. The cysteine residues which participate in zinc binding are boxed, as are identical amino acids of NIT2 and the two GF1 fingers; basic amino acids in the downstream 30 residues are circled. The region of high homology between GAL4 and the closely related LAC9 protein (12 amino acids) is underlined. In this region, replacement of amino acids in the LAC9 protein resulted in a deficiency in DNA binding (33).

part of the recognition site for the NIT2 protein. It is intriguing that a zinc finger DNA-binding motif of a protein responsible for regulation of genes of the nitrogen control circuit of the lower eucaryote *N. crassa* plays a major role in specific gene expression in the differentiated erythroid cell lineage of mammals.

The deletion analysis revealed that a *nit-2* protein lacking the carboxy-terminal 214 amino acids, approximately 21% of the protein, was still functional in activating gene expression. This seems to be a common feature of regulatory proteins, namely, that relatively large segments of control proteins such as the yeast GAL4 and GCN4 proteins can be deleted without noticeably altering their function (18, 21).

In the related filamentous fungus Aspergillus nidulans, a regulatory gene designated areA appears to be homologous to the N. crassa nit-2 gene (22). In fact, the nit-2 gene can complement an areA mutant of A. nidulans and turn on the expression of nitrate reductase, acetamidase, and related catabolic enzymes (7). This result implies that the activation function and DNA sequence specificity of the nit-2 and areA proteins must be quite similar; moreover, the upstream recognition elements that serve the nitrogen structural genes in A nidulans and N. crassa also must be remarkably alike. The *areA* gene encodes a protein which also possesses a single zinc finger domain with a central loop of 17 amino acids (4). On the other hand, A. nidulans lacks any counterpart of the N. crassa nmr gene. Thus, the nitrogen repression mechanism for these two fungi may be different, although it could also occur in a similar manner, perhaps with the areA protein playing the regulatory role carried out by a multimeric protein composed of *nit-2* and *nmr* subunits in N. crassa. A detailed comparison of the structure and function of these regulatory proteins should be informative and provide new insight concerning nitrogen control in lower eucaryotes.

ACKNOWLEDGMENTS

This research was supported by Public Health Service grant GM-23367 from the National Institutes of Health.

We acknowledge the valuable service of Richard Swenson and Jane Tolley, Ohio State Biochemical Instrument Center, who prepared the oligonucleotides used in this work. We thank James Young for assistance in computer analysis.

LITERATURE CITED

- 1. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA 69:1408–1412.
- Berk, A. J., and P. A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonucleasedigested hybrids. Cell 12:721-732.
- Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513–1523.
- 4. Caddick, M. X., P. Hooley, and H. N. Arst. 1989. Transcript analysis of the *Aspergillus nidulans* regulatory gene *areA*. J. Cell. Biochem. 13E:36.
- Chambers, J. A. A., S. M. Griffon, and G. A. Marzluf. 1983. Trans-nuclear action of the *nit-2* regulatory gene product and study of two additional nitrogen control genes in *Neurospora crassa*. Curr. Genet. 7:51–56.
- Corton, J. C., and S. Johnston. 1989. Altering the DNA binding specificity of GCN4 requires sequences adjacent to the zinc finger. J. Cell. Biochem. 13E:37.
- Davis, M. A., and M. J. Hynes. 1987. Complementation of areA⁻ regulatory mutants of Aspergillus nidulans by the heterologous regulatory gene nit-2 of Neurospora crassa. Proc. Natl.

Acad. Sci. USA 84:3753-3757.

- 8. DeBusk, R. M., and S. Ogilvie. 1984. Regulation of amino acid utilization in *Neurospora crassa*: effect of *nmr-1* and *ms5* mutations. J. Bacteriol. 160:656–661.
- 9. Dunn-Coleman, N. S., and R. H. Garrett. 1980. The role of glutamine synthetase and glutamine metabolism in nitrogen metabolite repression, a regulatory phenomenon in the lower eukaryote *Neurospora crassa*. Mol. Gen. Genet. 179:25–32.
- Dunn-Coleman, N. S., A. B. Tomsett, and R. H. Garrett. 1981. The regulation of nitrate assimilation in *Neurospora crassa*: biochemical analysis of the *nmr*-1 mutants. Mol. Gen. Genet. 182:234-239.
- Fu, Y. H., J. Y. Kneesi, and G. A. Marzluf. 1989. Isolation of nit-4, the minor nitrogen regulatory gene which mediates nitrate induction in Neurospora crassa. J. Bacteriol. 171:4067-4070.
- 12. Fu, Y. H., and G. A. Marzluf. 1987. Molecular cloning and analysis of the regulation of *nit-3*, the structural gene for nitrate reductase in *Neurospora crassa*. Proc. Natl. Acad. Sci. USA 84:8243-8247.
- Fu, Y. H., and G. A. Marzluf. 1987. Characterization of *nit-2*, the major nitrogen regulatory gene of *Neurospora crassa*. Mol. Cell. Biol. 7:1691–1696.
- 14. Fu, Y. H., and G. A. Marzluf. 1988. Metabolic control and autogenous regulation of *nit-3*, the nitrate reductase structural gene of *Neurospora crassa*. J. Bacteriol. **170**:657–661.
- Fu, Y. H., J. L. Young, and G. A. Marzluf. 1988. Molecular cloning and characterization of a negative-acting nitrogen regulatory gene of *Neurospora crassa*. Mol. Gen. Genet. 214:74–79.
- Grove, G., and G. A. Marzluf. 1981. Identification of the product of the major regulatory gene of the nitrogen control circuit of *Neurospora crassa* as a nuclear DNA-binding protein. J. Biol. Chem. 256:463–470.
- Hanson, M. A., and G. A. Marzluf. 1975. Control of the synthesis of a single enzyme by multiple regulatory circuits in *Neurospora crassa*. Proc. Natl. Acad. Sci. USA 72:1240–1244.
- Hope, I., and K. Struhl. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46:885–894.
- Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA 82:488–492.
- Laughton, A., and R. F. Gesteland. 1984. Primary structure of the Saccharomyces cerevisiae GAL4 gene. Mol. Cell. Biol. 4:260-267.
- Ma, J., and M. Ptashne. 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell 48:847– 853.
- Marzluf, G. A. 1981. Regulation of nitrogen metabolism and gene expression in fungi. Microbiol. Rev. 45:437–461.
- 23. Messenguy, F., E. Dubois, and F. Descamps. 1986. Nucleotide sequence of the ARGRII regulatory gene and amino acid sequence homologies between ARGRII, PPR1, and GAL4 regulatory proteins. Eur. J. Biochem. 157:77–81.
- Nahm, B. H., and G. A. Marzluf. 1987. Induction and de novo synthesis of uricase, a nitrogen-regulated enzyme in *Neurospora crassa*. J. Bacteriol. 169:1943–1948.
- 25. Paluh, J. L., M. J. Orbach, T. L. Legerton, and C. Yanofsky. 1988. The cross-pathway control gene of *Neurospora crassa*, *cpc*-1, encodes a protein similar to GCN4 of yeast and the DNA-binding domain of the oncogene v-jun-encoded protein. Proc. Natl. Acad. Sci. USA 85:3728-3732.
- Perrine, K. G., and G. A. Marzluf. 1986. Amber nonsense mutations in regulatory and structural genes of the nitrogen control circuit of *Neurospora crassa*. Curr. Genet. 10:677-684.
- Premakumar, R., G. J. Sorger, and D. Gooden. 1980. Physiological characterization of a *Neurospora crassa* mutant with impaired regulation of nitrate reductase. J. Bacteriol. 144: 542-551.
- Reinert, W. R., V. B. Patel, and N. H. Giles. 1981. Genetic regulation of the qa gene cluster in *Neurospora crassa*: induction of *qa* messenger ribonucleic acid and dependency on *qa*-1 function. Mol. Cell. Biol. 1:829-835.
- 29. Rigby, P. W., M. Dieckmann, C. Rhodes, and P. Berg. 1977.

Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113:**237–251.

- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- 31. Stewart, V., and S. J. Vollmer. 1986. Molecular cloning of *nit-2*, a regulatory gene required for nitrogen metabolite repression in *Neurospora crassa*. Gene **46**:291-295.
- 32. Thireos, G., M. D. Penn, and H. Greer. 1984. 5' Untranslated sequences are required for the translational control of a yeast regulatory gene. Proc. Natl. Acad. Sci. USA 81:5096-5100.
- 33. Tsai, S. F., D. I. Martin, L. I. Zon, A. D. D'Andrea, G. G. Wong, and S. H. Orkin. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature (London) 339:446–450.
- 34. Witte, M. W., and R. C. Dickson. 1988. Cysteine residues in the zinc finger and amino acids adjacent to the finger are necessary for DNA binding by the LAC9 regulatory protein of *Kluyvero*myces lactis. Mol. Cell. Biol. 8:3726–3733.
- 35. Zhou, K., P. R. G. Brisco, A. E. Hinkkanen, and G. B. Kohlhaw. 1987. Structure of the yeast regulatory gene LEU3 and evidence that LEU3 itself is under general amino acid control. Nucleic Acids Res. 15:5261–5273.