Mol Gen Genet (1993) 240:365–373

MGG

A tobacco cDNA clone encoding a GATA-1 zinc finger protein homologous to regulators of nitrogen metabolism in fungi

Françoise Daniel-Vedele, Michel Caboche

Laboratoire de Biologie Cellulaire, INRA, route de Saint-Cyr, F-78026 Versailles Cedex, France

Received: 25 January 1993 / Accepted: 17 March 1993

Abstract. In higher plants, the expression of the nitrate assimilation pathway is highly regulated. Although the molecular mechanisms involved in this regulation are currently being elucidated, very little is known about the *trans*-acting factors that allow expression of the nitrate and nitrite reductase genes which code for the first enzymes in the pathway. In the fungus *Neurospora crassa*, *nit-2*, the major nitrogen regulatory gene, activates the expression of unlinked structural genes that specify nitrogen-catabolic enzymes during conditions of nitrogen limitation. The *nit-2* gene encodes a regulatory protein containing a single zinc finger motif defined by the $C\text{-}X_2\text{-}C\text{-}X_{17}\text{-}C\text{-}X_2\text{-}C$ sequence. This DNA-binding domain recognizes the promoter region of *N. crassa* nitrogen-related genes and fragments derived from the tomato *nia* gene promoter. The observed specificity of the binding suggests the existence of a NIT2-like homolog in higher plants. PCR and cross-hybridization techniques were used to isolate, respectively, a partial cDNA from *Nicotiana plumbaginifolia* and a full-length cDNA from *Nicotiana tabacum*. These clones encode a NIT2-like protein (named NTL1 for *nit-2*-like), characterized by a single zinc finger domain, defined by the $C\text{-}X_2\text{-}C\text{-}X_{18}\text{-}C\text{-}X_2\text{-}C$ amino acids, and associated with a basic region. The amino acid sequence of NTL1 is 60% homologous to the NIT2 sequence in the zinc finger domain. The *Ntl1* gene is present as a unique copy in the diploid *N. plumbaginifolia* species. The characteristics of *Ntl1* gene expression are compatible with those of a regulator of the nitrate assimilation pathway, namely weak nitrate inducibility and regulation by light.

Key words: *Nicotiana plumbaginifolia* – *Nicotiana tabacum* – GATA-binding factor – Nitrate reductase – PCR

Communicated by H. Böhme

Correspondence to: F. Daniel-Vedele

Introduction

In higher plants, nitrate is first converted into nitrite by nitrate reductase (NR, EC 1.6.6.1), a cytoplasmic enzyme. Nitrite is then translocated to the chloroplast, where it is reduced by nitrite reductase (NiR, EC 1.7.7.1) to ammonium. Ammonium is then used for the biosynthesis of amino acids in the chloroplast (Guerrero et al. 1981). Nitrate assimilation is a highly regulated process. In higher plants, factors such as light, cytokinins and circadian rhythm affect NR expression (Crawford and Campbell 1990; Caboche and Rouzé 1990). In addition, nitrate induces, and ammonium and/or glutamine down-regulate, NR expression (Caboche and Rouzé 1990; Solomonson and Barber 1990). These controls appear to occur at the mRNA level (Vincentz and Caboche 1991; Vaucheret et al. 1992). Very little is known about the *trans*-acting factors that are required for the regulation of NR in higher plants. Many mutants defective in nitrate assimilation have been isolated in different species (reviewed in Kleinhofs et al. 1985; Wray 1988; Wray and Kinghorn 1989; Caboche and Rouzé 1990; Crawford and Campbell 1990). Among them, only one putative regulatory mutant, *chl2*, has been isolated, from *Arabidopsis thaliana*. Unfortunately, recent studies have demonstrated that the *chl2* mutant is not a regulatory mutant but is impaired in the biosynthesis of molybdopterin, a NR cofactor, or in stabilization of this cofactor (LaBrie et al. 1992). There are significant similarities between the NR regulation found in plants and fungi. In *Neurospora crassa* and *Aspergillus nidulans*, NR expression is highly regulated at the mRNA level and requires nitrogen derepression and simultaneous induction by nitrate. The regulation of the NR and NiR genes is governed on the one hand by the pathway-specific positive control genes, *nit-4* and *nirA*, involved in nitrate inducibility. On the other hand, *nit-2* and *areA*, the major positive control genes, are involved in N-metabolite repression. The nucleotide sequences of the *nit-2* and *areA* genes revealed that they encode transcription factors that contain a DNA-binding element consisting of a single

Cys2/Cys2 zinc finger and an adjacent basic region (Fu and Marzluf 1990a; Kudla et al. 1990). There is 98% amino acid identity between the DNA-binding domains of the NIT2 and AREA proteins. In vitro gel-band mobility shift and DNA footprinting studies showed that the NIT2 protein recognizes specifically sequences in the 5′ region of N-repressible genes of *N. crassa* and of the nitrate and nitrite reductase genes of *A. nidulans* (Fu and Marzluf 1990b). A core consensus sequence, TATCT (or on the complementary strand AGATA), was identified in all of these binding sites. The same consensus sequence, (A/T)GATA(A/G), was identified in the promoters and enhancers of all globin genes as the binding sequence for the major transcription factor, GATA-1, of the erythroid lineage (Evans and Felsenfeld 1989). These GATA-1 proteins have two C_4 zinc finger DNA-binding domains that are highly conserved among vertebrates including man (Trainor et al. 1990), mouse (Tsai et al. 1989) and chicken (Evans and Felsenfeld 1989), and invertebrates like *Caenorhabditis elegans* (Spieth et al. 1991) and *Spodoptora frugiperda* (Krappa et al. 1992). The $C\text{-}X_2\text{-}C\text{-}X_{17}\text{-}C\text{-}X_2\text{-}C$ zinc finger motif is also conserved in pleiotropic regulators affecting nitrogen pathway genes in yeast, such as GLN3, a positive regulator (Minehart and Magasanik 1991) or UGA43, a negative regulator (Coornaert et al. 1992). The DNA-binding domains of the NIT2 and AREA proteins are homologous to these GATA-1 and yeast zinc finger motifs.

Recent studies have demonstrated that two fragments of the tomato *nia* gene promoter region, which was shown to contain essential *cis*-acting regulatory elements (Dorbe et al. 1992), are specifically recognized and bound by the NIT2 protein (Jarai et al. 1992). Both fragments contain the putative GATA recognition sequence. These data suggest that a homolog of NIT2, belonging to the GATA-binding protein family, may exist in tomato and other higher plants and play a central role in controlling NR expression. No cross-hybridization was observed between the *nit-2* DNA-binding domain sequence and tobacco or *Nicotiana plumbaginifolia* genomic DNAs in Southern blotting experiments. The PCR method (polymerase chain reaction) has been successfully used to clone plant transcription factors by homology (reviewed in Katagiri and Chua 1992). Two degenerate oligonucleotides were derived from the $C\text{-}X_2\text{-}C$ amino acids of the NIT2 zinc finger and used in PCR experiments on *N. plumbaginifolia* cDNA. We show in this paper that a gene homologous to *nit-2* does exist in *N. plumbaginifolia* and in tobacco. We here describe the isolation and nucleotide sequence of a complete cDNA clone from tobacco, which encodes a zinc finger protein belonging to the GATA family. Its potential role in NR regulation will be discussed.

Materials and methods

Oligonucleotide synthesis. The oligonucleotides were synthesized on a 381A DNA synthesizer (Applied Biosystems). Four specific primers were used for cDNA synthesis and PCR amplification. ol1 and ol2 were derived from the amino acid sequence of the NIT2 zinc finger region. N corresponds to any nucleotide generated by adding an equal mixture of all four bases at each N position. At the two fold degenerated positions, equal amounts of the two nucleotides were included in the synthesis reaction: ol1 (22mer), 5′-GCGAATTCACNTG(C/T)ACNAA(T/C)TG-3′; ol2 (23mer), 5′-GCGGATCCNCC(G/A)CANGC(A/G)TT(G/A)CA-3′ (NIT2 sequences are underlined). RACE (Rapid amplification of cDNA ends) PCR experiments were performed using ol3, an oligo-d(T) adapted primer, and ol4 designed from the first partial sequence of *Ntl1-Npl1:* ol3 (32mer), 5′-CCTAGGCCACTGTGGCCT$_{(15)}$-3′; ol4 (20mer), 5′-ACACCTCAGTGGAGAGAGGG-3′ (underlined nucleotides in Fig. 1A). After deprotection, the oligonucleotides were ethanol precipitated and resuspended in water.

cDNA synthesis and PCR amplification. Total leaf RNA from the E56 *nia* mutant was used for cDNA synthesis. For first cDNA strand synthesis, 5 µg of RNA was mixed with 200 pmol of oligo d(T) or 10 pmol of specific primer (ol3) and 10 units of ribonuclease inhibitor (RNAsin, Pharmacia LKB Biotechnology) in a total volume of 12.4 µl, heated at 65° C for 5 min and cooled on ice. To this mixture, 10 units of RNAsin, 1 µg of nuclease-free bovine serum albumin (Promega), dNTPs (1 mM final concentration for each dNTP), and 200 units of M-MLV (Moloney mouse leukemia virus) reverse transcriptase (Bethesda Research Laboratories) were added in a total volume of 20 µl of the following buffer: 50 mM TRIS-HCl, pH 8.3, 75 mM KCl, 10 mM DTT, 3 mM $MgCl_2$. Control reactions for each sample contained the same components, except that the reverse transcriptase was omitted from the mixture. After 1 h incubation at 37° C, 80 µl of water was added and 5 µl of the diluted first-strand reaction mixtures were used for PCR amplification of the cDNA.

PCR amplification with the specific primers ol1 and ol2 was performed with *Taq* DNA polymerase (Promega) under standard conditions (Saiki et al. 1988) and as recommended by the supplier of the enzyme, with 30 pmol of each primer. Cycling conditions for PCR were the following: 3 min at 94° C, 3 cycles of DNA denaturation (1 min at 94° C), annealing (30 s at 37° C) and extension (2 min at 72° C), followed by 30 cycles of DNA denaturation (1 min at 94° C), annealing (30 s at 45° C) and extension (2 min at 72° C) and then 15 min at 72° C. Cycling conditions for the RACE PCR were the following: 3 min at 94° C, 1 cycle of DNA denaturation (1 min at 94° C), annealing (1 min at 55° C) and extension (15 min at 72° C) followed by 35 cycles of DNA denaturation (1 min at 94° C), annealing (1 min at 55° C) and extension (2 min at 72° C) and then 15 min at 72° C. One-tenth (10 µl) of the PCR reaction was then loaded on a 1% agarose gel and stained with ethidium bromide for analysis. PCR reactions were subsequently loaded on a 1% preparative agarose gel, the DNA bands were cut out of the gel and the amplified material was further purified using the GeneClean kit (Bio-101). DNA frag-

ments were blunt-ended by Klenow treatment and inserted into the *Eco*RV site of a pBluescript vector (Stratagene). The nucleotide sequence of the inserts was determined with an automated Applied Biosystems 373A DNA sequencer using dye primers and the Sequenase enzyme.

DNA and RNA extraction, Northern and Southern blot analysis. Total RNA extraction, Northern and Southern analysis were performed as previously described (Dorbe et al. 1992). Genomic DNAs from tobacco or *N. plumbaginifolia* leaves were extracted and purified according to the method of Dellaporta et al. (1983). The probe used in Northern and Southern experiments corresponds to the entire insert of *Ntl1-Np1B*. The *Eco*RV-*Hin*dIII fragment was isolated from an agarose gel, extracted and purified by the GeneClean method (Bio-101, La Jolla, Calif.). The probe was α-^{32}P-labelled by random priming (Feinberg and Vogelstein 1983). Washes of Northern and Southern blots were performed under low stringency conditions in 2 × SSC, 0.5% Sarkosyl at 65° C.

Plasmids, phages and bacterial strains. A tobacco λgt10 cDNA library (kindly provided by Dr. Klaus Palme, MPI, Köln) was amplified in the *Escherichia coli* strain RY 1083.

Phage plaque hybridization, using the *Ntl1-Np1B* cDNA as probe, was performed according to standard procedures (Maniatis et al. 1982). Washes were done under low stringency conditions in 2 × SSC, 0.5% sarkosyl at 65° C. DNA inserts of positive recombinant phages were subsequently cloned in the *Eco*RI site of a pBluescript vector (Stratagene).

Results

Isolation of a partial cDNA from N. plumbaginifolia

nia mutants expressing a non-functional NR protein accumulate NR and NiR mRNAs (Faure et al. 1991). This overexpression is a consequence of the absence of accumulation of N-metabolites such as glutamine, which repress the pathway (Vincentz et al. 1993). These physiological conditions were therefore chosen as the starting point of a search for transcripts homologous to *areA* and *nit-2,* putatively involved in N-metabolite regulation. Total RNA was extracted from the *N. plumbaginifolia* E56 *nia* mutant and used as a template in PCR experiments. The first cDNA strand was synthesized using oligo d(T) as primer (see Materials and methods). Two degenerate oligonucleotides (ol1 and ol2) corresponding, respectively, to the residues NH_2-T_{742}-C_{743}-T_{744}-N_{745}-C_{746}-COOH and NH_2-C_{764}-N_{765}-A_{766}-C_{767}-G_{768}-COOH of the NIT2 protein (Fu and Marzluf 1990a) were subsequently used during 33 cycles of amplification (see Materials and methods) on this first cDNA strand or on the pNit-2 plasmid, which contains the *nit-2* gene, as a positive control. A faint band of 100 bp, corresponding in size to the amplification product of the *nit-2* gene, was observed for the *N. plumbaginifolia* sample and not in the corresponding control (cDNA synthesis minus reverse transcriptase, data not shown). This DNA fragment was purified by GeneClean and used in a second PCR amplification, performed under the same conditions, in order to obtain enough material. After gel extraction and purification, the cDNA fragment was cloned into a pBluescript vector. The nucleotide sequence and the deduced amino acid sequence of this *Ntl1-Np1A* (*Ntl* for *nit-2*-like) partial cDNA is shown in Fig. 1A. A main open reading frame, covering the entire length of the fragment was found and comparison of the corresponding amino acid sequence with the sequence of the NIT-2 zinc finger loop (shown in bold face in Fig. 1A) revealed conservation of 7 out of 17 amino acids.

As a second step, we performed a 3′ RACE PCR experiment. A first cDNA strand was synthesized using the oligonucleotide ol3 (see Materials and methods) as primer and total RNA extracted from the E56 mutant as a template. This first strand was subsequently amplified between the same ol3 and ol4 oligonucleotides (corresponding to the underlined nucleotides in Fig. 1B). A 550 bp cDNA fragment was obtained and cloned as *Ntl1-Np1B*. The nucleotide and deduced amino acid sequences are shown in Fig. 1B. Only one large open reading frame of 100 amino acids is found in the entire sequence. The homology with the NIT2 zinc finger region extends to the succession C-N-A-C-G, which may represent the end of a hypothetical zinc finger in the *N. plumbaginifolia* protein. This homology is not maintained beyond the G residue. However, the region of the translated Ntl1-Np1B protein on the carboxy side of the zinc finger motif is highly basic (underlined residues in Fig. 1B) with a net charge of +9. This region has been shown to be part of the DNA-binding domain in the mouse GATA-1 protein (Martin and Orkin 1990) and the NIT2 protein (Fu and Marzluf 1990c). We used this partial cDNA as a probe for analyzing the expression pattern and genomic organization of the *Ntl1* genes and to isolate a full-length cDNA.

mRNA size and expression pattern

To determine the size and the expression pattern of the *Ntl1-Np1* cDNA, we performed Northern blot hybridizations on total RNA of *N. plumbaginifolia* plants grown under different conditions. A single 1.5 kb band was detected (Fig. 2). Expression levels were approximately fourfold lower than for NR.

Compared to the NR mRNA expression pattern, the regulation of the *Ntl1* gene expression is slightly different. It seems that no circadian rhythm influences the *Ntl1* mRNA pool: the steady-state level of mRNA is the same at the beginning and end of the day (Fig. 2, lanes 1 and 2). However, the amount of *Ntl1-Np1* transcript is lower in plants grown on ammonium than in plants grown on nitrate (Fig. 2, lanes 3 and 4), although the range of this variation is much lower than the degree of stimulation of NR mRNA synthesis by nitrate. The two- to threefold increase in the content of the *Ntl1* transcript under nitrogen-derepressed versus nitrogen-repressed conditions

A

```
(ol1) GCGAATTCACTTGTACAAATTGTCAAGTCACAAAGACACCTCAGTGGAGAGAGGGACCA
                                  Q  V  T  K  T  P  Q  W  R  E  G  P
                                  F  T  Q  T  T  P  L  W  R  R  N  P

      CTGGGGCCAAAGACACTATGTAACGCTTGTGGAGGATCC (ol2)

        L  G  P  K  T  L
        D  G  -  Q  P  L  C  N  A  C  G  (nit-2)
```

B

```
(ol4) CAGTGGAGAGAGGGACCACTGGGGCCAAAGACACTATGCAATGCCTGTGGAGTCCGCTAT

        Q  W  R  E  G  P  L  G  P  K  T  L  C  N  A  C  G  V  R  Y

      CGTTCAGGGCGTCTATTCCCAGAGTATCGACCTGCTGCTAGTCCTACATTTATTCCAACT

        R  S  G  R  L  F  P  E  Y  R  P  A  A  S  P  T  F  I  P  T

      TTGCACTCAAACTCTCACAGGAAGGTCGTAGAGATGAGAAAAAAAGCTATATATGGAGAA

        L  H  S  N  S  H  R  K  V  V  E  M  R  K  K  A  I  Y  G  E

      ACATCTGCACTTGAGGAACCTCACAATGTGATCGTAGAGGGCCCTCCAATGTCACCCGCG

        T  S  A  L  E  E  P  H  N  V  I  V  E  G  P  P  M  S  P  A

      CCAGAGTTTGTTCCAATGAGTAGCTATTTGTTTGATGTATACTGAAGAGATGATTATACT

        P  E  F  V  P  M  S  S  Y  L  F  D  V  Y  *

      TTTAGTTAGGAGCTTAATTTTAGTTTATTGTTTTATTTTCGATGGAGTCATGGGTGGTTC
      TGCTGGTATGGTGGCTACTCGCGTTATGCGTCGATGTACAAATGTTGAATCTTAGGTGTC
      TAGGTTAGATAATTGGTTGGAGGTCAGAAGGAAAAGCTTTGAATCAATGCTTGATGAAGC
      AGTTCCTGTTATGTAAAAACTGAACTAGATATAGTAGATGGTTAGGTTTCATTTTTAAAA
      AAAAAAAAAAAGGCCACAGTGGCCTAGG (ol3)
```

Fig. 1A, B. Nucleotide sequences of the *Nicotiana plumbaginifolia nit-2*-like cDNA. The nucleotide and deduced amino acid sequences of *Ntl1-Np1A* and *Ntl1-Np1B* are shown in **A** and **B**, respectively. Oligonucleotides ol1, ol2, ol3 and ol4 are *underlined*. The nucleotides written in *italics* in **B** correspond to the hybridization site of ol2 in the first PCR experiment. Amino acid sequence of the NIT2 zinc finger loop is in *bold face*. Basic residues are *underlined* in the amino acid sequence of *Ntl1-Np1*

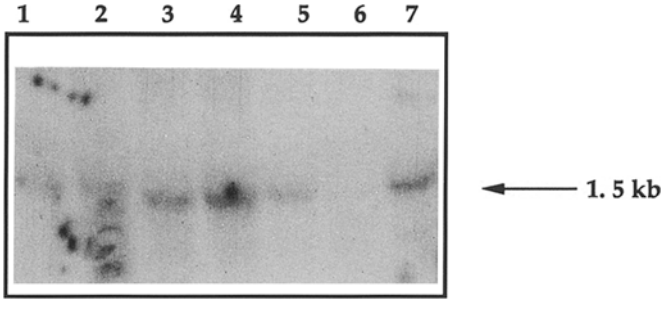

Fig. 2. RNA gel blot analysis of *N. plumbaginifolia Ntl1* mRNA. Total RNA (16 µg) from leaf tissues was probed with the radiolabelled *Ntl1-Np1B* cDNA insert. Lanes 1 and 2: plants were grown under a 16 h/8 h light-dark regime, leaves were harvested in the early morning (1) or at the end of the afternoon (2). Lanes 3 and 4: plants were grown in vitro on medium supplemented with ammonium (3) or nitrate (4), leaves were harvested at the begining of the day. Lanes 5, 6 and 7: plants were grown under a 16 h/8 h light-dark regime, kept in the dark for 72 h and then returned to the normal cycle. Leaves were harvested at the begining of a day period as a control (5), after 72 h of darkness (6) and 4 h after illumination (7). Equal loading of RNA was confirmed by ethidium bromide staining of the gel before the transfer. Size was estimated according to a size marker run in parallel

is comparable to the fluctuations in *nit-2* transcript levels under the same conditions (Fu and Marzluf 1987). On the other hand, the *Ntl1-Np1* transcript is light inducible. In dark-adapted plants, the *Ntl1* transcript is barely detectable (Fig. 2, lane 6) and 4 h illumination is sufficient to promote its synthesis (Fig. 2, lane 7), as is also the case for *nia* transcripts and other light-inducible transcripts. The characteristics of expression of the *Ntl1-Np1* mRNA are therefore compatible with those of a regulator of the nitrate assimilation pathway.

Genomic organization of N. plumbaginifolia nit-2-*like genes*

In higher plants, the genes isolated so far that encode proteins which contain zinc finger DNA-binding domains are present in one or two copies per genome (Takatsuji et al. 1991) or in multigene families (Baltz et al. 1992). We have examined the organization of the *nit-2*-related sequences in the *N. plumbaginifolia* genome. Results are shown in Fig. 3. DNA gel blot analysis of *N. plumbaginifolia* genomic DNA revealed the presence of only one *Eco*RI restriction fragment and two *Hin*dIII restriction fragments (Fig. 3, lanes 1 and 2) that show homology with the cDNA probe. These data suggest that only one copy of this *nit-2*-related gene exists in the *N. plumbaginifolia* genome.

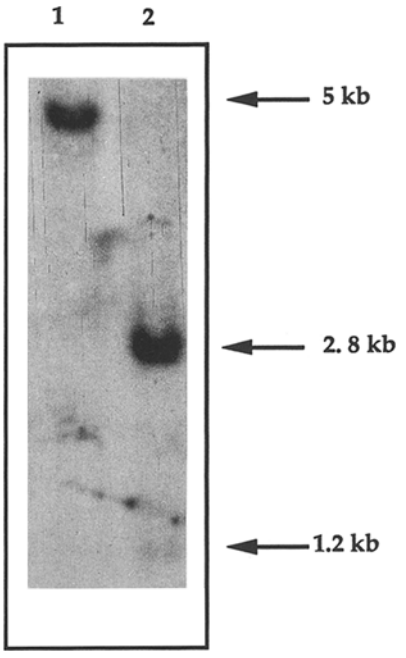

Fig. 3. DNA gel blot analysis of *N. plumbaginifolia* genomic DNA. *N. plumbaginifolia* genomic DNA (5 µg) was digested with *Eco*RI (lane 1) or *Hin*dIII (lane 2) and probed with the radiolabelled *Ntl1-Np1B* cDNA insert

Isolation of a tobacco cDNA encoding a NIT2-like protein

As attempts to isolate a full-length *N. plumbaginifolia Ntl1* cDNA were unsuccessful, we therefore screened for a full-length homolog of the partial cDNA in tobacco. Southern experiments revealed the presence in the tobacco genome of a gene homologous to the partial cDNA isolated from *N. plumbaginifolia* (data not shown). Using *N. plumbaginifolia* cDNA as probe, we screened 4×10^5 phages of a cDNA library made from young tobacco leaves, constructed in λgt10. From this screen we obtained two recombinant phages that showed either strong or faint hybridization signals (respectively, *Ntl1-Nt7* and *Ntl1-Nt2*). The nucleotide and deduced amino acid sequences of the *Ntl1-Nt7* insert (1.4 kb long) are shown in Fig. 4. The expected size of the corresponding *Ntl1-Nt7* mRNA is slightly less than the 1.5 kb mRNA detected in Northern blot experiments. This small difference can be explained by the lack of the entire poly(A) tail and of the entire leader sequence in the tobacco cDNA clone.

Only one open reading frame was found. The first AUG codon is closely followed by two additional in-frame AUG codons. It is not known which of these three in-frame AUG codons actually represents the initiation codon; however, the sequence surrounding the third AUG codon has the best match to the plant consensus translational start sequence, AACAAUGG (Lutcke et al. 1987). Translating from the initial AUG codon to the UGA stop codon yields a protein of 305 amino acids, with a molecular weight of approximately 33 000 Da. The polyadenylation of the *Ntl1-Nt7* transcript occurs 302 bp beyond the UGA stop codon, 20 bp downstream of a putative polyadenylation signal (underlined in Fig. 4). The *Ntl1-Nt7* nucleotide sequence is 98% identical to the partial cDNA clone from *N. plumbaginifolia*.

Analysis of the predicted protein sequence reveals the presence of several regions. First, a single putative zinc finger DNA-binding domain, covering 26 residues (amino acids 201–227) is located in the C-terminus of the protein. This domain is composed of two pairs of cysteine residues (circled in Fig. 4), separated by a loop of 18 amino acids (instead of 17 amino acids as in the NIT2 protein). Two highly basic regions occur immediately upstream and downstream of this putative zinc finger: amino acids 132 to 201 have a net charge of +17 and amino acids 230 to 266 have a net charge of +9 (basic amino acids are underlined in Fig. 4), respectively. This basic region, on the carboxy side of the zinc finger, is also found in the NIT2 protein and was shown to represent part of the DNA-binding domain (Fu and Marzluf 1990c). The other basic region, upstream of the zinc finger, contains the amino acid sequence KEKKRK (169–174). This sequence is very similar to the well-characterized SV40 nuclear localization signal, KKKRK (Kalderon et al. 1984). The presence of a putative nuclear localization signal in the NTL1-NT7 protein suggests the import of the protein from the cytoplasm into the nucleus. An acidic region (amino acids 12 to 48, with a net charge of −13) is localized in the N-terminus of the protein (acidic amino acids are boxed in Fig. 4). Such acidic regions have been shown to be responsible for transcriptional activation by the yeast GAL4 and GCN4 proteins (Hope and Struhl 1986; Ma and Ptashne 1987).

Computer analysis of the NTL1-NT7 amino acid sequence

A comparison of the predicted primary amino acid sequence of the large open reading frame of the *Ntl1-Nt7* cDNA with the GenEMBL database (GCG; TFASTA search) revealed that the zinc finger region is related to previously reported protein sequences. Best scores were obtained with the following proteins: AREA (Kudla et al. 1990), human GATA-1 (Trainor et al. 1990), mouse GATA-1 (Tsai et al. 1989), NIT2 (Fu and Marzluf 1990a), yeast GLN3 (Minehart and Magasanik 1991), and chicken GATA-1 (Evans and Felsenfeld 1989).

We have carried out a multiple sequence alignment (GCG; PILEUP program) between the amino acid sequences of the NIT2, AREA, yeast GLN3, chicken GATA-1 and the NTL1-NT7 proteins. Except for the zinc finger region (see below), only two amino acids, S_{141} and K_{259}, are conserved in all five proteins. The alignment which is obtained in the zinc finger region is shown in Fig. 5. In addition to the four cysteine residues, which are possibly involved in the coordination of a zinc atom and conserved in the five proteins, eight identical amino acids are found at equivalent positions in these proteins. In the consensus sequence derived from this alignment, these conserved amino acids are written in capital letters (Fig. 5).

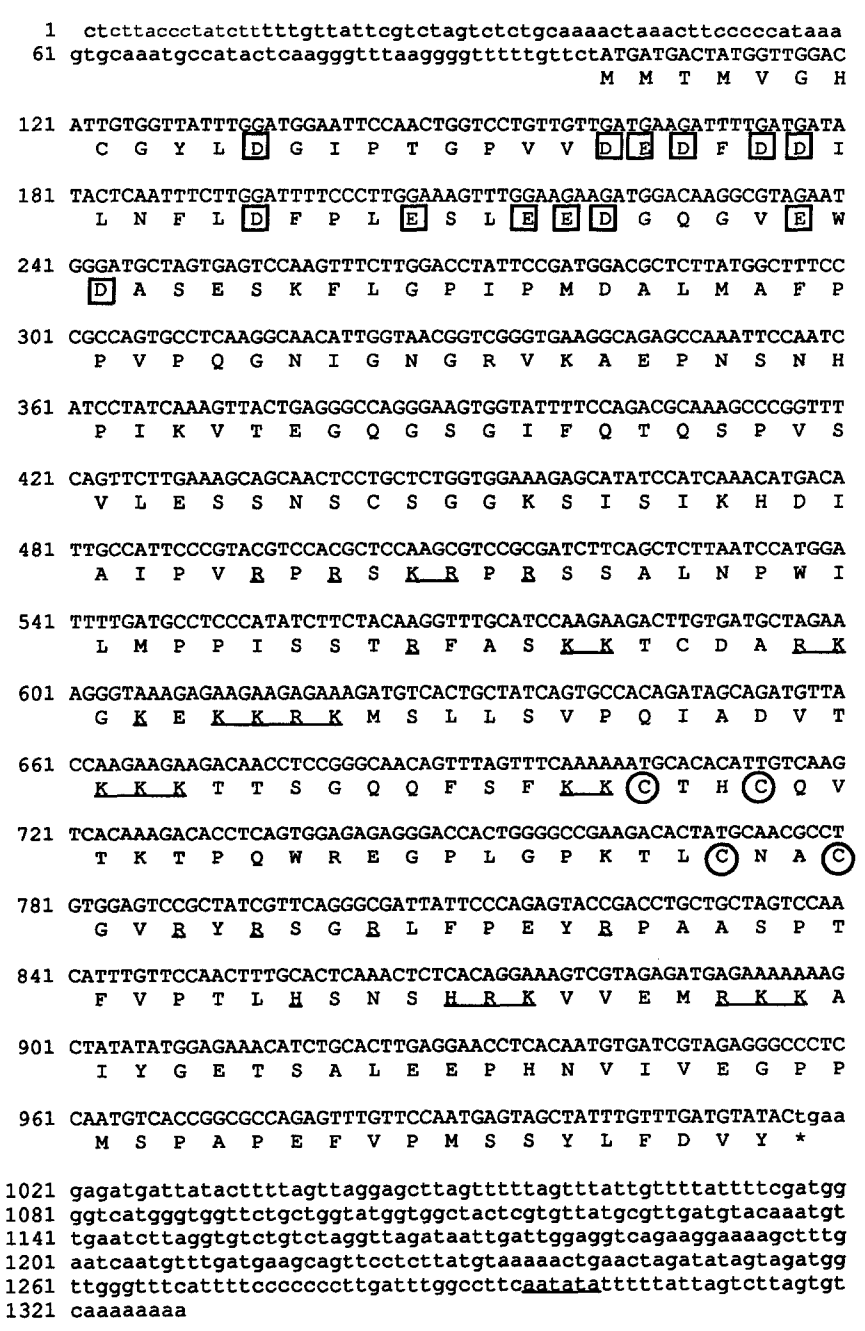

Fig. 4. Nucleotide and predicted amino acid sequences of the *Ntl1-Nt7* cDNA. The nucleotide sequence is numbered from the first nucleotide of the *Eco*RI linker added following cDNA synthesis. A putative polyadenylation site is *underlined*. The four cysteine residues of the zinc finger are *circled*. Acidic and basic residues are respectively, *boxed* and *underlined*

Discussion

The assimilation of nitrate is a highly regulated process both in fungi and higher plants. The first step in the assimilatory pathway, catalyzed by nitrate reductase, is the reduction of nitrate of nitrite. While the regulatory factors of fungal nitrate reductase expression have been characterized in some detail, very little is known about the *trans*-acting regulatory elements of higher plants, to a certain extent because of the lack of regulatory mutants. The NIT2 regulatory protein from *N. crassa* has been shown to bind to three different sites in the *nit-3* gene, which encodes the NR apoenzyme of this fungus, and also to recognize specific DNA fragments of the tomato *nia* gene promoter region. Here we demonstrate, by a PCR approach, that a *nit-2*-like gene exists in tobacco and *N. plumbaginifolia* and we will discuss its possible involvement in the regulation of NR expression. The complete tobacco cDNA that we have isolated is found to encode a protein which presents all the characteristics of a transcription factor. A $C\text{-}X_2\text{-}C\text{-}X_{18}\text{-}C\text{-}X_2\text{-}C$ zinc finger motif is located in the carboxy part of the protein. A highly basic region occurs in the 39 amino acid region immediately downstream of this putative zinc finger. This association of a zinc finger with a basic region has been found in a variety of transcription factors and serves as a DNA-binding domain (Corton and Johnston 1989; Fu and Marzluf 1990c). To exert their function,

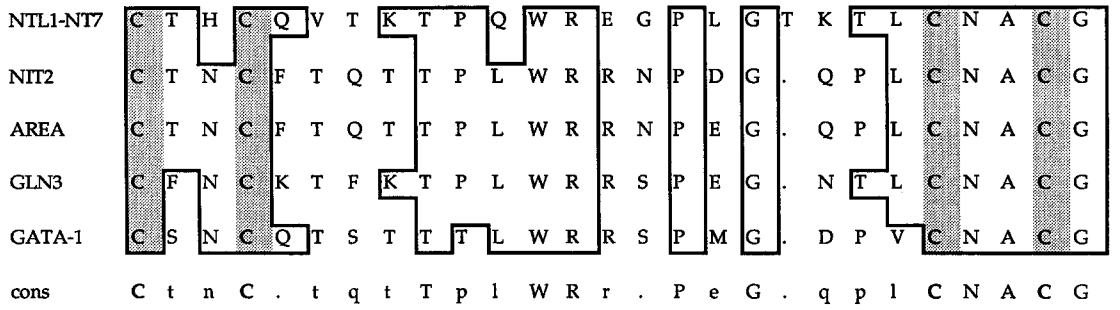

Fig. 5. Multiple sequence alignment of zinc finger domains. The predicted amino acid sequence of the NTL1-NT7 zinc finger region has been compared, using the PILEUP program of GCG, with the corresponding domain of NIT2 and AREA, the major nitrogen regulatory proteins of *Neurospora crassa* (Fu and Marzluf 1990a) and *Aspergillus nidulans* (Kudla et al. 1990), respectively, GLN3, the positive nitrogen regulatory protein of *Saccharomyces cerevisiae* (Minehart and Magazanik 1991), and GATA-1, the chicken erythroid transcription factor (Evans and Felsenfeld 1989). The four cysteine residues are *shaded* and amino acids conserved between NTL1-NT7 and at least one of the other sequences are *boxed*. The consensus sequence (cons) shows in *capital letters* the amino acids which are conserved in all five proteins

transcription factors have to be transported into the nucleus. The comparison of numerous nuclear localization signals shows no clear consensus amino acid sequence (reviewed in Silver 1991) and very little is known about nuclear transport in plants. However, a sequence homologous to the nuclear localization signal of the SV40 large T-antigen is found in the basic region upstream of the zinc finger domain. The function of the SV40 nuclear localization signal has been extensively characterized in the context of homologous proteins (Kalderon et al. 1984), heterologous proteins (Dingwall et al. 1988) or synthetic peptides (Breeuwer and Goldfarb 1990). Recently, this sequence has been shown to function as an efficient importer of β-glucuronidase into plant nuclei (van der Krol and Chua 1991). This suggests a possible nuclear localization of the NTL1-NT7 protein. Finally, an acidic region, located in the amino-terminal part of the protein, may be responsible for transcriptional activation.

A computer homology search has revealed that the NTL1-NT7 protein belongs to the large GATA-1 family, comprising the NIT2 and AREA nitrogen regulatory elements, the yeast GLN3 and UGA43 proteins and the GATA-1 erythroid transcription factors from chicken, mouse and man. These GATA-1 proteins appear to be a subclass of a large multigene family that includes several transcriptional activators such as the GATA-2 proteins of human endothelial cells (Wilson et al. 1990) and the GATA-3 proteins of the T-lymphocyte cell lineage (Ko et al. 1991). Multiple sequence alignment (Fig. 5) shows that the C-X_2-C-$X_{17\text{ or }18}$-C-X_2-C zinc finger motif is very similar in the selected proteins. This indicates a broad evolutionary conservation of this functional unit, which is used in individual organisms at the DNA-binding domain of various specialized transcription factors. More than 40 cDNA clones encoding putative transcription factors have been isolated from plants (reviewed in Katagiri and Chua 1992) but only four of them contain zinc-binding motifs: 3AF1 has a C-X_2-C-X_{10}-H-X_3-H motif (Lam et al. 1990), EPF1 a C-X_2-C-X_{12}-H-X_3-H motif (Takatsuji et al. 1991), SF3 has three possible configurations of the putative zinc finger (Baltz et al. 1992) and COP1 contains a zinc-binding motif more closely related to the glucocorticoid receptor and GAL4 types (Deng et al. 1992). Our NTL1 protein is the first C-X_2-C-X_{18}-C-X_2-C zinc finger protein to be characterized in higher plants.

Whether or not the NTL1 protein is involved in nitrate reductase expression remains to be demonstrated. Some arguments can be advanced in favor of a possible role of NTL1 in the nitrate assimilation pathway. Despite the low temperature of the first three PCR cycles and the use of degenerate oligonucleotides, we have isolated only one class of cDNA from total RNAs. If there is a gene homologous to *nit-2* in higher plants, then *Ntl1* may be the one. In addition, Northern and Southern hybridization experiments, performed under low stringency conditions, revealed only one copy per genome of the *Ntl1* gene in tobacco and *N. plumbaginifolia* genomes and only one homologous transcript in total *N. plumbaginifolia* RNAs. Furthermore, the expression pattern of the *Ntl1* gene is similar to that of *nit-2,* particularly under repressed and derepressed conditions. The size of the *Ntl1* transcript (1.5 kb) is smaller than the *nit-2* mRNA (3.5 kb, Fu and Marzluf 1987). However, deletion analysis has revealed that the NIT2 protein lacking the carboxy-terminal 214 amino acids, approximately 21% of the protein, was still functional in activating gene expression (Fu and Marzluf 1990a).

These results suggest that the NTL1-NT7 protein could be one of the regulatory elements of the nitrate assimilation pathway. In order to verify this hypothesis, we will study, by gel retardation assays, the interactions between the NTL1-NT7 zinc finger domain with the tomato *nia* gene promoter and compare the results with those of the NIT2/*nia* promoter interaction (Jarai et al. 1992). An antisense strategy will be developed to study effects of inhibiting expression of the *Ntl1* gene in *N. plumbaginifolia* and the resulting transgenic plants will be characterized with respect to the expression of the nitrate assimilatory pathway. This antisense approach should be fairly straightforward to perform due to the presence of a single copy gene, expressed at low level.

Acknowledgements. The authors wish to thank Dr. J. Paz-Ares for helpful discussion and advice and Dr. C. Meyer for his fruitful assistance in computer research and analysis. We also thank I. Small for reading the manuscript. This research was supported in part by BRIDGE program PL 890090 from the EEC.

References

Baltz R, Domon C, Pillay DTN, Steinmetz A (1992) Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein. Plant J 2:713–721

Breeuwer M, Goldfarb DS (1990) Facilitated nuclear transport of histone H1 and other small nucleophilic proteins. Cell 60:999–1008

Caboche C, Rouzé P (1990) Nitrate reductase: a target for molecular and cellular studies in higher plants. Trends Genet 6:187–192

Coornaert D, Vissers S, André B, Grenson M (1992) The *UGA43* negative regulatory gene of *Saccharomyces cerevisiae* contains both a GATA-1 type zinc finger and a putative leucine zipper. Curr Genet 21:301–307

Corton JC, Johnston S (1989) Altering the DNA binding specificity of GCN4 requires sequences adjacent to the zinc finger. J Cell Biochem 13E:37

Crawford NM, Campbell WH (1990) Fertile fields. Plant Cell 2:829–83

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

Deng X-W, Matsui M, Wei N, Wagner D, Chu AM, Feldmann KA, Quail PH (1992) *COP1*, an *Arabidopsis* regulatory gene encodes a protein with both a zinc-binding motif and a G_β homologous domain. Cell 71:791–801

Dingwall C, Robbins J, Dilworth SM, Roberts B, Richardson WD (1988) The nucleiplasmin nuclear location sequence is larger and more complex than the SV 40 large T-antigen. J Cell Biol 107:841–849

Dorbe MF, Caboche M, Daniel-Vedele F (1992) The tomato *nia* gene complements a *Nicotiana plumbaginifolia* nitrate reductase-deficient mutant and is properly regulated. Plant Mol Biol 18:363–375

Evans T, Felsenfeld G (1989) The erythroid-specific transcription factor Eryf1: a new zinc finger protein. Cell 58:877–885

Faure J-D, Vincentz M, Kronenberger J, Caboche M (1991) Co-regulated expression of nitrate and nitrite reductases. Plant J 1:107–113

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Fu YH, Marzluf GA (1987) Characterization of *nit-2*, the major nitrogen regulatory gene of *Neurospora crassa*. Mol Cell Biol 7:1691–1696

Fu YH, Marzluf GA (1990a) *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

Fu YH, Marzluf GA (1990b) *nit-2*, the major positive-acting nitrogen regulatory gene of *Neurospora crassa*, encodes a sequence-specific DNA-binding protein. Proc Natl Acad Sci USA 87:5331–5335

Fu YH, Marzluf GA (1990c) Site-directed mutagenesis of the 'zinc finger' DNA-binding domain of the nitrogen-regulatory protein NIT2 of *Neurospora*. Mol Microbiol 4:1847–1852

Guerrero MG, Vega JM, Losada M (1981) The assimilatory nitrate-reducing system and its regulation. Annu Rev Plant Physiol 44:1197–1199

Hope LA, Struhl K (1986) Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46:885–894

Jarai G, Truong HN, Daniel-Vedele F, Marzluf G (1992) NIT2, the nitrogen regulatory protein of *Neurospora crassa*, binds upstream of *nia*, the tomato nitrate reductase gene, in vitro. Curr Genet 21:37–41

Kalderon D, Roberts BL, Richardson WD, Smith AE (1984) A short amino acid sequence able to specify nuclear location. Cell 37:801–813

Katagiri F, Chua NH (1992) Plant transcription factors: present knowledge and future challenges. Trends Genet 8:22–27

Kleinhofs A, Warner RL, Narayanam KR (1985) Current progress towards understanding of the genetics and molecular biology of nitrate reductase in higher plants. In: Miflin B (ed) Oxford Surveys of Plant Molecular Biology, vol 2. Oxford University Press, New York, pp 91–121

Ko LJ, Yamamoto M, Leonard MW, George KM, Ting P, Engel JD (1991) Murine and human T-lymphocyte GATA-3 factors mediate transcription through cis-regulatory elements within the human T-cell receptor δ gene enhancer. Mol Cell Biol 11:2778–2784

Krappa R, Behn-Krappa A, Jahnel F, Doerfler W, Knebel-Mörsdorf D (1992) Differential factor binding at the promoter of early baculovirus gene PE38 during viral infection: GATA motif is recognized by an insect protein. J Virol 66:3494–3503

van der Krol A, Chua NM (1991) The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into the plant nuclei. Plant Cell 3:667–675

Kudla B, Caddick MX, Langdon T, Martinez-Rossi NM, Bennett CF, Sibley S, Davies RW, Arst HN Jr (1990) The regulatory gene *areA* mediating nitrogen metabolic repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. EMBO J 9:1355–1364

LaBrie ST, Wilkinson JQ, Tsay YF, Feldman KA, Crawford NM (1992) Identification of two tungstate-sensitive molybdenum cofactor mutants, *chl2* and *chl7*, of *Arabidopsis thaliana*. Mol Gen Genet 233:169–176

Lam E, Kano-Murakami Y, Gilmartin P, Niner B, Chua NH (1990) A metal-dependent DNA-binding protein interacts with a constitutive element of a light-responsive promoter. Plant Cell 2:857–866

Lutcke HA, Chow KC, Michel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG codons differs in plants and animals. EMBO J 6:43–48

Ma J, Ptashne M (1987) Deletion analysis of GAL4 defines two transcriptional activating segments. Cell 48:847–853

Maniatis T, Fritsch E, Sambrook J (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Martin DIK, Orkin SH (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf1. Genes Dev 4:1886–1898

Minehart PL, 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

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Silver PA (1991) How proteins enter the nucleus. Cell 64:489–497

Solomonson LP, Barber MJ (1990) Assimilatory nitrate reductase: functional properties and regulation. Annu Rev Plant Physiol Plant Mol Biol 41:225–253

Spieth J, Shim YH, Lea K, Conrad R, Blumenthal T (1991) *elt-1*, an embryonically expressed *Caenorhabditis elegans* gene homologous to the GATA transcription factor family. Mol Cell Biol 11:4651–4659

Takatsuji H, Mori M, Benfey PN, Ren L, Chua NH (1991) Characterization of a zinc finger DNA-binding protein expressed specifically in *Petunia* petals and seedlings. EMBO J 11:241–249

Trainor CD, Evans T, Felsenfeld G, Boguski MS (1990) Structure and evolution of a human erythroid transcription factor. Nature 343:92–96

Tsai SF, Martin DIK, Zorn LI, D'Andrea AD, Wong GG, Orkin SH (1989) Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature 339:446–451

Vaucheret H, Marion-Poll A, Meyer C, Faure J-D, Marin E, Caboche M (1992) Interest in and limits to the utilization of reporter genes for the analysis of transcriptional regulation of nitrate reductase. Mol Gen Genet 235:259–268

Vincentz M, Caboche M (1991) Constitutive expression of nitrate reductase allows normal growth and development of *Nicotiana plumbaginifolia* plants. EMBO J 10:1027–1035

Vincentz M, Moureaux T, Leydecker M-T, Vaucheret H, Caboche M (1993) Regulation of nitrate and nitrite reductases expression in *Nicotiana plumbaginifolia* leaves by nitrogen and carbon metabolites. Plant J, in press

Wilson DB, Dorfman DM, Orkin SH (1990) A non-erythroid GATA-binding protein is required for function of the human preproendothelin-1 promoter in endothelial cells. Mol Cell Biol 10:4854–4862

Wray JL (1988) Molecular approaches to the analysis of nitrate assimilation. Plant Cell Environ 11:369–382

Wray JL, Kinghorn JR (1989) Molecular and genetic aspects of nitrate assimilation. Oxford Science Publications, Oxford

Note added in proof

The nucleotide sequence data reported in this paper will appear in the EMBL, Genbank and DDBJ Nucleotide Sequence Databases under the accession number X73111 Ntl1-NT7