Isolation of two Arabidopsis cDNAs involved in early steps of molybdenum cofactor biosynthesis by functional complementation of Escherichia coli mutants

Tine Hoff, Kirk M. Schnorr, Christian C. Meyer, Michel M. Caboche

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 270, No. 11, Issue of March 17, pp. 6100–6107, 1995
Printed in U.S.A.

Isolation of Two *Arabidopsis* cDNAs Involved in Early Steps of Molybdenum Cofactor Biosynthesis by Functional Complementation of *Escherichia coli* Mutants*

(Received for publication, August 25, 1994)

Tine Hoff‡, Kirk M. Schnorr, Christian Meyer, and Michel Caboche

From the Laboratoire de Biologie Cellulaire, Institut National de la Recherche Agronomique, 78026 Versailles Cedex, France

Most organisms appear to have a molybdenum cofactor consisting of a complex of molybdenum and a pterin derivative. Very little is known about molybdenum cofactor biosynthesis in plants or other eukaryotes, because the instability of the cofactor and its precursors makes it difficult to analyze this pathway. We have isolated two cDNA clones from the higher plant *Arabidopsis thaliana* encoding genes involved in early steps of molybdenum cofactor biosynthesis. The cDNAs were obtained by functional complementation of two *Escherichia coli* mutants deficient in single steps of molybdenum cofactor biosynthesis. The two cDNAs, designated *Cnx2* and *Cnx3*, encode proteins of 43 and 30 kDa, respectively. They have significant identity to the *E. coli* genes, *moaA* and *moaC*, involved in molybdenum cofactor biosynthesis. Both genes have N-terminal extensions that resemble targeting signals for the chloroplasts or the mitochondria. Import studies with the translated proteins and purified mitochondria and chloroplasts did not show import of these proteins to either of these organelles. Northern analysis show that *Cnx2* is expressed in all organs and strongest in roots. *Cnx3* is not expressed in abundant levels in any tissue but roots. For both genes there is no detectable difference in the expression level from plants grown with nitrate or with ammonium. The *Cnx2* gene has been mapped to chromosome II. Southern analysis suggests that both genes exist as single copies in the genome.

All molybdoenzymes studied to date, except for molybdodinitrogenases, contain a common molybdenum cofactor. This molybdenum cofactor (MoCo)[1] is essential for the activity of several key enzymes like nitrate reductase, xanthine dehydrogenase, sulfite oxidase, aldehyde oxidase, and formate dehydrogenase (1). These enzymes are important in both prokaryotic and eukaryotic pathways such as nitrogen assimilation, sulfur and purine metabolism, and hormone biosynthesis.

The basic structure of the molybdenum cofactor appears to be the same in all organisms investigated (2). The molybdenum cofactor consists of a 6-alkyl pterin (named molybdopterin) proposed to be complexed with molybdenum through the sulfur atoms of a dithiolene moeity on its 6-alkyl side chain (3). In plants the MoCo is thought to be in the pterin-base form, but in some organisms such as *Escherichia coli* MoCo exists in a nucleotide form, where a guanine dinucleotide is linked to the 4-carbon side chain of the moeity (2). Fig. 1 shows the proposed structure of MoCo in plants and in *E. coli*.

Very little information is available on MoCo biosynthesis in plants or other eukaryotes. The instability of the cofactor and its precursors makes it difficult to establish the biosynthetic pathway by direct detection and analysis of the intermediates. The isolation and analysis of mutants affected in MoCo biosynthesis have made it possible to classify them into several complementation groups corresponding to different steps in MoCo biosynthesis. Chlorate resistance has been the basis for selecting MoCo mutants in bacteria, fungi, algae, and higher plants (for reviews see Ref. 4). The mutants are pleiotropically defective in nitrate reductase and other molybdoenzyme activities. In plants, MoCo-deficient mutants have been classified into six different complementation groups (*cnxA–cnxF*), indicating that at least 6 genes are involved in MoCo biosynthesis (see Refs. 5 and 6 for reviews). The genes and their precise functions in MoCo biosynthesis are not known. In *Nicotiana plumbaginifolia*, the *cnxA* complementation group shows a characteristic different from the other groups; their phenotype can be partially restored when grown on high levels of molybdate (7, 8). This led to the hypothesis that the *CnxA* gene product is involved in the insertion of molybdenum in the molybdopterin (7). MoCo deficiency in plants leads to the inability to use nitrate as a nitrogen source due to a lack of active nitrate reductase enzyme. *N. plumbaginifolia* MoCo mutants can be partially rescued either by grafting, where they show a chlorotic and wilty phenotype, or by growth *in vitro* on a reduced N source, where they show developmental abnormalities such as reduced leaf size and long internodes (9). The wilty phenotype is due to a defect in abscisic acid biosynthesis, linked to the absence of the MoCo enzyme aldehyde oxidase[2] (10).

Several cases of MoCo deficiency have been reported in humans. It causes mental retardation and other severe symptoms and leads to death in infancy (11). By co-culturing cells from affected individuals, it has been possible to identify two complementation groups in humans (11). No treatment is yet possible, and the diagnostics and prenatal testing are difficult. MoCo mutants have been studied in *Drosophila melanogaster*, where such mutants are easily identified by eye color variations. This variation is a result of reduced or deficient xanthine dehydro-

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

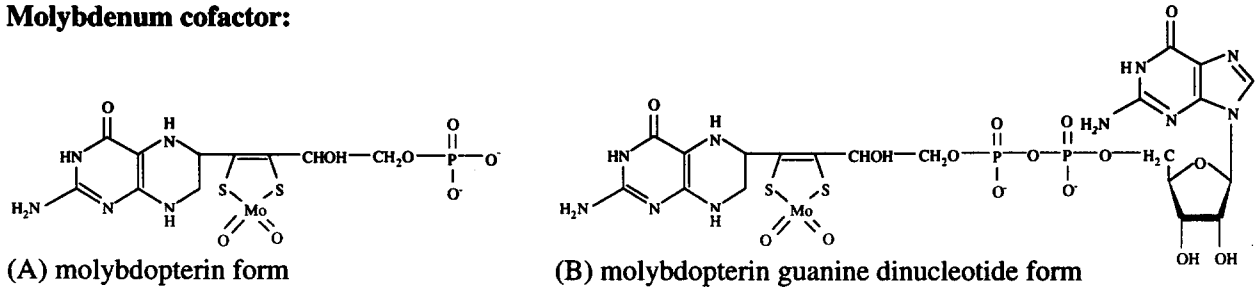
The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) Z48046 (Cnx3) and Z48047 (Cnx2).

‡ Supported by the Carlsberg Foundation. To whom correspondence should be addressed: Laboratoire de Biologie Cellulaire, INRA, F-78026 Versailles Cedex, France. Tel: 33-1-30-83-30-71; Fax: 33-1-30-83-30-99.

[1] The abbreviations used are: MoCo, molybdenum cofactor; bp, base pair(s); IPTG, isopropyl-β-D-thiogalactoside; PCR, polymerase chain reaction; PQQ, pyrroloquinoline-quinone; RFLP, restriction fragment length polymorphism; RACE, rapid amplification of cDNA ends.

[2] Leydecker, M.-T., Moureaux, T., Kraepiel, Y., Schnorr, K. M., Caboche, M. (1995) *Plant Physiol.*, in press.

Molybdenum cofactor:



(A) molybdopterin form (B) molybdopterin guanine dinucleotide form

FIG. 1. **The structure of MoCo in plants (*A*) and in *E. coli* (*B*).** Structure is modeled after that done by Rajagopalan and Johnson (2).

genase (12). Recently a gene involved in molybdenum cofactor biosynthesis has been isolated from *Drosophila* (13).

MoCo and its biosynthesis have been studied extensively in *E. coli* (for a review, see Ref. 2). Five loci involved in the synthesis of a mature MoCo have been identified by screening for chlorate resistance. Each locus encodes one or more enzymes in the MoCo biosynthetic pathway. Two loci, *moa* and *moe*, are required for early steps of MoCo biosynthesis: the synthesis of molybdopterin. The sequence has been determined for both loci (14, 15). The *moa* locus contains five genes in an operon (*moaA–moaE*). The *moaA–moaC* gene products are thought to be involved in molybdopterin precursor production, while the *moaD* and *moaE* encode two proteins forming a heteromeric "converting factor" that is responsible for the conversion of the precursor into molybdopterin (15–17). The *moe* locus is probably involved in the sulfur donation to the converting factor (18).

The aim of this study was to obtain information on MoCo biosynthesis in plants by cloning genes involved in its biosynthesis. As the basic structure of MoCo is probably the same in all organisms, its biosynthesis might share common steps as well. Our approach was to complement *E. coli* MoCo mutants with plant cDNAs by selecting for restored ability of an *E. coli* MoCo mutant to use nitrate as the terminal electron acceptor in anaerobic conditions (restored nitrate reductase activity). Here we report the isolation, characterization, and functional expression of two genes involved in early steps of MoCo biosynthesis in *Arabidopsis thaliana*. These cDNAs specifically supressed the *E. coli moaA* and *moaC* mutations.

EXPERIMENTAL PROCEDURES

Plant Material—Plants of *Arabidopsis* ecotype Columbia were used in all experiments. Three- to 6-week-old plants grown with a 12-h photoperiod at 20 °C/18 °C day/night temperature were used for DNA and RNA isolation. NO_3^--grown and NH_4^+-grown plants were obtained *in vitro* after 3 weeks of culture on solid medium (19) with either 9 mM KNO_3 or 1 mM ammonium succinate and 2 mM glutamine as the nitrogen source. For RNA analysis plant material was harvested in the beginning of the light period. The subcultured suspension cells, initiated by Michele Axelos, was a gift from Dr. Yves Mathieu (CNRS, Gif sur Yvette, France).

Bacterial Strains—The *E. coli* strain MC4100 (*araD139 Δ(argF-lac)U169 rpsL150 relA1 flbB3501 deoC1 ptsF25 rbsR*) was used as wild type in the *in vivo* nitrate reductase assay. The *E. coli* mutants KB2039 and KB2066 (F^- thr, leu his pro arg thi ade gal lacY malE xyl ara mtl str T^r, λ^r) were used for the complementation studies. KB2039 has a point mutation in the moa locus in gene *A*, and KB2066 has a point mutation in the *moa* locus gene *C*. The mutants were provided by G. Giordano and C. Iobbi-Nivol, CNRS, Marseille, France with permission from D. H. Boxer, University of Dundee, United Kingdom.

Complementation—The *A. thaliana* cDNA expression library used, provided by R. W. Davis (Stanford University), was an amplification of the original library reported by Elledge *et al.* (20). A plasmid library was made from the λYES phage library as described by Elledge *et al.* (20) except that the DNA was purified on Qiagen columns according to the manufacturer's instructions (Qiagen Inc., Chatsworth, CA) instead of CsCl purification. Preparation of electrocompetent *E. coli* cells and electroporation of hosts were performed according to the methods of Dower *et al.* (21) using a Bio-Rad Gene Pulser. For functional complementation studies, 100-ng λYES plasmid library was electroporated into 40 μl of frozen electrocompetent mutant cells in chilled electroporation cuvettes. The Bio-Rad Gene Pulser was set at 12.5 kV cm^{-1}, 200 microfarads capacitance, and 200 ohms resistance. After electroporation, 1 ml of LB medium containing 0.5 mM IPTG was added to the cuvette, the cells were transferred to a sterile test tube and then incubated in a rotary shaker at 195 rpm for 2 h at 28 °C. Cells were then washed in sterile water and plated at a density between 10^6 and 10^7 transformants/plate on the following solid medium (IK medium: 25 mM Na_2HPO_4, 7 mM KH_2PO_4, 0.25 mM $MgCl_2$, 9 mM NH_4Cl, 10 mM KNO_3, 0.4% glycerol, 5% LB medium) containing 50 μg/ml ampicillin and 0.5 mM IPTG. The plates were incubated at 28 °C anaerobically for 3–5 days. The anaerobic condition was obtained in an anaerobic jar using the "Anaerobic System" (Difco Laboratories, Detroit, MI) to generate H_2 and CO_2.

In Vivo Assays for Nitrate Reductase Activity—*E. coli* cultures used for inoculation were prepared as follows: cultures were grown to an optical density of 0.6 (A_{600}) in LB medium, pelleted at 4000 × *g* for 10 min, washed three times with 10 mM $MgSO_4$, and resuspended in IK medium at a cell density of 10 A_{600} units/ml. *In vivo* nitrate reductase assays were performed on cell cultures adjusted to A_{600} = 0.2 and grown anaerobically at 37 °C in liquid IK medium containing 50 μg/ml ampicillin and 0.5 mM IPTG. After 2 h of growth, the accumulation of NO_2^- was measured directly in an aliquot of the culture as described by MacGregor *et al.* (22). At the same time point, an aliquot of the culture was taken and used to prepare dilutions for plating in order to determine the cell numbers in each culture.

DNA Sequence Determination—DNA manipulations were performed by the general methods of Sambrook *et al.* (23). Template DNA for sequence analysis were made by restriction enzyme subcloning into pBluescript (Stratagene, La Jolla, CA). Nucleotide sequences were determined in both directions by a modification of the dideoxy method of Sanger *et al.* (24) using an Applied Biosystems model 373A automated sequencer and the PRIZM® sequencing system (Applied Biosystems, Foster City, CA). Sequence analysis was performed with the Wisconsin Genetics Computer Group software suite (25).

RACE—The 5′ ends of both gene transcripts were cloned by the RACE technique (rapid amplification of cDNA ends) using a 5′-Amplifinder RACE kit (Clontech Laboratories, Palo Alto, CA). First strand cDNA were made from poly(A)-enriched RNA from a cell suspension culture by the method recommended by the manufacturer using gene-specific primers (for p39A5: 5′-CAACATGTCAGAGACAGGATTGTC-3′; for p66A4: 5′-TCAGAGAAATCCATTGTCCCAG-3′). An anchor oligonucleotide was ligated to the 3′ends of the single-stranded cDNA. This cDNA was template for PCR amplification using a primer complementary to the anchor and a nested gene-specific primer (for p39A5: 5′-ACTTGATGAGCTGCATAACTACTG-3′ and for p66A4: 5′-ACTGCATCTCCTGATTCAGCTC-3′). The PCR products were cloned into pBluescript and sequenced.

Northern and Southern Analysis—Total RNA was isolated as described previously (26), and poly(A)-enriched RNA was isolated using poly(T) columns according to the manufacturer's instructions (Pharmacia LKB Biotechnology, Uppsala, Sweden). The RNA was denatured in formamide/formaldehyde, electrophoresed in 1.5% (w/v) agarose-formaldehyde gels, and transferred to Hybond N filters according to Sambrook *et al.* (23). Total DNA was isolated from 4-week-old plants according to the the methods of Dellaporta *et al.* (27). DNA samples were digested with restriction enzymes and electrophoresed on 0.8% agarose gels. DNA gels were denatured, neutralized, and transferred to Hybond N membranes according to the manufacturers instructions (Amersham, Buckinghamshire, United Kingdom). DNA and RNA blots

were UV treated in a Stratalinker oven at 120 mJ/cm^2. Probes consisted of full-length complementing cDNA fragments randomly primed according to the methods of Feinberg and Vogelstein (1983) (28). The *Nia2* probe, encoding nitrate reductase from *Arabidopsis*, was kindly provided by Nigel Crawford, University of California, San Diego, CA. Nucleic acid blots were prehybridized, hybridized, and washed as described previously (29). Filters were autoradiographed at −80 °C using intensifying screens.

Chloroplast and Mitochondrial Import Studies—The *Xho*I whole fragment inserts from p39A5 and p66A4 were subcloned into a pBluescript vector so that insert messages could be transcribed from the T3 RNA polymerase promoter. The PsaD construct, encoding a chloroplast-localized photosystem I protein was kindly provided by V. Skovgaard Nielsen, The Agricultural University, Copenhagen, and the mitochondria-localized superoxide dismutase cDNA (SOD3) was kindly provided by J. G. Scandalios, North Carolina State University, Raleigh, NC. *In vitro* transcription/translation was performed with 1 μg of each plasmid and with [^{35}S]methionine using the TNT Coupled Reticulocyte Lysate System (Promega Corp., Madison, WI). Chloroplasts and mitochondria were isolated from leaves of 2-week-old pea plants. Chloroplasts were isolated according to Cline *et al.* (30) and mitochondria were isolated according to White and Scandalios (31). Chloroplast import assays were done essentially as described by Cline *et al.* (30). Chloroplasts equivalent to 25–35 μg of chlorophyll were incubated with 25 μl of the translation reaction mix in HS buffer (50 mM Hepes-KOH (pH 8.0) and 330 mM sorbitol) containing 5 mM methionine and 8 mM Mg-ATP in a total volume of 300 μl. The reactions were incubated at 25 °C for 30 min in light. After incubation the reactions were pelleted, resuspended in 500 μl of HS buffer, and divided in two parts. One part was treated with 25 μg of thermolysin for 40 min on ice. The proteolysis was terminated by adding EDTA to a final concentration of 50 mM. The chloroplasts of the thermolysin-treated and the untreated samples were pelleted, resuspended in Laemmli buffer, and analyzed on SDS-polyacrylamide gels (14%) according to Laemmli (32), followed by fluorography (33). Mitochondrial import studies were done as follows. Purified mitochondria equivalent to 150 μg of protein were incubated with 15 μl of the translation reaction mix in HM buffer (50 mM Hepes-KOH (pH 7.4) and 0.4 M mannitol) containing 1 mM Mg-ATP, 5 mM GTP, 1 mM $MgCl_2$, 1 mM dithiothreitol, 40 mM KCl, and 1 mM methionine in a total volume of 200 μl. The reactions were incubated 1 h at 26 °C. After incubation the mitochondria were pelleted, resuspended in HM buffer, and divided in two parts. One part was treated with thermolysin as described for the chloroplast experiment. The samples were analyzed on SDS-polyacrylamide gels as described above.

RESULTS

Complementation of the E. coli moaA and moaC Mutation by Arabidopsis cDNAs—Approximately 5 × 10^6 total transformants were obtained after four electroporations of an *Arabidopsis* cDNA plasmid library into the *E. coli* mutants KB2039 (affected in gene *A* in the *moa* locus). After growth for 5 days under selective conditions (described under "Materials and Methods"), eight colonies appeared. Plasmids were prepared from all eight colonies. Upon retransformation of KB2039, four of the eight plasmids conferred the ability to complement the mutation. These four plasmids had nearly the same size insert, and identical restriction pattern of the insert DNA suggested these plasmids to code for the same gene. The four colonies that contained plasmids that did not re-complement were probably revertants, and they contained plasmids with varying insert size.

The *E. coli* mutant KB2066 (affected in gene *C* in the *moa* locus) was transformed with the cDNA plasmid library in the same manner. 8 × 10^6 total transformants were obtained upon four electroporations. After 5 days of growth under selective conditions two groups of colonies appeared: small and large (29 small colonies and 55 large colonies). Plasmids from six small colonies and 12 large colonies were prepared, and upon retransformation of the mutant only the plasmids of the group giving rise to small colonies were able to complement the mutant. The group consisting of large colonies were probably revertants. The six complementing plasmids had approximately the same size insert and similar DNA restriction pattern.

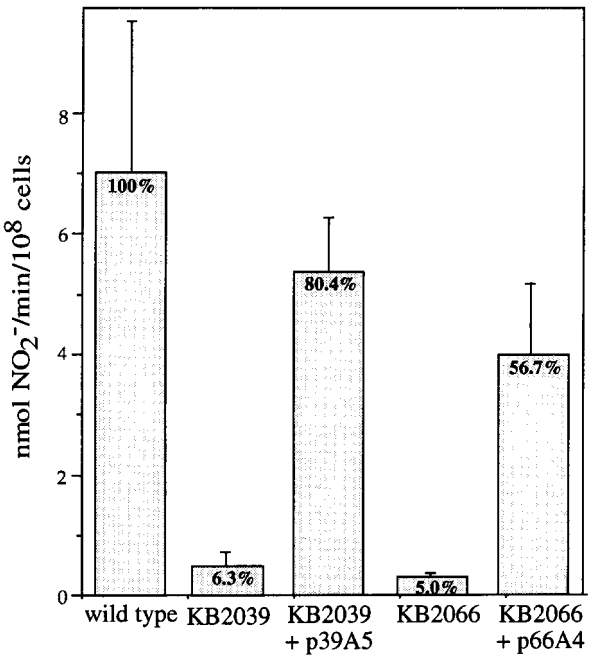


FIG. 2. ***In vivo* nitrate reductase assays of the non complemented and the complemented *moa* mutants.** The mutants KB2039 and KB2066 and the mutants containing the *Arabidopsis* cDNAs: p39A5 (*Cnx2*) and p66A4 (*Cnx3*), respectively, were used for the assay. The wild type *E. coli* strain MC4100 was used as control. Each value is the mean of three independent experiments. The NO_2^- produced are expressed in nmol/min/10^8 cells, and for each experiment the percentage of the wild type level is indicated.

Plasmid p39A5 complementing the KB2039 mutant and plasmid p66A4 complementing the KB2066 mutant were selected for further analysis.

In Vivo Assay of Nitrate Reductase Activities in the Complemented moa E. coli Mutants—The presence of nitrate reductase activity in the complemented *E. coli* mutants was confirmed by an *in vivo* nitrate reductase assay performed as described under "Materials and Methods." The mutants and a wild type *E. coli* carrying λYES plasmids with no inserts were used as controls. The cells were grown in minimal medium under nitrate reductase inducing conditions: anaerobiosis in the presence of nitrate. After 2 h of growth, the accumulation of NO_2^- in the medium was measured and the number of cells were determined. The results presented in Fig. 2 clearly show that both complemented mutants produce NO_2^-, while the mutants carrying an empty plasmid produce only background levels of NO_2^-. The level of NO_2^- produced by the *moaA* mutant complemented with p39A5 is 80% of the wild type, while the *moaC* mutant complemented with p66A4 produced 58% of the wild type level.

The lower level of NO_2^- produced by the p66A4 complemented *moaC* mutant, and thereby the lower level of nitrate reductase activity, when compared to the wild type, may explain the slow growth of the complemented *moaC* mutant on solid selection medium in the original complementation experiment, where only small colonies where obtained. It might simply reflect poor complementation by the *Arabidopsis* protein, or it could be a question of optimizing expression of the construct. We know that the coding region of this clone is not in frame with the ATG in the expression vector, so, upon expression under the *lac* promoter, the proper translation of the product must arise from translation re-initiation probably at the first ATG in the clone. Furthermore, p66A4 is not a full-length clone (see below), but the missing part is in the N-terminal extension that is not present in the corresponding *E.*

coli gene.

Characterization of the Arabidopsis Homolog to E. coli moaA—The 1550 bp insert of plasmid p39A5 was completely sequenced. Analysis of the derived sequence indicated an open reading frame of 1170 nucleotides encoding a protein of 390 amino acids with an estimated molecular mass of 43.4 kDa (Fig. 3*A*). The clone contained 26 bp upstream of the first ATG and a 351-bp 3′-untranslated region, including a poly(A) tail of 24 residues. Several putative polyadenylation signals close to the motif AATAAA can be found upstream from the poly(A) tail. The clone contained no stop codons upstream of the first start codon in the same reading frame, so in theory the reading frame could be longer. We decided therefore to clone the 5′ end of the gene transcript by the RACE technique to determine if our clone was full-length. Ten RACE clones were sequenced. Two of the longest clones contained an additional 97 nucleotides when compared to p39A5. The additional 97 nucleotides in the 5′ end of the cDNA sequence do not change the start of the reading frame but changes the length of the leader sequence to a total of 123 bp, including a stop codon 81 bp upstream of the methionine start codon in the reading frame. The 123-bp long untranslated region is AT-rich (59% A+T). The full-length transcript of this cDNA is therefore at least 1647 bp long including the poly(A) tail. This corresponds well with the transcript size on a Northern analysis.

We have named this cDNA *Cnx2* according to the guidelines of Caboche *et al.* (34). The mnemonic *cnx* (cofactor of nitrate reductase and xanthine dehydrogenase) is the same as used for plant MoCo-deficient mutants. Even though this is the first *Cnx* gene cloned from plants, it is given the number 2, because *Cnx1* is assigned to the molybdenum-repairable locus (*CnxA*) once it is identified (34).

Genomic Organization of Cnx2—In order to reveal the *Cnx2* genomic organization in *Arabidopsis*, a Southern blot analysis of *Arabidopsis* genomic DNA, digested with different restriction enzymes, was performed with the whole insert of p395A as probe (Fig. 4*A*). Both at high and low stringency hybridization and wash conditions, *Cnx2* appears to exist as a single copy in the genome. In order to map the gene, the insert of p39A5 was used as probe for RFLP (restriction fragment length polymorphism) analysis of recombinant inbred lines of *Arabidopsis thaliana* (Lister and Dean, 1993). The RFLP analysis indicates that *Cnx2* is located on chromosome II about 16.3 centimorgan below the m336 marker.

Expression Analysis of Cnx2—Northern blot analysis of total RNA shows that the *Cnx2* gene is expressed in all organs, and the strongest expression is found in roots and cell suspension cultures (Fig. 5). There is no detectable difference in the message level from plants grown with or without nitrate (Fig. 5).

Amino Acid Analysis of Cnx2—The predicted amino acid sequence of *Cnx2* was compared by the Blast program (Altschul *et al.* 1990) to available peptide and nucleotide sequences in the data bases. The highest level of identity is found to the *E. coli moaA* sequence encoding one of the first steps of MoCo biosynthesis. The aligned peptide sequences show 32% identity (Fig. 6*A*). The Cnx2 peptide sequence also shows similarity to NifB proteins from several nitrogen-fixing bacteria such as *Azotobacter vinelandii* and to FixZ proteins from *Rhizobium leguminosarum* and other bacteria. *NifB* and *fixZ* gene products are involved in the biosynthesis of the nitrogenase iron-molybdenum cofactor. Furthermore, identity can be established to a protein involved in the biosynthesis of the cofactor pyrroloquinoline-quinone (PQQ), which is a cofactor for dehydrogenases. NifB, FixZ, and PQQ synthesis protein show 20–24% overall identity (45–49% similarity) to Cnx2. Among other positive matches above the Blast score of 55, the *Arabidopsis* sequence matches with a region in guanine-binding proteins (26% identity over 53 amino acids). The *Arabidopsis* sequence also shows identity (30% identity over a 200-amino acid overlap) to a reading frame of unknown function located upstream of the genes encoding the molybdo-enzyme formate dehydrogenase of *Methanobacterium formicium*. This could be a homolog of MoaA and Cnx2. A multiple alignment of Cnx2, moaA, PQQ synthesis protein, FixZ, NifB, and a possible homolog to Cnx2 from *M. formicium* is shown in Fig. 7 together with a derived consensus sequence. Alignment of these proteins with the *Cnx2* and *moaA* translations reveals two domains. One domain shows significant conservation between all the aligned proteins (the domain shown in Fig. 7) and another domain in the C-terminal region where Cnx2 and MoaA are conserved, but there is no consensus with the remaining proteins.

Alignment of the *E. coli* MoaA and the *Arabidopsis* Cnx2 sequences reveals that the *Arabisopsis* cDNA has a 60-amino acid extension in the N-terminal region of the peptide translation. This N-terminal extension has some characteristics of transit peptides for either the chloroplast or for the mitochondria: a high percentage of hydroxylated amino acids and a low level of acidic amino acids (35, 36). To analyze the function of the N-terminal extension, import studies with purified chloroplasts and mitochondria were performed. The results are described in a separate section below.

Characterization of the Arabidopsis Homolog to E. coli moaC—One plasmid (p66A4) complementing the KB2066 mutant was characterized by sequencing. The 1135-bp insert had an open reading frame of 699 bp encoding a putative 233-amino acid protein. The clone contained no stop codons in the 105-bp-long sequence upstream of the start codon in the translated reading frame. We decided therefore to clone the 5′ end of the gene transcript by the RACE technique to determine if p66A4 contained the full-length clone. Ten RACE clones were sequenced, and the longest contained 19 bp more than the p66A4 clone in the 5′ end. The additional 19 bp included a methionine start codon and, upstream of this, a stop codon in the same reading frame. This changed the length of the reading frame to 810 bp encoding a protein of 270 amino acids with an estimated molecular mass of 29.5 kDa (Fig. 3*B*). Isolation of a genomic clone using the p66A4 insert as probe and sequencing of the 5′ end of the corresponding gene confirm the RACE results (results not shown). The 5′-untranslated region apparently consists of at least 14 bp, and the 3′-untranslated region is 328 bp long, including a 78-residue poly(A) tail. We have named this cDNA *Cnx3*.

Genomic Organization of Cnx3—Southern blot analysis of restriction-digested *Arabidopsis* DNA using the p66A4 insert as probe indicates a single copy gene even under low stringency hybridization and wash conditions (Fig. 4*B*). Chromosome mapping of the *Cnx3* gene by RFLP analysis of recombinant inbred *Arabidopsis* lines was not succesful, as no polymorphism could be established between the two parental *Arabidopsis* lines using the *Cnx3* cDNA as probe and testing with 25 different restriction enzymes.

Expression Analysis of Cnx3—Northern blot analysis of total RNA from different organs using *Cnx3* cDNA as probe shows a very weak signal in leaves, little signal in flowers and siliques, and a stronger signal in roots and in cell suspension cultures (Fig. 5). There is no detectable difference in the expression level of the *Cnx3* gene from plantlets grown with or without nitrate (Fig. 5).

Amino Acid Analysis of Cnx3—The deduced amino acid translation of *Cnx3* has significant homology to the *E. coli* MoaC peptide sequence (51% identity in 150-amino acid overlap). Alignment of the two sequences shows that the *Arabidop-*

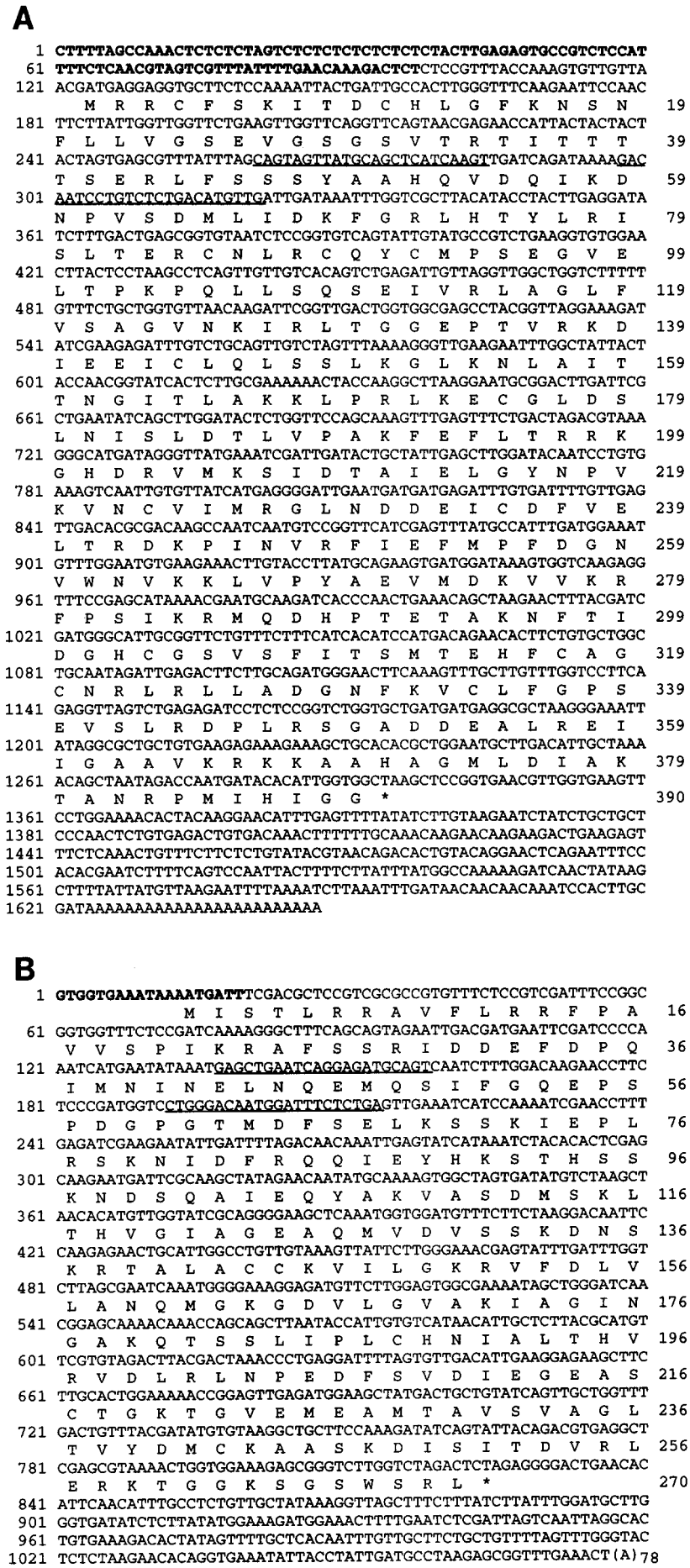


FIG. 3. **Nucleotide sequence and the deduced amino acid translation of the *Arabidopsis* cDNAs: *Cnx2* from plasmid p39A5 (*A*) and *Cnx3* from plasmid p66A4 (*B*).** *Numbers* on the *left* and on the *right* correspond to the nucleotide base pair and amino acid positions, respectively. Sequences corresponding to the primers used for amplification of the 5′ ends are *underlined*. The 5′ sequences provided from the RACE clones are shown in *bold*.

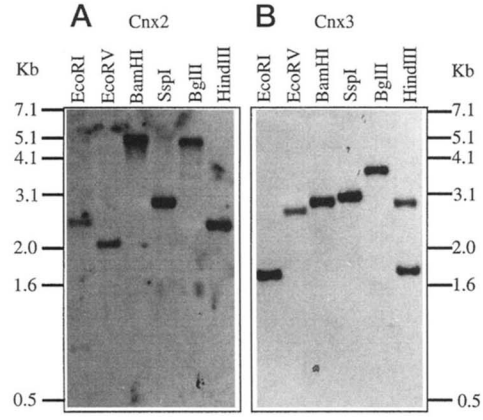


FIG. 4. **Southern blot analysis of *Arabidopsis* genomic DNA.** 2 μg of genomic DNA was digested with each of the restriction enzymes indicated on the figure. In *A* the *Cnx2* cDNA was used as probe, and in *B* the *Cnx3* cDNA was used as probe.

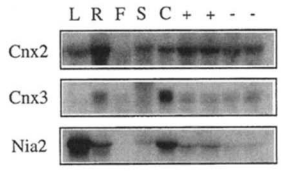


FIG. 5. **Northern blot analysis of total RNA from different plant tissues.** The lanes were loaded with approximately 15 μg of total RNA from leaves (*L*), roots (*R*), siliques (*S*), flowers (*F*), cell suspension culture (*C*), and from whole plantlets grown *in vitro* with KNO_3 (+) or with NH_4-succinate (−). The same filter was probed with three different probes: the entire *Cnx2* cDNA, the entire *Cnx3* cDNA, or the *Arabidopsis Nia2* cDNA (nitrate reductase gene).

sis cDNA has an additional 112 amino acids in the N-terminal region (Fig. 6*B*). As for Cnx3, the N-terminal sequence has some of the characteristics of chloroplastic and mitochondrial transit peptides (35, 36). The data bases were searched for other sequence similarities, and high similarity was found to two open reading frames with no known function. One is an incomplete open reading frame in a cluster containing three genes involved in tryptophan biosynthesis in *Azospirillum* (47% identity in a 87-amino acid overlap). The other open reading frame showing homology to Cnx3 is located upstream of an adhesin gene from *Helicobacter pylori* and has 38.5% identity in 104-amino acid overlap. Both reading frames could be homologs of Cnx3 and MoaC.

Chloroplast and Mitochondrial Import Studies—The two cDNA clones p39A5 and p66A4 were used for *in vitro* transcription and translation. The labeled translation product were incubated with either purified chloroplasts or mitochondria. One aliquot of each reaction was treated with thermolysin to digest proteins on the surface of the chloroplasts or mitochondria and only the proteins inside the organelles should be protected. Fig. 8 shows the fluorography of SDS-polyacrylamide gels loaded with the chloroplast and mitochondrial importation reactions. For the chloroplast importation experiment, PsaD, a photosystem I protein, was used as a positive control. The PsaD protein is imported and processed as expected (Fig. 8*A*, *lanes 7–9*) but neither the Cnx2 (*lanes 1–3*) nor the Cnx3 (*lanes 4–6*) are imported to the chloroplasts. The result of the mitochondria importation study is seen in Fig. 8*B*. A mitochondria-localized superoxide dismutase cDNA (SOD3) was used as the positive control and is imported as expected (Fig. 8*B*, *lanes 7–9*). But the Cnx2 and the Cnx3 peptides are not imported or processed by the mitochondria (*lanes 1–6*).

A

```
Cnx2    1                                                    MRRCFSKITDC

Cnx2   12 HLGFKNSNFLLVGSEVGSGSVTRTITTTTSERLFSSSYAAHQVDQIKDNP

Cnx2   62 VSDMLIDKFGRLHTYLRISLTERCNLRCQYCMPSE..GVELTPKPQLLSQ
          :   | | |:|   |||:|:|: ||:|| ||:| :    ::| |  :|
MoaA    1 MASQLTDAFARKFYYLRLSITDVCNFRCTYCLPDGYKPSGVTNKG.FLTV

Cnx2  110 SEIVRLAGLFVSAGVNKIRLTGGEPTVRKDIEEICLQLSSLKGLKNLAIT
           || |:   |   | :|:|||||| :|:|: :|   :    :::.:|:|
MoaA   50 DEIRRVTRAFARLGTEKVRLTGGEPSLRRDFTDIIAAVRENDAIRQIAVT

Cnx2  160 TNGITLAKKLPRLKECGLDSLNISLDTLVPAKFEFLTRRKGHDRVMKSID
          |||  | : :: ::: || ::|:|:| | :  |  :|     : || :||
MoaA  100 TNGYRLERDVASWRDAGLTGINVSVDSLDARQFHAITGQDKFNQVMAGID

Cnx2  210 TAIELGYNPVKVNCVIMRGLNDDEICXFVESTRDKPINVRFIEFMPFDGN
           |:| |:: |||| |:||::|  ::  |::    :|| :||||:|  ::
MoaA  150 AAFEAGFEKVKVNTVLMRDVNHHQLDTFLNWIQHRPIQLRFIELMETGEG

Cnx2  260 VWNVKKLVPYAEVMDKVVKRFPSIKRMQDHPTETAKNFTIDGHCGSVSFI
              :|    ::|:   : |  | : ::  : |  |   :  | :::|
MoaA  200 SELFRKHHISGQVLRDELLRRGWIHQLRQRSDGPAQVFCHPDYAGEIGLI

Cnx2  310 TSMTEHFCAGCNRLRLLADGNFKVCLFGPSEVSLRDPLRSGADDEALREI
               ||| |||||:   | : :|||| ::| ||| |  : :::||
MoaA  250 MPYEKDFCATCNRLRVSSIGKLHLCLFGEGGVNLRDLLEDDTQQQALEAR

Cnx2  360 IGAAVKRKKAAHAGMLDIAKTANRPMIHIGG 39 0
          |:||:: ||  |  : :         :  |||
MoaA  300 ISAALREKKQTHF.LHQNNTGITQNLSYIGG 329
```

B

```
Cnx3    1                                                    MISTLRRAVFLR

Cnx3   13 RFPAVVSPIKRAFSSRIDDEFDPQI MNINELNQEMQSIFGQEPSPDGPGI

Cnx3   63 MDFSELKSSKIEPLRSKNIDFRQQIEYHKSTHSSKNDSQAIEQYAKVASD

Cnx3  113 MSKLTHVGIAGEAQMVDVSSKDNSKRTALACCKVILGKRVFDLVLSNQMG
          || |||:  ||||:||:|| | :  | | |   | :    : :::
MoaC    1 MSQLTHINAAGEAHMVNVSAKAETVREARAEAFVTMRSETLAMIIDGRHH

Cnx3  163 KGDVLGVAKIAGINGAKQTSSLIPLCHNIALTHVRVDLRLNPEDFSVDIE
          ||||:: |:|||| :|| |  |||||| : |  | |:|  :  : |  ||
MoaC   51 KGDVFATARIAGIQAAKRTWDLIPLCHPLMLSKVEVNLQAERAQ.SGRIE

Cnx3  213 GEASCTGKTGVEMEAMTAVSVAGLTVYDMCKAASKDISITDVRLERKTGG
                |||||||||:|| |||:||:|||||  ||: |  |||  | ||
MoaC  100 TLCRLTGKTGVEMEALTAASVAALTIYDMCKAVQKDMVIGPVRLLAKSGG

Cnx3  263 KSGSWSRL 270
          |||  :
MoaC  150 KSGDFKVEADD 160
```

FIG. 6. **Peptide alignment of Cnx2 and Cnx3 from *Arabidopsis* with their *E. coli* homologs.** *A*, alignment of the deduced amino acid sequence of *Cnx2* from *Arabidopsis* and of MoaA from *E. coli*. *B*, alignment of the deduced amino acid sequence of *Cnx3* from *Arabidopsis* and MoaC from *E. coli*. *Vertical lines* indicate identical residues, and *double points* indicate conservative changes. Sequences were obtained from GenBank and have the following accession numbers: P30745 and P30747 for *moaA* and *moaC*, respectively.

DISCUSSION

We have cloned and characterized two *Arabidopsis* cDNAs able to restore nitrate-dependent anaerobic growth in two *E. coli* mutants affected in early steps of MoCo biosynthesis. The two MoCo-deficient, and therefore NR-deficient, *E. coli* mutants had at least partially restored nitrate reductase activity when complemented with the cDNAs. Both cDNA peptide translations have significant homology to the *E. coli* molybdenum biosynthetic genes *moaA* and *moaC*, respectively. On the basis of their function and sequence similarities, we suggest that the *Arabidopsis* cDNAs correspond to enzymes involved in MoCo biosynthesis. To our knowledge these are the first characterized cDNAs encoding enzymes involved in MoCo biosyn-

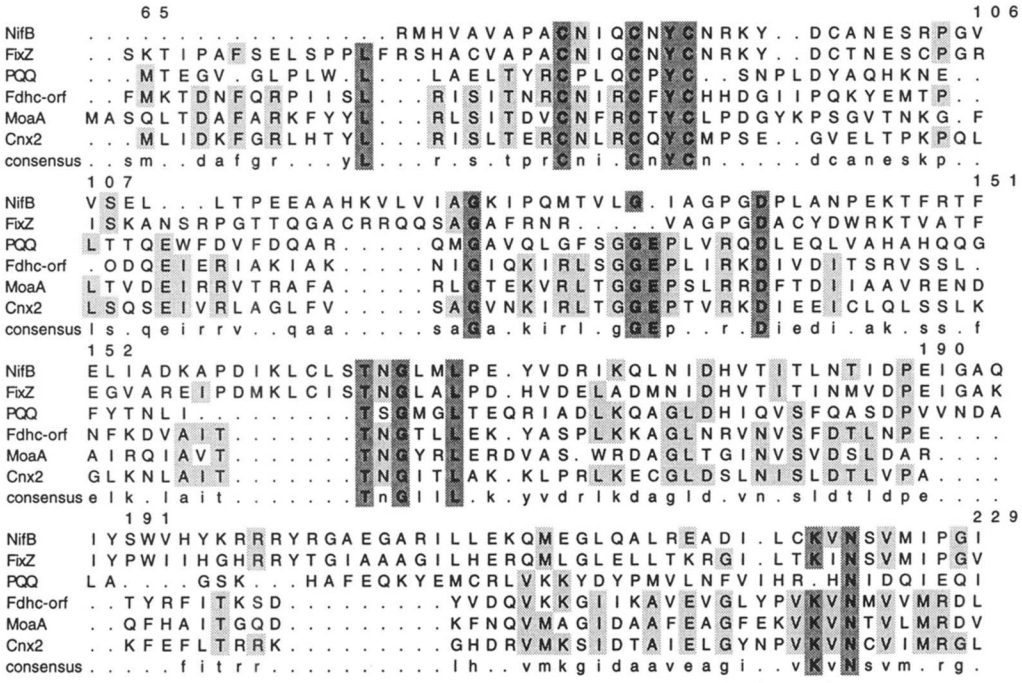


FIG. 7. **Multiple alignment of the predicted amino acid sequence of *Cnx2* with related proteins.** The following proteins are aligned with Cnx2: MoaA from *E. coli* (accession number P30745), Nifb from *Anabaena sp.* (accession number J05111), Fixz from *Rhizobium leguminosarum* (accession number P07748), PQQ synthesis protein III from *Acinetobacter calcoaceticus* (accession number X06452) and Fdhc-orf, an open reading frame from *Methanobacterium formicium* (accession number M64798). The numbering is according to the *Arabidopsis* Cnx2 peptide sequence. *Dark gray shaded boxes* represent sequence identity, and *light gray shaded boxes* represent at least 3 identities in the alignment. The alignment was made with Pileup (GCG program) and adjusted by hand.

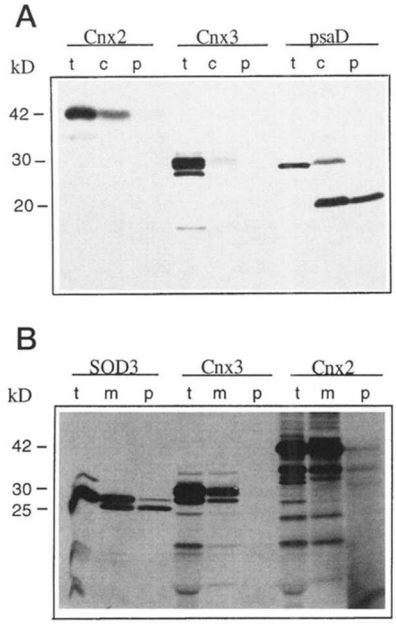


FIG. 8. **Chloroplast and mitochondrial import studies with p39A5 (*Cnx2*) and p66A4 (*Cnx3*) translation products.** The positive controls used are psaD and superoxide dismutase 3 (*SOD3*) translation products, respectively localized in the chloroplast and in the mitochondria. *A* shows the SDS-PAGE fluorography of the chloroplast import experiment, and *B* shows the SDS-PAGE fluorography of the mitochondrial import experiment. The *t lanes* are loaded with 1 μl of the translation mix, the *c* and *m lanes* are loaded with import reaction (corresponding to one-sixth of the total import reaction). The *p lanes* are loaded with protease-treated import reaction (corresponding to one-sixth of the total import reaction).

thesis in plants. Very little information is available on this pathway in any eukaryote. Cloning the plant cDNAs by functional complementation of *E. coli* mutants indicates that some steps of the MoCo biosynthesis are identical in prokaryotes and plants. With the gene sequences from *Arabidopsis* and *E. coli*, it might now be possible, by a PCR approach, to clone the human genes to obtain diagnostic tools for MoCo deficiency in humans. It would be interesting to see if the conservation between these early steps of MoCo biosynthesis is true for all eukaryotes.

Recently a gene involved in MoCo biosynthesis has been cloned from *Drosophila*. This gene appears to encode a multifunctional protein because different regions show homology to three different MoCo genes from *E. coli* (13). A similar gene has been identified in plants.[3] *Cnx2* and *Cnx3* appear to encode monofunctional proteins, although we cannot exclude the possibility that the N-terminal extensions could encode other functions. The N termini do not show homology with the sequenced *E. coli* MoCo genes or any other gene upon a search in the GenBank data base. In the purine biosynthetic pathway, we find monofunctional enzymes in *E. coli* and in plants, while the homologous enzymes are multifuntional in other eukaryotes (29). It remains to be shown if MoCo biosynthesis in plants is dominated by monofunctional genes as in *E. coli*.

We have tried to complement two other *E. coli* MoCo mutants, affected in the *moa* locus gene *D* and *E*, respectively. We failed to complement these mutations with plant cDNAs. The reason for the unsuccessful complementation could be the fact that these proteins, encoded by *moaD* and *moaE*, are part of a heteromeric complex (17). This means that the proteins not only need the proper enzymatic function to complement, they also need a specific structure to fit into the *E. coli* complex.

The *Cnx2* and *Cnx3* products probably have the same function as the *E. coli* proteins encoded by *moaA* and *moaC*, which are suggested to be responsible for molybdopterin precursor production (16, 17). The precursor is a 6-alkyl pterin with a 4-carbon phosphorylated side chain. It has been shown that the

[3] R. Mendel, personal communication.

phosphate is bound in diester linkage to C-2′ and C-4′ of the side chain to form a six-membered ring and that the precursor does not contain any of the sulfurs present in the mature MoCo (37, 38). It is believed that molybdopterin synthesis is different from the synthesis of other known pterins like folic acid and tetrahydrobiopterin, since their precursor, dihydroneopterin triphosphate, has a 3-carbon alkyl side chain (37). The substrate for the molybdopterin precursor synthesis is not yet defined, but studies with labeled guanosine in *E. coli* indicate that the substrate may be a derivative of GTP (2).

Some of the sequence similarities that were established for the MoaA peptide (15) can also be found for the Cnx2 peptide. Cnx2 has regions of similarity with guanine-binding proteins. Moreover, similarity can be established between Cnx2 and a protein in PQQ synthesis, a cofactor whose structure resembles a pterin moeity. The homology of Cnx2 to NifB and FixZ proteins was not expected as these proteins are involved in early steps of the iron-molybdenum cofactor biosynthesis:; the cofactor of dinitrogenase. The two cofactors have different structures; the iron-molybdenum cofactor is homocitrate-based, while the MoCo is a pterin derivative. The similarity that was established between the *moaA* gene product from *E. coli* and the molybdoenzyme formate dehydrogenase-H (15) cannot be found to the plant *Cnx2* gene product.

An interesting feature of the two plant proteins is their N-terminal extension when compared to the *E. coli* homologs. They have some of the characteristics of transit peptides for the chloroplasts and mitochondria. The fact that guanine, a likely precursor in MoCo biosynthesis, probably is made in the plastids (39), supports a possible plastid localization of early steps of MoCo biosynthesis in plants. Although there are differences between targeting signals for chloroplast and mitochondrial proteins (35), these are not obvious and it was not possible for us to predict if the putative transit peptides are for the chloroplasts or the mitochondria. We decided therefore to do import studies with both organelles. The inabilty to demonstrate import of these peptides to chloroplasts or mitochondria in our experiments are not necessarily conclusive. The Cnx2 and Cnx3 proteins might need factors for importation that are not present in standard *in vitro* import experiments. The *Cnx3* clone might not contain all the signals for import and processing as it was not complete in length. Other possible roles for the N-terminal extensions should not be ignored; they could, for example, encode other enzyme activities so the genes simply encode bifunctional enzymes as mentioned previously. The N-terminal extensions could also be needed for specific plant structural features of these enzymes.

Expression analysis of these genes show that the *Cnx2* gene is expressed in all organs and with the strongest expression in roots and cell suspension culture. *Cnx3* is not strongly expressed in any tissue except in roots and in suspension culture cells. It is surprising that we detect nearly no *Cnx3* transcript in leaves. This observation suggests that MoCo biosynthesis has no direct links with photosynthesis, this is supported by the fact that both genes are expressed in etiolated seedlings (results not shown). The plant molybdo-enzyme nitrate reductase is active in leaves, the major site of nitrate reduction. The low expression of *Cnx2* and *Cnx3* in leaves suggest that nitrate reductase is probably not a major sink for MoCo utilization. It remains to be understood why there is a high level of expression of both genes in roots. This may reflect a MoCo requirement for another molybdo-enzyme located in the roots. For both genes there is no detectable difference in the mRNA level from plants grown with nitrate or with ammonium. These experiments were done to investigate if the mRNA levels of these genes are regulated by some of the same factors as nitrate reductase. The fact that other key enzymes in the plant need MoCo as well may explain why the nitrogen source of the plant have no influence on the two genes' expression level.

Further studies of these genes will include complementation experiments to determine which of the six complementation groups of *N. plumbaginifolia*, if any, the *Cnx2* and *Cnx3* genes correspond to. It will also be interesting to know were in the cell these early steps of MoCo biosynthesis take place. Therefore we intend to use other approaches to establish the location of the two proteins: by immunolocalization studies and by analyzing transgenic plants containing fusions of a reporter gene with the putative targeting signals.

Acknowledgments—We thank G. Giordano and C. Iobbi-Nivol (CNRS, Marseille, France) for providing us with the *E. coli* mutants and for advice about selection conditions. From our laboratory we thank B. Courtial and D. Bouchez for help with the mapping, J. Drouaud for running our sequences, and M. Laloue for encouraging us to do these experiments. We also thank V. Skovgaard Nielsen (Agricultural University, Copenhagen) for advice about the chloroplast experiments.

REFERENCES

1. Coughlan, M. P. (ed) (1980) *Molybdenum and Molybdenum-containing Enzymes*, Pergamon Press, Oxford
2. Rajagopalan, K. V., and Johnson, J. L. (1992) *J. Biol. Chem.* **267,** 10199–10202
3. Kramer, S. P., Johnson, J. L., Ribeiro, A. A., Millington, D. S., and Rajagopalan, K. V. (1987) *J. Biol. Chem.* **262,** 16357–16363
4. Wray, J. L., and Kinghorn, J. R. (eds) (1989) *Molecular and Genetic Aspects of Nitrate Assimilation*, Oxford Science Publications, Oxford
5. Hoff, T., Truong, H. N., and Caboche, M. (1994) *Plant Cell Environ.* **17,** 489–506
6. Kleinhofs, A., and Warner, R. L. (1990) in *The Biochemistry of Plants* (Miflin, B. J., and Lea, P. J., eds) pp. 89–120, Academic Press, San Diego
7. Mendel, R. R., and Müller, A. J. (1985) *Planta* **163,** 370–375
8. Gabard, J., Pelsy, F., Marion-Poll, A., Caboche, M., Sallbach, I., Grafe, R., and Müller, A. J. (1988) *Mol. & Gen. Genet.* **213,** 206–213
9. Pelsy, F., and Caboche, M. (1992) *Adv. Genet.* **30,** 1–40
10. Walker-Simmons, M., Kudrna, D. A., and Warner, R. L. (1989) *Plant Physiol.* **90,** 728–733
11. Johnson, J. L., Wuebbens, M. M., Mandell, R., and Shih, V. E. (1989) *J. Clin. Invest.* **83,** 897–903
12. Warner, R. C., and Finnerty, V. (1981) *Mol. & Gen. Genet.* **184,** 92–96
13. Kamdar, P. K., Shelton, M. E., and Finnerty, V. (1994) *Genetics* **137,** 791–801
14. Nohno, T., Kasai, Y., and Saito, T. (1988) *J. Bacteriol.* **170,** 4097–4102
15. Rivers, S. L., Mcnairn, E., Blasco, F., Giordano, G., and Boxer, D. H. (1993) *Mol. Microbiol.* **8,** 1071–1081
16. Pitterle, D. M., and Rajagopalan, K. V. (1989) *J. Bacteriol.* **171,** 3373–3378
17. Pitterle, D. M., and Rajagopalan, K. V. (1993) *J. Biol. Chem.* **268,** 13499–13505
18. Pitterle, D. M., Johnson, J. L., and Rajagopalan, K. V. (1993) *J. Biol. Chem.* **268,** 13506–13509
19. Estelle, M. A., and Sommerville, C. (1987) *Mol. & Gen. Genet.* **206,** 200–206
20. Elledge, J. E., Mulligan, J. T., Ramer, S. W., Spottswood, M., and Davis, R. W. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 1731–1735
21. Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) *Nucleic Acids Res.* **16,** 6127–6145
22. MacGregor, C. H., Schnaitman, C. A., Normansell, D. E., and Hodgins, M. G. (1974) *J. Biol. Chem.* **249,** 5321–5327
23. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
24. Sanger, F., Niklen, S., and Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
25. Devereaux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395
26. Verwoerd, T. C., Dekker, B. M. M., and Hoekema, A. (1989) *Nucleic Acids Res.* **17,** 2362
27. Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983) *Plant Mol. Biol. Rep.* **1,** 19–21
28. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13
29. Schnorr, K. M., Nygaard, P., and Laloue, M. (1994) *Plant J.* **6,** 113–121
30. Cline, K., Werner-Washburne, M., Lubben, T. H., and Keegstra, K. (1985) *J. Biol. Chem.* **260,** 3691–3696
31. White, J. A., and Scandalios, J. G. (1987) *Biochim. Biophys. Acta* **926,** 16–25
32. Laemmli, U. K. (1970) *Nature* **227,** 680–685
33. Skinner, M. K., and Griswold, M. D. (1983) *Biochem. J.* **209,** 281–284
34. Caboche, M., Campbell, W., Crawford, N. M., Fernandez, E., Kleinhofs, A., Ida, S., Mendel, R., Omata, T., Rothstein, S., and Wray, J. (1994) *Plant Mol. Biol. Rep.* **12,** S45–S49
35. von Heijne, G., Steppuhn, J., and Herrmann, R. G. (1989) *Eur. J. Biochem.* **180,** 535–545
36. von Heijne, G., and Nishikawa, K. (1991) *FEBS Lett.* **278,** 1–3
37. Johnson, J. L., Wuebbens, M. M., and Rajagopalan, K. V. (1989) *J. Biol. Chem.* **264,** 13440–13447
38. Wuebbens, M. M., and Rajagopalan, K. V. (1993) *J. Biol. Chem.* **268,** 13493–13498
39. Senecoff, J. F., and Meagher, R. B. (1993) *Plant Physiol.* **102,** 387–399