Biochem. J. (1991) 278, 203–209 (Printed in Great Britain)

Expression of a cDNA clone encoding the haem-binding domain of *Chlorella* nitrate reductase

Andrew C. CANNONS, Naoko IIDA and Larry P. SOLOMONSON
Department of Biochemistry and Molecular Biology, USF College of Medicine, 12901 Bruce B. Downs Boulevard, Tampa, FL 33612, U.S.A.

A partial cDNA clone coding for the haem-binding domain of NADH:nitrate reductase (EC 1.6.6.1) (NR) from the unicellular green alga *Chlorella vulgaris* has been isolated, sequenced and expressed. A 1.2 kb cDNA (pCVNR1) was isolated from a λgt11 expression library produced from polyadenylated RNA extracted from nitrate-grown *Chlorella* cells. pCVNR1 hybridized to a 3.5 kb mRNA transcript that was nitrate-inducible and absent from ammonium-grown cells. The entire sequence of pCVNR1 was obtained and found to have a single uninterrupted reading frame. The derived amino acid sequence of 318 amino acids has a 45–50% similarity to higher-plant NRs, including *Arabidopsis thaliana*, spinach (*Spinacia oleracea*) and tobacco (*Nicotiana tabacum*). A comparison with the putative domain structure of higher-plant nitrate reductases suggested that this sequence contains the complete haem-binding domain, approximately one-third of the Mo-pterin domain and no FAD-binding domain. A 32% sequence similarity is evident when comparing the *Chlorella* NR haem domain with that of calf cytochrome b_5. Expression of pCVNR1 in a pET vector synthesized a 35 kDa protein that was antigenic to anti-(*Chlorella* NR) antibody. The spectral properties of this protein (reduced and oxidized) in the 400–600 nm region are identical with those of native *Chlorella* NR and indicate that haem is associated with the protein.

INTRODUCTION

Higher plants, algae, filamentous fungi, yeasts and bacteria can all assimilate nitrate as a form of inorganic nitrogen [1–3]. Biological conversion of nitrate into ammonium ion is an eight-electron reduction process requiring two enzymic steps: the reduction of nitrate to nitrite by nitrate reductase (NR), followed by the reduction of nitrite to ammonium ion via nitrite reductase. Assimilatory NR is considered to be the regulated step in the process of nitrate assimilation [4] and consequently is important for growth and development of all nitrate-assimilating organisms. The biological importance of NR has prompted a wealth of studies aimed at understanding its catalytic properties and regulation. In eukaryotes, NR is a multi-centre protein containing the prosthetic groups molybdopterin, FAD and haem in a 1:1:1 stoichiometry [5] that pass electrons from NAD(P)H to nitrate. The regulation of NR is complex, since a number of factors, including light and nitrogen source, can substantially influence the level of NR expression. The potential for understanding the regulation of NR at the molecular level has improved recently with the isolation of NR cDNAs and genes from a variety of species, including barley (*Hordeum vulgare*) [6], squash (*Cucurbita maxima*) [7], *Arabidopsis thaliana* [8], tobacco (*Nicotiana tabacum*) [9], the fungus *Aspergillus* [10] and the alga *Chlamydomonas* [11]. Generally, ammonium ion is the repressor of NR activity and nitrate addition and/or removal of ammonium ion is required for expression [3]. Such changes in NR levels appears to primarily be at the level of transcription [3,6,8]. Sequencing of these clones has revealed a high degree of similarity at the amino acid level for NRs [10] and has identified the functional domains associated with the prosthetic groups as well as the hinge regions that link the domains together [12].

The NR from *Chlorella* has been intensively studied, and the biophysical characteristics of this enzyme have been well analysed, such that the *Chlorella* NR could be considered a 'model' for NRs and indeed molybdo-enzymes. The midpoint potentials of the prosthetic groups have been determined [13,14], as well as spectroscopic and kinetic properties [5]. Various approaches have been employed to obtain information on the individual domains of NR, including limited proteolysis [15], radiation-inactivation analysis [16] and specific inhibitors [5]. Although the NADH/FAD-binding domain can be isolated from the other domains after limited proteolysis, this approach has not been successful for the separation of the haem- or molybdenum-cofactor (Mo-Co)-binding domains. Towards this end we report the successful isolation and expression of a cDNA clone for *Chlorella* NR that maps to the haem-binding domain of NR, compare the nucleotide and deduced amino acid sequences with those of other NRs, as well as with cytochrome b_5, and compare the spectral properties of the expressed protein with those of native NR.

EXPERIMENTAL

Growth of cells

Chlorella vulgaris cells were grown as previously described [17], using 20 mM-KNO_3 or 20 mM-NH_4Cl as the nitrogen source. *Escherichia coli* strains were maintained and grown on Luria broth (LB) [18] unless otherwise described.

RNA extraction and purification

RNA was extracted from *Chlorella* cells using the method of Bascomb & Schmidt. [19], except that the cells were disrupted using a French pressure cell (Aminco) at 104 MPa (15000 lbf/in^2). Polyadenylated [poly(A)$^+$] RNA was isolated as described by Aviv & Leder [20], using oligo(dT)–cellulose.

Electrophoresis and hybridization

Equal quantities of poly(A)$^+$ RNA (5 μg) were electrophoresed on Mops/formaldehyde gels and transferred to nitrocellulose [18]. The blots were hybridized with pCVNR1 insert, random-prime-labelled (Boehringer-Mannheim) with ^{32}P, for 16 h at

Abbreviations used: NR, nitrate reductase; IPTG, isopropyl thio-β-D-galactoside; Mo-Co, molybdenum cofactor; PMSF, phenylmethanesulphonyl fluoride; poly(A)$^+$, polyadenylated; LB, Luria broth; p.f.u., plaque-forming unit.

These sequence data have been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X56771.

45 °C in 50 % (v/v) formamide then washed in 2 × SSC (20 × SSC is 3 M-NaCl/300 mM-sodium citrate, pH 7.0) at 55 °C for 30 min and in 0.1 × SSC/0.1 % SDS at 65 °C for 1 h.

cDNA synthesis, library construction and screening

Poly(A)$^+$ RNA was used as a template to synthesize oligo(dT)-primed cDNA by the ribonuclease H method [21]. The double-stranded DNA was blunt-ended with T4 DNA polymerase [18], followed by the addition of *Eco*R1/*Not*1 linkers (Invitrogen, San Diego, CA, U.S.A.). Linkered cDNA was size-fractionated on a 1 %-agarose gel, and cDNA > 1 kb and < 5kb was isolated. This step also removed unincorporated linkers. Size-fractionated linkered cDNA was ligated into λgt11, packaged and used to infect *E. coli* Y1090 cells [18]. The amplified library was screened, as described by Mierendorf *et al.* [22], with using antibodies to purified *Chlorella* NR [17] pretreated with *E. coli* extract [22], and recombinant protein was detected on nitrocellulose plaque lifts using goat anti-rabbit IgG·alkaline phosphatase conjugate as described by Mierendorf *et al.* [22].

PCR of cDNA

PCR was performed on the synthesized cDNA to ensure NR cDNAs had been made. PCR was done using *Taq* polymerase according to the manufacturer's (Cetus/Perkin–Elmer) instructions. Template DNA (10 ng) was denatured at 94 °C for 2 min, followed by annealing at 37 °C for 2 min and chain extension at 72 °C for 2 min. In the next cycle the annealing temperature was raised to 42 °C, and in the third and subsequent 32 cycles the annealing temperature was 50 °C. The primer concentration was 1.0 μM.

Bacteriophage DNA isolation and subcloning

DNA from a pure λgt11 clone was isolated by the method of Steffens & Gross [23]. The *Eco*R1 insert was subcloned into the *Eco*R1 site of the Bluescript II KS vector (Stratagene, La Jolla, CA, U.S.A.) by using standard techniques [18].

Sequencing

The 5′ and 3′ ends of the λgt11 clone were sequenced using oligonucleotides constructed to the forward and reverse *Eco*R1 sites of λgt11 as primers [23] in the dideoxy-chain-termination sequencing method [24]. The full sequence of the clone was obtained from sequencing single-stranded DNA after subcloning the insert into M13mp18 and 19 and using Sequenase ver 2.0 (U.S. Biochemicals, Cleveland, OH, U.S.A.) according to the manufacturer's recommendations. Both DNA strands were sequenced, and nucleotide data were analysed using Genepro ver 4.0 (Riverside Scientific, Seattle, WA, U.S.A.).

Expression of cDNA in *E. coli*

The *Eco*R1 insert was ligated into the *Eco*R1 site of the pET 5b expression vector [25] kindly provided by Dr. F. W. Studier (see Fig. 6 below). Ligated DNA was transformed into competent HMS174 *E. coli* and clones containing plasmids were isolated. Plasmids containing the insert in the correct orientation were then transformed into B121(DE3) *E. coli* for expression. To express the cDNA, transformed cells were inoculated into TBM9 medium [25] and grown overnight at 37 °C. The overnight culture was used to inoculate TBM9 medium and the culture grown to an A_{600} of 0.6. At this point isopropyl thio-β-D-galactoside (IPTG) was added to a concentration of 0.4 mM and the cells grown further at 37 °C. Production of the protein was measured by taking 30 μl samples at timed intervals, electrophoresing the samples on an SDS/10 %-(w/v)-polyacrylamide gel [26] and blotting the gel on to nitrocellulose by means of an ABN polyblot electroblotter (American Bionetics, Hayword, CA, U.S.A.). The blot was then probed with a polyclonal antibody to *Chlorella* NR as described previously [17] using goat anti-rabbit IgG·alkaline phosphatase conjugate as the second antibody.

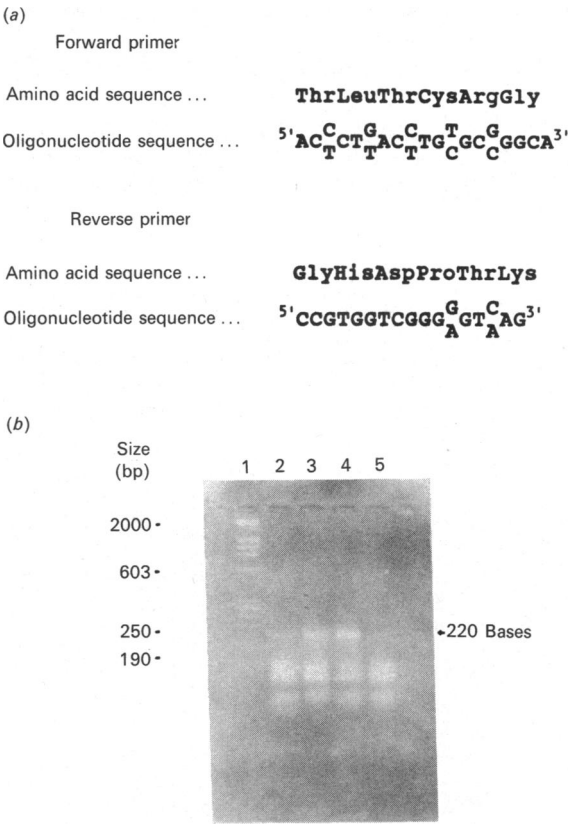

Fig. 1. Detection of NR cDNA using PCR

cDNA was made from poly(A)$^+$ RNA extracted from *Chlorella* grown on nitrate or ammonium ion as a sole nitrogen source. These cDNAs were used as a template for specific DNA amplification by PCR. Primers were designed from limited amino-acid-sequence data [5]. Products of the amplification were run on a 1.5 %-agarose gel. (*a*) Oligonucleotide constructs. Oligonucleotides were synthesized on the basis of the amino acid sequences available for *Chlorella* NR. (*b*) Agarose gel of PCR reactions. Lane 1, molecular-mass markers; lane 2, amplified products of cDNA synthesized from poly(A)$^+$ extracted from ammonium-ion-grown cells; lanes 3 and 4, amplified products of cDNA transcribed from poly(A)$^+$ RNA extracted from cells grown on nitrate; lane 5, amplified products using λ-DNA as a template. All samples were amplified under the same conditions using the same amount of DNA as template, and equal volumes (10 μl) were loaded on each lane.

Haem analysis

Expressed protein, antigenic to anti-*Chlorella* NR, was partially purified from 1 litre of expressing cells. The cells were harvested, washed in 10 mM-potassium phosphate, pH 7.5, and resuspended in buffer A [10 mM-potassium phosphate, pH 7.5, containing 1 mM-EDTA, 0.1 % Triton X-100 and 1 mM-phenylmethanesulphonyl fluoride (PMSF)] at 1 g/5 ml. The extract was passed through a French pressure cell at 69 MPa (10000 lbf/in^2), and the homogenate was clarified by centrifugation at 10000 *g* for 10 min. The supernatant was subjected to 50 %-$(NH_4)_2SO_4$ fractionation, and the final pellet was resuspended in 1 ml of buffer B (10 mM-potassium phosphate, pH 7.5, containing 1 mM-EDTA and 1 mM-PMSF). Partial purification of the protein was achieved on a Pharmacia f.p.l.c.

system using a Mono Q column and an NaCl gradient of 0 to 0.5 M. Elution of the expressed protein from the column was monitored by measuring the A_{280}/A_{413} ratio, and purity was determined by SDS/PAGE and blotting of samples as described above. Spectroscopic analysis of the oxidized and reduced forms of the haem associated with the expressed protein was as previously described [15], a Beckman DU70 spectrophotometer being used.

RESULTS

Isolation and characterization of a *Chlorella* NR cDNA

A cDNA expression library was constructed from cDNA synthesized from poly(A)$^+$ RNA extracted from cells utilizing KNO_3 as a sole nitrogen source. This cDNA population was shown to contain cDNA to the NR message by employing the PCR. Degenerate primers, shown in Fig. 1(*a*), were designed on the basis of limited amino-acid-sequence data of *Chlorella* NR [5] and the codon usage of *Chlamydomonas* [27]. An amplified product of approx. 220 bases would be predicted from alignment of the *Chlorella* NR amino acid sequences with the full *Arabidopsis* sequence [8]. As shown in Fig. 1(*b*), little or no amplification of a 220-base product was detected using cDNA made from mRNA extracted from ammonium-grown (NR-repressed) cells or λ DNA (Fig. 1*b*, lanes 2 and 5). Amplification of the 220-base product was only observed with cDNA made from mRNA extracted from nitrate-grown (NR-induced) cells (Fig. 1*b*, lanes 3 and 4).

The cDNA synthesized from mRNA extracted from cells grown on nitrate was used to construct a λgt11 expression library. The amplified library produced was in excess of 10^{10} plaque-forming units (p.f.u.)/μg of cDNA. Screening of 1 million independent clones with anti-(*Chlorella* NR) antibody identified one clone, which was subsequently isolated and designated 'pCVNR1'. This clone was approx. 1.2 kb in length, indicating it was not full length (result not shown).

Northern-blot analysis of mRNA blots using ^{32}P-labelled pCVNR1 showed that a transcript of 3.5 kb was recognized by pCVNR1. This transcript, of the expected size for a NR mRNA, is specific to cells using nitrate and absent from cells utilizing ammonium (Fig. 2).

Sequencing of pCVNR1 and comparison with other NR clones

The sequence of pCVNR1 is shown in Fig. 3. This partial cDNA is 954 nucleotides long, has a G/C content of 62% and contains no poly(A)$^+$ sequences, suggesting an internal clone has been isolated. The codon usage suggests a similarity to *Chlamydomonas* rather than to higher plants. The single uninterrupted reading frame would encode a polypeptide of 318 amino acids with a molecular mass of 35 kDa, which would represent approximately one-third of the full NR subunit.

Comparison of the open reading frame of pCVNR1 with the predicted amino acid sequences of tomato (*Lycopersicon*) and *Arabidopsis* NRs, shown in Fig. 4, reveals an overall sequence similarity of approx. 46%. The sequence lines up with an internal portion of higher-plant NR, beginning at amino acid 316 and ending at position 646. There are some amino acid deletions in the sequence of *Chlorella* NR compared with higher-plant NR, notably at residue numbers 383–387 and 536–541 (*Arabidopsis* numbering). The polypeptide encoded by pCVNR1 would correspond to the haem domain of NR located between positions 540 and 620 of *Arabidopsis* NR.

A significant degree of similarity (32%) was also observed between the haem region of calf cytochrome b_5 and amino acids 214–306 of the pCVNR1-derived sequence (Fig. 5). The invariant residues suggested by Matthews [28] were also present in the *Chlorella* NR sequence (shaded residues). Regions of similarity between the *Chlorella* NR sequence and residue numbers 25–30, 37–42, 53–60, 62–66 and 76–80 of cytochrome b_5 correspond

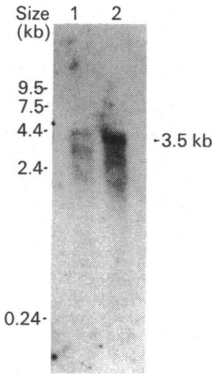

Fig. 2. Analysis of NR mRNA levels in *Chlorella* cells

Northern-blot analysis using ^{32}P-labelled pCVNR1 as a probe was carried out as described in the Experimental section, using 5 μg of oligo(dT)-selected RNA loaded per lane. Lane 1, RNA extracted from cells grown on ammonium ion as a sole nitrogen source; lane 2, RNA extracted from cells grown on nitrate as a sole nitrogen source.

```
CGG CCG CCG GTC ACC GAG GTG GAG AGC CAG AAC TAC TAC CAC TTC      45
R   P   P   V   T   E   V   E   S   Q   N   Y   Y   H   F

CAC GAC AAC AGA GTG CTG CCC TCG CAC GTT GAT GAG GCG CTG GCC      90
H   D   N   R   V   L   P   S   H   V   D   E   A   L   A

AAC TCC GAA GGC TGG TGG TAC AAG CCG GAC TTC ATC ATC AAT GAC     135
N   S   E   G   W   W   Y   K   P   D   F   I   I   N   D

CTC AAT GTC CAG TCA GCC ATA GGC TAC CCG GCA CAT GAA GAG GTG     180
L   N   V   Q   S   A   I   G   Y   P   A   H   E   E   V

GTG CCC CTC GTG GCC GGC ACC TAC GCC GTA CGA GGC TAC GCG CGC     225
V   P   L   V   A   G   T   Y   A   V   R   G   Y   A   R

GGG CAC GGC AAC AAG ATC ATT CGG TGC GAG GTT TCG CTG GAC GAC     270
G   H   G   N   K   I   I   R   C   E   V   S   L   D   D

GGC AAG AGC TGG CGG CTG GGC TCT GTG ACG CAC GAG GGG CAA CCT     315
G   K   S   W   R   L   G   S   V   T   H   E   G   Q   P

ACT GAG TAC GGC AAG CAC TGG GGC TGG GTG TGG TGG AGC TTG GAG     360
T   E   Y   G   K   H   W   G   W   V   W   W   S   L   E

GTG CCC ATT GCT GAG CTG CTT ACC ACT CCT GAG ATC ATT TGC CGA     405
V   P   I   A   E   L   L   T   T   P   E   I   I   C   R

GCC TGG GAC AGC TCC ATG AAC ACA CAG CCC AAC ACG TTC ACC TGG     450
A   W   D   S   S   M   N   T   Q   P   N   T   F   T   W

AAC GTG ATG GGG ATG ATG AAC AAC TGC TGC TAC CGC GTC AAG ATC     495
N   V   M   G   M   M   N   N   C   C   Y   R   V   K   I

CAC CCC CGC CAG ACC ACC GAC GGC CGC TTT GCG CTG CAG TTT GAG     540
H   P   R   Q   T   T   D   G   R   F   A   L   Q   F   E

CAC CCC ACC ATT GCC GGC CCC ACT GTC GGC GGC TGG ATG AAC CGA     585
H   P   T   I   A   G   P   T   V   G   G   W   M   N   R

GCA GAG GAC GTG GCG GCG GCG GCA GCA GTG ACG GTG GCG CCG CCA     630
A   E   D   V   A   A   A   A   A   V   T   V   A   P   P

CCC GCG CCC GCA GGT GCC AAG AGC TTC ACC ATG GCA GAA GTG GAG     675
P   A   P   A   G   A   K   S   F   T   M   A   E   V   E

ACG CAC ACG ACC ATG GAG AGC GCG TGG TTT GTT GTT GAT GGA AAG     720
T   H   T   T   M   E   S   A   W   F   V   V   D   G   K

GTG TAC GAC GCA ACA CCC TTC CTG AAG GAC CAC CCG GGT GGC GCC     765
V   Y   D   A   T   P   F   L   K   D   H   P   G   G   A

GAC TCG ATC CTG CTT GTG GCT GGC ATA GAC GCC ACT GAC GAG TTT     810
D   S   I   L   L   V   A   G   I   D   A   T   D   E   F

AAT GCC ATC CAC TCG CTC AAA GCC AAG AAG CAG TTG CTG GAG TAC     855
N   A   I   H   S   L   K   A   K   K   Q   L   L   E   Y

TAC ATT GGA GAG CTG GCG GAA GAG GGG CAG GAG GCA GCA GCG AGC     900
Y   I   G   E   L   A   E   E   G   Q   E   A   A   A   S

GAC CGG GCG ACC CCG GGG CCA GCG GCG GCA ATC GGC ACG GCG GTG     945
D   R   A   T   P   G   P   A   A   A   I   G   T   A   V

CCA GTT GCC                                                    954
P   V   A
```

Fig. 3. Complete nucleotide sequence (upper line) and predicted amino acid sequence (lower line) of pCVNR1

Nucleotide numbers are given on the right of the sequence.

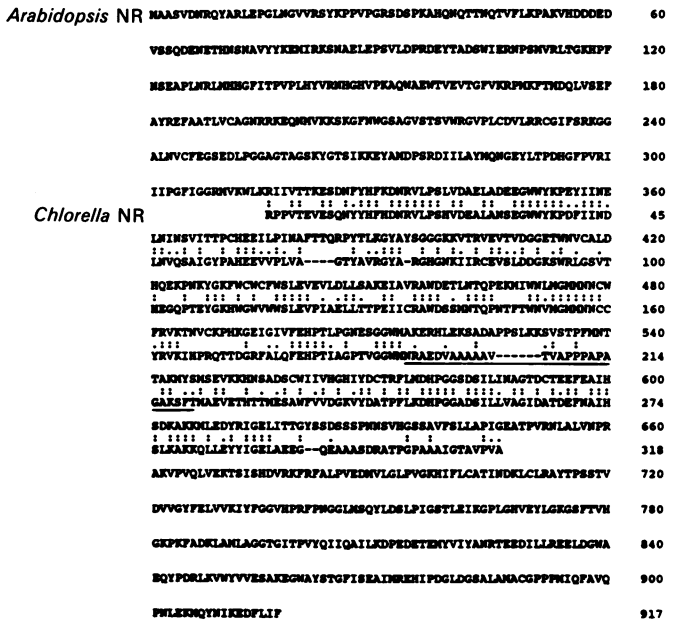

Fig. 4. Alignment of the predicted amino acid sequence of pCVNR1 (lower line) with the sequence of *Arabidopsis* NR (upper line)

Identity between residues is shown by a colon (:). The single dot (.) indicates a conservative substitution. Sequences underlined have been verified by partial amino acid sequencing of *Chlorella* NR fragments after digestion with CNBr.

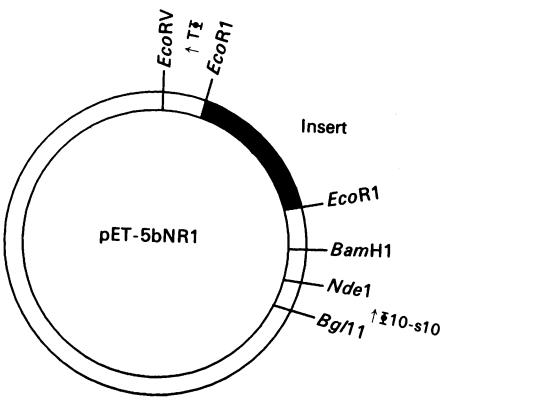

Fig. 6. Construction of plasmid for expressing the haem-binding domain of NR

*Eco*R1-digested insert of pCVNR1 was ligated into the *Eco*R1 site of the expression plasmid pET-5b. The construct was transformed into competent HMS174 *E. coli* cells and expressing clones were further transformed into Bl21 *E. coli* from expression studies. '↑T' is the termination signal, and '↑10-s10' indicates the promoter plus initiation region for gene-10 protein.

to zones of α-helices or β-sheets important in the structural arrangement of the b_5 crevice [28]. Histidine residues 39 and 63, determined as the iron ligands [28], are located in these areas of sequence identity.

Expression of pCVNR1 in *E. coli*

pCVNR1 was inserted into the pET5b vector, which contains the necessary elements for transcription initiation and termination and translation of cDNA inserts (Fig. 6). The IPTG induction of B121 *E. coli* cells transformed with pET5b-NR initiated the synthesis of a 35 kDa protein, as evidenced by SDS/PAGE (Fig. 7*a*, lanes 11–15). Induction of non-transformed cells (Fig. 7*a*, lanes 1–5) or cells transformed with the NR clone in the reverse orientation (Fig. 7*a*, lanes 6–10), did not result in the production of a 35 kDa protein. This expressed protein was shown to be a part of nitrate reductase by probing a Western blot of the gel with a specific polyclonal antibody to *Chlorella* NR. The 35 kDa protein was recognized by anti-(*Chlorella* NR) antibody, whereas no cross-reacting material was present in the non-transformed cells or cells transformed with the reverse-oriented cDNA (Fig. 7*b*).

The expressed product of pET5b-NR contains haem

Induction of pET5b-NR cells with IPTG resulted in the production of cells exhibiting a distinctive pink colour (results not shown). This suggests that the expressed protein was also binding haem. Partial purification of the protein using f.p.l.c. (Mono Q) resulted in a protein preparation with an A_{280}/A_{413} ratio of < 4. Since this ratio is 1.8 for pure *Chlorella* NR, then a value of less than 1.8 would be expected for the pure haem-binding domain. From SDS/PAGE we estimate the level of expression to be in the range of 5% of the total cell protein. A haem scan of this protein revealed a spectrum typical of cytochrome b_5 (Fig. 8*b*) that is essentially identical with that of authentic *Chlorella* NR (Fig. 8*a*). An oxidized Soret band at 413 nm is evident that can be shifted to 423 nm after reduction with sodium dithionite. Reduction of the haem also resulted in the production of the typical α and β bands at 527 nm and 557 nm.

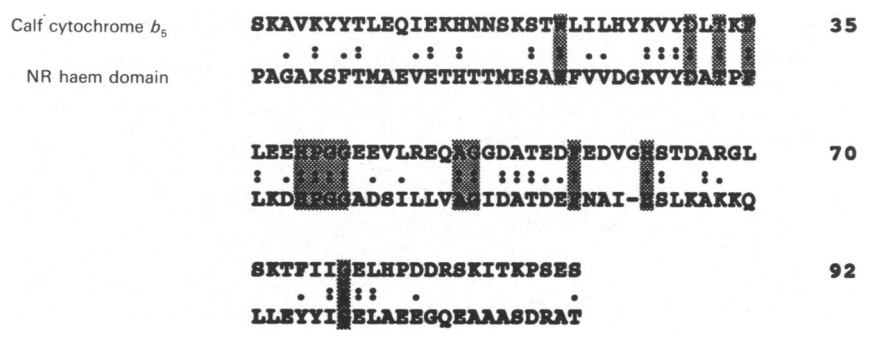

Fig. 5. Sequence comparison of the haem-binding domain

The amino acid sequence of *Chlorella* NR haem-binding region was aligned with the sequence of calf cytochrome b_5 to maximize similarities. Identical amino acid positions are identified (colon), as well as the invariant residues (shaded) determined for the b_5-type fold. The single dot indicates a conservative substitution.

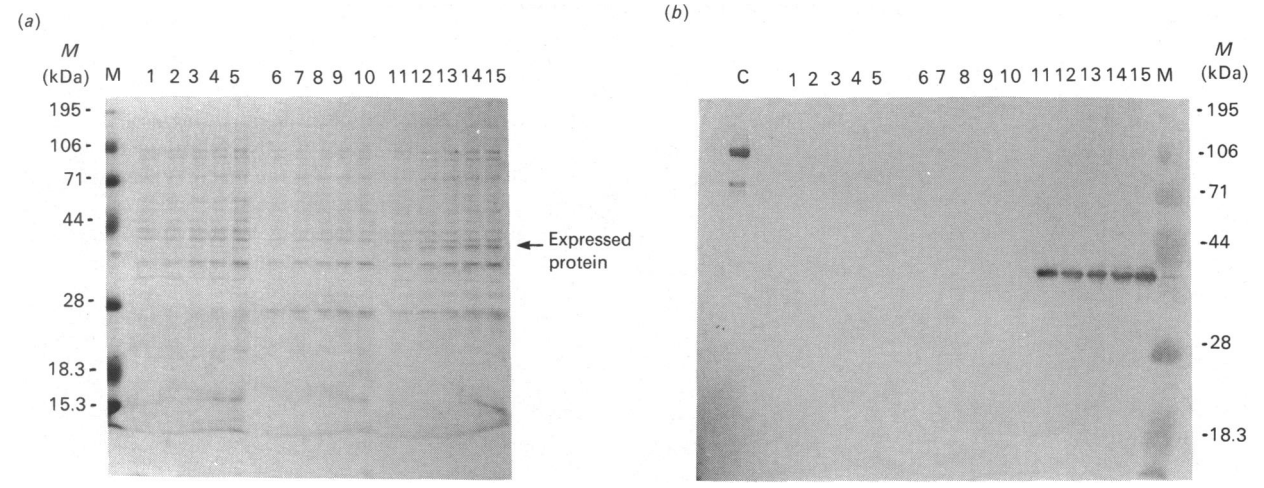

Fig. 7. Expression of pET-5b-NR1

Clones expressing NR cDNA as a protein were identified by SDS/PAGE. Transformed cells were grown to an A_{600} of 0.6 and then induced with 0.4 mM-IPTG. Samples (30 μl) were taken at 0, 15, 30, 60 and 120 min after addition of IPTG, then electrophoresed on an SDS/12%-(w/v)-polyacrylamide gel. The gel was either stained for protein with Coomassie Blue or blotted on to nitrocellulose for immunoblotting (Western) with anti-(*Chlorella* NR) antibody. Lanes 1–5 are samples from non-transformed control cell taken at 0, 15, 30, 60 and 120 min respectively after the addition of IPTG. Lanes 6–10 are samples from cells transformed with the insert in the reverse orientation, and lanes 11–15 are samples from cells containing the pET5b-NR1 plasmid taken at the same respective time points as for the control cells. (*a*) Coomassie Blue stain of gel; the arrow indicates an expressed protein at 35 kDa induced by IPTG. (*b*) Western blot of similar gel to that shown in (*a*). The blot was hybridized with anti-(*Chlorella* NR) antibody as detailed in the Experimental section. Lane C refers to purified *Chlorella* NR, which has a size of 115 kDa. Abbreviations: *M*, molecular mass; M, molecular-mass markers.

DISCUSSION

The authenticity of a partial cDNA clone to *Chlorella* NR was established by RNA hybridization and sequence comparisons with other NRs. Northern-blot analysis showed that the clone, pCVNR1, hybridized to a single poly(A)$^+$ RNA transcript of 3.5 kb that is inducible by nitrate. This size of NR message is in agreement with previous results for NR mRNA from plants, fungi and algae [3,6–9,11].

pCVNR1 contained one single open reading frame encoding a polypeptide of 35 kDa. The codon usage suggests a similarity to that of *Chlamydomonas* rather than higher plants or chloroplast-encoded proteins in *Chlorella* [29]. Also, a strong preference for codons ending in C/G (80%) is evident, and this is characteristic of *Chlamydomonas* nuclear genes rather than *Chlorella* chloroplastic genes and higher-plant genes [29]. This may be why higher-plant cDNAs for NR (barley and squash) were not effective as probes in the isolation of a *Chlorella* NR cDNA (results not shown). A comparison of the predicted amino acid sequence of pCVNR1 has revealed a degree of similarity to the published sequences for higher-plant NRs [8–10]. The sequence is 46% identical with the *Arabidopsis* and tobacco sequences [8,9]. However, there are regions where this similarity approaches 100% and areas where there is little or no identity. Specific alignment of the amino acid sequence of pCVNR1 with higher-plant NR indicated that pCVNR1 codes for an internal portion of NR. The clone starts at amino acid 316 and reads through to amino acid 646 of *Arabidopsis* NR. Thus pCVNR1 encodes the entire haem-binding domain and a part (about one-third) of the molybdopterin-binding domain [8,12]. Between these two domains is the proposed hinge 1 region [12], where the sequence similarity with higher-plant NR is low. Some of the predicted amino acid sequence has been confirmed by amino acid sequencing of CNBr-digested *Chlorella* NR, including the alanine/proline-rich region located between the molybdopterin and haem-binding domains.

Chlorella NR contains a b_5-type cytochrome [5], and a comparison of the sequences of the haem-binding domain of *Chlorella* NR with calf cytochrome b_5 indicated 32% similarity. Comparable sequence similarity to cytochrome b_5 has been reported for other NRs, including those of *Arabidopsis* and *Neurospora crassa* [8,30]. All 13 invariant residues of the b_5-type fold [28] were present in *Chlorella* NR. One noticeable change is the deletion of glycine-62 (cytochrome b_5 numbering) just before the second axial histidine residue. This glycine residue is present in microsomal and mitochondrial b_5s [31], yet absent from all the NRs that have been sequenced [8–10]. Flavocytochrome b_2, another member of the b_5 superfamily, also lacks this glycine residue [32]. Glycine-62 in cytochrome b_5 introduces a tight bend in the backbone as a result of Lα helical dihedral angles. In flavocytochrome b_2, proline-64 serves the same function. Neither of these residues is conserved in NR. However, in the haem-binding domain of *Chlorella* NR, an asparagine residue preceding the second axial histidine is possibly responsible for this tight turn, since asparagine is considered a conservative replacement for glycine in roles that involve unusual conformations [33].

Isolation and expression of a cDNA that encoded the haem domain of NR permits the study of the biophysical characteristics of this domain as a separate entity. A construct was made of pCVNR1 in the transcription/translation expression vector pET5b, developed by Studier *et al.* [25]. The pET5b-NR1 plasmid that was engineered expressed a soluble protein that was antigenic to a polyclonal antibody directed against *Chlorella* NR. This protein was 35 kDa, the expected size as deduced from the open reading frame of the sequence of pCVNR1. Overexpression of this NR domain resulted in the cells turning pink. This suggested that overproduction of haem was occurring to supply an expressed protein requiring haem. Similar results have been reported for the expression of mammalian cytochrome b_5 and yeast flavocytochrome b_2 [34,35]. The expressed haem protein was shown to be a *b*-type cytochrome with spectral features apparently identical with those of native *Chlorella* NR [15]. No enzyme activity was associated with the protein, suggesting the lack of any other cofactors. No FAD-associated activity would

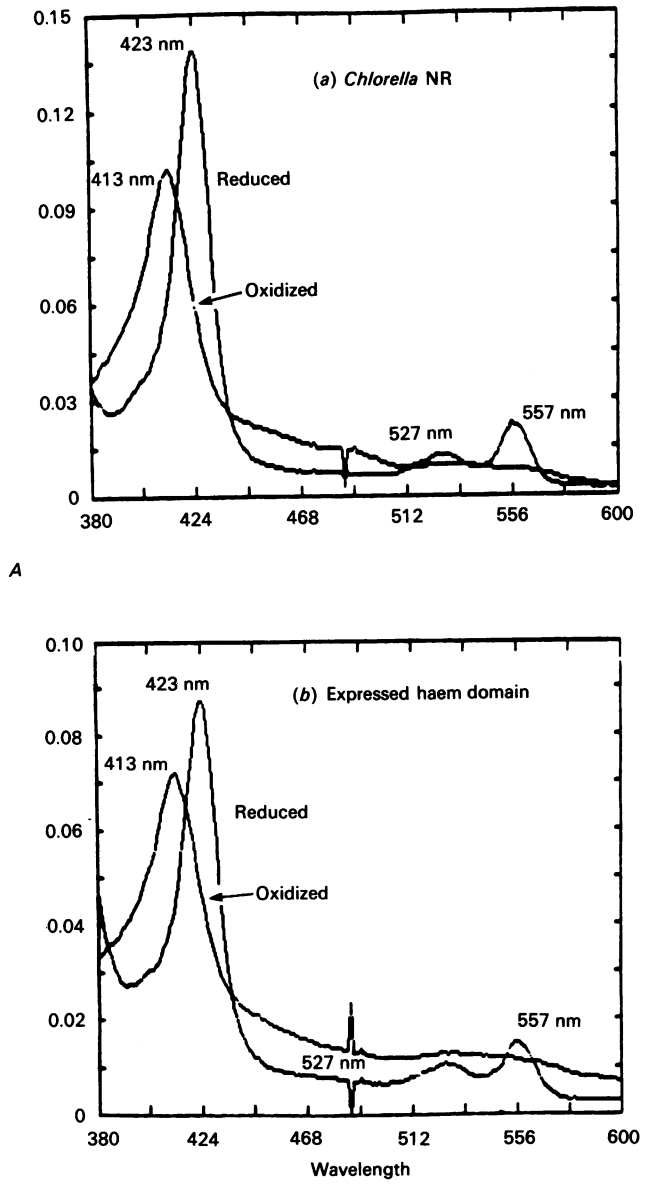

Fig. 8. Haem spectral properties of expressed protein of pET5b-NR1 in *E. coli* and *Chlorella* NR

Partially purified expressed protein and pure *Chlorella* NR were scanned between 380 and 600 nm in a Beckman DU70 spectrophotometer. The reduced spectra were run after the addition of a small amount of dithionite. (*a*) *Chlorella* NR; (*b*) expressed haem domain.

be expected, since the FAD domain is missing from the clone. However, as yet the exact boundaries of the Mo-pterin-binding domain have not been identified, and it is possible that pCVNR1 contains enough of the domain to bind Mo-Co. No nitrate-reducing partial activity (using Methyl Viologen as an electron donor) was detected, indicating a lack of a functional Mo-Co associated with the expressed protein. This does not mean that pCVNR1 coded protein lacks the Mo-Co-binding domain, since incorrect folding of the expressed protein may occur in *E. coli*, preventing binding, or *E. coli* may lack the correct Mo-Co factor required for *Chlorella* NR [36]. The putative hinge region in the expressed domain does not contain a trypsin-sensitive site. This site is evident in higher-plant NR [37] and is located between the haem- and Mo-Co-binding domains. This is consistent with the apparent differences in limited proteolysis products reported for *Chlorella* [15] and higher plants [37].

The availability of an expressed haem-binding domain will open the way to a number of future studies. For example, a comparison of the mid-point potential and c.d. spectrum of the isolated domain with that of the haem domain of native *Chlorella* NR will identify the effect, if any, of other domains on the physicochemical properties of the haem group of NR. Such studies, together with mutagenesis *in vitro*, will aid in the identification of residues responsible for the unusual negative state of the haem mid-point potential [14] of *Chlorella* NR.

This work was supported by National Science Foundation grant DCB 8615836. We are grateful to Dr. Peter J. Neame for providing us with the partial amino acid sequences of *Chlorella* NR, and Dr. C. J. Kay for helpful discussion. We also thank Dr. H. Goodman and Dr. N. Crawford for generously donating the NR cDNAs from barley (pbNRp1) and squash (pCmc-2).

REFERENCES

1. Dunn-Coleman, N. S., Smarelli J., Jr. & Garrett, R. H. (1984) Int. Rev. Cytol. **92**, 1–50
2. Thomsett, A. B. (1989) Microbiol. Rev. **2**, 31–55
3. Wray, J. L. & Kinghorn, J. R. (eds.) (1989) Molecular and Genetic Aspects of Nitrate Assimilation, Oxford Science Publications, Oxford
4. Beevers, L. & Hageman, R. H. (1969) Annu. Rev. Plant Physiol. **20**, 495–522
5. Solomonson, L. P. & Barber, M. J. (1990) Annu. Rev. Plant Physiol. Plant Mol. Biol. **41**, 225–253
6. Cheng, C.-L., Dewdney, J., Kleinhofs, A. & Goodman, H. M. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 6825–6828
7. Crawford, N. M., Campbell, W. H. & Davis, R. W. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 8073–8076
8. Crawford, N. M., Smith, M., Bellisimo, D. & Davis, R. W. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 5006–5010
9. Calza, R., Huttner, E., Vincente, M., Rouze, P., Galangau, F., Vauchert, H., Cherel, I., Meyer, C., Kroneberger, J. & Caboche, M. (1987) Mol. Gen. Genet. **209**, 552–562
10. Kinghorn, J. R. & Campbell, E. I. (1989) in Molecular and Genetic Aspects of Nitrate Assimilation (Wray, J. L. & Kinghorn, J. R., eds.), pp. 385–403, Oxford Science Publications, Oxford
11. Fernandez, E., Schnell, R., Ranum, L. P. W., Hussey, S. C., Silflow, C. D. & Lefebvre, P. A. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 6449–6453
12. Campbell, W. H. & Kinghorn, J. R. (1990) Trends Biochem. Sci. **15**, 315–319
13. Kay, C. J., Barber, M. J. & Solomonson, L. P. (1988) Biochemistry **27**, 6142–6149
14. Kay, C. J., Solomonson, L. P. & Barber, M. J. (1986) J. Biol. Chem. **261**, 5799–5802
15. Solomonson, L. P., Barber, M. J., Robbins, A. P. & Oaks, A. (1986) J. Biol. Chem. **261**, 11290–11294
16. Solomonson, L. P., McCreery, M. J., Kay, C. J. & Barber, M. J. (1987) J. Biol. Chem. **262**, 8934–8939
17. Zeiler, K. G. & Solomonson, L. P. (1989) Arch. Biochem. Biophys. **269**, 46–54
18. Sambrook, J., Fritsch, E. F. & Maniatis, R. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
19. Bascombe, N. F. & Schmidt, R. R. (1987) Plant Physiol. **83**, 75–84
20. Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 1408–1412
21. Gubler, U. & Hoffman, B. J. (1983) Gene **25**, 263–269
22. Mierendorf, R. C., Percy, C. & Young, R. A. (1987) in Guide to Molecular Cloning Techniques (Berger, S. L. & Kimmel, A. R., eds.), pp. 458–469, Academic Press, Orlando, FL
23. Steffens, D. L. & Gross, R. W. (1989) Biotechniques **7**, 674–678
24. Sanger, F., Nickler, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
25. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) Methods Enzymol. **185**, 60–89
26. Laemmli, U. K. (1970) Nature (London) **227**, 680–685
27. Aota, S-i., Gojobori, T., Ishibashi, F. & Ikemura, T. (1988) Nucleic Acids. Res. **16s**, 315–402
28. Matthews, F. S. (1985) Prog. Biophys. Mol. Biol. **45**, 1–56
29. Campbell, W. H. & Gowri, G. (1990) Plant Physiol. **92**, 1–11

30. Lê, K. H. D. & Lederer, F. (1983) EMBO J. **2**, 1909–1914
31. Lederer, F., Ghrir, R., Guiard, B., Cortial, S. & Ito, A. (1983) Eur. J. Biochem. **132**, 95–102
32. Lederer, F., Cortial, S., Becam. A. M., Haumont, P. Y. & Perez, L. (1985) Eur. J. Biochem. **152**, 419–423
33. Richardson, J. S. & Richardson, D. C. (1989) in Prediction of Protein Structure and the Principles of Protein Conformation (Fasman, G. D., ed.), pp. 1–98, Plenum Press, New York
34. Beck von Bodman, S., Schuler, M. A., Jollie, D. R. & Sligar, S. G. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 9443–9447
35. Black, M. T., White, S. A., Reid, G. A. & Chapman, S. K. (1989) Biochem. J. **258**, 255–259
36. Johnson, J. L., Bastian, N. R. & Rajagopalan, K. V. (1990) FASEB J. **4**, A1534
37. Kubo, Y., Ogura, N. & Nakagawa, H. (1988) J. Biol. Chem. **263**, 19684–19689

Received 17 December 1990/20 February 1991; accepted 4 March 1991