Respiratory nitrate reductase of *Escherichia coli*

Sequence identification of the large subunit gene

Michael J. McPherson, Andrew J. Baron, Darryl J.C. Pappin+ and John C. Wootton*

Department of Genetics and +Department of Biochemistry, University of Leeds, Leeds LS2 9JT, England

Received 28 August 1984

The nucleotide sequence is presented of part of the transcription initiation end of the *nar* operon of *Escherichia coli* K12, which encodes the respiratory nitrate reductase complex. The first coding sequence transcribed is the *narG* gene, encoding the large catalytic molybdoprotein of the complex. This sequence was assigned unambiguously by automated N-terminal amino acid sequencing of the purified large subunit. The deduced partial amino acid sequence of this polypeptide is hydrophilic and rich in basic residues. Membrane insertion does not involve N-terminal proteolytic processing of this subunit.

Nitrate reductase Nar *operon* *DNA sequence* *Amino acid sequence* *Molybdoprotein* (Escherichia coli)

1. INTRODUCTION

Respiratory nitrate reductase (EC 1.7.99.4) of *Escherichia coli* is a membrane-bound enzyme complex that enables nitrate to function as a terminal electron acceptor in cells growing anaerobically. Its synthesis is induced by nitrate in the absence of oxygen and is repressed in aerobic growth [1]. The purified active nitrate reductase complex contains a large α-subunit (~150 kDa, the catalytic polypeptide containing the molybdenum cofactor and non-haem iron), a β-subunit (~60 kDa, possibly involved in membrane binding) and in some preparations a γ-subunit (~19 kDa, a *b*-type cytochrome) [2–5] (the subunits are named, respectively, A, B and C in [2,] [5]). Genetic evidence based on insertions of transposons Tn*10* [6,7], Tn*5* [8] and bacteriophage μ [9] has demonstrated that these three polypeptides, and possibly other regulatory or uptake functions [6,7], are encoded by the *nar* operon at 27 units on the *E. coli* linkage map (previous names for this gene complex include *chlC* and *narC*) [10]. The genes encoding the α, β- and γ-polypeptides are named, respectively, *narG, narH* and *narI* [8]. The properties of transposon insertion mutants [6,8,9] indicate, but do not prove, a direction of transcription *narG*→*H*→*I*.

Molecular studies of these interesting genes and proteins at the sequence level are desirable but have previously been hindered by the large size of the α-subunit and by deleterious effects and genetic instabilities associated with cloned *nar* DNA. These problems have recently been overcome by our isolation of a reasonably stable cosmid clone, pNAR10, that carries functional *nar* DNA and other linked genes (submitted). We report here the nucleotide sequence of part of *nar* DNA that corresponds to the 5′-end of the *nar* transcript and the sequence coding for the N-terminal 147 amino acids of the α-subunit. This coding sequence was assigned definitively using solid phase automated amino acid sequencing of the N-terminal 10 residues of purified α-polypeptide. These sequences demonstrate that the *narG* gene is the promoter proximal structural gene of the *nar* operon, and that the α-polypeptide does not undergo post-translational proteolytic processing at its N-terminal end except for removal of *N*-formylmethionine. The region of the polypep-

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V.

tide sequenced is relatively hydrophilic and rich in basic residues.

2. MATERIALS AND METHODS

2.1. *Bacterial strains and plasmids*

A *nar*$^+$ strain of *E. coli* K12, X5119 (*pro-AB-lacDX111*, spc^r, *supE*$^+$), was obtained from J. Scaife. Plasmid pAF5, a gift from A. Fimmel [11], has a 1.7 kb *Bam*HI–*Hin*dIII insert that includes part of *nar* DNA, defined in fig.1, fused to bacteriophage λ DNA (this DNA was cloned from a λ-*nar-lacZ* fusion that placed *lacZ* under nitrate and anaerobic control [11]). Plasmid pNAR10 (fig.1) was cloned from a library of DNA of *E. coli* strain 803 in cosmid pBHB3030 (its isolation and properties are described separately (submitted).

2.2. *DNA sequencing*

The dideoxynucleotide method [12], supplemented by reverse sequencing [13], was used to sequence fragments of *nar* DNA in bacteriophage M13 vectors M13mp8 and M13mp9 [14,15], employing general methods previously used in this laboratory [16]. The 1.7 kb *Bam*HI–*Hin*dIII insert of pAF5 was recloned into M13mp8 and this clone was used as starting material for the preparation of deleted derivatives by restriction cleavage followed by fill-in using Klenow fragment and circularisation by blunt-end ligation. Other deleted derivatives were made by digestion by exonuclease Bal-31 following cleavage at the upstream *Pst*I site in *nar* DNA (fig.1), followed by *Hin*dIII cleavage, fill-in and recircularisation. Overlapping inserts were subcloned into M13mp10 from restriction fragments of pNAR10 DNA and identified by plaque hybridisation. The sequences obtained from these pNAR10 fragments (by J.M. Cock in this laboratory) extended and confirmed the sequences obtained from the pAF5 fragments and demonstrated that pAF5 and the corresponding part of pNAR10 are colinear in sequence.

2.3. *Purification of the nitrate reductase large subunit*

Nitrate reductase was purified from 38 g cells of *E.coli* strain X5119 harvested by centrifugation from cultures in L broth containing 0.2% (w/v) D-glucose and 0.1% (w/v) KNO_3, anaerobically induced for 4 h. Purification and assay methods

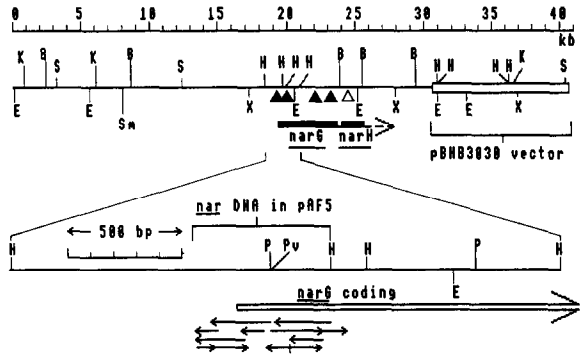

Fig.1. Restriction map of the *nar* region of *E.coli* DNA and the location of the *narG* (α-subunit gene) coding sequence. The location and direction of transcription of the *narG* gene and the approximate location of the *narH* gene are shown by blocks on the labelled right-pointing arrow. The upper map shows some restriction cleavage sites of the cosmid clone pNAR10, mapped from fragment sizes generated by single and double, complete and partial digests using the enzymes *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Kpn*I (K), *Sal*I (S), *Sma*I (Sm) and *Xho*I (X). Other *Hin*dIII sites exist between coordinates 0 and 17.5 but have not been mapped accurately. The lower map shows restriction sites at higher resolution around the control region and part of the *narG* coding sequence, as deduced from the DNA sequence reported here and from Southern blots of restriction digests of pNAR10 DNA hybridised with single-stranded DNA probes of this region subcloned into phage M13 vectors (hybridisation method of [21]). This map includes *Pst*I (P) and *Pvu*II (Pv) sites. Left- and right-pointing unlabelled arrows below the lower map show the extent and direction of nucleotide sequences determined from different clones in M13 vectors. Southern blots of restriction digests of chromosomal DNA of several strains of *E.coli* K12 were also hybridised with these single-stranded DNA probes and demonstrated colinearity of at least coordinates 8.7–28 of the pNAR10 restriction map (between *Bam*HI and *Xho*I sites at these positions) with the *E.coli* chromosome. Closed and open triangles show approximate sites of insertion of transposon Tn5 into the *nar* operon of the *E.coli* chromosome [8], as discussed in the text.

were as in [2] with the following modifications: (i) 0.02% (w/v) phenylmethylsulphonyl fluoride was added to the extraction buffer to inhibit proteolysis; (ii) for solubilisation of nitrate reductase, the washed envelope fraction was suspended in 10 mM sodium phosphate buffer (pH 7.9) containing 1 mM KNO_3 and 0.25% (w/v) Na deoxycholate, and heated for 15 min at 60°C–70% of the activity

was recovered in the supernatant following this treatment; (iii) this supernatant was fractionated on a column of Sepharose CL-4B (Pharmacia) in the above buffer lacking sodium deoxycholate. This column replaced the DEAE-cellulose and Bio-Gel steps of [2] and gave enzyme pure by criteria of protein-stained and activity-stained [4] polyacrylamide gels except for one minor contaminating band. Active fractions, which contained 4.2 mg protein and enzymic activity in NO_2^- production of 68 μmol · min^{-1}, were concentrated by dialysis against 15% (w/v) polyethylene glycol PEG6000, and subunits were dissociated by dialysis against 100 mM Na phosphate buffer (pH 7.1) containing 1% (w/v) SDS and 1% (v/v) 2-mercaptoethanol for 3 h followed by dialysis against the same buffer but containing 0.1% (w/v) SDS and 0.1% (v/v) 2-mercaptoethanol for 18 h. Two ml of this mixture were fractionated on a column of Ultrogel ACA34 (LKB) (63 × 1.6 cm) in the latter buffer. Fractions monitored by electrophoresis in 0.1% (w/v) SDS–3% (w/v) polyacrylamide gels gave three separated peaks of protein, the first eluted containing 0.28 mg (1.9 nmol) pure large subunit, ~150 kDa. This was dialysed against 0.1% (w/v) SDS and lyophilised.

2.4. *N-terminal amino acid sequence determination*

One nmol of lyophilised α-subunit was dissolved in 0.2 ml of 0.1 M $NaHCO_3$ containing 4% (w/v) SDS and added to 100 mg *p*-phenylenediisothiocyanate derivatised glass prepared as in [17]. To improve the efficiency of coupling of a large polypeptide in the presence of SDS micelles, glass of 170 Å pore size was used instead of the commonly used 75 Å size [18]. The glass was incubated at 56°C for 90 min under N_2, then washed with water and then methanol, both containing 0.5% (v/v) *n*-propylamine. The coupled protein was sequenced for 10 cycles using a Rank-Hilger APS 240 solid sequencer (modified by A. Auffret of the SERC protein sequencing group, University of Leeds) and a sequencing buffer of 5% *N*-methyl morpholine–45% methanol–25% water–25% propan-1-ol –0.1% *n*-propylamine (v/v) adjusted to pH 8.6 with trifluoroacetic acid. Anilinothiazolinone amino acid derivatives were converted in 0.07 ml of 20% (v/v) aqueous trifluoroacetic acid for 12 min at 72°C under N_2. Phenylthiohydantoin amino acid derivatives were identified by reverse-phase HPLC [19] and quantified by peak integration.

3. RESULTS AND DISCUSSION

Fig.2 shows the sequence of 509 nucleotides of *nar* DNA. Nucleotide 1 corresponds to the first nucleotide of a major *nar* transcript identitfied by nuclease S1 mapping (transcription studies will be reported separately, unpublished). Nucleotide 509 is the boundary of the *Hin*dIII site at coordinate 19.7 (fig.1). Table 1 shows the results of quantitative N-terminal automated amino acid sequencing of the α-subunit of nitrate reductase. This amino acid sequence corresponds exactly to that deduced from the DNA (fig.2), and thus assigns the deduced coding sequence unambiguously to the α-polypeptide. The 147 amino acids deduced as the N-terminal part of this major molybdoprotein are rich in hydrophilic residues, especially arginine and lysine. Computer searches have not revealed any notable homology to other protein sequences. The codon bias in this coding sequence is typical of *E.coli* genes expressed at a medium to high level, according to the analysis of [20], in that a clear preference is shown for CTG(leucine), CCG(proline), ACC(threonine), CAG(glutamine), AAC (asparagine), AAA(lysine), CGT and CGC (arginine), and an absence of CTA(leucine) and ATA (isoleucine).

The coding sequence is preceded by a strong Shine-Dalgarno sequence (AGGAG, fig.2). The N-terminal amino acid of the purified α-polypeptide is the serine immediately following the initiating *N*-formylmethionine (table 1, fig.2). Therefore, there is no evidence from the sequence for post-translational proteolytic processing of the α-polpeptide at the N-terminal end, apart from removal of the *N*-formylmethionine. Since the nitrate reductase complex was released from membranes during purification, modifications of the α-subunit reported to be associated with membrane insertion [5] do not involve cleavage of a conventional N-terminal signal peptide, consistent with the suggestion [5] of covalent attachment of a cofactor.

This precise localisaiton of the *narG* coding sequence is consistent with an earlier approximate assignment based on the effects of transposon Tn*5* insertions [8], as indicated in fig.1. The leftmost insertion mapped [8] (leftmost solid triangle, fig.1)

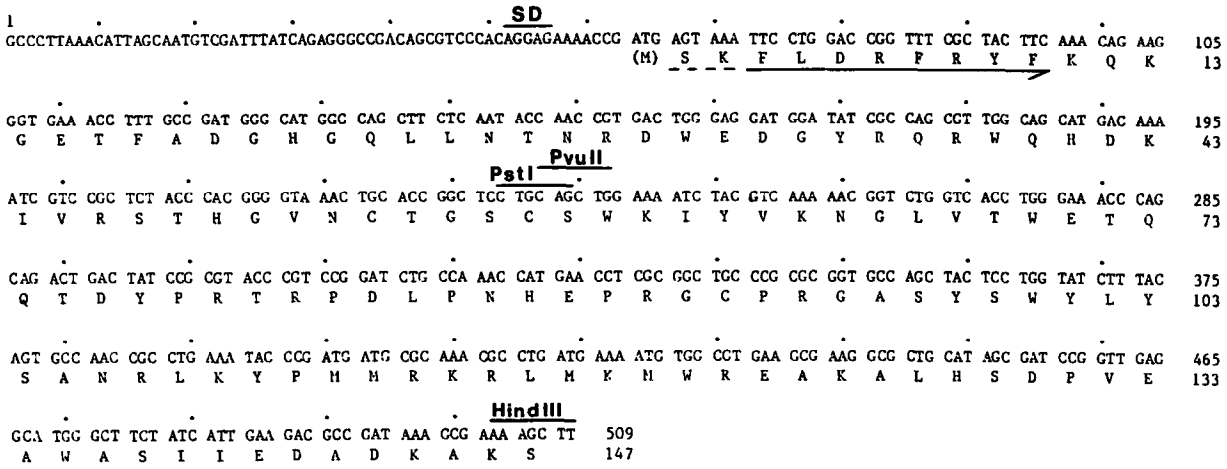

Fig. 2. Nucleotide sequence of *nar* DNA from the transcription initiation point into the *narG* coding sequence. The DNA sequence was determined unambiguously from both strands, except for nucleotides 145–235 which were determined from several independent unambiguous gels of one strand only. The deduced partial amino acid sequence of the nitrate reductase α-polypeptide is shown in 1-letter code below the nucleotide sequence. The amino acids determined by automated sequencing (table 1) are underlined (right-pointing arrow). SD, Shine-Dalgarno (ribosome-binding) sequence. The *Pst*I, *Pvu*II and *Hin*dIII sites shown permit localisation and orientation of the sequence on the high resolution restriction map shown in fig. 1.

prevents synthesis of both the α- and β-polypeptides [8]. Our sequence assignment places this insertion just upstream of the *narG* coding sequence, approximately in the promoter region, where it might be expected to disrupt transcription of the whole operon. Insertions that give truncated α-polypeptides [8] fall within the DNA we assign to the α-subunit coding sequence (other solid black triangles, fig. 1). An insertion that disrupts the β-subunit coding sequence but leaves the α-subunit intact [8] (open triangle, fig. 1) suggest that the *narH* gene, encoding the β-polypeptide, is likely to be located immediately downstream of *narG*, as shown in fig. 1.

Table 1

Automated solid-phase sequencing of the nitrate reductase α-subunit

Cycle	Residue identified	Amount (pmol)
1	–[a]	–
2	–[b]	–
3	Phe	170
4	Leu	210
5	Asp	210
6	Arg	200
7	Phe	150
8	Arg	170
9	Tyr	150
10	Phe	130

Repetitive yield (Phe-3 to Phe-10) was 96.2%

[a] Residue 1 cannot be identified as it remains covalently attached to the glass support via the N-terminal amino group

[b] Residue 2 is a Lysine (from the DNA sequence, fig. 2) and is not expected to be identified because it remains coupled to the glass via the ϵ-amino group

Present evidence suggests that the promoter region of the *nar* operon, complete with sequences necessary for control by nitrate and anaerobic growth, is located within the adjacent 101 nucleotides upstream of the transcribed sequence shown in fig. 2, since this is the total extent of *nar* DNA in the pAF5 insert as far as the junction with bacteriophage λ DNA (unpublished). The λ-*nar–lacZ*fusion from which the pAF5 insert was cloned [11] showed control of *lacZ* characteristic of the normal anaerobic control of *nar*, and showed nitrate control that qualitatively resembled normal *nar* control but with quantitative differences [11]. The *narG* coding sequence is therefore clearly the promoter proximal coding sequence of the *nar* operon. The existence of functional *nar* DNA on clone pNAR10 will now permit extensive detailed

studies of the structure and function of this important set of genes and proteins at the sequence level.

ACKNOWLEDGEMENTS

We thank J.M. Cock, A.D. Auffret and A.L. Fimmel for the contributions noted in the text. Financial support was from an SERC Research Grant (GR/C/45836 to J.C.W.) and SERC funding of the University of Leeds Protein Sequencing Unit.

REFERENCES

[1] Haddock, B.A. and Jones, C.W. (1977) Bacteriol. Rev. 41, 47–99.
[2] MacGregor, C.H., Schnaitman, C.A., Normansell, D.E. and Hodgins, M.G. (1974) J. Biol. Chem. 249, 5321–5327.
[3] Forget, P. (1974) Eur. J. Biochem. 42, 325–332.
[4] Lund, K. and DeMoss, J.A. (1976) J. Biol. Chem. 251, 2207–2216.
[5] Chaudry, G.R. and MacGregor, C.H. (1983) J. Bacteriol. 154, 387–394.
[6] Stewart, V. and MacGregor, C.H. (1982) J. Bacteriol. 151, 788–799.
[7] Stewart, V. (1982) J. Bacteriol. 151, 1320–1325.
[8] Edwards, E.S., Rondeau, S.S. and DeMoss, J.A. (1983) J. Bacteriol. 153, 1513–1520.
[9] Bonnefoy-Orth, V., Lepelletier, M., Pascal, M.-C. and Chippaux, M. (1981) Mol. Gen. Genet. 181, 535–540.
[10] Bachmann, B.J. (1983) Microbiol. Rev. 47, 180–230.
[11] Fimmel, A.L. and Haddock, B.A. (1981) FEMS Microbiol. Lett. 12, 125–129.
[12] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[13] Hong, G.F. (1981) Biosci. Rep. 1, 243–252.
[14] Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309–321.
[15] Messing, J. and Vieira, J. (1982) Gene 19, 269–276.
[16] McPherson, M.J. and Wootton, J.C. (1983) Nucleic Acids Res. 11, 5257–5266.
[17] Machleidt, W. and Hofner, H. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, C.H.R., ed.) pp. 35–48, Elsevier, Amsterdam, New York.
[18] Wachter, E., Machleidt, W., Hofner, H. and Otto, J. (1973) FEBS Lett. 35, 97–102.
[19] Zimmerman, C.L., Appella, E. and Pisano, J.J. (1977) Anal. Biochem. 77, 569–573.
[20] Grosjean, H. and Fiers, W. (1982) Gene 18, 199–209.
[21] Hu, N. and Messing, J. (1982) Gene 17, 271–277.