Nucleotide sequence of the *narL* gene that is involved in global regulation of nitrate controlled respiratory genes of *Escherichia coli*

Robert P.Gunsalus, Lisa V.Kalman and Ramsey R.Stewart

Department of Microbiology and the Molecular Biology Institute, University of California, Los Angeles, CA 90024, USA

Received November 7, 1988; Revised and Accepted January 31, 1989

ABSTRACT

The DNA sequence was determined for the *narL* gene of *Escherichia coli*. This gene is involved in global regulation of a number of nitrate controlled genes including *frdABCD*, *tor*, *narGHJI*, and *adhE* which are associated with bacterial respiration and fermentation. Comparison of the deduced amino acid sequence of *narL* to that of other bacterial genes revealed significant homologies to the *phoB*, *ompR*, and *virG* gene products based on the presence of similar protein domains. These DNA binding proteins are members of two-component regulatory systems. The similarities suggest that *narL* may also participate in such a two-component regulatory system and that the *narR* gene, which lies upstream of *narL*, may encode a second component required for nitrate control of gene regulation. In vitro protein synthesis experiments using a *narL* plasmid identified a putative NarL protein of 29 kDa in size consistent with the DNA sequence analysis. Primer extension experiments revealed the presence of two 5' termini for *narL* mRNA, and indicates that transcription may be complex.

INTRODUCTION

The *narL* gene of *Escherichia coli* is necessary for controlling both positive and negative regulation of a number of operons involved in anaerobic cell growth. Strains containing mutations in *narL* are unable to repress transcription of fumarate reductase, alcohol dehydrogenase and trimethylamine-N-oxide dehydrogenase or induce expression of nitrate reductase synthesis in the presence of nitrate (1, 2, 3). These enzymes are involved in electron transfer reactions that allow energy to be generated by anaerobic respiration or by fermentation (4). Their production is carefully regulated in response to availability of alternative cellular electron acceptors, including nitrate, to insure optimal

energy production when cells are growing on a variety of fermentable and non-fermentable carbon substrates.

The narL gene is located at 27 min on the E. coli genetic map adjacent to narGHJI (1,5). It has been cloned by complementation of narL (=frdR2) mutants using a frdA-lacZ indicator system that responds to the presence of nitrate in the culture medium (3) and by restoration of narGHJI gene function (6). The narL gene product is required to form an active repressor protein for controlling frdABCD expression and an activator for induction of narGHJI (2, 3, 7, 8). We recently proposed a model for regulation of frdABCD gene expression whereby the narL and narR gene products act to sense nitrate and repress frdABCD gene expression (7, 9). Thus, when cells are grown in the presence of nitrate, RNA polymerase is unable to bind at the frd promoter and initiate transcription.

In this paper, we present the nucleotide sequence of the narL gene. This sequence was analyzed and compared to other bacterial regulatory proteins.

MATERIALS AND METHODS

Bacterial strains, bacteriophage and plasmids

The E. coli K-12 strains and bacteriophage used were: strain JM103, M13 phage manipulations (10); P90C, plasmid isolation (11); A19 for preparation of bacterial S-30 extracts (12); and LK4100 (as MC4100/λJ100 but recA), for mRNA preparation (7). M13mp18 and M13mp19 (New England Biolabs, Inc., Beverly, MA) were used for DNA sequence analysis. Plasmids pLK92, pLK63, pLK633 and pLK634 are described elsewhere (7) as is pACYC184 (13).

Bacteriophage and plasmid procedures

General procedures for use of M13 phage and the preparation of DNA are as described (10).

DNA sequence analysis

DNA sequencing was performed by the Sanger dideoxy chain-termination method using Sequenase (14). Urea (7M) denaturing

polyacrylamide gels (6 or 8%) were used to separate the chain termination reaction products (15). DNA sequence data were stored and analyzed by computer using the University of Wisconsin Genetics Computer Group (16) and Staden software (17). Protein homologies were found using the FASTA routine searching the PIR data base (release 11).

In vitro protein synthesis

Bacterial S-30 extracts were used for in vitro transcription-translation reactions with the indicated plasmid DNA as template (12). Proteins were labeled with ^{35}S -methionine (Amersham) and separated by SDS PAGE (18).

Primer extension reactions

Total bacterial mRNA was prepared as described (19). The primer extension reactions used to locate the 5' ends of the narL mRNA were performed as previously described (19). The DNA oligonucleotide TTCGGAATAAAGGTGACC which is complementary to the region at the start of narL translation was used as a primer for DNA synthesis. The reaction products were separated using an 8% 7M urea polyacrylamide gel (15).

Materials

The universal M13 sequencing primer (#1212) was obtained from New England Biolabs, Inc., Beverly, MA, while additional primers were synthesized using an Applied Biosystems Synthesizer. Deoxyadenosine 5'-(α -[^{35}S]thio) triphosphate was obtained from Amersham Corp., Arlington Heights, IL. Sequenase was obtained from U.S. Biochemicals. Other DNA sequencing reagents and enzymes were obtained from Bethesda Research Laboratories, Inc., Gaithersburg, MD, and Pharmacia, Inc., Piscataway, NJ. All other reagents were commercial products of the highest grade available.

RESULTS AND DISCUSSION

Sequencing strategy

The 1.16 kb NsiI fragment from plasmid pLK2 (narR⁺L⁺,3) that contains the narL gene region was isolated and cloned into both

Figure 1. DNA sequencing strategy and partial restriction map of the region containing the *narL* gene of *E. coli*. The horizontal line at the top of the figure shows a partial restriction map with distances indicated in 200 bp increments. The open reading frame predicted for *narL* is represented by the open box. The direction of transcription is from left to right as drawn. Heavy arrows indicate the position and extent of DNA sequencing reactions. The sites that correspond to the 5' termini of mRNA (Fig. 3) are represented by the carrots (▲).

M13mp18 and M13mp19 to facilitate sequencing of both strands. The complete DNA sequence was determined for both strands by the dideoxy chain termination method (Fig. 1). Either the Universal primer or synthetic oligonucleotide sequencing primers designed from *narL* sequence data were used. There were no ambiguities in the final sequence (Fig. 2).

Sequence and properties of the *narL* gene

The DNA sequence of the region that complements the *narL* (=frdR2) mutation (3) was analyzed by computer and revealed a single open reading frame of greater than sixty codons in length that starts with either an ATG or GTG. This open reading frame is of sufficient size (254 codons) to encode the putative NarL protein as revealed by gene complementation experiments and confirmed by in vitro protein synthesis from plasmid DNA (see below). The deduced amino acid sequence, shown below the DNA sequence (position 102 to 866), is predicted to give a polypeptide of 28,177 daltons in size. A GTG initiation codon is predicted for initiation of *narL* translation based on the lack of

-78	TGTTGCAAATTGCCCGTGAGGCATTAAGTAACGCCCTCAAACATTCGCAAGCGAGTGAAG	-19
	+1	+33
-18	TCGTGGTACGGTGGCGCAAAACGATAATCAGGTCAAACCTGACCGTCGAGGATAACGGCT	41
	▲	▲
42	GCGGCGTGCCTGAAATGCCATCCGACGCAATCACTACGGCATGATAATAATGCGCGATC	101
102	GTGCGCAAAGTTTACGAGGCGATTGCCGCGTCCGCGTCGTGAATCAGTGGCACCGAAG	161
	MetArgLysValTyrGluAlaIleAlaAlaSerAlaValValAsnGlnValAlaProLys	
162	TGGTGGTACGCTTTATTCGCGAAAAAATTTCACAGACGTCCAAGGAGATACCGATGAGT	221
	TrpTrpSerProLeuPheProLysLysLeuSerGlnThrSerLysGluIleProMetSer	
222	AATCAGGAACGGCTACTATCCTGCTGATTGACGATCACCCGATGTCGGAATGCGCGTA	281
	AsnGlnGluProAlaThrIleLeuLeuIleAspAspHisProMetLeuArgThrGlyVal	
282	AAACAGCTTATCAGTATGGCACCGATATCACCGTGGTTGGCGAAGCGAGTAATGGCGAA	341
	LysGlnLeuIleSerMetAlaProAspIleThrValValGlyGluAlaSerAsnGlyGlu	
342	CAGGGTATTGAATGGCGGAGTCTCTTGATCCCGATCTGATCCTGTTAGATCTCAATATG	401
	GlnGlyIleGluLeuAlaGluSerLeuAspProAspLeuIleLeuLeuAspLeuAsnMet	
402	CCCGCATGAACGGTCTGGAACGGCTGGATAAACTGCGCGAAAAAGTCCCTCTCAGGCGCC	461
	ProGlyMetAsnGlyLeuGluThrLeuAspLysLeuArgGluLysSerLeuSerGlyArg	
462	ATTGTGGTATTACGGCTCTCTAACCATGAAGAAGATGTGGTCACCGCACTGAAACGCGGC	521
	IleValValPheSerValSerAsnHisGluGluAspValValThrAlaLeuLysArgGly	
522	GCGGATGGCTATCTGTTAAAAGATATGGAACCGGAAGATCTGCTGAAAGCATTGCATCAG	581
	AlaAspGlyTyrLeuLeuLysAspMetGluProGluAspLeuLeuLysAlaLeuHisGln	
582	GCAGCTGCTGCGGAAATGGTATTAAACCGAAGCATTAAACGCTGTTCTGGCGCCAGCTTG	641
	AlaAlaAlaGlyGluMetValLeuSerGluAlaLeuThrProValLeuAlaAlaSerLeu	
642	CGCGCTAACCGTGCCACTACTGAGCGCGATGTTAACCACTTAACCCACGCGAGCGCGAT	701
	ArgAlaAsnArgAlaThrThrGluArgAspValAsnGlnLeuThrProArgGluArgAsp	
702	ATTCTCAAGCTGATTGCCCGAGGGTTGCCGAACAAGATGATTGCCGCGCGCTGGATATC	761
	IleLeuLysLeuIleAlaGlnGlyLeuProAsnLysMetIleAlaArgArgLeuAspIle	
762	ACCGAAAGCACAGTAAAGTGCACGTCGAAGCACATGCTGAAGAAAAATGAAGCTCAAGTCT	821
	ThrGluSerThrValLysValHisValLysHisMetLeuLysLysMetLysLeuLysSer	
822	GCGGTGGAAGCAGCGGTATGGGTGCATCAGGAGCGCATTCTTCTGA	866
	ArgValGluAlaAlaValTrpValHisGlnGluArgIlePheEnd	

Figure 2. The nucleotide sequence and deduced amino acid sequence of the nontranscribed strand of the *narL* gene of *E. coli*. The complete DNA sequence of the region is shown. The +1 position indicates the location of the first of the two 5' mRNA termini (Fig. 3). Lines above the nucleotide sequence indicate possible -10 and -35 *narL* promoter sequences. The sites that correspond to the 5' termini of mRNA (Fig. 3) are represented by the carrots (▲).

any ATG codons which would produce a protein of sufficient size. Additionally, multiple termination codons are present in the other reading-frames (Fig 2). Thus it seems reasonable that the open reading frame shown is *narL*.

Figure 3. Location of two 5' termini of *narL* mRNA. An 8% acrylamide gel was used to resolve the primer extension reaction products using mRNA prepared from strain LK4100 with pACYC184 (lane 1) or pLK63(*narL*⁺) (lane 2) shown adjacent to a M13 sequencing ladder of the region preceding *narL*. The nucleotide sequence is shown to the right of the panel. Locations of the 5' ends are indicated by the arrows (◄).

Identification of two 5' mRNA termini for *narL*

To establish that the open reading frame identified for *narL* was transcribed, primer extension reactions were performed using total cellular mRNA prepared from *E. coli* LK4100 containing either plasmid pLK63(*narL*⁺) or the plasmid vector pACYC184. As shown in Fig. 3, two 5' ends were revealed (lane 2). The mRNA ends correspond to the nucleotides at positions +1 and +33 (Fig. 2) which are 102 and 70 nucleotides before the predicted translational initiation site for *narL* (Fig. 2). Both 5' termini were observed in mRNA preparations from strains containing either a plasmid with *narL*⁺ (pLK63) or without (pACYC184), although the

Figure 4. Identification of the NarL polypeptide by in vitro protein synthesis. Lane 1 represents the reaction products from pLK63 (narR⁺L⁺), Lane 2, pLK633 (narR⁺), and Lane 3, pLK634 (narL⁺). The sizes are indicated, and given in kDa (NarL, 29 kDa; NarR, 66 kDa).

two bands seen in the strain lacking multiple copies of narL⁺ were of weak intensity. This is apparently due to low abundance of the narL mRNA that is produced from the chromosome.

Identification of the NarL polypeptide

In vitro protein synthesis reactions were performed using bacterial S-30 extracts to catalyze transcription and translation products from narL using plasmid DNA templates (Fig. 4). When a narL⁺ plasmid, pLK634, that complements narL but not narR mutants was used in the reaction, a polypeptide of 29 kDa was produced (lane 3). This protein was absent when a plasmid (pLK633) was used which lacks the right half of the narL open reading frame but contains the 3.4 kb region to the left of the BglII sites located in narL (Fig. 1). When either plasmid pLK63 that contains the entire narR⁺L⁺ region, or pLK633 (narR⁺) that contains the region 5' to narL was used, a 66 kDa protein was observed. This protein apparently corresponds to the narR gene product that is immediately 5' to narL (Fig. 1; 6, 7, unpublished). All lanes also contain a polypeptide of 24 kDa that corresponds to the chloramphenicol acetyl transferase encoded by the plasmid vector.

Figure 5. Primary amino acid sequence homology alignment of NarL to OmpR, VirG, and PhoB proteins from procaryotic microorganisms. The sequences are aligned to show maximum similarity. Vertical lines indicate identity to narL while asterisks (*) indicate conserved amino acids. The numbering is relative to the N-terminus of NarL.

Features of the DNA sequence

The proposed narL gene is preceded by a potential ribosome binding site (AATGcG) located at position +91 to +96 (Fig. 2). Because two 5' termini were observed for narL, it is not clear whether two promoters exist for transcribing narL, or if mRNA processing or decay may be occurring. These termini correspond to nucleotide positions +1 and +33 (Fig. 2) and occur 102 and 70 nucleotides before the predicted start for NarL. Potential -35 and -10 regions for a RNA polymerase binding site are indicated at positions -35 to -6 (Fig. 2). No identifiable promoter sequence was found for the second 5' terminus.

Primary sequence homology to other proteins

Computer analysis was performed with the predicted narL amino acid sequence by comparing it to other bacterial proteins.

Homology was found between the phoB and ompR genes of E. coli, and the virG gene of Agrobacterium tumefaciens (20, 21, 22). The primary amino acid sequences were aligned by computer and optimized by hand to give the best apparent fit (Fig. 5). Using the same criteria used previously by others (21), we compared NarL to PhoB, OmpR and VirG. Comparison of the A domains of PhoB, OmpR and VirG to a similar amino terminal region in NarL (positions 41-161) showed 27%, 29% and 24% identity respectively. There were an additional 23%, 15% and 16% conserved residues between these regulatory proteins and NarL. The PhoB, OmpR and VirG A domains showed a similar to somewhat higher identity (28-45%) when compared to one another (21). The B domains of the proteins were less well conserved and showed a 10%, 14% and 16% identity when compared to NarL (positions 162-254). The PhoB, OmpR and VirG proteins appear to be more related to one another than to NarL as they have 34%-40% identity in the B domain. Another related regulatory protein, NtrC completely lacks the B domain, but shows a 20% identity to NarL with respect to its A domain.

Gene products (PhoB, OmpR, VirG, NtrC) are all members of two-component regulatory systems (23). Sequence relatedness of these proteins to narL suggests that it may have a similar function. This possibility is supported by observations that narR mutants of E. coli are also defective in nitrate control of genes involved in anaerobic electron transfer, including frdABCD and narGHJI (6, 7, unpublished). It may be inferred from the above studies that the product of the narR gene, which is 66 kDa in size (Fig. 4), may be the second component. If narL functions in a manner similar to the ompR, phoB, and virG genes, then it may code for a DNA binding protein. NarL apparently acts as a modulator for repression or activation of gene expression in response to the action of the narR gene product (a putative nitrate sensor). Experiments are in progress to test this possibility.

In sequence comparisons with proteins from other organisms, narL also revealed similarity to a number of kinases. These include creatine kinase (human and chicken), adenylate kinase

(chicken) and 6-phosphofructokinase (*E. coli*). The explanation for this observation is yet unclear but may indicate a role for NarL in phosphorylation/dephosphorylation.

ACKNOWLEDGEMENTS

We thank Gene Bryant, Steve Melville and Casey Case for assistance and advice in these studies. This work was supported by Public Health Service Grant AI-21678 from the National Institutes of Health. L.V.K. was supported in part by the Public Health Service Cellular and Molecular Biology Training Grant GM-07185.

REFERENCES

1. Stewart, V. and MacGregor, C.H. (1982) *J. Bacteriol.* **151**, 788-799.
2. Iuchi, S. and Lin, E.C.C. (1987) *Proc. Natl. Acad. Sci.* **84**, 3901-3905.
3. Kalman, L.V. and Gunsalus, R.P. (1988) *J. Bacteriol.* **170**, 623-629.
4. Ingledew, W.J. and Poole, R.K. (1984) *Microbiol. Rev.* **48**, 222-271.
5. Bachmann, B.J. (1983) *Microbiol. Rev.* **47**, 180-230.
6. Stewart, V. and Parales, J. (1988) *J. Bacteriol.* **170**, 1589-1597.
7. Kalman, L.V. and Gunsalus, R.P. (1989) *J. Bacteriol.* (submitted).
8. Stewart, V. and Berg, B.L. (1988) *J. Bacteriol.* **170**, 4437-4444.
9. Jones, H.M. and Gunsalus, R.P. (1987) *J. Bacteriol.* **169**, 3340-3349.
10. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
11. Grove, C.L. and Gunsalus, R.P. (1987) *J. Bacteriol.* **169**, 2158-2164.
12. Zalkin, H., Yanofsky, C. and Squires, C.L. (1974) *J. Biol. Chem.* **249**, 465-475.
13. Chang, A.C.Y. and Cohen, S.N. (1978) *J. Bacteriol.* **134**, 1141-1156.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
15. Sanger, F. and Coulson, A.R. (1978) *FEBS Lett.* **87**, 107-110.
16. Devereux, J., and Haeberli, P. (1983) *Program Library of the University of Wisconsin Genetics Computer Group*, Madison, Wisconsin.
17. Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673-3694.
18. Laemmli, U.K. (1970) *Nature* **227**, 680-685.

19. Case, C.C., Roels, S.M., Gonzalez, J., Simons, E.C. and Simons, R.S. (1988) *Gene* (in press).
20. Wurtzel, E.T., Chou, M.Y. and Inouye, M. (1982) *J. Biol. Chem.* **257**, 13685-13691.
21. Winans, S.C., Ebert, P.R., Stachel, S.E., Gordon, M.P. and Nester, E.W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8278-8282.
22. Makino, K., Shinagawa, H., Amemura, M. and Nakata, A. (1986) *J. Mol. Biol.* **190**, 37-44.
23. Ronson, C.W., Nixon, B.T. and Ausubel, F.M. (1987) *Cell* **49**, 579-581.