Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4294–4297, July 1984
Biochemistry

Primary structure of the *Escherichia coli* ribonucleoside diphosphate reductase operon

(*nrdA* gene/*nrdB* gene/M13 dideoxy sequencing/gene structure)

JOHN CARLSON, JAMES A. FUCHS, AND JOACHIM MESSING

Department of Biochemistry, University of Minnesota, St. Paul, MN 55108

Communicated by Peter Reichard, March 26, 1984

ABSTRACT **The nucleotide sequence of the *Escherichia coli* K-12 DNA comprising the operon for the structural genes of the subunits of ribonucleotide diphosphate reductase has been determined. The DNA sequenced maps at 48.5 minutes on the *E. coli* chromosome and includes a total length of 8557 nucleotides. An open reading frame between nucleotides 3506 and 5834, encoding a 776-amino acid polypeptide chain with a molecular weight of 87,532, has been identified as the *nrdA* gene. An open reading frame between nucleotides 6012 and 7139, encoding a 375-amino acid polypeptide with a molecular weight of 43,466, has been identified as the *nrdB* gene. The sequences reveal not only the primary structures for both subunits, but also some interesting aspects of potential regulatory sites.**

Ribonucleotide reductase catalyzes the enzymatic reduction of ribonucleosides to deoxyribonucleotides, the first step in the pathway unique to DNA replication (1). Although ribonucleotide reductases have been isolated and characterized from many sources (2), no primary structure for any has been determined. Ribonucleoside diphosphate reductase isolated from *Escherichia coli* B is the best characterized enzyme, and thus it serves as the model system for the enzyme from other organisms. Ribonucleoside diphosphate reductase activity as well as substrate specificity is elegantly regulated by allosteric effectors (1). Ribonucleoside diphosphate reductase is composed of two nonidentical subunits designated B1 and B2. The B2 subunit is composed of two identical polypeptides (B_2) and contains two ferric ions and a free radical located on a tyrosine residue (1). The B1 subunit is composed of two polypeptides (α and α') of similar size that have identical COOH termini but different NH_2 termini (1). Both polypeptides are encoded by one structural gene (3).

Since *E. coli* has no alternative pathway for the production of deoxyribonucleotides, mutants in the *nrd* genes were obtained in *E. coli* K-12 as conditional mutants. The *nrdA* and *nrdB* genes encode the polypeptides of subunits B1 and B2, respectively (4, 5). Mutations in *nrdA* and *nrdB* map at 48.5 minutes on the *E. coli* chromosome (6). Derivatives of bacteriophage λ that could transduce the *nrd* genes were isolated (7), and the *nrd* genes were transferred into ColE1 (8) and subcloned into pBR322 to give a plasmid designated pPS2 (9). Complementation of *nrdA* and *nrdB* mutants by pPS2 and pPS2 subclones were used to localize the *nrd* genes (9). The λd*nrd* lysogen is capable of producing large quantities of ribonucleoside diphosphate reductase, and most recent studies of *E. coli* ribonucleoside diphosphate reductase have used the enzyme produced from this K-12 strain rather than the enzyme from a B strain (7). No differences have been found in ribonucleoside diphosphate isolated from a B strain or a K-12 strain (7).

The synthesis of ribonucleoside diphosphate reductase is controlled at the level of transcription (10). The *nrdA* and *nrdB* genes direct the synthesis of a 3.2-kilobase (kb) polycistronic mRNA (3). Perturbations in DNA replication, either a shift-up in growth conditions or an inhibition of DNA synthesis, leads to increased synthesis of *nrd* mRNA (10). Protein synthesis during perturbation of DNA replication is required for expression of the increased *nrd* expression (11).

To determine the amino acid sequence of ribonucleoside diphosphate as well as to gain insight into the regulation of *nrd* expression, we determined the nucleotide sequence of an 8557-nucleotide region of DNA that includes the *nrd* region. Fragments of plasmid pPS2 were generated with seven different restriction enzymes, cloned in M13mp9, and sequenced. Computer analysis of this DNA sequence was used to generate the amino acid sequence of ribonucleoside diphosphate reductase.

MATERIALS AND METHODS

Materials. The nucleotides and DNA polymerase I (Klenow fragment) were obtained from P-L Biochemicals, [^{32}P]-ATP was from New England Nuclear, and restriction enzymes and T4 DNA ligase were from New England Biolabs and Bethesda Research Laboratories, respectively.

Construction of M13mp9 Clones Used in Sequencing. The 16-kb plasmid pPS2 was digested with restriction enzymes *Rsa* I, *Fnu*DII, *Alu* I, *Hae* III, *Taq* I, *Hpa* II, and *Sau*3a, and fragments of 300–800 base pairs (bp) were isolated after agarose gel electrophoresis. These fragments were cloned into either the *Hin*cII, *Acc* I, or *Bam*HI sites of M13mp9 (12). To identify clones containing sequences from the 8.5-kb *Bam*HI/*Pst* I region of pPS2 (see Fig. 1*A*, fragments B and C), four master probes containing either a 3.5-kb *Bam*HI/*Eco*RI (Fig. 1*A*, fragment C) or a 5.0-kb *Eco*RI/*Pst* I fragment (Fig. 1*A*, fragment B) in both orientations in M13mp8 or M13mp9 were used (13).

Nucleotide Sequence Determination. The DNA sequence was determined using the dideoxynucleotide termination method (14) as described by Messing (15) with minor modifications. Analysis of the nucleotide sequences was accomplished on an Apple II computer using the programs by Larson and Messing (16).

RESULTS

Fig. 1*A* gives a partial restriction map of plasmid pPS2 (9). The *Bam*HI/*Pst* I region (Fig. 1*A*, fragments B and C), which encodes the information for *nrdA* and *nrdB* as well as regions both 5′ and 3′ to the coding region (9) was sequenced. The nucleotide sequences of the 137 clones (Fig. 1*B*) were sufficient to determine >95% of both strands of the 8557 nucleotides (Fig. 2).*

***nrdB*.** When the DNA sequence was scanned for open reading frames, an open reading frame between nucleotides

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

Abbreviations: kb, kilobase(s); bp, base pair(s).

*DNA sequence can be obtained from Genbank (Cambridge, MA).

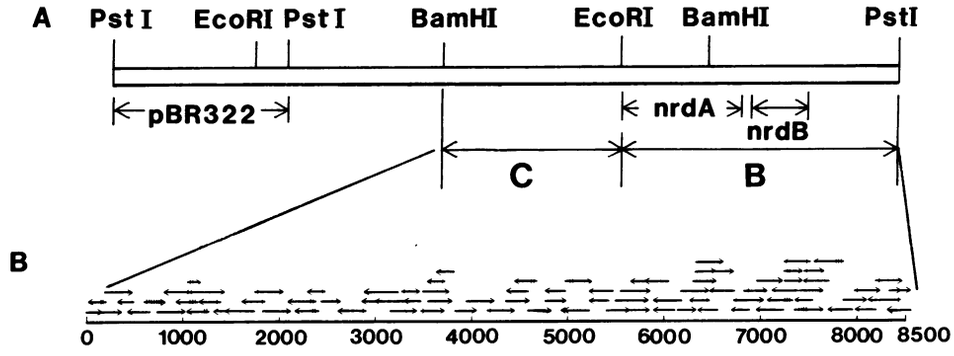

FIG. 1. Subclones of plasmid pPS2 used to determine the DNA sequence. (*A*) Partial restriction map of pPS2 (9) indicating location of *nrdA* and *nrdB* genes. Regions B and C have been forced cloned into M13mp8 and M13mp9 (12) to produce the master probes used to identify the clones to be sequenced as well as their orientation (13). (*B*) the distribution and location of all the gel readings are shown after their map position had been reconstructed from overlaps. All readings are represented by their polarity and their length. Lines with more than one arrowhead represent clones that were analyzed in stretch reaction to obtain additional sequence information. Sequences from apparent duplicate clones are marked.

6012 and 7139 (Fig. 2) of sufficient size to code for B2 protein was observed. The derived amino acid sequence has an alanine following the NH_2-terminal methionine and a COOH-terminal leucine. This protein minus the methionine contains 374 amino acid residues and has a molecular weight of 43,355. The sequence of the first 31 amino acid residues minus the methionine is in complete agreement with the 31 amino acid residues determined by sequencing the NH_2-terminal region of the protein isolated from a K-12 strain (B. M. Sjöberg and H. Jörnvell, personal communication). Protein B2 isolated from a B strain of *E. coli* was reported to contain an NH_2-terminal alanine and a COOH-terminal leucine (17). The comparison of the amino acid composition of the B2 protein derived from the K-12 DNA sequence is compared to the amino acid composition of the B2 protein from a B strain of *E. coli* in Table 1. The amino acid compositions are compared as reported (17) and differ in molecular weight (43,355 vs. 39,000).

***nrdA*.** An open reading frame (3506–5834; Fig. 2) was found that is the size required to encode a B1 polypeptide. The derived amino acid sequence has a Met-Asn-Gln at the NH_2 terminus and a leucine at the COOH terminus. The COOH-terminal residue of the B1 protein isolated from a K-12 strain was found to be a leucine residue (B. M. Sjöberg and M. Bachmann, personal communication). This protein contains 776 amino acid residues and has a molecular weight of 87,532. The B1 subunit purified from either a B or a K-12 strain contains two polypeptides that appear to be similar but that differ in molecular weight. Both polypeptides are encoded by the *nrdA* gene (3). Protein B1, isolated from a B strain, contains two different NH_2-terminal residues, glutamic acid (or glutamine) and aspartic acid (or asparagine) but one COOH-terminal residue, isoleucine (17). The composition of two chymotryptic peptides containing the NH_2 termini was also determined. Peptide 1 contains an NH_2-terminal glutamic acid (or glutamine) and 8–11 amino acid residues. In Fig. 2, region 3511–3544 encoding 9 amino acid residues in common with the composition of peptide 1 are underlined. A region following the apartic acid (3585; Fig. 2) has 11 residues of 18 in common with polypeptide 2 that Thelander found (17). If we assume that one B1 polypeptide found in B1 protein from a K-12 strain has an NH_2-terminal glutamine (3511; Fig. 2) and that the other polypeptide has an NH_2-terminal aspartic acid (3585; Fig. 2), then the polypeptide with an NH_2-terminal glutamic acid would contain 774 amino acid residues and have a molecular weight of 87,287, and the polypeptide with an NH_2-terminal aspartic acid would contain 750 amino acid residues and have a molecular weight of 84,490. As shown in Table 1, there is a good agreement of the published amino acid composition of the protein from a B strain and the composition derived from the K-12 DNA sequence.

The codon usage in the *nrdA* and *nrdB* genes is nonrandom and correlates well with the codon used in nonregulatory proteins of *E. coli* (18). At positions 3493 (*nrdA*) and 6000 (*nrdB*) (Fig. 2) are nucleotides showing close agreement with the ribosome binding sites (19), further justifying the placement of the NH_2 terminals of the two polypeptides.

***nrd* mRNA.** The region 90 bp 5′ to the ATG for *nrdA* containing a −35 and a −10 region shows close correspondence with the consensus sequences involved in binding of RNA polymerase (20). The probable start of transcription would be approximately at position 3434 (Fig. 2), 72 bp 5′ of the ATG start codon of *nrdA*. A possible transcription termination site is located between nucleotides 7335 and 7350 and is underlined in Fig. 2.

There is an apparently untranslated region of 182 bases found between the COOH terminus of B1 protein and the NH_2 terminus of B2 polypeptide. This can be compared to the intergenic regions of −1, 14, 3, and −1 in the *trp* operon and a 45-nucleotide intergenic region between *lacY* and *lacZ*. Interestingly, there are two 63-base direct repeats that differ by a one-base insertion (Fig. 2, dot above inserted base). The first repeat, 5806–5869, includes the COOH terminus of *nrdA* protein and the second repeat, 5892–5955, is in a noncoding region. In addition, there is a partial 11-base direct repeat (Fig. 2; 5975–5985).

The DNA sequence was scanned for inverted repeats (including stem–loop structures) using the computer program of Larson and Messing (16). A stem–loop structure with a stability of −6.4 kcal (1 cal = 4.184 J) (21) (−5.2 kcal without the two base pairs at position 3373) was found at position 3366 (Fig. 2). The location of this structure is adjacent to the −35 region of the possible promoter and may be a site of regulation of gene activity.

DISCUSSION

Using the M13 dideoxy sequencing technique, we have determined the nucleotide sequence of the *E. coli* operon for the ribonucleoside diphosphate reductase genes. The derived amino acid sequence of both subunits is in good agreement with the amino acid composition as determined from the proteins directly (17). There are approximately twice as many cysteine residues (22 vs. 10.5) and tryptophan residues (9 vs. 5) in the composition of the derived sequence from a K-12 strain compared to the published composition of the B1 protein from a B strain. These differences may be due to the

FIG. 2. DNA sequence of the 8557 nucleotides from the *Bam*HI/*Pst* I region (Fig. 1, fragments B and C) of plasmid pPS2 and the derived amino acid sequence of the *nrdA* and *nrdB* genes. Regions underlined sequentially include the −35 and −10 region, a stem–loop structure near the −35 region, a ribosome binding site, the amino acid corresponding to two different NH_2-terminal chymotryptic peptides previously reported (17) in a B strain, COOH-terminal amino acid of B1, two direct repeats as well as a partial repeat, a ribosome binding site, NH_2-terminal amino acid of B2, COOH-terminal amino acid of B2, and a possible termination of mRNA transcription. The dots above the direct repeat represent nucleotide insertion found in one repeat but not in the other.

different analytical procedures used or to strain differences. Since the NH_2 termini for the α and α' subunits of *nrdA* differ from the NH_2 terminus of the predicted sequence, it seems likely that *nrdA* is translated as a precursor that is processed into either α or α' by the removal of the methionine and asparagine or the first 25 amino acid residues from the NH_2 terminus. The only processing of the B2 appears to be the removal of the NH_2-terminal methionine.

Table 1. Comparison of the amino acid composition of subunit B1 (α α') and subunit B2 (β) of ribonucleoside diphosphate reductase derived from the DNA sequence of a K-12 strain and published amino acid analysis of the protein from a B strain (17)

Amino acid residue	K-12		B	K-12	B
	α	α'	$1/2(\alpha + \alpha')$	β	β
Ala	63	63	56.5	22	25.5
Cys	22	22	10.5	5	5
Asp	41	39	76	25	37.5
Asn	30	28		17	
Glu	40	39	72.5	32	50
Gln	27	26		23	
Phe	27	27	25	17	15
Gly	38	37	43.5	13	16.5
His	16	15	17	8	7.5
Ile	48	48	44	28	24
Lys	43	41	38	16	15
Leu	72	68	67	37	34.5
Met	17	17	13	8	7.5
Pro	32	32	30	13	13.5
Arg	56	53	35.5	17	17
Ser	57	56	41.5	26	21
Thr	53	51	32.5	20	18
Val	50	48	35	25	24
Trp	9	9	5	7	6
Tyr	33	33	35	16	13

Polypeptides α and α' were analyzed in the heterodimer form. Glutamine and glutamic acid were determined as glutamic acid, and asparagine and aspartic acid were analyzed as aspartic acid.

The sequence presented here is an important step in the long-term goal of understanding the catalytic mechanism and the allosteric regulation of ribonucleoside diphosphate reductase at the molecular level. Eriksson (22) has found that the allosteric sites could be photoaffinity-labeled, while Thelander *et al.* (23) found that the inhibitor 2'-deoxy-2'-chloro CDP causes a modification of the oxidation–reduction dithiols of protein B1. Isolation and sequencing of the appropriate peptide from a modified enzyme would allow identification of important regions of the primary sequence presented in this paper. In addition, the sequence itself is important for the comparison to the structure of ribonucleotide reductase genes from other organisms. The B1 polypeptides contain two thiols that are oxidized to a disulfide during the enzymatic reaction. Reduction of the disulfide is catalyzed by either thioredoxin or glutaredoxin (1). The active site of thioredoxin and glutaredoxin as well as thioredoxin reductase (24) contain two cysteines separated by two amino acids, which allows the formation of a stable 15-member ring structure containing the disulfide. The B1 polypeptide should have a similar structure. The only place this amino acid sequence occurs is near the COOH terminus at position 5505 (Fig. 2). Site-directed mutagenesis (25) will be useful in confirming that this sequence is part of the active site.

In addition to obtaining the primary sequence of the proteins encoded by the *nrdA* and *nrdB* genes, the DNA sequence will be useful in the long-term goal of understanding the regulation of expression of the genes. Work is in progress to determine the region 5' to the structural genes that is involved in regulation and to identify nucleotide changes in mutants that have altered regulation.

We thank Kris Kohn and Bonnie Allen for their help in preparing the manuscript. This work has been supported by Grant DE-AC02-81ER 10901 from the Department of Energy (J.M.), National Institutes of Health Training Grant ST32 GM07094 (J.C.), and Public Health Grant GM20884 (J.A.F.).

1. Thelander, L. & Reichard, P. (1979) *Annu. Rev. Biochem.* **48,** 133–158.
2. Lammers, M. & Follmann, H. (1983) *Struct. Bonding (Berlin)* **54,** 27–91.
3. Hanke, P. D. & Fuchs, J. A. (1983) *J. Bacteriol.* **156,** 1192–1197.
4. Fuchs, J. A., Karlström, H. O., Warner, H. R. & Reichard, P. (1972) *Nature (London) New Biol.* **238,** 69–71.
5. Fuchs, J. A. & Karlström, H. O. (1973) *Eur. J. Biochem.* **32,** 457–462.
6. Fuchs, J. A. & Karlström, H. O. (1976) *J. Bacteriol.* **128,** 810–814.
7. Eriksson, S., Sjöberg, G. M., Hahne, S. & Karlström, O. (1977) *J. Biol. Chem.* **252,** 6132–6138.
8. Collins, J., Johnson, M., Jorgensen, P., Valenlin-Hansen, P., Karlström, H. O., Gautier, F., Lindenmaier, W., Mayer, H. & Sjöberg, B. M. (1978) in *Microbiology—1978*, ed. Schlessinger, D. (American Society of Microbiology, Washington, DC).
9. Platz, A. & Sjöberg, B. M. (1980) *J. Bacteriol.* **143,** 561–568.
10. Hanke, P. D. & Fuchs, J. A. (1983) *J. Bacteriol.* **154,** 1140–1145.
11. Hanke, P. D. & Fuchs, J. A. (1984) *Mol. Gen. Genet.* **193,** 327–331.
12. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
13. Hu, N.-T. & Messing, J. (1982) *Gene* **17,** 271–277.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
16. Larson, R. & Messing, J. (1983) *DNA* **2,** 31–35.
17. Thelander, L. (1973) *J. Biol. Chem.* **248,** 4591–4601.
18. Konigsberg, W. & Godson, G. N. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 687–691.
19. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254,** 34–38.
20. Siebentist, U., Simpson, R. B. & Gilbert, W. (1980) *Cell* **20,** 269–281.
21. Tinoco, I. Jr., Borer, P. N., Dewgler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. & Gralla, J. (1973) *Nature (London)* **246,** 40–41.
22. Eriksson, S. (1983) *J. Biol. Chem.* **258,** 5674–5678.
23. Thelander, L., Larsson, B., Hobbs, J. & Eckstein, F. (1976) *J. Biol. Chem.* **251,** 1398–1405.
24. Hoog, J.-O., Jörnvall, H., Holmgren, A., Carlquist, M. & Persson, M. (1983) *Eur. J. Biochem.* **136,** 223–232.
25. Hutchison, C. A., III, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P. & Smith, M. (1978) *J. Biol. Chem.* **253,** 6551–6560.