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 NH₂ 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 $\lambda dnrd$ 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, [³²P]-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, *Fnud*II, *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 *Hinc*II, *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. 1A, fragments B and C), four master probes containing either a 3.5-kb *Bam*HI/*Eco*RI (Fig. 1A, fragment C) or a 5.0-kb *Eco*RI/*Pst* I fragment (Fig. 1A, 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. 1A gives a partial restriction map of plasmid pPS2 (9). The BamHI/Pst I region (Fig. 1A, 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. 1B) 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 NH₂-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 NH2-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 NH₂-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 NH₂ 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 NH₂-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 NH₂ termini was also determined. Peptide 1 contains an NH2-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 NH₂-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 NH₂-terminal glutamic acid would contain 774 amino acid residues and have a molecular weight of 87,287, and the polypeptide with an NH2-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 NH₂ 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 NH₂ 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

t MICCSGCETSSGATCAGGATCEGGATCEGGITESGATCAGTGGCCGATGGCGGCGTTACTTCTIGCGATCGCAGATACCAGTCGTTSTTATCTICCACCAGCGTGTAGTCATACGCICC 123	4839 LYS PROLEU ASM ASP VAL ASM ASP GLU ASM GLY GLU ILE ALA LEU CYS THR LEU SER ALA PHE ASM LEU GLY ALA ILE ASM ASM LEU ASP AGA CCS CYS AAC SAC SYC AAC GAC GAG GAG CASC GGT GAA AYC GCG CYS TST AGS CYG TCY GCT YCT AAC CYG GGC GAA TI AAT AAC CYG GA
CATATISACEIRECESCIESCIESCABACIBAACIESEATISIIIIBAAACIBOBIBAACISCBAATCASCIESEBAATCAACCACIIIAATESCIEISEBAIGISEBCIACAATECCIETAAT 243 Canatisaceire and an	4729 BLU LEU GLU LEU ALA ILE LEU ALA VAL ARG ALA LEU ASP ALA LEU LEU ASP TYR GLK ASP TYR PRO ILE PRO ALA ALA LYS ARG GLY GLU LEU GLU LEU ALA ILE LEU ALA VAL ARG ALA LEU ASP ALA LEU LEU ASP TYR GLK ASP TYR PRO ILE PRO ALA ALA LYS ARG GLY
UNA THE CHECK SHARE OF THE SHARE AND THE SHARE AND	SAR LIS DRA BHG LIS GLA HIT LIS GLE STI CEI SCH CHI SHC GLE LIS CIG GHI THI CHO GHI THE CHO HIT SCHOOL CHI SHC GLE CHO GHI THE CHO HIT SCHOOL CHI SHC GLE CHO CHO HIT CHI CHO HIT SCHOOL CHI SHC GLE CHO CHO HIT CHI CHO CHI SHC GLE CHO CHO CHO HIT CHI CHO CHI SHC GLE CHO
493 S1166TA11FAGECTACTG1CAACG1CACCATCG6166CAT1CACCAGC61CCCATC6CC51TAAATAG16A66CATTAA666C6AC66AAGTATCCT6CCC66TCA66CSCAAAACC	ACK HE OF AND AND HE CO GET ALL CELLINE TE AND THE ACT TO GET TAC THE CTS AND CAC GET ANA CEC TAC TEC GAC GEC AGE GEC.
503 Stacca (de tester ta tecce) terra tarcatear tercter ter ter terececcatarcearcettecccateaccarcecale tercerarca	ASM ASM LEU THR HIS LYS THR PHE GLU ALA ILE GLN TYR TYR LEU LEU LYS ALA SER ASM GLU LEU ALA LYS GLU GLN GLY ALA CYS PRO AMC AMC CIG ACB CAT AMA ACC TIC GMA GCC ATT CAG TAT TAC CIS CIG AMA GCC TCT AMT GMG CIG GCG AMA GMG CAM GBC GCG TGC CCG
723 ATATCAGEARCATEACCTITEASSEISECTICITEGACAGETACEGECEGACIGEGITETACEGETSSEGTEAGCTECAGTETACCAGECECEGITITAGTCAATSTGETACCCTCATECT	5199 TRP PHE ASM GLU THR THR TYR ALA LYS GLY ILE LEU PRO ILE ASP THR TYR LYS LYS ILE TRP ILE PRO SER LEU MET SER ARG CYS ILE
943 GATESCEGECTECTECTECEGECATCAGTEGESCECCACTECETETETETECECCECATCAACTECCACTTECACSTEGEGEGECATTTECAATATCACETECETECEECATCTEGAAC P. 1	TIGE TTI AAC GAA ACC ACT TAC SCB AAA GGE ATC CTE CCS ATC GAT ACC TAT AAG AAG ATC TEG ATA CCA TEG CTA ATE AEL CEC TEC ATT 5289
TOS SERVICES AND A CONTRACTOR AND A CONTRACT AND A CONTRACT AND	THE THE ELV LYS LED CTS VAL SEN BLW SEN LYS AND THE VAL LYS VAL THE YAR THE LED SEN ALL LED HET FAD SEN DED THE SEN ALL THE SEN ALL DATE THE SEN ALL DATE ALL DATE ALL THE SEN ALL DATE ALL THE SEN ALL DATE ALL DATE ALL THE SEN ALL DATE ALL DATE ALL THE SEN ALL DATE ALL THE SEN ALL DATE ALL DATE ALL THE SEN ALL DATE ALL THE SEN ALL DATE ALL THE SEN ALL DATE ALL DATE ALL THE SEN ALL DATE ALL THE SEN ALL DATE ALL DATE ALL THE SEN ALL DATE ALL DATE ALL THE SEN ALL DATE ALL DATE ALL DATE ALL THE SEN ALL DATE ALL DATE ALL THE SEN ALL DATE AL
SECSATTISTESTISTEGECCAMATAACTATTATTETTECCAGACTTACCTECCEGTCECTCECATCGACCACCACETAEGTAGACECCATCTTCCCETCECCEGTATTECAMATTECCEC 1203	SUTT BLN ILE SER ASM ALA THR ASM BLY ILE BLU ARG ARG ALA VAL THR SER ALA SER LYS ARB ARG LYS THR VAL PHE CYS ALA ARG TRP CYS CAR ATC TCT AAC SEC ACT AAC SET ATT SAA CBE CBE BEE STT ACS TCA SEA TCA AAG CBT CBA AAG ACG GTA TIT TBE GEC AGG
ACC DECARGATETACCACATECTEACTERCESATAACCETICCECTECTECCACTERCTICAAGEECCATEGTISTATEATETECECECTECCAGETESAGAAATCAGABABATCEAGEAC 1323	S469 Arg the the see the cvs the the pro met see cvs cvs gly lys cvs arg val the met val ile cvs asm trp trp val see cvs arg
	C66 ACT AC6 ABC ACC TEC AC6 AC5 CCT AT6 ABC TEC TET 666 AAA TEC C66 ETA AC6 AT6 ETT ATC TEC AAC TEG TAG ETA TCA TEC AGA 5359
	ASN LEU SER ILE SER ARB SER LEU PRO THR PRO THR THE ILE ARB HIS ALA SER ARB BUN BLU LTS THE ARD HIS ALA SER ARB An ITA ICS ATC AGT CSA ICT CTG CCA ACA CCA ACT ACG ATC LGT CAC GCT TCC CGT CAG GAA AAG TSC CCG ATG CAG CAG ITG CTG AAA
1883 CITECEBATESTCCT65CAA16C5TATESECEACATTCATCAS66ABTTCATCAS66ABTTCAEC6CCA666TAACTTCAEC6TTTTCAATCISCATCTCAECASTAAA6TCATT6TTATCT6CATTAA	SOMY ASP LEU LEU THR ALA TYR LYS PHE GLY WAL LYS THR LEU TYR TYR GLN THR PRO VAL THR ALA VAL LYS ASP ALA GLN ASP ASP LEU GLY Gar TTE FIT ATT FAT AGA TTE SEG GTC AGA ATA TTE TAT TAT GAS ACA CEC GTG ACS GEC GTG AGA GAC GCA CAA GAC GAT CTT GGT
1903 BTACCAGATCBCC6BAACCTGTTTTG6T6ATTAACCC6GTACCAGEAATAGAGTCAACAGETCCGTCATTCTCTGTATTGCCAATAACCAGCGTTTTTCCGTCGGCAATATCAAAGGTAA	ALA VAL ASN PRO GLY ARG TRP LEU GLU THR ALA HIS VAL ARG SER ASP ILE GLU HET ARG ILE GLY LYS ARG LEU ILE PRO VAL ARG LEU
1923 CITEACTIRABECESABATACATARABECERCEGECGEGEAGAGBARGBACEATETECEATAGECTGEGECTATTGITETEATEGACTARAGECETEEGETGTAGETGTAACA	GCC GIC AAT CCA SGA CGA TGG CTC GAA ACG GCG CAT GTA AGA TCT GAT ATT GAG ATG CGG ATC GGT <u>AAA CGC CIT ATC CCG GTA CGG CTC</u> 5829
2043 Gara ta transfarge aggesecs i caced tatise tata ta i csclcege cata cicege aggeste tatista aggeses tatista ctatega tageste citects i Data ta transformed aggeste cata constant a csclcege cata cicege aggeste cata constant source cata constant sourc	BLY <u>LEU *** SCCTSATAASACECECCASESTEGEATEASEC</u> TEEGGATECAGEATECAGEGTE <u>AAGECTTATEEGGTAEGGETEGGATTETAGGEETGATAAGAEGEEGEAGEG</u> Get TTS TAG
2183 CECTATTATTATCEGTAACATCEATTECCCCECCATCECCATCECCATCACTETAACTTETTEATECTTEGTTETTATCCAAAAATAACATCACCTTAAATAACETCATTATTAATEETAT 2981	5946 <u>Cecatcagec</u> acageategegeteraaat <u>eccttatccgeca</u> ttaaactcccaa <u>cageba</u> cacagtc <u>met ale</u> tyr thr thr thr phe ser glu thr lys asn asp glu ate sea tat are are to the tag cas are and an tat cas
AAT 15050CACCT 1 150CATCA T TABOBATGI TA 1 1505AAACA 165051 TAG 16ACACSTAAA 1054056CACCAG TATOSTAGTACCAGAAGAA TAGAT 16050CACCA TASOC6C 2403	6051 IFILIYS FILIPRO MET PHE PHE BEY SIN PRO VAL ASN VAL ALA ARS TYR ASP BLN GLN LYS TYR ASP FLE PHE BLU LYS LEU ILE BLU LYS
CIBCGACGITACCGGAAAAAATAACATCAGTAAGATTCAGCGITGAGTTTTCTTTAGCAAATATGGCCCCGCCGTTATTATATTCTCCCTGTAACGGTGTTATTGGCAAACAGGTCATTC 2523	CTC ARA GRA CCG ATG TTC TTT SGT CAG CCG GTC AAC GTG GCT CGC TAC GAT CAG CAA ARA TAT GAC ATC TTC GAA AAG CTG ATC GAA AAG 6141
	GLM LEU SER PHE PHE TRP ARG PRO GLU GLU VAL ASP VAL SER ARG ASP ARG ILE ASP TYR GLM ALA LEU PRO GLU HIS GLU LYS HIS ILE CAG CICI CICI TIC TIC 16G CGT CCG GAA GAA GIT GAC GIC CCC CGC CAC CGT ATA GAT TAC CAG GCG CIG CCG GAG CAC GAA AAA CAC ATC
	6231 PHE ILE SER ASM LEU LYS TYR GLM THR LEU LEU ASP SER ILE GLM GLY ARG SER PRO ASM VAL ALA LEU LEU PRO LEU ILE SER ILE PRO
2883 11 TISTAAATTISTCAACCCSSICSA TSSCTATTSSTATTASA TSAATSAATAATTICASTSCSTCATAATTCAASTAATTCASSITAATACCTASSTATATTAACTAAATTAASTCATS	111 ALC ABLE ARE LIE ARK THI LANS ALS LIE LIE DAT TEL HIT DAE DET LET MEL ELE MAL DIE DUE LIN THE LEE ATT LEE 5321 CILITECHT TWO TOO VAL ETH TWO TOO ALA DHE SED EIH THO TIE HIS SED ARG SED TWO THO HIS THE TIE ARG ASM THE VAL ASM ASP
3003 Artartitestritatartargectaratiracararteectitaccatitarcartarcegartagaraaccatticeccatcarcargictcttacaticectiatatte	SAN CIE SAN ACC IGG GIC SAN ACC IGG GCG ITC ICA SAN ACG ATT CAT ICC CGI ICC TAT ACT CAT ATC ATT CGI AAT ATC GIT AAC GAT 6411
3123 Carlisa tratatarsa tiatstga tgactog tecttagatcaattittecaatcattageaaaaga taataageca tota tateaattitatetaactatta tseccotteaagaaat Tateaa	PRO SER VAL VAL PHE ASP ASP ILE VAL THR ASN GLU GLN ILE GLN LYS ARG ALA GLU GLY ILE SER SER TYR TYR ASP GLU LEU ILE GLU CCB TCT GTT GTE TIT SAC GAT ATC GTC ACC AAC GAG CAS ATC CAG AAA CGT SCG GAA GGG ATC TCC AGC TAT TAC GAT GAG CTG ATC GAA
	6501 NET THR SER TYR TRP HIS LEU LEU GLY GLU GLY THR HIS THR VAL ASN GLY LYS THR VAL THR VAL SER LEU ARG GLU LEU LYS LYS LYS NED DE NER LEG THE AND THE THE THE THE AND THE
ACTI <u>BCARGAGEGE</u> CATTITCAC <u>AC</u> TA <u>TCTIBCA</u> BTGAA:CCCCAAACATACCCCC <u>TATATAB</u> GIGTTCTAAGCAGCTTCCCGTACTACAGGTAGTCIGCATGAAACTATTGCGGAAAGA 3483	NTO NUC HEL THE LIT LIE LIS DEL DIN BET NUC SHE HEL SHE HEL SHE HEL DE HEN BET DIN HEL DIT HEL DIT HEL DIE DE UN DIE HEN HEN HEN 6591 JET TYD JET TYD JET HET DER VAN ASM ALA JET STU ALA THE ARE PHE TYR VAL SER PHE ALA CYS SER PHE ALA PHE ALA GEU
ATICCRAMAAC <u>ABGTA</u> GGACATAC <u>HET</u> ASM <u>ELU ASM LEU LEU YAL THR L'IS ARG ASP ELY SER</u> THR GLU ABG LLE ASM LEU ASP L'IS LIE HIS ANG Ats art cas art cis cits ets aca abg coc bac bet age aca abg coc atc ant cic gac aba atc cat coc 7578	CTG TAT CTC TEC CTS ATE AEC STT AAC GCS CTS BAA SCS ATT CGT TTC TAC GTC AGC TTT ECT TET TCC TTC GCA TTT SCA GAA CGC GAA
VAL LEU <u>HSP</u> THP <u>Ala Ala Blu bli leu</u> his ash <u>yal ser ile ser</u> gun val glu <u>leu</u> arg <u>ser</u> his ile glu phe tyr ASP glv ile lis Sti tisg gat tgg ges sea gaa sea ets eat aac git tee att tee ere gag etg ges etg ere ere tit ere tit tag tit tag	LEU MET BLU BLY ASH ALA LYS ILE ILE AMS LEU ILE ALA AMS ASP BLU ALA LEU HIS LEU THR BLY THR BLN HIS MET LEU ASH LEU LEU TIS ATG GAA GAC AAC GCC AMA ATT ATT CSC CTG ATT GCC CGC GAC GAA GCC CTG CAC CTG ACC GAC CAG CAT ATG CTG AAT CTG CTG 6/71
JOOT THR SER ASP ILE HIS GLU THR ILE LIS ALK ALA ALA ASP LEU ILE SER ARG ASP ALA PRO ASP TIR GLW THR LEU ALA ALA ARG LEU Thr Gen ato can san acc ati ato ang got got got got got oto tot oto got gat got can tat can bat cit got got gog	ANG SER GLY ALA ASP ASP PRO BLU MET ALA BLU ILE ALA BLU BLU CYS LYS BLN GLU CYS TYR ASP LEU PME VAL GLN BLA ALA GLN GLN CSC ARC SEC BCS GAC GAT CCT GAG ATS BCS GAA ATT BCC GAA GAG TST AMG CAG GAG TSC TAT GAC CTG TTT GTT CAG GCA GCT CAL
7759 Ala ile PNE his leu Arg lis lis ala thr ala ser leu arg PRO Ala leu TYR ASP his VAL VAL Lis Net Val suu net Gli Lis Seg arg iil car che get ann ana get arg bec agt tis agg cog cot ged ets tar gar car etg gtg and atg gtg gar atg	GUILYS ASP TRP ALA ASP TYR LEU PME ARG ASP GLY SER MET ILE GLY LEU ASM LYS ASP ILE LEU CYS GLN TYR VAL GLU TYR ILE THA Gar ama gac tgs gcg gat tat CTS tic CGC gac ggt tcg atg att ggt ctg aat aaa gac att ctc tgc cag tac gtt gaa tac atc acc
3849 Tra Asp asm mis leu leu blu asp tra thr glu glu glu blu phe lrs glm met asp thr phe ile asp mis asp arg asp met thr phe ser Tra gat ant cat cits cits saa gac tac acs gaa gaa gaa gag cat cats gac acc tit aic gat cac gac gat ats acc tit t	AYSI ILE ARG MET GUN ALA VAL SULY LEU ASP LEU PRO PME GUN THR ARG SER ASM PRO ILE PRO TRP ILE ASM THR TRP LEU VAL SER ASP Ami atc csi atg cag gca gic ggt tig gat ctg ccg itc cag acg cgc tcc aac ccg atc ccg tgg atc aac act tgg ctg stg tct gat
1939 The Ala Ala Val Lys Gin Leu Glu Giy Lys Tyr Leu Val Gin Arm Arm Val The Giy Glu Ile Tyr Glu Ser Ala Gin Phe Leu Tyr Ile Tat gen gen gyt aga car eys gag gen gaat at org sta car age cer beta aft gen gaat at tat gag age set car tit gi	7041 Ash wal gan wal ala pro blin glu val glu val ser ser tyr leu val gly glin tie asp ser glu val asp thr asp asp leu ser asm Arg gtg cag gtt gct gcg gan gtg gan gtg gan gtg rat tag to gg cag att gac tos gan gtg gac acc gac gat tig ast aa
4029 LEU WAL ALA CYS LEU PHE SER AGN TYR PRO ARG GLU GLY LEU GLN TYR VAL LYS ARG PHE TYR ASP ALA VAL SER THR PHE LYS ILE CYS SYT GYG GRU THE TYR TYR TYR AGA TAR TYR PRO ARG GLU GLY FR GAS TAR TYR AGA FRI TYT TAR GAR RYS SYT TYR AMA AT	7151 Phe Gui L <u>eu nee</u> teseccecettaccteceatactecegacacaactectecegeateaacaccettecette
THE STATE DEG THE THE THE THE THE ARE THE DEG DET WAR DEC DEG WAR AND DE WARD DET THE THE DEG DEG OF THE MEM THE 4117 SER 15U PRO THE PRO ILE MET SER GLY VAL ARG THE PRO THE ARG GLN PHE SER SER CYS VAL LEU ILE GLU CYS GLY ASP SER LEU ASP	7246 Caststeecbaadsttactscebsttctstescacacbect6sttbcabsteaastbact6sattbccbaaccsttabctttattča <u>eccbs6sbaantttt</u> 6cct5ttstbccb
TES ETS CER AES CEA ATE ATE TEE SEC BTE EST AEE CEB AET CEB CAE TEE ABE TEE ATE ATE ATE ATE ARE TEE SAE AGE C'A GAT 4209	7366 Benamages bata teanatear teattest tetrateccasa tacaacect tatece tettateecct acaacea tacatese stata tacatese tata Januari
SER ILE ASM ALA THR SER SER ALA ILE VAL LYS TYR VAL SER GLM ARG ALA GLY ILE ALA SER THR PRO GLY VAL PHE VAL ARG TRP VAL TOC ATC AMC GOC ACC TOC AGO GOG ATT GTT AMA TAC ST! TOC CAG GGT GOC GGG ATC GOA ATCA AGO GOG GGA GTA TOC GIG GGA	7400 TCCSCITISSEGATISSECTCCAGATATTTTTCCABCTGECSGAASTCATGATTATCSCTTTATCSCBACGACGGCGACGACGTTTTCTCCAGGTCAAGAAAACCAGCTTCTGCATTA 7A0A
4299 Ala arg phe ala val lys arg ser ile pro ala ala phe arg ser thr thr phe pro asp ser gly glu ser cys ser arg arg cys	CETTETETTTECATABATATBECEMACATACAACAGECATECTGACGATTAATGCTATECATTTETTAAACGECAGEGECACTECTTTCAACATGECTTGECGTACTTCGTCAGA 7726
DEL COR ITE DED STE DIE ARD DET IEE RIK EED DET SER THE EDT IET ALR ALR THI LER DAL ADE DET DAR TEE TOE TEE ADE 4389 Af Ally ADE ADE CYE ELV ACH VAL DUE TYD DOE NET TOD VIE TET ETH VAL ETH CED TETH TET VAL TETH IVE ACH ADE ADE CH VAL ETH	TANGGEGATACTECATECTEGECATACCAETCAECAATECTEATGAACCCCECCATATCTTCAETCACCAECAACGCTEGECATTCACCCTCAATTTCACCECTTCACCAAAAACGATU 7846
BC5 66C C66 C66 C66 C6C AAC 615 TIC TAC CC5 ATE T65 CAT C15 6AA 615 6AA ABC C15 C16 615 TIS AAA AAC AAC C61 651 515 6AA 4479	I ICEBECAUSA ISALACCEBECCEGI ICENSI ICE I TAATI ACCEBECANCE I CACEGANEANTESI I SEGECEBANCESA TAACETACESAAN TAAA 1966 Taanas i teri teri teri teri teri teri teri te
ELY ASM ARG VAL ARG HIS MET ASP TYR GLY VAL GLM ILE ASM LYS LEU MET TYR THR ARG LEU LEU LYS GLY GLU ASP ILE THR LEU PHE Geg aac coc gig cgi cat atg sac tac sgg gia caa atg aac aag atg aat is atg tat acc cgi cig cig aag gat aga gat atg acc cig tic	10000011011000011000011000001001100110
4564 SER PRO SER ASP WAL PRO BLY LEU TYR ASP ALA PHE PHE ALA ASP GLN GLU GLU PHE GLU ARG LEU TYR THR LYS TYR GLU LYS ASP ASP ASP CRE TER ASP AND THE THE ASP AND THE THE ASP AND	8206 BCTTBACACCSCTTTTACCCTTCATTTECAGESCATAGATTCCACCAGCCACTACCACTAATAGTSCECCCAGCAGATAAAATGTSAATTTAATCAGSCSGATACAGCGATSAATTTTA
mee use the one off led dog its the one due the the led but led bar bar bar fit bar CST CTS TRI ACC man the bar bar bar bar bar 4859 SER ILE ARE LYS BLN ARE WAL LYS ALE WAL BLU LEU PRE SER LEU HET NET RIN GILL ARE ALE GER THAR RIV ARE TIF TVO THE CHIN ACM	8326 Teancabbattabcatesecttacaaceagattaacsetsaaageattseagetsaatt6t6tteetteetteetaaeeaaetttetatesaaeaattteettsetsetse Taanaasaa
AGC ATC CBC AAG CAG CAT GTG AAA GCC GTT GAS CTG TTC TCG CTG ATS ATG CAG GAA CGT GCG TCT ACC GGT CGT ATC TAT ATT CAG AAC 4749	8446 Margelebitagaagbeagileactgittitactacaan tgibeatggaacaat tagiggebattggaanaattigatatgggecabatggeebatggaegelateree Communications and the second
VAL ASP HIS CYS ILE THR HIS SER PRO PHE ASP PRO ALA ILE ALA PRO VAL ARG GLN SER ASM LEU CYS LEU GLU ILE ALA LEU PRO THR Git gac cac tgc ata acc cat agc ccs tit gat ccs gcc atc gcg cca gtg cgs cag tct agc ctg tgc ctg gag ata gcc ctg cc	

FIG. 2. DNA sequence of the 8557 nucleotides from the BamHI/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 NH2-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, NH₂-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 NH₂ termini for the α and α' subunits of *nrdA* differ from the NH₂ 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 NH₂ terminus. The only processing of the B2 appears to be the removal of the NH₂-terminal methionine.

Table 1. Comparison of the amino acid composition of subunit B1 ($\alpha \alpha'$) 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	В	K-12	В
	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.).

- Thelander, L. & Reichard, P. (1979) Annu. Rev. Biochem. 48, 133-158.
- Lammers, M. & Follmann, H. (1983) Struct. Bonding (Berlin) 54, 27-91.
- Hanke, P. D. & Fuchs, J. A. (1983) J. Bacteriol. 156, 1192– 1197.
- 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.
- Fuchs, J. A. & Karlström, H. O. (1976) J. Bacteriol. 128, 810– 814.
- Eriksson, S., Sjöberg, G. M., Hahne, S. & Karlström, O. (1977) J. Biol. Chem. 252, 6132–6138.
- 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.
- 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.
- 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.
- 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.
- Thelander, L., Larsson, B., Hobbs, J. & Eckstein, F. (1976) J. Biol. Chem. 251, 1398-1405.
- Hoog, J.-O., Jörnvall, H., Holmgren, A., Carlquist, M. & Persson, M. (1983) Eur. J. Biochem. 136, 223-232.
- Hutchison, C. A., III, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P. & Smith, M. (1978) J. Biol. Chem. 253, 6551–6560.