

Nucleotide sequence of the gene coding for the large subunit of ribonucleotide reductase of *Escherichia coli*. Correction

Olle Nilsson, Anders Åberg, Tomas Lundqvist and Britt-Marie Sjöberg

Departments of Molecular Biology, University of Stockholm S-106 91 Stockholm and Swedish University of Agricultural Sciences, Uppsala Biomedical Center, S-75124 Uppsala, Sweden
Submitted February 9, 1988 Accession no.X06999

The nucleotide sequence of the nrdA gene of Escherichia coli, coding for the large subunit of ribonucleotide reductase, was originally published by Carlson et al. (1). However, the published nucleotide sequence of the nrdA gene is afflicted with scattered errors leading to stretches of frame-shifts, which have severe consequences for the deduced protein sequence. The difference in amino acid composition between the corrected and the previously deduced sequence is 20%. We have determined the nucleotide sequence of the nrdA gene in 3 different random point mutants (2), and 5 different oligonucleotide-directed mutants (Åberg, A., in preparation), in addition to the wild-type gene (Fig. 1). The corrected length of the gene is 2286 bp. The deduced polypeptide is 761 amino acid residues, i.e. 17 residues shorter than earlier stated, because the corrected sequence has an in-frame UGA stop-codon 52 nucleotides upstream the previously published (1) UAG stop-codon. In addition the deduced amino acid composition (Fig. 1) corresponds within 5% to that obtained previously with the intact protein (3). The deduced composition of the former sequence (1) differed by 19% from the analytical data.

1 M N O N L L V T K R D D G S T E R I M L D K I H R V L D M A A E G L C 33
11 ATG AAT V S I C T G T G A C A A G C C G A G G C C T A T C A T C T G A C A A T S S C A T C G C T G G G C A G A A G G A C T G 39
34 F H V S I C T G T G A C A A G C C G A G G C C T A T C A T C T G A C A A T S S C A T C G C T G G G C A G A A G G A C T G 46
100 CAT AAC GGT TGG ATT TCC GAG CTC GAG CTT GGC TCC CAC ATT CAG TTT TAT GAC GGT ATC AAG ACC TCT GAC ATC CAC GAA ACC ATT AGC GCT GCC 198
67 A D L I S S C G G A T Y Q Y L A A R L A A I F B L R K K A I F B G G O F G F E P 99
139 GCA GAC CTG ATC TCC GGT GAT GAG CCG GAT TAT CAG TAT CTC GCC CCG GCG ATC TTC CAC CTG GGT AAA AAA GCC TAC GGC CAG CCG 297
100 P A L Y D H V C M K M V E M G X Y D M H R L L E E D Y Z E E F K Q M D 132
238 CCT CCG CTG TAC GAG CTT GAG AAA ATG GTC GAG ATG GGC AAA TAC D M A T C A T C T G C T G G A A G C A C T A C A C G A A G A A G A A G C A T G C A T G 396
133 T F I D E D R D H T F S Y R A V E Q L E G K Y L V Q H R V Z G E I 165
397 ACC TTT ATC GAT CAC GGC GGT GAT ATG ACC TTC TCT TAT ACT GGC GGT AMG CAG CTG GAA GGC AAA TAT CTG GTA CAG AAC CCG ACC GGC GAA ATC 495
166 Y E S A G L V A A C L F S H Y P R E Z A L Q Y V K R F Y D 198
496 TAT GAG ACG GCC CAG TTC GTT TAT ATT CTA GTT GCC GCG TCC TGC TCC AAC TAC CCG GGT GAA CGC CCG CTG GAA TAT GTG AAG CTT TTT TAC GAC 594
195 A V S B T F K I S L P T F I M S G Y R Z F T R Q F S C V I I M A G R A L 631
595 GCG GTT TCC ACA TTT AAA ATG TCC ACC ATC ATG TCC GCC GTG COT ACC CCG ACT GGT CAG TTC AAC TCC TGC GTA ATC GAG TCC GGT 693
232 D S L N A T S R I V K Y Y S Q R A D I G I M A G R A L 693
694 GAC AAC CTG GAT TCC AAC AAC GCC ACC TCC ACG GCG ATT GTT AAA TAC GTT TCC CAC GGT GCC GGG ATC GGC ATC AAC CCG GCG GGT ATT GGT GCG 792
265 G S P I R G E A F H T G C I P F K H G E D I T L F S P S D V F G 891
752 GGT ACG CCG ATT CCG GGT GGT GAA CCG TTC CAG ACC GGC TCC ATT CCG TTC TAC AAA CAT TTC CAG ACA CCG GCG GGT ATT GGT GCG 891
298 R G G A A L F Y F M N H L E V I K H G V L R K V 891
892 CCG GCG GGT GCG GCA ACC CTG TTC TAC CCG A TG TCG CAT CTG GAG GAG CTG GCG ATC AAA CAT CTG GCG GGT GGT GGT GGT GGT GGT GGT 990
321 R H L K G E D I T L F S P S D V F G 990
991 COT GAT AAC GAC TAC GGG GTA CAA ATC AAC AAA CTG ATG TAT ACC GGT CTG CTG AAA GGT GAA GAT ATC ACC CTG TTC ACC CCG TCC GAC GTA CCG 1089
364 L Y D A F A D O E E F R L Y T K Y E K D D S I R K Q R V X A V 396
1090 CTG TAC GAC GCG TTC TCC GAT CAG CAG GAA CCG GCG TTT GAA CTT CTG TAT ACC AAA TAT GAG AAA GAC GAC AGC ATC CCG AAG CCG GGT GGT AAA GCC GTT 1188
397 E L F S L M H M Q L R A T G R I I O N V D H C H T H S F D P A 429
1189 GAG CTG TTC TCG ATG ATG CAG GAA COT GCG TCT ACC GGT GGT ATC TAT ATT CAG AAC CTG GCG GAC TGC ATT ACC CAT ACC CCG TTT GAT CCG GCC 1287
420 I A F V R G S H L C L E I R L F Z K F L M D H N G E I A L C 462
1288 ATC GCG CCA GTG COT CAG TCT AAC CTG TCC CTG GAG ATA GCC CTG CCG ACC AAA CCG CTG AAC GAC GTC AAC CAG GAG ACC GGT GAA ATC GCG CTG TGT 1386
463 T L S A I N M L G A I N M L D E L A I D V R A L D A L D A L D Y 495
1387 ACG CTG TCT GGT TTC P A A C T G C G C G A A T T A M C C T G G A T G A A G A G C T G C A A T T C T G C G G G T T C C T G C A T T G A T G A T 1485
496 Q D I E S I K T R G E R L G I H F A Y L A W D G 528
1486 CAG GAT TAC CCG ATC CCG GCC CCG AAA COT GGA CCG ATG GGT COT GGT ACG CTG GGT ATT GGT GGT ATT GGT ATT GGT ATT GGT ATT GGT ATT GGT 1584
529 K R Y S D S A N M L T T H R T F A A I O K G A T Y T C T G G L L K A S H M E L A K E 561
1585 AAA CCG TAC TCC GAC GCG ACC CCG AAC AAC CTG ACG CAT AAA ACC TTC GAA GCC ATC CAG TAT C T G G L L K A S H M E L A K E 563
562 G A C P F S A N E T Z Y A K G I L F I D T Y K K D T A I A N E F 594
1684 CAA CCG GCG TCC GCG TGG TTT AAC GAA ACC ACT TAC AAC AAC GAG ACT CTG CCG ATC GAT ACC TAT AAG AAA GAT CTG ATC ACC ATC GCT AAT GAG CCG 1782
595 L R Y D N E A L E F S I K T R G E R L G I H F A Y L A W D G 528
1783 CTG TAC GAC TCG GAG GCT CTG COT GAT CAG CCG AAC ACC CAG GCT CTG COT AAC TCC ACC TCT CTT GGT GGT ATT GGT GGT ATT GGT GGT ATT GGT 1881
628 I S H A T G I E F P R G Y V S K A S K D G I L R Q V V F G D Y Z 660
1882 ATC TGT AAC GCC ACT AAC GGT ATG GAA CCG CCG GCG GGT TAC CTC ACC ATC AAA GCC TCG AAA GAC GGT ATT GGT CCG CAG GGT GCG GAC TAC GAG 1980
661 R L H D A Y E L L M E M F G N D G Y L Q L V F G I H Q K I D Q S I 692
1981 CAC CTG CAG CAC GCC TAT GAG CTG CTG TCG GAA ATG ACC GAT GGT TAT CTG CAA CTG GGT ATC ATG CAG AAA TTT ATC CAG CAG TCC ATC 2079
694 S A T N V D F S R G K V S H Q Q L K D L T R Y R F G V 726
2080 TCT GCC AAC ACC AAC CAG CAG CTC CCG TCA GCG TTC CCG TCA GAA GAT CCG ATG CAG CAG TTT CTG AAA GAC CTG ATC ACC GTC AAC TTC GCG ATC 2178
727 K T L Y D M T R D G A E D A G D D L V F S I G D G G C E S G A C 759
2178 AAA ACA CTG TAT TAT CAG AAC ACC CTT GAC GGC GCT GAA GCA CAA GAT CTG CCG TCA ATC CAG GAT GGC TCC GAA GCC GCA TGT 2277
760 K I + 762
2278 AAG ATC TGA 2286

Figure 1. Corrected nucleotide sequence and derived polypeptide sequence for the E. coli nrdA gene.

1 Parts of the correction presented here has been available at request from J. A. Fuchs, Dept. of Biochemistry, University of Minnesota, S:1 Paul, Minnesota 55108, USA.

REFERENCES

- 1. Carlson, J. et al. (1984) Proc. Natl. Acad. Sci. USA 81, 4294-4297.
2. Platz, A. et al. (1985) J. Bacteriol. 164, 1194-1199.
3. Sjöberg, B.-M. et al. (1985) Eur. J. Biochem. 150, 423-427.