The EMBO Journal vol.5 no.8 pp.2031-2036, 1986

The bacteriophage T4 gene for the small subunit of ribonucleotide reductase contains an intron

Britt-Marie Sjöberg, Solveig Hahne, Catherine Z.Mathews[1,3], Christopher K.Mathews[1,3], Keith N.Rand[2,4] and Michael J.Gait[2]

Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala Biomedical Center, S-751 24 Uppsala, [1]Department of Biochemistry, Medical Nobel Institute, Karolinska Institute, Box 60400, S-10401 Stockholm, Sweden, [2]Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

[3]Permanent address: Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA
[4]Present address: Biotechnology Australia, 28 Barcoo Street, Roseville, Sydney, Australia

Communicated by C.-I.Brändén

The bacteriophage T4 gene *nrdB* codes for the small subunit of the enzyme ribonucleotide reductase. The T4 *nrdB* gene was localized between 136.1 kb and 137.8 kb in the T4 genetic map according to the deduced structural homology of the protein to the amino acid sequence of its bacterial counterpart, the B2 subunit of *Escherichia coli*. This positions the C-terminal end of the T4 *nrdB* gene approximately 2 kb closer to the T4 gene 63 than earlier anticipated from genetic recombinational analyses. The most surprising feature of the T4 *nrdB* gene is the presence of an approximately 625 bp intron which divides the structural gene into two parts. This is the second example of a prokaryotic structural gene with an intron. The first prokaryotic intron was reported in the nearby *td* gene, coding for the bacteriophage T4-specific thymidylate synthase enzyme. The nucleotide sequence at the exon-intron junctions of the T4 *nrdB* gene is similar to that of the junctions of the T4 *td* gene: the anticipated exon-intron boundary at the donor site ends with a TAA stop codon and there is an ATG start codon at the putative downstream intron-exon boundary of the acceptor site. In the course of this work the *denA* gene of T4 (endonuclease II) was also located.

Key words: prokaryotic intron/bacteriophage T4/ribonucleotide reductase/endonuclease II

Introduction

Ribonucleotide reductase catalyses the essential conversion of ribonucleotides to their corresponding deoxyribonucleotides (Lammers and Follmann, 1983). After bacteriophage T4 infection of *Escherichia coli* cells a new, phage-specific ribonucleotide reductase is produced (Berglund *et al.*, 1969). T4 ribonucleotide reductase and most of the enzymes involved in dNTP biosynthesis are organized in a multienzyme complex, which probably serves as a substrate channel for DNA replication in phage-infected cells (Mathews and Allen, 1983). Like its bacterial counterpart, T4 ribonucleotide reductase consists of two non-identical subunits, denoted proteins B1 and B2. A one-to-one complex of these two subunits constitutes the catalytically active enzyme, in which protein B1 contributes redox-active sulfhydryl groups (Berglund, 1975) and protein B2 contributes a unique tyrosyl radical (Berglund, 1975; Sahlin *et al.*, 1982), which is essential for activity.

The bacteriophage T4 genome contains two adjacent genes, *nrdA* and *nrdB*, coding for proteins B1 and B2, respectively, of the phage specific ribonucleotide reductase (Yeh *et al.*, 1969; Berglund, 1975). The two loci were mapped genetically in the region between the thymidylate synthase gene (*td*) and gene 63 coding for T4 RNA ligase and tail fibre attachment. Later the gene coding for endonuclease II (*denA*) was also localized to this area, between *nrdB* and gene 63 (Ray *et al.*, 1972). Previous attempts by us (B.-M.Sjöberg, S.Hahne, C.Z.Mathews and C.K. Mathews) and others (R.B.Greenberg, personal communication) at cloning of T4 *nrdB* have been unsuccessful. These strategies all relied upon the current bacteriophage T4 map (Kutter and Rüger, 1985), in which T4 *nrdA* and *nrdB* genes are positioned very close to the *td* gene. Here we report the complete nucleotide sequence of T4 *nrdB*, obtained on a 3.1 kb *Sph*I/*Cla*I fragment reaching 2.8 kb upstream of and including a short N-terminal part of gene 63. The structural gene coding for T4 protein B2 comprises two exons and one intron.

Results and Discussion

Determination of the genomic nucleotide sequence of T4 nrdB

Nucleotide sequences for the *td* gene and gene 63, surrounding the *nrd* region, were recently determined. The *td* data (Chu *et al.*, 1984) reached approximately 250 nucleotides downstream, towards the *nrdA* region. The gene 63 data (Rand and Gait, 1984) reached approximately 1.7 kb upstream, towards the *denA-nrdB* region but only the sequence of gene 63 itself was published at the time. The upstream region of gene 63 included three open reading frames (ORF 1, exon 2 and ORF 2 in Figure 1). To our surprise, a deduced protein sequence very similar to the sequence of the *E. coli nrdB* product (Carlson *et al.*, 1984) was found only 0.5 kb upstream of the gene 63 start codon. This region (denoted exon 2 in Figure 1) comprised 159 amino acid residues with 48% homology to the C-terminal part of the *E. coli* B2 protein. However, there was no obvious continuation of the T4 *nrdB* coding sequence immediately upstream of this open reading frame. On the other hand, 625 bp further upstream the original gene 63 data included another open reading frame (denoted exon 1 in Figure 1) with 77% homology to a 26 amino acid residue stretch of the middle part of the *E. coli nrdB* protein. This arrangement of two separate coding regions suggested to us that the T4 *nrdB* gene contains an intron.

We have now continued the DNA sequencing further upstream of gene 63. The complete nucleotide sequence of a 3.1 kb *Sph*I/*Cla*I fragment from this region of the T4 genome is shown in Figure 1. It includes the *nrdB* coding sequences (exons 1 and 2), one unassigned open reading frame (ORF 1) within the *nrdB* intron sequence, one open reading frame (ORF 2) probably representing the *denA* gene (see below), and the N-terminal part of gene 63.

There is an AGGA ribosome-binding site six nucleotides up-

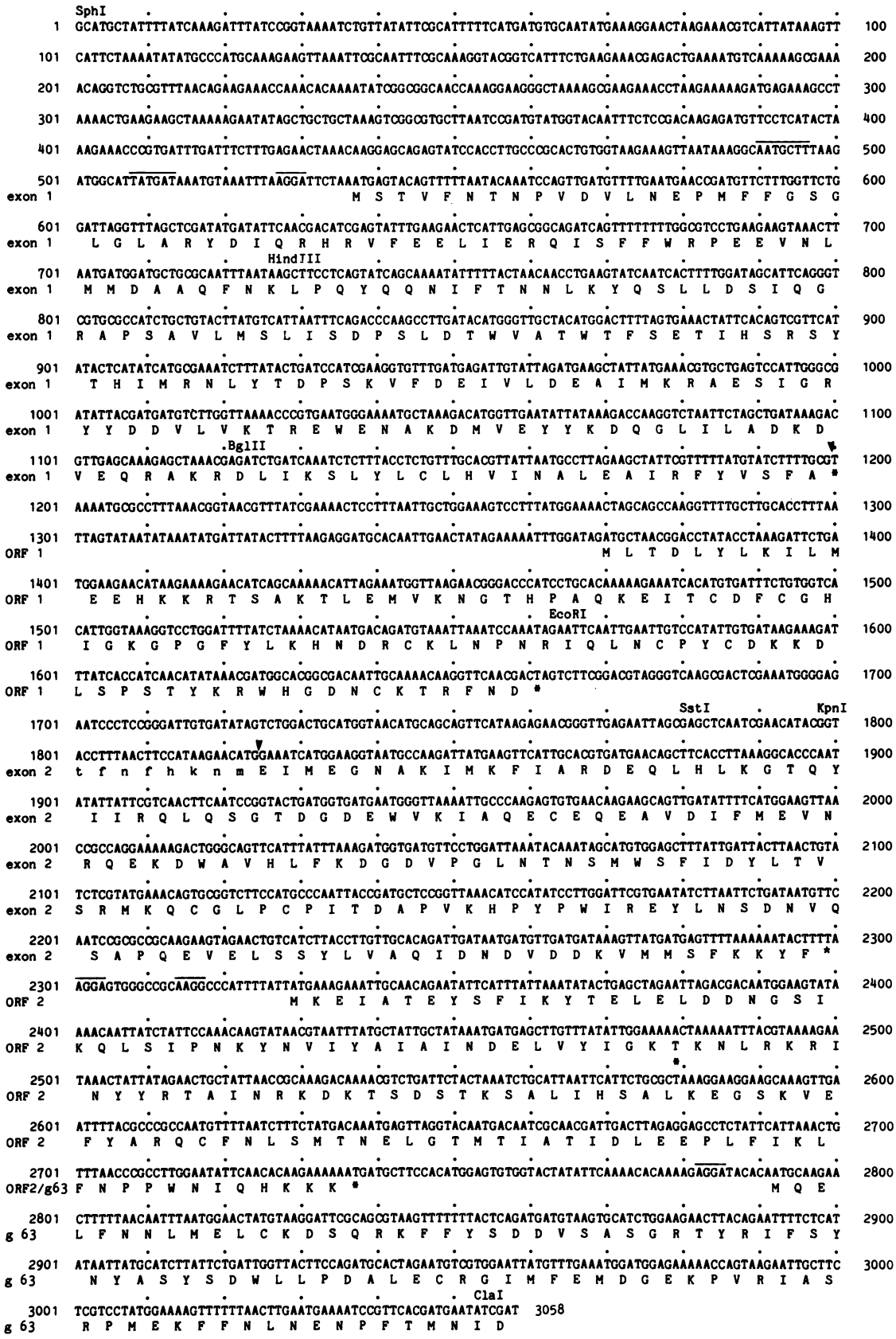

Fig. 1. DNA and protein sequence of the *nrdB* gene of bacteriophage T4 coding for the small subunit of ribonucleotide reductase. Exon 1 and 2 refer to protein B2, ORF 1 is an unassigned open reading frame within the protein B2 intron, ORF 2 most likely refers to endonuclease II (the *denA* protein), g63 refers to RNA ligase (the gene 63 protein). The presumed splice-junctions of protein B2 have been indicated with arrows above the nucleotide sequence. The asterisk at nucleotide 2579 denotes the presumed position of the *nd*28 mutation (T→C transition) of the *denA* gene. Some possible −10 and −35 promoter regions and ribosome-binding sites have been underlined. The cleavage sites of restriction endonucleases *Bgl*II, *Cla*I, *Eco*RI, *Hin*dIII, *Kpn*I, *Sph*I and *Sst*I have been indicated.

stream of the initiator ATG of exon 1 of T4 *nrdB* (Figure 1). Interestingly, a perfect *mot* promoter sequence (Brody *et al.*, 1983) AATGCTT is also present 30 nucleotides upstream of the ribosome-binding site. There are at least two plausible −10 promoter regions, TATGAT and TAAATG, which would result in RNA transcripts starting 7 or 1 nucleotides upstream of the ribosome-binding site. However, none of these −10 regions have a corresponding plausible −35 region, apart from the already mentioned T4 middle promoter region. There are no obvious promoter regions or ribosome-binding sequences immediately upstream of exon 2 or ORF 1. Close to ORF 2 there is an AGGA sequence 24 bp upstream and an AAGG sequence 11 bp upstream of the ATG start codon. Both are far from optimal, and there are no obvious promoter sequences further upstream.

Comparison of the T4 nrdB intron to other prokaryotic and eukaryotic introns

The first example of an intron within a prokaryotic gene is the nearby T4 *td* gene, coding for the bacteriophage-specific thymidylate synthase enzyme (Chu *et al.*, 1984). In this case a 1017 bp intron sequence divides the gene into two parts, exon 1 and exon 2. The junction boundaries in the *td* sequence do not conform to the common eukaryotic consensus splice sequences, but show some similarities to the junction sequences of *Tetrahymena* rRNA and class I mitochondrial mRNA of fungi (Cech, 1983). A specific feature of the *td* intron is that it starts with a TAA stop codon in the reading frame of exon 1 and ends with an ATG start codon in frame with exon 2. The methionine at the 3′ end of the intron is not incorporated into the translation product.

Comparison of the putative exon−intron junction sequences of the *nrdB* gene with the corresponding sequences of the known *td* gene shows several similarities. It is obvious from Figure 1 that exon 1 of *nrdB* may end at a TAA stop codon as was found for exon 1 of the *td* gene. Similarly the 3′ end of the T4 *nrdB* intron is most likely the ATG codon at position 1822 in Figure 1, in analogy with the *td* intron. A splice junction at this point would result in a T4 *nrdB* protein having an eight amino acid deletion at the splice junction as compared with the *E. coli nrdB* protein (see below).

Unlike the *td* intron, the anticipated *nrdB* intron does not conform to the T consensus of *Tetrahymena* and mitochondrial introns at the 5′ splice site (Figure 2). However, there are 16 identities out of 34 nucleotides between T4 *nrdB* and *td* at the 5′ splice sites. A most remarkable observation is that the second nucleotide in the conserved GCCT sequence is the site of a known transition mutant in the *td* gene, the *tdN57* lesion (Povinelli *et al.*, 1985). At the 3′ splice site there are nine out of 23 identities, and these include a G residue at the extreme end of the intron, as for the *Tetrahymena* consensus sequence (Figure 2).

The intron sequence of the *td* gene contained an open reading frame of 735 bp, but there was no detectable expression of such a protein in an *in vitro* translation system (Chu *et al.*, 1985; Belfort *et al.*, 1986). It is interesting to note that the T4 *nrdB* intron includes a 291 nucleotide open reading frame (denoted ORF 1 in Figure 1). It may not be expressed, since it lacks promoter region as well as ribosome-binding site. It remains to be seen if the presence of an open reading frame within an intron is a general feature of the bacteriophage T4 genome. An open reading frame is frequently found in the *Tetrahymena*/mitochondrial introns (Waring and Davies, 1984) and was also recently reported for the 622 bp intron of the 23S rRNA gene in archaebacteria (Kjems and Garrett, 1985).

```
                          exon 1▼intron                          intron▼exon 2
nrdB                 ATGTaTctTTtGcgTAAaatgcGCCTttaaAcgg.....TtcCatAAgAscATGgaAatcaT
td                   ATGTtTtcTTgGgTTAAttgagGCCTgagtAtaa.....TatCtgAAcAtaATGctAccgtT
Tetrahymena consensus     ...Tx.....................................Gx...
```

Fig. 2. Nucleotide sequence alignment of splice junctions of the *nrdB* gene (putative), the *td* gene (Chu *et al.*, 1984) and the consensus splice junctions of *Tetrahymena* introns (Cech, 1983). Conserved nucleotides have been capitalized. The splice points have been indicated with arrows.

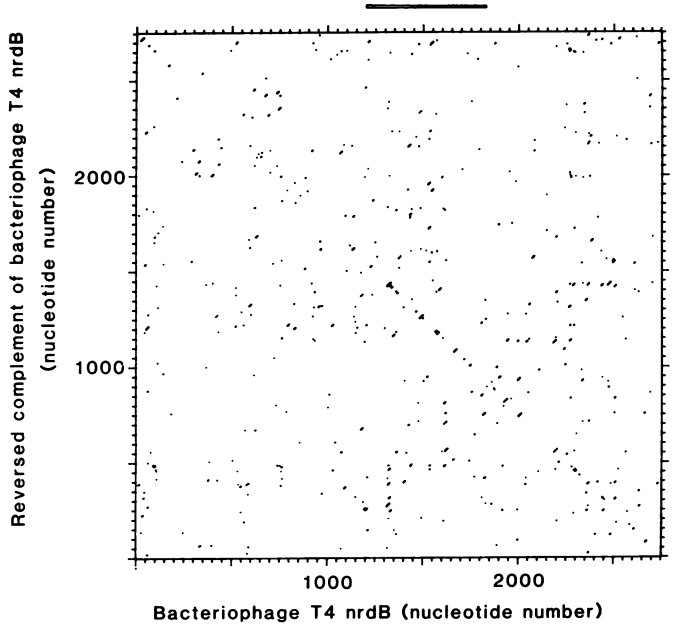

Fig. 3. DOTPLOT (Devereux *et al.*, 1984) comparison of the forward nucleotide sequence of the *nrdB* gene and its reverse complement. The window was 21 and the stringency 14. The bar at the top indicates the position of the intron.

The *nrdB* intron shows ample possibilities for secondary structure (Figure 3). This has also been reported for the *td* intron (Chu *et al.*, 1985) and clearly shown in the case of *Tetrahymena*/mitochondrial intron class (Waring and Davies, 1984). There is no obvious internal guide sequence (IGS) within the *nrdB* intron to align accurately the two splice-junction points. The published sequence information on the *td* intron also does not show an obvious IGS (Chu *et al.*, 1984; Belfort *et al.*, 1985). Perhaps the juxtapositioning mechanism of bacteriophage T4 splice points differs from the mechanism observed in *Tetrahymena*. It should be taken into account that transcription and translation are tightly coupled in *E. coli*. Most likely the common features we observe in bacteriophage T4 introns, the TAA stop codon, the open reading frame and the ATG start codon, are all necessary constituents of a proper folding of the pre-mRNA to enable a correct splicing of the intron sequence.

Comparison of protein B2 from bacteriophage T4 and E. coli

The deduced amino acid sequence of T4 protein B2 has a total length of 379 residues. This implies that the nascent T4 B2 polypeptide is 3−4 residues longer than that of *E. coli* B2. From its migration in SDS−polyacrylamide gels, the native T4 B2 polypeptide was estimated as 3−4 kd shorter than the *E. coli* B2 polypeptide (Berglund, 1975; Cook and Greenberg, 1983). This may suggest that T4 B2 is post-translationally modified by proteolysis. However, the mobility of proteins in SDS−gel electrophoresis may be influenced also by amino acid composition, as was recently reported for pre-albumin of different eukaryotic

```
            1                                                          100
T4 B2       MSTVFNTNPVDVLNEPMFFGSGLGLARYDIQRHRVFEELIERQISFFWRPEEVNLMMDAAQFNKLPQYQQNIFTNNLKYQSLLDSIQGRAPSAVLMSLIS
COLI B2     AYTTFSQTKNDQLKEPMFFGQPVNVARYDQQKYDIFEKLIEKQLSFFWRPEEVDVSRDRIDYQALPEHEKHIFISNLKYQTLLDSIQGRSPNVALLPLIS
Consensus   --t-f-----d-l-epmffg-----aryd-q----fe-lie-q-sffWrpeev----D------Lp-----iF--nlkyq-lldsiqgr-p---l--lis

          101                                                          200
T4 B2       DPSLDTWVATWTFSETIHSRSYTHIMRNLYTDPSKVFDEIVLDEAIMKRAESIGRYYDDVLVKTREWENAKDMVEYYKDQGLILADKDVEQRAKRDLMKS
COLI B2     IPELETWVETWAFSETIHSRSYTHIIRNIVNDPSVVFDDIVTNEQIQKRAEGISSYYDELIEMTSYWH.....LLGEGTHTVNGKTVTVSLRELKKK
Consensus   -p-l-twv-tw-fsEtiHsrsYthi-rn---dps-vfd-iv--e-i-krae-i--yyd-----t--w----------g----------r-l-k-

          201                                                          300
T4 B2       LYLCLHVINALEAIRFYVSFA.....EIMEGNAKIMKFIARDEQLHLKGTQYIIRQLQSGTDGDEWVKIAQECEQEAVDIFMEVNRQEKDWAVHLFK
COLI B2     LYLCLMSVNALEAIRFYVSFACSFAFAERELMEGNAKIIRLIARDEALHLTGTQHMLNLLRSGADDPEMAEIAEECKQECYDLFVQAAQQEKDWADYLFR
Consensus   lylcl---nalEairFyvSFa--------e-megnaki---IaRDE-lHl-gtq-----l-sg-d--e---ia-ec-qe--d-f-----qekdwA--lf-

          301                                                   387
T4 B2       DGDVPGLNTNSMWSFIDYLTVSRMKQCGLPCPITDAPVKHPYPWIREYLNSDNVQSAPQEVELSSYLVAQIDNDVDDKVMMSFKKYF
COLI B2     DGSMIGLNKDILCQYVEYITNIRMQAVGLDLP..FQTRSNPIPWINTWLVSDNVQVAPQEVEVSSYLVGQIDSEVDTDDLSNFQL..
Consensus   dg---gLn---------y-t--rm---gl--p--------p-pwi---l-sdnvq-apqEve-ssYlv-qid--vd------f----
```

Fig. 4. Alignment of the deduced protein sequences of the small subunits of ribonucleotide reductase from bacteriophage T4 and *E. coli*. The bottom line indicates residues conserved between T4 and *E. coli* (lower case) and among all published B2 homologues (upper case). The presumed splice-junctions in T4 protein B2 have been indicated by a bar above the sequence.

origins (Sundelin *et al.*, 1985). Because of the lack of amino acid sequence and compositional analyses, it is premature to propose a native form of T4 B2 from the present data.

An alignment of the deduced sequences of protein B2 from T4 and *E. coli* is shown in Figure 4. There is an overall homology of 47%; exon 1 shows 52% homology and exon 2 shows 48% homology. From a similar alignment of *E. coli* B2 with homologues in clam, Epstein–Barr virus and herpes simplex virus a total of 20 conserved residues were found (Sjöberg *et al.*, 1985). All of these are also present in T4 B2 (Figure 4, upper case letters in bottom line), and a recently deduced amino acid sequence of protein B2 from mouse (Thelander and Berg, 1986). These conserved residues include the catalytically essential radical tyrosine residue at position 122 (Figure 4; Sjöberg *et al.*, 1985; Larsson and Sjöberg, 1986).

The splice junctions indicated in Figure 1 result in a T4 B2 sequence, which is eight residues shorter at the splice point than the corresponding *E. coli* protein B2 (Figure 4) and the other four *nrdB* gene products of eukaryotic origins (Sjöberg *et al.*, 1985; Thelander and Berg, 1986). Such a deletion may be tolerated in the B2 protein, since it occurs in an area of rather low homology in the alignment. Secondary structure predictions (Garnier *et al.*, 1978) suggested an α-helix/loop/β-strand structure in this region (Sjöberg *et al.*, 1985) and deletions may readily be accommodated in the loop area. On the contrary continuation of exon 2 another seven codons upstream of the ATG start codon would result in an additional amino acid sequence with only 22% homology to *E. coli nrdB* protein. This possibility is thus considered unlikely in the light of the known splice junction sequence of the T4 *td* gene. In addition, a DOTPLOT comparison of the nucleotide sequence of the T4 *nrdB* gene and the *E. coli nrdB* gene shows very few sequence similarities upstream and including the ATG codon (Figure 5). However, the definite assignment of the splice point of the T4 *nrdB* gene has to await direct protein sequencing data.

The T4 B2 polypeptide has an eight residue insertion in exon 1 as compared with *E. coli* B2 (Figure 4). This discrepancy occurs at a region where *E. coli* B2 differs from its eukaryotic counterparts in having an insertion of 26 residues. This part of the general B2 structure may be a region of functional importance. In fact residues 169–176 in T4 B2 may take the place of residues 222–229 in *E. coli* B2 in a similar three-dimensional structure.

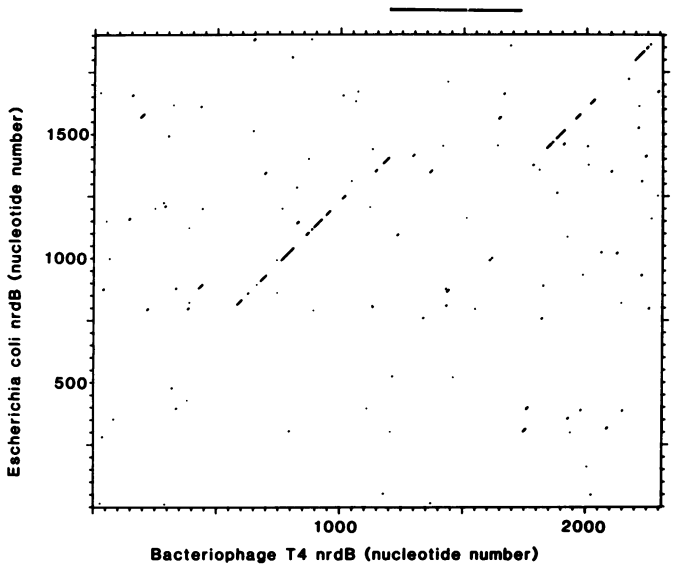

Fig. 5. DOTPLOT (Devereux *et al.*, 1984) comparison of bacteriophage T4 and *E. coli* (Carlson *et al.*, 1984) *nrdB* gene nucleotide sequences. The window was 21 and the stringency 14. The bar at the top indicates the position of the T4 *nrdB* intron.

Amino acid sequence alignments of B1 subunits from *E. coli* (Carlson *et al.*, 1984), mouse (Caras *et al.*, 1985), Epstein–Barr virus (Gibson *et al.*, 1984) and herpes simplex virus (Swain and Galloway, 1986) revealed the same extent of similarities (Caras *et al.*, 1985; Swain and Galloway, 1986; B.-M.Sjöberg, unpublished results) as in the alignment of B2 subunits. To check if the C-terminal end of T4 *nrdA* was present in our cloned fragment its nucleotide sequence was compared with 540 nucleotides from the C-terminal end of the *E. coli nrdA* gene. There was no indication of a *nrdA*-related region upstream of T4 *nrdB* at the nucleotide (Figure 5) or at the amino acid (data not shown) sequence levels. Part of an open reading frame with extensive homology to the *E. coli nrdA* gene was recently found by Chu *et al.* (unpublished) shortly downstream of the *td* gene. The distance of approximately 2.7 kb between the end of the *td* gene and our *SphI*/*ClaI* fragment leaves more than enough space for the T4 B1 polypeptide of approximately 84 kd molecular mass (Berglund, 1975; Cook and Greenberg, 1983).

Location of the denA gene coding for endonuclease II

The location of the *nrdB* gene is approximately 2 kb closer to gene 63 than earlier anticipated from genetic recombination analyses (Yeh *et al.*, 1969). Our present finding leaves only a little space for the *denA* gene (Ray *et al.*, 1972). However, the nucleotide sequence was obtained from a *denA*$^-$ phage (the *nd28* mutant) (Warner *et al.*, 1970). The *nd*28 mutation has been mapped to be very close to the gene 63 M69 mutation and *denA* therefore placed between gene 63 and *nrdB* (Ray *et al.*, 1972). The only open reading frame in this location is ORF 2. To see if ORF 2 might correspond to *denA*, we resequenced the region using a *denA*$^+$ source of T4 DNA. The only difference found was a T at position 2579 (see Figure 1) in the *denA*$^+$ strain instead of a C. This would correspond to a leucine to proline mutational change at amino acid 84 in the *nd*28 phage. On this basis we tentatively assign ORF 2 to be the *denA* gene. Since prolines are known to be good at disrupting α-helices and β-sheets it may be that *nd*28 mutation gives rise to an improperly folded and inactive endonuclease. However, since no mol. wt or protein sequence data are known for *denA*, further work will be necessary to determine if ORF 2 codes for the complete *denA* gene product (endonuclease II).

Relevance of the present data to earlier results

Genetic mapping of *nrdA* and *nrdB* mutants revealed three complementation groups. Two of these were shown by enzymological means to affect the *nrdB* gene (Yeh and Tessman, 1972). Although one of the *nrdB* mutations was later classified as an *nrdA* mutation (Cook and Greenberg, 1983), the *nrdB* locus still comprises two complementation groups with at least as high recombination values in two factor crosses as between *nrdA* and *nrdB* mutants. In light of the findings reported here of an intron in the *nrdB* gene these two-factor crossover values are readily explained as *nrdB* mutations occurring in exon 1 and exon 2, respectively.

The map position of 136.1−137.8 kb reported here for the *nrdB* gene explains a series of unpublished observations (B.-M.Sjöberg and S.Hahne, unpublished results) obtained with a bacteriophage λ recombinant (NM1037) encompassing the T4 map region 136.8−143.25 kb (Mileham *et al.*, 1980). NM1037 was expected to comprise the T4 *nrdA* and *B* genes, but only the *nrdA* product was found in SDS−polyacrylamide gel electrophoresis (Mileham *et al.*, 1980). On the other hand, we (B.-M. Sjöberg and S.Hahne, unpublished results) observed in NM1037 cells a bacteriophage T4-specific tyrosyl radical EPR signal (Sahlin *et al.*, 1982), indicative of a *nrdB* product. Several subclones of NM1037 DNA in different plasmid vectors also showed a T4-specific EPR signal but there was no plasmid-encoded band of the expected size on SDS−polyacrylamide gels. On the other hand, both crossed immunoelectrophoresis (Chua and Blomberg, 1979) and direct immunoprecipitation of extract from such recombinant plasmids with polyclonal anti-T4 ribonucleotide reductase antisera precipitated a protein band of approximately 20 kd molecular mass. Since the *Eco*RI site at 136.8 kb divides the *nrdB* gene into exon 1 and exon 2, only exon 1 can be contained in the NM1037 recombinant (Mileham *et al.*, 1980). The cross-reactivity of T4 *nrd* antisera with a shorter than expected polypeptide fits with exon 1 encoding a product of at the most 221 amino acid residues. It was recently observed that a truncated form of the *E. coli nrdB* product, missing 30 C-terminal residues, is capable of forming and stabilizing the *E. coli*-specific tyrosyl radical (Sjöberg, 1986).

T4 ribonucleotide reductase was found to belong to the delayed early group of T4 gene products (Berglund *et al.*, 1969). The phage-specific ribonucleotide reductase activity appears approximately 5 min after infection, similarly to what has been found for the T4 thymidylate synthase (Trimble *et al.*, 1972). In the latter case the presence of the intron sequence was considered indicative of this rather late appearance of a product expected to be needed early in infection. The delayed occurrence of active ribonucleotide reductase may be caused by the intron in the *nrdB* gene or the *mot* promoter upstream of the *nrdB* gene. In both cases the *nrdA* and the *nrdB* genes would be differentially expressed. In accordance with this Cook and Greenberg (1983) observed that there was no polar effect on *nrdB* gene expression of the *nrdA*am67 mutation. Similar differential expression of the two subunits of ribonucleotide reductase has also been reported in eukaryotic organisms such a clams (Standart *et al.*, 1985), sea urchins (Evans *et al.*, 1983) and mouse cells (Eriksson *et al.*, 1984).

Materials and methods

T4 phages and vectors

The T4 phages used for making cytosine-containing DNA were JW800 [*am*N55x-(g42), *am*E51(g56), *nd*28(*denA*), D2a23x5(*denB*), *alc*10] (Clark *et al.*, 1980) and T4*alc*7 [*am*51(g56), NB5060(*denB-rII*), *am*87(g42), *alc*7] (Wilson *et al.*, 1977). The M13 derivatives mp8 (Messing and Vieira, 1982), mp12 (Karn *et al.*, 1984), mp18 and mp19 (Yanisch-Perron *et al.*, 1985) were used as cloning vectors.

Enzymes

Restriction endonucleases *Sma*I, *Sph*I, *Cla*I, *Sst*I, *Hin*dIII and *Pst*I were from Boehringer and New England Biolabs. T4 DNA ligase was obtained as gifts from Göran Magnusson and David Bentley. DNA polymerase (Klenow) was from Boehringer and Amersham. [α-^{35}S]Thio-dATP was from Amersham. General cloning methods were according to Maniatis *et al.* (1982).

Sequencing strategy

A 15 kb *Sma*I fragment was isolated from digested JW800 DNA by phenol extraction (Niggemann *et al.*, 1981) of DNA from pieces cut out of 0.5% agarose gels. The *Sma*I fragment was further digested by *Sph*I and *Cla*I endonucleases and fragments were cloned into M13mp19. To obtain the reverse orientation, the 3.1 kb fragment was cloned in M13mp18. The nucleotide sequence from the *Sph*I site to the *Bgl*II site was obtained from these two derivatives by the chain termination method (Biggin *et al.*, 1983; Olsson *et al.*, 1984). The nucleotide sequence from the *Hin*dIII site to past gene 63 of the *denA*$^-$ (*nd*28) mutant JW800 was obtained by shotgun cloning of a 9 kb *Hin*dIII/*Pst*I fragment into M13mp8 as previously described (Rand and Gait, 1984). The wild-type nucleotide sequence of ORF 2 was obtained from the *denA*$^+$ strain T4*alc*7 by cloning of a 1.9 kb *Sst*I fragment (from position 1782 to a position in the RNA ligase gene beyond the sequence shown in Figure 1) into M13mp12 (Karn *et al.*, 1984) and sequencing by the chain termination method (Biggin *et al.*, 1983).

Computer programs

DNA sequences were compiled and analysed using the programs of Staden (1982, 1984) and University of Wisconsin Genetics Computer Group (Devereux *et al.*, 1984). DOTPLOT analyses of the latter program package was used for Figures 3 and 5 (Devereux *et al.*, 1984).

Acknowledgements

The authors wish to thank Mohinder Singh for help in DNA sequencing, Noreen Murray for a gift of T4*alc*7 DNA and NM1037, and Gunilla Englund for synthesizing oligonucleotides. K.N.R. gratefully acknowledges a MRC Training Fellowship award. C.K.M. and C.Z.M. gratefully acknowledge an Eleanor Roosevelt International Cancer Fellowship, which supported a sabbatical leave in the laboratory of Professor Peter Reichard. This work was supported by grants from the Swedish Medical Research Council and Magn. Bergvall Foundation.

References

Belfort,M., Pedersen-Lane,J., West,D., Ehrenman,K., Maley,G., Chu,F. and Maley,F. (1985) *Cell*, **41**, 375−382.

Belfort,M., Pedersen-Lane,J., Ehrenman,K., Chu,F.K., Maley,G.F., Maley,F., McPheeters,D.S. and Gold,L. (1986) *Gene*, **41**, 93−102.

Berglund,O. (1975) *J. Biol. Chem.*, **250**, 7450−7455.

Berglund,O., Karlström,O. and Reichard,P. (1969) *Proc. Natl. Acad. Sci. USA*, **62**, 829−835.

Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
Brody,E., Rabussay,D. and Hall,D.H. (1983) In Mathews,C.K., Kutter,E.M., Mosig,G. and Berget,P.B. (eds.), *Bacteriophage T4*. American Society for Microbiology, Washington DC, pp. 174–183.
Caras,I.W., Levinson,B.B., Fabry,M., Williams,S.R. and Martin,D.W., Jr. (1985) *J. Biol. Chem.*, **260**, 7015–7022.
Carlson,J., Fuchs,J.A. and Messing,J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4294–4297.
Cech,T.R. (1983) *Cell*, **34**, 713–716.
Chu,F. and Maley,F. (1985) *Cell*, **41**, 375–382.
Chu,F.K., Maley,G.F., Maley,F. and Belfort,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3049–3053.
Chu,F.K., Maley,G.F., Belfort,M. and Maley,F. (1985) *J. Biol. Chem.*, **260**, 10680–10688.
Chua,M. and Blomberg,F. (1979) *J. Biol. Chem.*, **254**, 215–223.
Clark,R.W., Wever,G.H. and Wiberg,J.S. (1980) *J. Virol.*, **33**, 438–448.
Cook,K.S. and Greenberg,G.R. (1983) *J. Biol. Chem.*, **258**, 6064–6072.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
Eriksson,S., Gräslund,A., Skog,S., Thelander,L. and Tribukait,B. (1984) *J. Biol. Chem.*, **259**, 11695–11700.
Evans,T., Rosenthal,E., Youngblom,J., Distel,D. and Hunt,T. (1983) *Cell*, **33**, 389–396.
Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97–120.
Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B. (1984) *Nucleic Acids Res.*, **12**, 5087–5099.
Karn,J., Matthes,H.W.D., Gait,M.J. and Brenner,S. (1984) *Gene*, **32**, 217–224.
Kjems,J. and Garrett,R.A. (1985) *Nature*, **318**, 675–677.
Kutter,E.[1] and Rüger,W.[2] (1985) *Bacteriophage T4 genomic map.* ([1]Evergreen State College, Olympia, WA 98505; [2]Ruhr Universität, D-4630 Bochum 1, FRG).
Lammers,M. and Follmann,H. (1983) *Struct. Bonding*, **54**, 27–91.
Larsson,Å. and Sjöberg,B.-M. (1986) *EMBO J.*, **5**, 2037–2040.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Mathews,C.K. and Allen,J.R. (1983) In Mathews,C.-K., Kutter,E.M., Mosig,G. and Berget,P.B. (eds.), *Bacteriophage T4*. American Society for Microbiology, Washington, DC, pp. 59–70.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.
Mileham,A.J., Revel,H.R. and Murray,N.E. (1980) *Mol. Gen. Genet.*, **179**, 227–239.
Niggemann,E., Green,I., Meyer,H.-P. and Rüger,W. (1981) *Mol. Gen. Genet.*, **184**, 289–299.
Olsson,A., Moks,T., Uhlén,M. and Gaal,A. (1984) *J. Biochem. Biophys. Methods*, **10**, 83–90.
Povinelli,C.M., Hall,D.H., Ehrenman,K., Pedersen-Lane,J. and Belfort,M. (1985) Data presented at the 1985 Evergreen International T4 Meeting, Evergreen State College, Olympia, WA.
Rand,K.N. and Gait,M.J. (1984) *EMBO J.*, **3**, 397–402.
Ray,P., Sinha,N.K., Warner,H.R. and Snustad,D.P. (1972) *J. Virol.*, **9**, 184–186.
Sahlin,M., Gräslund,A., Ehrenberg,A. and Sjöberg,B.-M. (1982) *J. Biol. Chem.*, **257**, 366–369.
Sjöberg,B.-M. (1986) In Holmgren,A., Brändén,C.-I., Jörnvall,H. and Sjöberg, B.-M. (eds), *Thioredoxin and Glutaredoxin Systems. Structure and Function*. Raven Press, NY, pp. 199–207.
Sjöberg,B.-M., Eklund,H., Fuchs,J.A., Carlson,J., Standart,N.M., Ruderman, J.V., Bray,S.J. and Hunt,T. (1985) *FEBS Lett.*, **183**, 99–102.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 4731–4751.
Staden,R. (1984) *Nucleic Acids Res.*, **12**, 499–503.
Standart,N.M., Bray,S.J., George,E.L., Hunt,T. and Ruderman,J. (1985) *J. Cell. Biol.*, **100**, 1968–1975.
Sundelin,J., Melhus,H., Das,S., Eriksson,U., Lind,P., Trägårdh,L., Peterson, P.A. and Rask,L. (1985) *J. Biol. Chem.*, **260**, 6481–6487.
Swain,M.A. and Galloway,D.A. (1986) *J. Virol.*, **57**, 802–808.
Thelander,L. and Berg,P. (1986) *Mol. Cell. Biol.*, in press.
Trimble,R.B., Galivan,J. and Maley,F. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1659–1663.
Waring,R.B. and Davies,R.W. (1984) *Gene*, **28**, 277–291.
Warner,H.R., Snustad,D.P., Jorgensen,S.E. and Koerner,J.F. (1970) *J. Virol.*, **5**, 703–708.
Wilson,G.G., Tanyashin,V.I. and Murray,N.E. (1977) *Mol. Gen. Genet.*, **156**, 203–214.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
Yeh,Y.-C. and Tessman,I. (1972) *Virology*, **47**, 767–772.
Yeh,Y.-C., Dubovi,E.J. and Tessman,I. (1969) *Virology*, **37**, 615–623.

Received on 9 May 1986

Note added in proof

Gott *et al.* recently found evidence for multiple self-splicing introns in bacteriophage T4 by autocatalytic GTP labeling of bacteriophage T4 RNA *in vitro*. An ~0.6 kb intron mapped in the *nrd*B region (J.M.Gott, D.A.Shub and M.Belfort, submitted).