FEBS Letters 400 (1997) 94–98 FEBS 17967

Evolution of *dna*Q, the gene encoding the editing 3′ to 5′ exonuclease subunit of DNA polymerase III holoenzyme in Gram-negative bacteria

Yiping Huang, Dan K. Braithwaite, Junetsu Ito*

Department of Microbiology and Immunology, College of Medicine, The University of Arizona, Tucson, AZ 85724, USA

Received 25 October 1996; revised version received 11 November 1996

Abstract The nucleotide sequences of the *dna*Q genes from *Salmonella typhimurium* and *Buchnera aphidicola*, encoding the ε-subunit of the DNA polymerase III holoenzyme, have been determined. The *Salmonella typhimurium dna*Q protein consists of 243 amino acid residues with a calculated molecular weight of 27 224. The *Buchnera aphidicola dna*Q protein contains 233 amino acid residues with a calculated molecular weight of 27 170. A multiple sequence alignment of the amino acid sequences of the *dna*Q proteins and those of DNA polymerase IIIs from Gram-positive bacteria produced six homologous segments. These homologous segments contain highly conserved amino acid sequence motifs involved in catalytically important metal ion bindings (ligands 1, 2 and 3). However, metal ligand 4 is found to be altered in the 3′-5′ exonuclease domain of the family C DNA polymerases and *dna*Q proteins in Gram-negative bacteria. From these results, we propose that the last common ancestor of the *dna*Q gene of Gram-negative bacteria and the DNA polymerase III gene (pol C gene) of Gram-positive bacteria was a single gene containing both 3′-5′ exonuclease and DNA polymerase domains and then the *dna*Q gene separated from the polymerase gene in Gram-negative bacteria.

*Key words: Salmonella dna*Q; *Buchnera aphidicola dna*Q; 3′ to 5′ editing exonuclease; DNA polymerase III

1. Introduction

In bacteria, chromosomal DNA replication is carried out by DNA polymerase III (pol III) which contains a 3′ to 5′ exonucleolytic proofreading activity [1]. DNA pol IIIs have been globally classified as family C DNA polymerases [2,3]. While the 3′ to 5′ exonuclease resides in the polymerase catalytic polypeptide in Gram-positive bacterial pol IIIs, it exists as a separate subunit in Gram-negative bacterial pol IIIs [1,4,5]. In *E. coli* DNA pol III holoenzyme, the α-subunit encoded by the *dna*E gene has polymerase catalytic activity and the ε-subunit encoded by the *dna*Q gene has proofreading 3′ to 5′ exonuclease activity [1,6–8]. The *dna*Q genes of *Salmonella typhimurium* and *Buchnera aphidicola*, like that of *E. coli*, specify the ε-subunits of their pol III holoenzymes and have been partially sequenced previously [9,10]. It has been shown that the primary sequence of the *E. coli dna*Q protein has weak but significant homology with that of the N-terminal region of *Bacillus subtilis* pol III [4,5]. Subsequently, site-directed mutagenesis has localized the 3′ to 5′ exonuclease domain within a 200 amino acid segment in the N-terminal of *B. subtilis* DNA pol III [11]. Our interest in understanding the evolutionary relationships between Gram-positive and Gram-negative family C DNA polymerases has led us to perform comparative analyses between *dna*Q proteins from Gram-negative bacteria and DNA pol IIIs from Gram-positive bacteria. For these analyses, we have determined the complete nucleotide sequences of the *dna*Q genes of *S. typhimurium* and *B. aphidicola* and compared them to other reported sequences of *dna*Q genes of Gram-negative bacteria and DNA pol IIIs of Gram-positive bacteria.

2. Materials and methods

A recombinant plasmid, pFF1, carrying the wild type *dna*Q gene of *S. typhimurium* in a 3.1 kb *Hin*dIII-*Eco*RI fragment was prepared from a strain, RM616 [9]. For sequencing purposes, the 3.1 kb *Hin*dIII-*Eco*RI fragment was then subcloned into a pCRII vector (Invitrogen Corp., San Diego, CA, USA). The DNA sequence was determined with double-stranded DNA templates and the Sequenase version 2.0 DNA sequencing Kit (US Biochemical, Cleveland, OH, USA). Both strands of the template DNA were sequenced. The purified genomic DNA of *B. aphidicola* was generously provided by Dr. Paul Baumann at the University of California at Davis.

3. Results and discussion

The nucleotide and deduced amino acid sequences for the *dna*Q genes of *S. typhimurium* and *B. aphidicola* are shown in Figs. 1 and 2, respectively. The *Salmonella dna*Q gene consists of 732 nucleotides encoding 243 amino acid residues with a calculated molecular weight of 27 224. The *Aphidicola dna*Q gene consists of 699 nucleotides specifying 233 amino acid residues with a calculated molecular weight of 27 170. Comparisons of the *S. typhimurium dna*Q gene sequence to the nucleotide sequences of homologous genes from *E. coli* [8,12], *Haemophilus influenzae* [13], and *B. aphidicola* show identities of 83%, 61.8%, and 58.1%, respectively. In addition, the *B. aphidicola dna*Q gene shows sequence identities to *E. coli* and *H. influenzae* of 58.8% and 65.1%, respectively. Similar comparisons of the deduced amino acid sequences of *Salmonella* and *Aphidicola dna*Q proteins show the following identities: *Salmonella–E. coli*, 93.4%; *Salmonella–H. influenzae*, 60.7%; *Salmonella–aphidicola*, 52.2%; *Aphidicola–E. coli*, 47.4%; and *Aphidicola–H. influenzae*, 50.2%.

Fig. 3 shows that the amino acid sequences of the *dna*Q proteins clearly align with those of the DNA polymerase IIIs. Six blocks of homology (A–F) in a linear array are evident in this analysis. Since the sequence similarity stretches over the major N-terminal portion of the *dna*Q proteins and only small gaps are necessary for optimal alignment, it is unlikely that these similarities arose by convergent evolution. Rather, the *dna*Q genes of Gram-negative bacteria and the DNA polymerase III genes of Gram-positive bacteria must have evolved divergently from a common ancestor.

*Corresponding author. Fax: (1) (520) 626-2100.
E-mail: ito@aruba.ccit.arizona.edu

PII S0014-5793(96)01361-0

```
   1  gtagacttcctgtaattgaaatcgataacaaaacgcaagtctgacataaatgaccgatATGAGCACTGCAATTACACGACAGATTGTCCT   90
   1                                                          M  S  T  A  I  T  R  Q  I  V  L    11

  91  CGATACCGAAACCACCGGTATGAATCAGATAGGCGCGCACTATGAAGGTCACAAGATTATTGAGATCGGTGCGGTTGAGGTGATAAACCG  180
  12   D  T  E  T  T  G  M  N  Q  I  G  A  H  Y  E  G  H  K  I  I  E  I  G  A  V  E  V  I  N  R   41

 181  TCGTCTGACCGGCAACAATTTTCATGTTTACCTGAAGCCCGATCGCCTTGTCGATCCAGAGGCTTTTGGCGTACACGGTATTGCCGATGA  270
  42   R  L  T  G  N  N  F  H  V  Y  L  K  P  D  R  L  V  D  P  E  A  F  G  V  H  G  I  A  D  E   71

 271  GTTTCTGCTGGATAAGCCGGTTTTTGCTGATGTGGTCGATGAGTTTCTTGATTATATCCGCGGCGCGGAGCTGGTCATCCATAACGCATC  360
  72   F  L  L  D  K  P  V  F  A  D  V  V  D  E  F  L  D  Y  I  R  G  A  E  L  V  I  H  N  A  S  101

 361  GTTCGATATCGGCTTTATGGATTATGAGTTTGGTCTGCTTAAACGCGATATTCCTAAAACCAATACTTTCTGCAAAGTTACCGACAGCCT  450
 102   F  D  I  G  F  M  D  Y  E  F  G  L  L  K  R  D  I  P  K  T  N  T  F  C  K  V  T  D  S  L  131

 451  GGCGTTGGCGCGGAAAATGTTCCCCGGCAAGCGTAACAGCCTTGATGCACTGTGTTCGCGTTATGAGATAGATAATAGCAAACGTACTTT  540
 132   A  L  A  R  K  M  F  P  G  K  R  N  S  L  D  A  L  C  S  R  Y  E  I  D  N  S  K  R  T  L  161

 541  GCACGGCGCATTGCTCGATGCCCAGATCCTTGCTGAAGTGTATCTGGCGATGACGGGCGGACAAACGTCCATGACGTTTGCGATGGAAGG  630
 162   H  G  A  L  L  D  A  Q  I  L  A  E  V  Y  L  A  M  T  G  G  Q  T  S  M  T  F  A  M  E  G  191

 631  AGAGACGCAACGGCAGCAAGGTGAGGCGACCATTCAGCGAATCGTTCGCCAGGCCAGCCGGTTACGGGTCGTTTTTGCCTCTGAGGAAGA  720
 192   E  T  Q  R  Q  Q  G  E  A  T  I  Q  R  I  V  R  Q  A  S  R  L  R  V  V  F  A  S  E  E  E  221

 721  GCTGGCTGCGCATGAATCGAGGCTTGATCTGGTGCAGAAAAAGGGCGGAAGTTGCCTTTGGCGGGCGTAAtttattccttttttaggctat  810
 222   L  A  A  H  E  S  R  L  D  L  V  Q  K  K  G  G  S  C  L  W  R  A  *                        244

 811  aaaaatcatcctttcgggcgatttttgcagcaactgattcaaaaggtgagaaaaagcgttgacgggcgacaccgcaaaccgtaatattccn  900
 901  nctcgttcccagcggaacatcacggagcggtagttcagttggttagaatacctgcctgtcacgcagggggtt                    972
```

Fig. 1. Nucleotide sequence of the *dnaQ* gene of *Salmonella typhimurium*. The predicted amino acid sequence is given in one-letter code (aligned with first nucleotide of each codon) (GenBank accession number U77465).

Based on crystallographic as well as site-directed mutagenesis studies on the 3′ to 5′ exonuclease domain of the *E. coli* DNA pol I Klenow fragment, Steitz and coworkers have proposed a two-metal-ion model for phosphodiester hydrolysis [14,15]. One metal ion (metal A) is considered to activate a coordinated H_2O molecule to yield an attacking nucleophile, while the other metal ion (metal B) facilitates the departure of the 3′ oxyanion and stabilizes the pentacovalent transition state. Three aspartic acid side chains in the Klenow fragment have been identified as binding metal ions and deoxynucleotides [16]. These aspartic acids are Asp^{355} (metal A and B binding), Asp^{424} (metal B binding), and Asp^{501} (metal A binding) and correspond to ligands L1, L3, and L4, respectively, in Fig. 3. In addition, two other important amino acid side chains are Glu^{357} (corresponding to ligand L2 in Fig. 3) and Tyr^{497}, both of which bind metal A and orient the attacking nucleophile [15,16]. Assuming the two-metal-ion catalysis operates in family C DNA polymerases, motifs containing the

```
   1  ttaataattattccaatattaaataacttaaatagatcatttatgATGAATAATACACAAAGAATAATTGTTTTAGATACTGAAACAACA   90
   1                                              M  N  N  T  Q  R  I  I  V  L  D  T  E  T  T    15

  91  GGAATGAATAGTGTAGGTCCACCTTATCTTAATCATAGAATTATTGAAATTGGTGCCATTGAAATAATTAATCGTCGTTTTACAGGGAAA  180
  16   G  M  N  S  V  G  P  P  Y  L  N  H  R  I  I  E  I  G  A  I  E  I  I  N  R  R  F  T  G  K   45

 181  AAATTTCATACTTATATTAAACCTAATAGATTGATAGAATCTGACGCTTCAAAAATTCATGGTATTACTGATGATTTTTTATCAGATAAA  270
  46   K  F  H  T  Y  I  K  P  N  R  L  I  E  S  D  A  S  K  I  H  G  I  T  D  D  F  L  S  D  K   75

 271  CCATCTTTTAAAGACATAGCTAAAGATTTTTTTAATTATATAAAAAATTCAGAATTAATAATTCATAATGCATCTTTTGATGTAGGTTTC  360
  76   P  S  F  K  D  I  A  K  D  F  F  N  Y  I  K  N  S  E  L  I  I  H  N  A  S  F  D  V  G  F  105

 361  ATAAATCAAGAATTTTCAATGTTAACTAAAAAAATACAAGATATATCAAATTTTTGTAATATTATAGATACATTAAAGATAGCTAGAAAA  450
 106   I  N  Q  E  F  S  M  L  T  K  K  I  Q  D  I  S  N  F  C  N  I  I  D  T  L  K  I  A  R  K  135

 451  TTATTTCCTGGTAAAAAAAATACCTTAGACGCATTATGTATGCGTTATAAAATAAAAAATTCTCATAGAGTTTTACATGGTGCTATTTTA  540
 136   L  F  P  G  K  K  N  T  L  D  A  L  C  M  R  Y  K  I  K  N  S  H  R  V  L  H  G  A  I  L  165

 541  GATGCTTTTCTATTAGGTAAATTATATCTTTTAATGACTAGTGGTCAAGAATCAATTATATTTAATAAAAATATCCAAAATGAAAGAAAT  630
 166   D  A  F  L  L  G  K  L  Y  L  L  M  T  S  G  Q  E  S  I  I  F  N  K  N  I  Q  N  E  R  N  195

 631  TTTAGATATATTAAAAAATCAATAACAAAAAAACATCGTTTTTTAAAAATAATAAAAGCAAATAAAACAGAATTAAAACTTCATAATGAA  720
 196   F  R  Y  I  K  K  S  I  T  K  K  H  R  F  L  K  I  I  K  A  N  K  T  E  L  K  L  H  N  E  225

 721  TATTTAAAATTTTTAAAAGAAAAATAAtgtctatttcgtagtaaattatgaataactataaaaagatgattgactcatttttttaaaata  810
 226   Y  L  K  F  L  K  E  K  *                                                                 255

 811  Tgtacaatgtataaatattaataataaaggtgcggtagttcagtcggttagaatacctgcctgtca                          876
```

Fig. 2. Nucleotide sequence of the *dnaQ* gene of *Buchnera aphidicola*. The predicted amino acid sequence is given in one-letter code (aligned with first nucleotide of each codon) (GenBank accession number U77464).

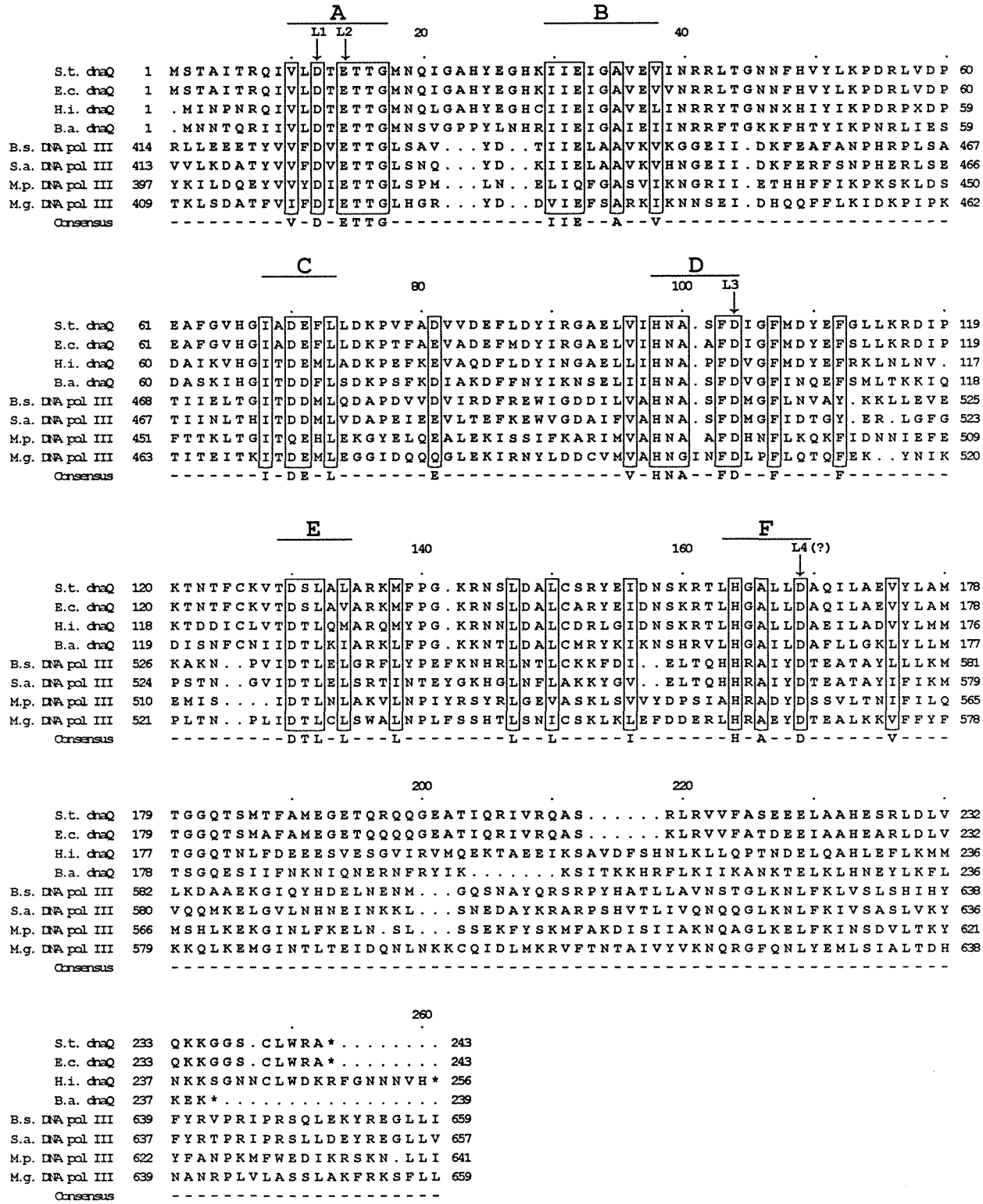

Fig. 3. Alignment of the amino acid sequences of the *dnaQ* proteins of *Salmonella typhimurium* and *Buchnera aphidicola* with the sequences of homologous proteins of Gram-negative bacteria and sequences of DNA polymerase IIIs from Gram-positive bacteria. Six blocks of conserved regions are identified (labeled A–F). Identical and similar residues are boxed. Possible metal binding ligands are indicated by the arrows and labeled L1–L4 (see text). *Salmonella typhimurium* (S.t.) *dnaQ* ([9] and this study); *E. coli* (E.c.) *dnaQ* [8,12]; *Haemophilus influenzae* (H.i.) *dnaQ* [13]; *Buchnera aphidicola* (B.a.) *dnaQ* ([10], and this study); *Bacillus subtilis* (B.s.) DNA polymerase III [5]; *Staphylococcus aureus* (S.a.) DNA polymerase III [22]; *Mycoplasma pulmonis* (M.p.) DNA polymerase III [23]; *Mycoplasma genitalium* (M.g.) DNA polymerase III [24]. The multiple sequence alignment was performed using the PILEUP program from the Genetics Computer Group (GCG). Some mismatches were hand-corrected to include obvious short signatures and the boxes and notation were added using a word processing program.

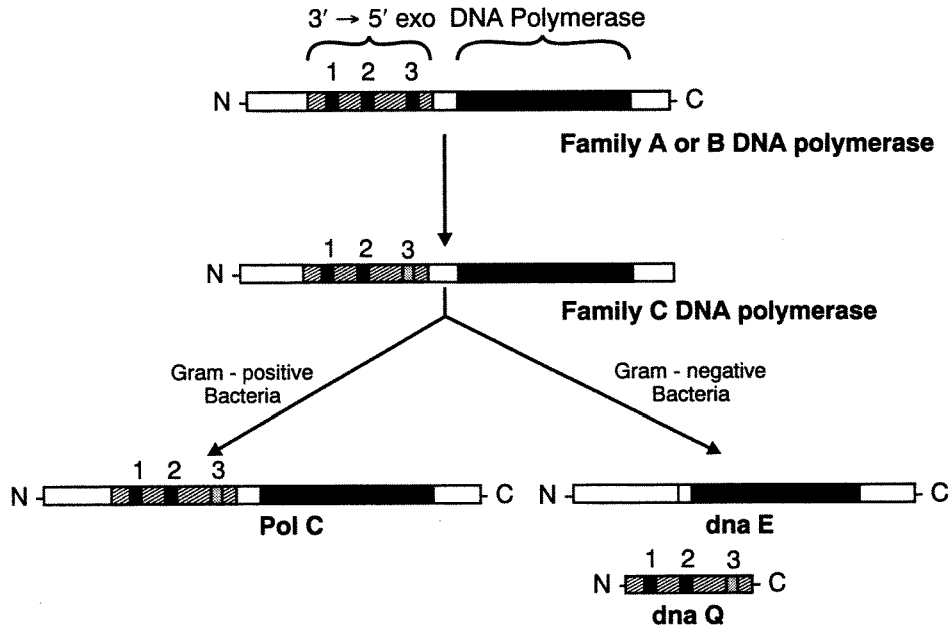

Fig. 4. An evolutionary scenario for the *dna*Q gene of Gram-negative bacteria. Numerals 1, 2, 3 in the 3′ to 5′ exonuclease domain represent 3′-exo I, 3′-exo II, and 3′-exo III motifs, respectively. The 3′-exo III motif of family C DNA polymerases is different from that common to family A and family B DNA polymerases. It has been suggested that Gram-positive and Gram-negative bacteria diverted 1.2–1.5 billion years ago [17].

three carboxylic acids L1 and L2 and L3 were found to be conserved in all *dna*Q and pol III proteins, but a YxxxD motif containing L4 [15] could not be found. We have found, instead, a highly conserved HxAxxD motif in these proteins. This motif must have arisen from a common ancestor sequence which evolved before the Gram-positive and Gram-negative bacteria diverged some 1.2 billion years ago [17]. We suggest that this motif is equivalent to the conserved YxxxD motif found among family A and B DNA polymerases [15,18–20]. Recently, Brown and coworkers have also identified this HxAxxD motif and used site-directed mutagenesis on His^{565} and Asp^{570} in *B. subtilis* pol III to show the motif is critical for 3′ to 5′ exonuclease activity [21]. Therefore, our observations and those of Brown and coworkers [21] raise the strong possibility that a metal ligand (L4) was altered during the course of evolution.

Based on the phylogenetic analyses with the heat-shock chaperonin proteins (Hsp60 and Hsp70), Gupta has suggested that the Gram-positive bacteria have more ancient lineages than Gram-negative bacteria [25]. Studying the secondary structure of 5S ribosomal RNA, Hori and Osawa earlier suggested a similar view [17]. It seems likely that a common progenitor of the *dna*Q gene of Gram-negative bacteria and the DNA pol III gene of Gram-positive bacteria was a single gene containing both 3′ to 5′ exonuclease and DNA polymerase domains and that the *dna*Q gene then separated from the polymerase gene in Gram-negative bacteria, as shown in Fig. 4. When did the editing exonuclease domain separate from the DNA polymerase domain in Gram-negative bacteria? It has been estimated that *E. coli* and *S. typhimurium* separated around 100 million years ago [26]. Obviously, the two genes separated well before that time. *Buchnera aphidicola* is an interesting Gram-negative bacterium which is an obligate intracellular symbiont of aphids [27]. Like *E. coli* and *S. typhimurium*, *B. aphidicola* is a member of the γ-subdivision of Proteobacteria [27]. Evolutionary studies with 16S rRNA sequences indicated that the endosymbiont-aphid association originated 200–250 million years ago, at which time the *dna*Q gene was already separated from the polymerase gene [28]. Therefore, one can infer that the *dna*Q and *dna*E genes became apart in Gram-negative bacteria 0.25–1.2 billion years ago. This time frame could be considerably refined by studying the *dna*Q and DNA pol III genes from various groups of microorganisms.

Acknowledgements: We are deeply indebted to Dr. Russell Maurer at the Case Western Reserve University in Cleveland, OH for his generous gift of a recombinant plasmid, pFF1, carrying the wild type *dna*Q gene *Salmonella typhimurium* and we are also grateful to Dr. Paul Baumann at the University of California, Davis for his kind gift of genomic DNA of *Buchnera aphidicola*. This work was supported by National Institutes of Health Grant GM28013 to J.I.

References

[1] Kornberg, A. and Baker, T.A. (1992) DNA Replication, 2nd edn. Freeman, San Francisco, CA.
[2] Ito, J. and Braithwaite, D.K. (1991) Nucleic Acids Res. 19, 4045–4057.
[3] Braithwaite, D.K. and Ito, J. (1993) Nucleic Acids Res. 21, 787–802.
[4] Sanjanwala, B. and Ganesan, A.T. (1989) Proc. Natl. Acad. Sci. USA 86, 4421–4424.
[5] Hammond, R.A., Barnes, M.H., Mack, S.L., Mitchener, J.A. and Brown, N.C. (1991) Gene 98, 29–36.
[6] Scheuermann, R.H. and Echols, H. (1984) Proc. Natl. Acad. Sci. USA 81, 7747–7751.
[7] Maki, H., Horiuchi, T. and Sekiguchi, M.C. (1983) Proc. Natl. Acad. Sci. USA 80, 7137–7141.
[8] Cox, E.C. and Horner, D.L. (1986) J. Mol. Biol. 190, 113–117.
[9] Lancy, E.D., Lifsics, M.R., Kehres, D.G. and Maurer, R. (1989) J. Bacteriol. 171, 5572–5580.
[10] Munson, M.A., Baumann, L. and Baumann, P. (1993) Gene 137, 171–177.

[11] Barnes, M.H., Hammond, R.A., Kennedy, C.C., Mack, S.L. and Brown, N.C. (1992) Gene 111, 43–49.
[12] Takano, K., Nakabeppu, Y., Maki, H., Horiuchi, T. and Sekiguchi, M. (1986) Mol. Gen. Genet. 205, 9–13.
[13] Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M. et al. (1995) Science 269, 496–512.
[14] Beese, L.S. and Steitz, T.A. (1991) EMBO J. 10, 25–33.
[15] Joyce, C.M. and Steitz, T.A. (1994) Annu. Rev. Biochem. 63, 777–822.
[16] Derbyshire, V., Grindley, N.D.F. and Joyce, C.M. (1991) EMBO J. 10, 17–24.
[17] Hori, H. and Osawa, S. (1979) Proc. Natl. Acad. Sci. USA 76, 381–385.
[18] Leavitt, M.C. and Ito, J. (1989) Proc. Natl. Acad. Sci. USA 86, 4465–4469.
[19] Bernad, A., Blanco, L., Lazaro, J.M., Martin, G. and Salas, M. (1989) Cell 59, 219–228.
[20] Zhu, W. and Ito, J. (1994) Nucleic Acids Res. 22, 5177–5183.
[21] Barnes, M.H., Spacciapoli, P., Li, D.H. and Brown, N.C. (1995) Gene 165, 51–56.
[22] Pacitti, D.F., Barnes, M.H., Li, D.H. and Brown, N.C. (1995) Gene 165, 45–50.
[23] Barnes, M.H., Tarantino, P.M.Jr., Spacciapoli, P., Brown, N.C., Yu, H. and Dybvig, K. (1994) Mol. Microbiol. 13, 843–854.
[24] Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M. et al. (1995) Science 270, 397–403.
[25] Gupta, R.S. (1995) Mol. Microbiol. 15, 1–11.
[26] Ochman, H. and Wilson, A.C. (1987) in: *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 1649–1654. ASM Press, Washington, DC.
[27] Munson, M.A., Baumann, P. and Kinesey, M.G. (1991) Int. J. Syst. Bacteriol. 41, 566–568.
[28] Munson, M.A. Baumann, P., Clark, M.A., Baumann, L., Moran, N.A., Voegtlin, D.J. and Campbell, B.C. (1991) J. Bacteriol. 173, 6321–6324.