T7 and *E.coli* share homology for replication-related gene products

Hiroyuki Toh

Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Received 7 October 1985

Recently, the complete nucleotide sequence of the bacteriophage T7 genome was determined and 50 genes were identified on the genome. We compared amino acid sequences of all the gene products of T7 and replication-related gene products of *E.coli*. As a result, we found that T7 and *E.coli* share homology for each pair of exonuclease, DNA primase and helix-destabilizing protein. For *E.coli*, these gene products are known to be involved in the process of discontinuous DNA replication. These observations suggest that T7 and *E.coli* have a common origin for a part of their replication systems.

Bacteriophage T7 *Homology* *5′–3′-Exonuclease* *DNA primase* *DNA polymerase I* *Helix-destabilizing protein*

1. INTRODUCTION

Bacteriophage T7 is a double-stranded DNA phage, whose particle contains a linear DNA genome of about 40000 bp long. Recently, the complete nucleotide sequence of T7 was determined and 50 known and potential genes were identified on the genome [1].

Ollis et al. [2] reported that the amino acid sequence predicted from the DNA polymerase gene of T7 is homologous to that of DNA polymerase I (Pol I) of *Escherichia coli*, which is known to play important roles in the repair of damaged DNA and the processing of Okazaki fragments. This observation suggests that these 2 proteins have evolved from a common ancestor.

Motivated by this observation, we carried out further comparison of the amino acid sequence between 50 gene products of T7 and replication-related proteins of *E. coli*. As a result, we found that T7 and *E. coli* share homology for 3 pairs of replication-related gene products, i.e., 5′–3′-exonuclease, DNA primase and helix-destabilizing protein. These observations, together with the previous work, delineate more clearly the evolutionary relationship between bacteriophage T7 and *E. coli*.

2. MATERIALS AND METHODS

Amino acid sequence comparison is one of the most useful methods for investigating the evolution and function of protein [3–6]. Therefore, our analysis described here was carried out by means of computer-assisted sequence comparison.

Amino acid substitution of a gene product that occurred during the evolutionary process is conservative and the degree of conservation of an essential gene product is high [7]. Therefore, we compared amino acid sequences between all the gene products of T7 and replication-related gene products of *E. coli*, because replication-related gene products are thought to be essential for any organisms.

The complete nucleotide sequence of T7 genome was taken from the paper of Dunn and Studier [1]. We selected *polA, rnh, ssb, dnaA, dnaG, dnaK* and *dnaQ* as replication-related genes and the nucleotide sequences of these genes were taken from the following papers: Joyce et al. [8] for

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

polA; Kanaya and Crouch [9] and Maki et al. [10] for *rnh*; Sancar et al. [11] for *ssb*; Hansen et al. [12] for *dnaA*; Smiley et al. [13] for *dnaG*; Bardwell and Craig [14] for *dnaK*; Maki et al. [10] for *dnaQ*.

Initial homology search by the matrix method was made with the program of Toh et al. [5]. According to the information from the homology matrices, the regions which share homology were aligned. Significance of the similarity was calculated with the program of Toh et al. [5], which subjects a pair of aligned sequences to the jumbling test [15]. In the calculation of percent homology and averaged scores for jumbling test, a continuous gap is counted as one substitution regardless of its length, whose score is given as -60.

3. RESULTS AND DISCUSSION

Amino acid sequence comparisons between replication-related gene products of *E. coli* and known and potential 50 gene products of T7 were made by means of the computer-assisted matrix method. As a result of the comparison, we found that each of 3 compared pairs of amino acid sequences shares extensive homology. Fig.1 shows the homology matrices of the 3 pairs. According to the matrices, each of the 3 homologous pairs was aligned (see fig.2).

Fig.1 (a1) shows the amino acid sequence comparison between T7 exonuclease and *E. coli* Pol I. For reference, amino acid sequence comparison of T7 DNA polymerase and *E. coli* Pol I is also shown in fig.1 (a2). The latter homology was reported by Ollis et al. [2]. The homologous region indicated by fig.1 (a1) was aligned (fig.2a). The degree of homology of this region is 34%. The significance of this homology is 7.02 SD, which indicates that the probability of occurrence of the homology by chance is less than 10^{-9}. Thus the N-terminal region of about 100 residues long of *E. coli* Pol I shares homology with exonuclease of T7. *E. coli* has 3 DNA polymerases; Pol I, Pol II and Pol III [16]. Pol I, which is encoded by the *polA* gene of *E. coli*, is a multifunctional single-subunit enzyme, which is involved in the repair of damaged DNA and the processing of Okazaki fragment. Three enzymatic activities have been assigned for this molecule; a DNA polymerase, a 3′–5′-exonuclease and a 5′–3′-exonuclease. Pol I is cleaved into 2 fragments by proteolysis [17,18]. The large fragment is derived from the C-terminal region of Pol I and is called 'Klenow fragment'. This large fragment has a DNA polymerase activity and a 3′–5′-exonuclease activity. The 3′–5′-exonuclease activity is seen in all 3 *E. coli*

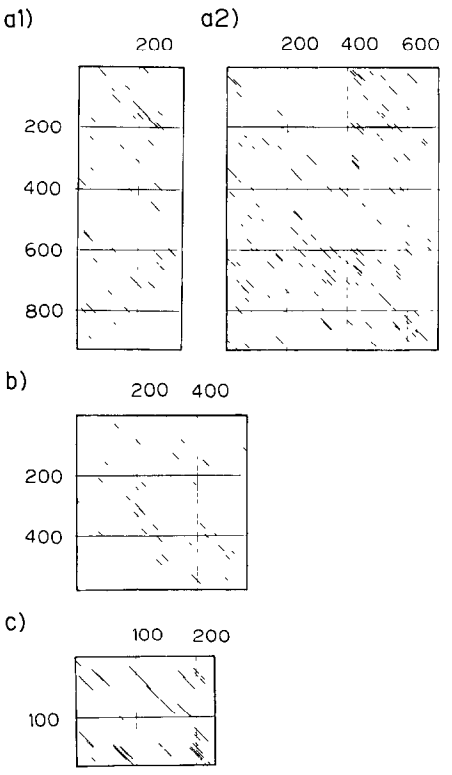

Fig.1. Homology matrix comparisons of amino acid sequences. Matrices (a) show the comparisons of *E. coli* Pol I (ordinate) against T7 DNA exonuclease (abscissa, a1) and T7 DNA polymerase (abscissa, a2). Matrix (b) shows the comparison of DNA primase between *E. coli* (ordinate) and T7 (abscissa). Matrix (c) shows the comparison of helix-destabilizing protein between *E. coli* (ordinate) and T7 (abscissa). Each diagonal line in matrix (a1) indicates a segment of 25 residues long that shows similarity above a threshold level, at which diagonal lines were generated with the probability of occurrence by chance of less than 6.4×10^{-4}. In matrix (a2), a diagonal line indicates a segment of 25 residues long and the probability of occurrence of a diagonal line by chance is less than 9.7×10^{-4}. In matrix (b), the segment is 25 residues long and the probability is less than 4.4×10^{-4}. In matrix (c), the segment is 20 residues long and the probability is 4.3×10^{-3}.

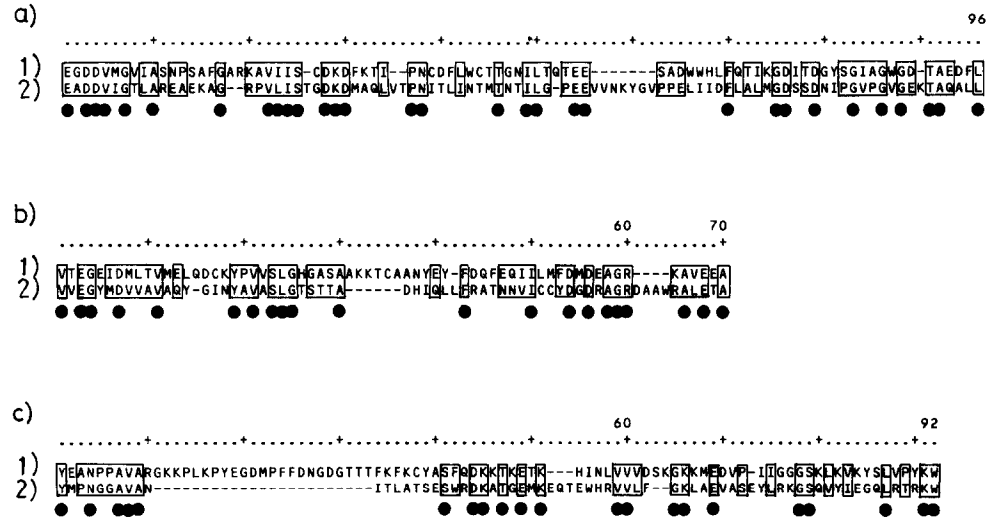

Fig.2. Alignments of amino acid sequences. Amino acid sequences were aligned according to the homology matrices in fig.1. In each alignment, the upper sequence is derived from T7 gene products and the lower from *E. coli*. (a) Alignment of DNA exonucleases between T7 (182–266) and *E. coli* (113–205). The numbers in parentheses indicate the residue sites from the N-terminal of the aligned region. (b) Alignment of DNA primases between T7 (155–219) and *E. coli* (262–324). (c) Alignment of helix-destabilizing protein between T7 (73–160) and *E. coli* (23–89). Gaps (-) were inserted to increase the homology. Positions occupied by identical or chemically similar amino acid residues are boxed. (●) Sites occupied by identical residues.

DNA polymerases and is involved in editing out mismatched bases during the DNA polymerization process [16]. Ollis et al. [2] reported that T7 DNA polymerase shows homology to the Klenow fragment of Pol I. Fig.1 (a2) shows the homology matrix of the comparison between DNA polymerase of T7 and Pol I. Diagonal lines observed in the C-terminal region on the matrix correspond to the homology reported by Ollis et al. [2]. Based on this homology, they suggested a possibility of common ancestry of these 2 polymerases. On the other hand, the smaller fragment is derived from the N-terminal region of Pol I and has a 5′–3′-exonuclease activity. 5′–3′-Exonuclease activity exists only in Pol I and is involved in the excision repair and the removal of RNA primer from Okazaki fragments during discontinuous replication [16]. Our finding is consistent with these observations. Therefore, we speculate that the DNA polymerase of T7 may take a subunit structure with the exonuclease, which may act in the processing of Okazaki fragment during the replication of T7.

Fig.1b shows the amino acid sequence comparison between T7 DNA primase and *E. coli dnaG* gene product. The matrix shows extensive but limited homology shared between the compared pair. The homologous region was aligned (see fig.2b). The degree of homology of this region is 34% and the significance of this similarity is 5.95 SD, which indicates that the probability of the homology occurring by chance is 2.0×10^{-9}. The *dnaG* gene encodes DNA primase of *E. coli* [13]. DNA primase is involved in the synthesis of primer RNA during the discontinuous replication process [19]. The homologous region may be important for the primase activity.

Fig.1c shows amino acid sequence comparison between helix-destabilizing proteins of T7 and *ssb* gene product of *E. coli*. The alignment of the homologous region is shown in fig.2c. The degree of homology of this region is 31% and the significance of this homology is 3.09 SD. The probability of occurrence by chance corresponding to the standard deviation is 1.0×10^{-3}. The *ssb* gene encodes helix-destabilizing protein, which is

also called single-strand binding protein, and is involved in DNA replication. In addition, this protein is known to participate in recombination, repair and SOS response [20]. The helix-destabilizing protein of *E. coli* can be divided into 3 domains based on the secondary structure predicted from its amino acid sequence [11]. The homologous region is included in the first domain, which has a high degree of secondary structure and is suggested to be important for DNA-binding activity [11].

These 3 cases of homology, together with the work of Ollis et al. [2], strongly suggest that a part of the replication system of bacteriophage T7 has a common origin with that of *E. coli*. Further comparison of other gene products between *E. coli* and T7 may uncover the evolutionary relationship between them. On the other hand, the functions of the some of the T7 gene products remain unknown. Therefore, sequence comparison between gene products, whose functions are already known, and potential gene products of T7 may provide information on the function of the T7 potential gene products.

REFERENCES

[1] Dunn, J.J. and Studier, F.W. (1983) J. Mol. Biol. 166, 477–535.
[2] Ollis, D.L., Kline, C. and Steiz, T.A. (1985) Nature 313, 818–819.
[3] Barker, W.C. and Dayhoff, M.O. (1982) Proc. Natl. Acad. Sci. USA 79, 2836–2839.
[4] Doolittle, R.F., Hunkapiller, M.W., Hood, L.E., Devare, S.G., Robbins, K.C., Aaronson, S.A. and Antoniades, H.N. (1983) Science 221, 275–277.
[5] Toh, H., Hayashida, H. and Miyata, T. (1983) Nature 305, 827–829.
[6] Toh, H., Kikuno, R., Hayashida, H., Miyata, T., Kugimiya, W., Inouye, S., Yuki, S. and Saigo, K. (1985) EMBO J. 4, 1267–1272.
[7] Kimura, M. (1983) in: The Neutral Theory of Molecular Evoltuion, pp.149–193, Cambridge University Press, Cambridge.
[8] Joyce, C.M., Kelley, W.S. and Grindley, N.D.F. (1982) J. Biol. Chem. 257, 1958–1964.
[9] Kanaya, S. and Crouch, R.J. (1983) J. Bacteriol. 154, 1021–1026.
[10] Maki, H., Horiuchi, T. and Sekiguchi, M. (1983) Proc. Natl. Acad. Sci. USA 80, 7137–7141.
[11] Sancar, A., Williams, K.R., Chase, J.W. and Rupp, W.D. (1981) Proc. Natl. Acad. Sci. USA 78, 4274–4278.
[12] Hansen, E.B., Hansen, F.G. and Von Meyenburg, K. (1982) Nucleic Acids Res. 10, 7373–7385.
[13] Smiley, B.L., Lupski, J.R., Svec, P.S., McMacken, R. and Godson, G.N. (1982) Proc. Natl. Acad. Sci. USA 79, 4550–4554.
[14] Bardwell, J.C.A. and Craig, E.A. (1984) Proc. Natl. Acad. Sci. USA 81, 848–852.
[15] Schwartz, R.M. and Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, suppl.3, pp.353–358, National Biomedical Research Foundation, Washington.
[16] Kornberg, A. (1980) in: DNA Synthesis, pp.67–121, Freeman, San Francisco.
[17] Klenow, H. and Henningsen, I. (1970) Proc. Natl. Acad. Sci. USA 65, 168–175.
[18] Brutlag, D., Atkinson, M.R., Setlro, P. and Kornberg, A. (1969) Biochem. Biophys. Res. Commun. 37, 982–989.
[19] Rowen, L. and Kornberg, A. (1978) J. Biol. Chem. 253, 758–764.
[20] Coleman, J. and Oakley, J. (1979) Crit. Rev. Biochem. 7, 247–289.