Evolution and structure of two ADP-ribosylation enterotoxins, *Escherichia coli* heat-labile toxin and cholera toxin

Tatsuo Yamamoto, Teruko Nakazawa*, Takashi Miyata+, Akira Kaji° and Takeshi Yokota

Department of Bacteriology, School of Medicine, Juntendo University, 2-1-1 Hongo, Bunkyo-ku, Tokyo,
**Department of Biochemistry, Yamaguchi University, School of Medicine, Ube, Yamaguchi,*
+Department of Biology, Kyushu University, Faculty of Science, Fukuoka, Japan, and
°Department of Microbiology, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Received 27 February 1984

Nucleotide sequence comparisons of the heat-labile enterotoxin (LTh) genes of *E. coli* pathogenic for humans with cholera toxin (CT) genes suggest that the two toxin genes have evolved from a common ancestry by a series of single selective base changes, while conserving the catalytic fragment A1 (ADP-ribose transferase). Based on the local hydrophilicity profiles of LTh and CT peptides, a transmembrane segment appears to be present in A1 in both toxins.

Escherichia coli *heat-labile enterotoxin* *Cholera toxin*
Sequence comparison *ADP-ribose transferase* *Transmembrane*

1. INTRODUCTION

E. coli heat-labile enterotoxin (LT) and *Vibrio cholerae* enterotoxin (cholera toxin, CT) are two members of the ADP-ribosylation toxin family, and cause watery diarrhea. The two toxins are structurally very similar. In both cases, the holotoxin consists of one molecule of subunit A and 5 molecules of subunit B [1–3]. Subunit B binds to the receptor (GM1-ganglioside) of the target cells [4–6]. Subunit A undergoes post-translational processing to generate A1 and A2 fragments, which are linked to each other by a single disulfide bond [1,2,7–9]. The A2 fragment presumably facilitates the binding of the A1 fragment to subunit B [1,2]. The A1 fragment has the catalytic function of transferring the ADP-ribose moiety of NAD to the GTP-binding regulatory component of adenylate cyclase, resulting in activation of adenylate cyclase and the subsequent elevation of intracellular cyclic AMP levels [1,9–11]. LT is divided into two distinct classes [12–14]: LTh produced only by *E. coli* pathogenic for humans, and LTp produced only by *E. coli* pathogenic for piglets.

The entire nucleotide sequences of the genes of subunits A and B of LTp [15,16], of CT [17,18] and of LTh [19–21] have been determined, and the entire amino acid sequences of subunits A and B of those toxins have been predicted. In those reports, comparisons of complete nucleotide sequences (and of predicted complete amino acid sequences) have also been made between CT and LTp [17], and between LTh and LTp [20,21]. However, those comparisons unexpectedly showed a significant heterogeneity in the catalytic fragment, A1, in terms of both amino acid number and amino acid sequence.

We show here that the LTh and CT genes comprise the same number of nucleotides (and thus, LTh and CT subunits share the same number of amino acids), and that the LTh and CT genes could have evolved from a common ancestry by a series of single base changes, while conserving the A1 fragment. Local hydrophilicity and predicted secondary structures of LTh and CT subunits are also compared for structure analyses.

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

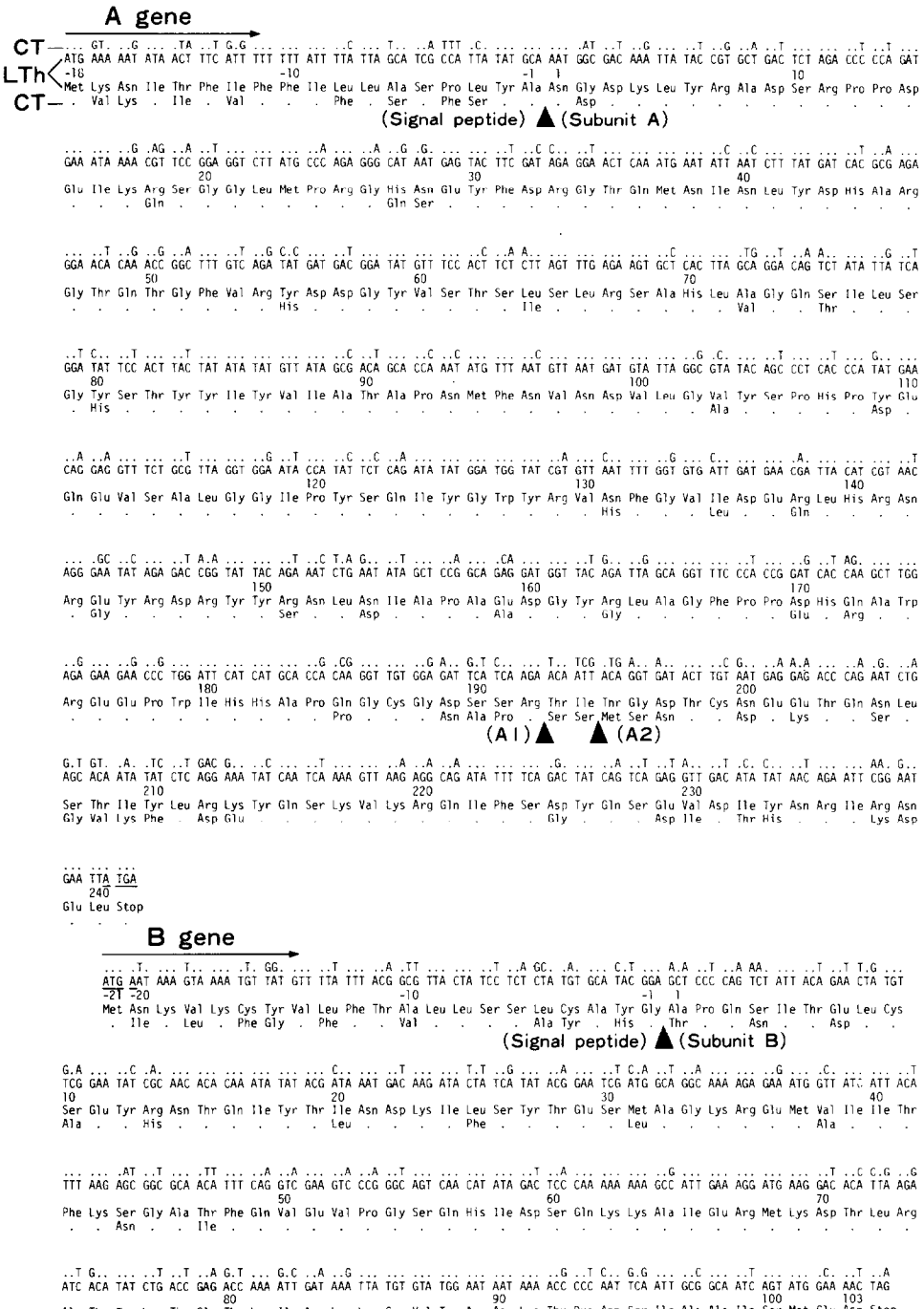


Fig.1. Sequence comparison of the LTh and CT genes and the predicted gene products. The LTh A and LTh B gene sequences are from [21] and [20], respectively. The CT A and CT B gene sequences are from [17] and [18], respectively. Nucleotides or amino acid residues of CT that are identical to those of LTh are shown as dots. (▲) Post-translational processing site. Nucleotides underlined overlap between the A and B genes.

2. METHODS

A local hydrophilicity profile of the peptides was obtained as in [22]. An averaged hydrophilicity value at residue i was calculated as follows [21]: sum of values of a running hexapeptide ($i-2$ to $i+3$)/6. The secondary structures of the peptides were predicted by computer as in [23]. Frequency (f) of use of optimal codons was calculated as in [24]: f = sum of numbers of optimal codons/sum of numbers of optimal and nonoptimal codons.

3. RESULTS AND DISCUSSION

The LTh A and LTh B genes lie on the operon that is presumably transcribed as a single mRNA; the LTh A gene is proximal to the LTh B gene [25]. The CT A and CT B genes also lie on a similar operon structure [17,26]. In both cases, the gene products are precursors of a subunit A or B [17,18,21,27,28].

3.1. Sequence comparison of the LTh and CT genes

As shown in fig.1, the LTh A and LTh B genes shared the same number of nucleotides with CT A and CT B genes, respectively. In the entire A and B regions, the homology between LTh and CT was 77.9% at the nucleotide level and 78.8% at the amino acid level. This sequence comparison strongly suggests that the LTh and CT genes have evolved from a common ancestry by a series of single base changes. The close resemblance of the LTh and CT genes was further confirmed with respect to codon usage and G + C content (mol%): the frequency of usage of optimal codons, which is closely related to the amount of gene product (protein) in *E. coli* [24], was 0.41 for the LTh A gene, 0.44 for the LTh B gene, 0.44 for the CT A gene and 0.41 for the CT B gene; G + C content of the LTh A and LTh B genes was 38.2 and 37.1%, respectively [20,21], and G + C content of the CT A and CT B genes was 38.5 and 32.8%, respectively (G + C contents of the B genes were slightly heterogenous).

3.2. Evolutionally conserved and diverged regions between LTh and CT

Table 1 shows the percent homology in various regions. Compared to nucleotide sequence homologies, amino acid sequence homologies varied markedly from region to region. The amino acid sequence homology was most apparent in the catalytic fragment, A1. The GM1-binding subunit B was also highly homologous. In marked contrast, signal peptides and A2 fragments showed less homology.

Fig.2 summarizes base changes between the LTh and CT genes and predicted peptide structures. The amino-terminal sequence of 130 amino acids in A1 (positions 1–130) was extremely homologous between LTh and CT (fig.2a). The homology was 91.5% (82.1% at the nucleotide level), and reached 94.6% when homologous amino acids were included. Four long and identical (conserved) sequences (C1–C4) existed (positions: C1, 3–17; C2, 29–54; C3, 81–102; and C4, 110–130). This highly homologous segment of 130 amino acids contained large and strongly hydrophobic regions (positions 69–106 and 113–134, fig.2a) as described previously with LTh [21]. No such strongly hydrophobic regions were present in the A2 fragment of subunit B. Those hydrophobic regions must represent a transmembrane segment (discussed later).

Recently, autoADP-ribosylation was reported to occur at Arg in CT A1 [29]. The Arg residue probably corresponded to Arg at position 146 (fig.1,2a). The Arg was located on a predicted random coil structure of the highest hydrophilicity (fig.2a), suggesting that the autoADP-ribosylation site sequence is looped out on the outer surface of the molecule. The corresponding Arg (position 146) in LTh A1 was also located on a predicted random coil structure of the highest hydrophilicity (fig.2a). Thus, the Arg (position 146) of LTh A1 could also be the site of autoADP-ribosylation. Like the sequence surrounding the autoADP-ribosylation site, the amino-terminal C1 region in A1 (of both LTh and CT) was also rich in charged amino acids and located on a predicted random coil structure of high hydrophilicity. This C1 region may also participate in the catalytic function.

As indicated earlier [21], there are markedly divergent regions (D1 and D2) in subunit A (positions 189–197 and 205–213, respectively, fig.2a). Of these regions, D1 manifested a unique base change pattern; codon latter changes occurred predominantly at the first position. Based on this, we speculate that the divergence is a result of adaptation of the D1 region to *E. coli* or *V. cholerae*

which may have a 'distinct' proteolytic enzyme to generate A1 and A2 fragments. The D1 sequence is:

187		189					↓		↓		197		199	
Cys	Gly	Asp	Ser	Ser	Arg	Thr	Ile	Thr	Gly	Asp	Thr	Cys		(LTh A)
Cys	Gly	Asn	Ala	Pro	Arg	Ser	Ser	Met	Ser	Asn	Thr	Cys		(CT A)

(D1 sequence is marked with a box; arrows represent the cleavage sites; the two Cys residues form a disulfide bond). The D1 region was located on a predicted random coil structure. The D1 may be looped out on the protein surface, and may create an appropriate structure which is more accessible to the active site of the proteolytic enzymes of *E. coli* and *V. cholerae*. In contrast, base changes in the other significantly divergent region, D2, seem to be at random, suggesting that the amino acid sequence in this region is variable. Signal peptides were also significantly divergent, however, the hydrophobic characteristics that are required for membrane penetration were well conserved (fig.2a,b).

Trp (position 88) of subunit B is involved in the GM1-binding reaction [30]. As shown in fig.2b, the Trp was located in the very strictly conserved C2 region (positions 84–93); this entire sequence may function in the binding reaction. Another highly conserved region, C1 (in subunit B), was the longest of all conserved sequences found. Therefore, this C1 sequence (in subunit B) must play an important functional role.

Table 1

Homology between LTh and CT

Region	Homology (%)	
	Amino acid level	DNA level
A		
Precursor (268)[a]	79.1 (83.3)[b]	78.2
Signal peptide (18)[a]	55.6 (61.1)[b]	72.2
A1 fragment (192)[a]	87.0 (90.1)[b]	81.1
A2 fragment (46)[a]	58.7 (65.2)[b]	70.3
(Subunit A [240][a]	80.8 [85.0][b]	78.6)
B		
Precursor (124)[a]	78.2 (84.7)[b]	77.4
Signal peptide (21)[a]	57.1 (71.4)[b]	74.6
Subunit B (103)[a]	82.5 (87.4)[b]	78.0

[a] Numbers represent residue numbers of peptides
[b] Homologous amino acids are included

3.3. *Amino acid sequence divergence between LTh and LTp*

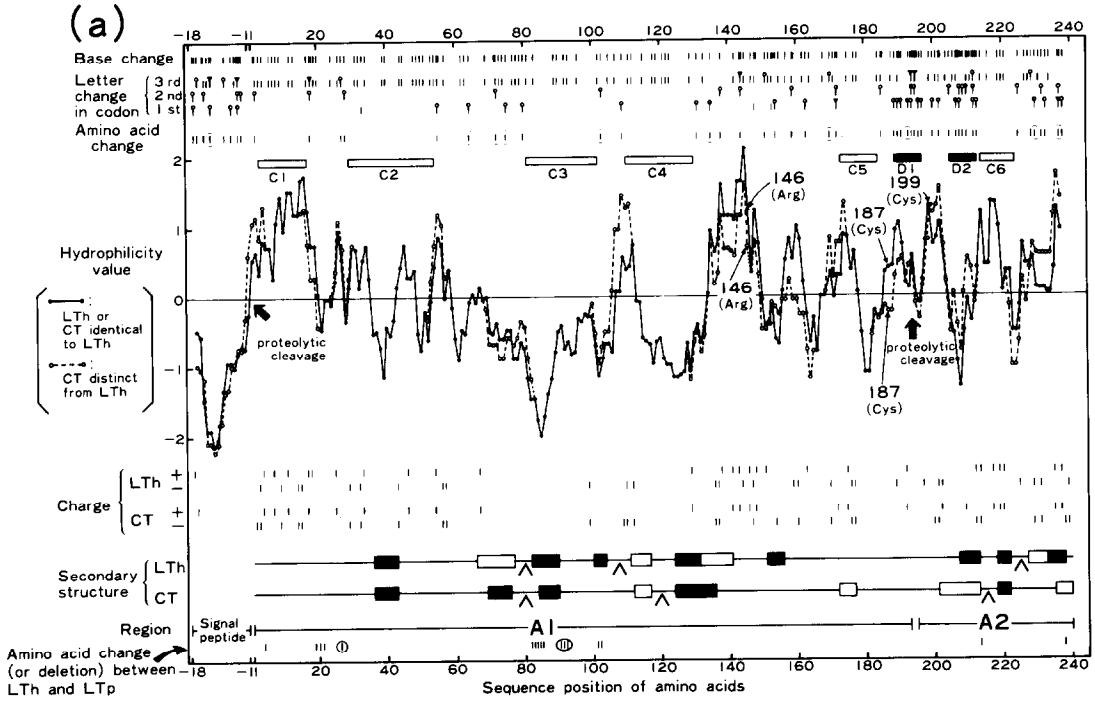
Amino acid sequence positions where amino acid changes or deletions (in LTp) occur are indicated in fig.2 (bottom). With the exception of the C3 region in the A1 fragment, the highly conserved regions discussed above, between LTh and CT, and the D1 sequence of LTh were also seen in LTp. The reason why so many amino acid changes and deletions (in LTp) are unusually compressed in the A1-C3 region is not known.

3.4. *Transmembrane model*

CT A1 (ADP-ribose transferase) is most probably transferred across the membrane of the target cells into the cytosol during the intoxication process [1,31]. Gill presented a possible model of the entry of CT A1 in which initially CT B enters into the membrane and then creates a channel to allow the entry of CT A1. However, our present and previous data [21] clearly show that, in both LTh and CT, strongly hydrophobic regions, which are required for entry into layers, do not exist in subunit B, but do exist within the A1 fragment. Based on this, we present a new possible entry model for LTh and CT (fig.3). The outline of the model is that:

(i) In the holotoxin, the A1 fragment is packed into a subunit B pentamer with the aid of the A2 fragment to shield the hydrophobicity; and

(ii) After the binding of subunit B to the GM1 receptor, the subunit B pentamer changes conformation so as to open the rigid structure, resulting in the entry of the A1 fragment from its leader segment (hydrophobic sequences).



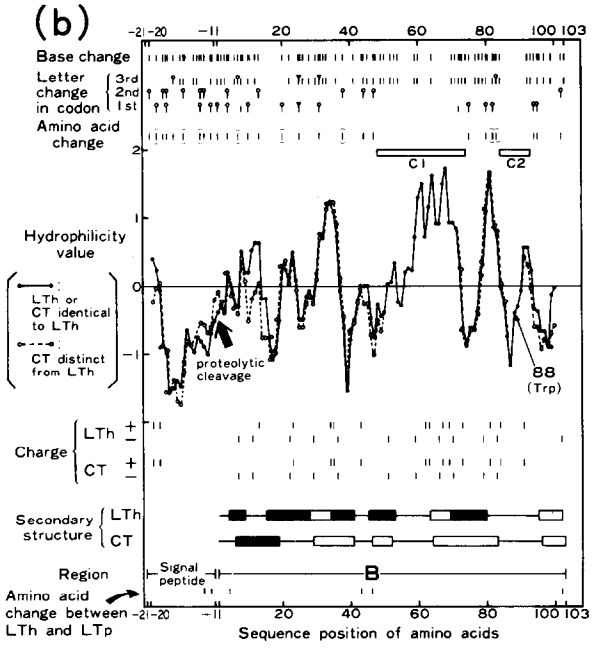


Fig.2. Base changes between the LTh and CT genes and predicted peptide structures. (a) LTh A- and CT A-precursors, (b) LTh B- and CT B-precursors. Sequence numbers are the same as those in fig.1. Symbols used in codon-letter change: (|) base change without amino acid change, (♀) base change with amino acid change, (♂) base change with amino acid change in case. Symbols used in predicted secondary structures: (▭) α-helix, (▬) β-sheet, (⌒) β-turn, (——) random coil. C1–C6 represent conserved regions of more than 10 residues, and D1 and D2 represent divergent regions of 9 residues. A local hydrophilicity profile, a charged amino acid distribution and predicted secondary structures of LTh are from [21]. Amino acid sequence differences between LTh and LTp, shown at the bottom, are from [20,21]: (|) positions of amino acid change; positions of amino acid deletion in LTp are marked with a circle.

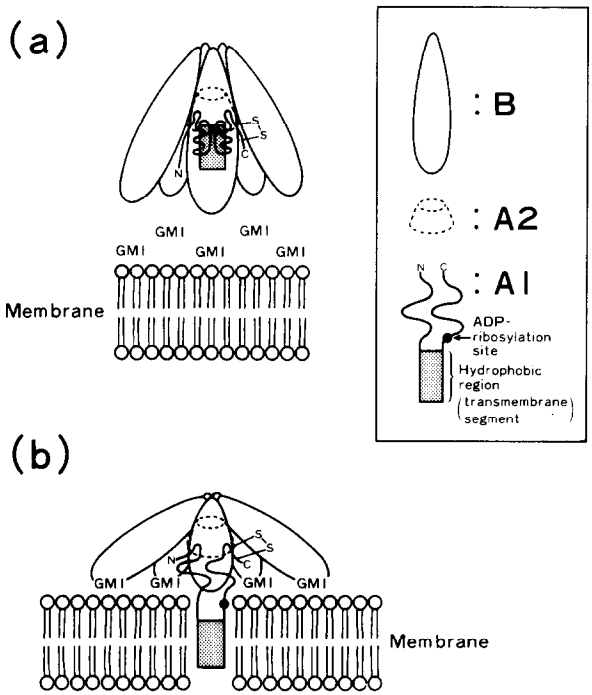


Fig.3. A possible model of the entry of LTh A1 and CT A1 across the membrane. (a) Holotoxin, (b) entry of the A1 fragment across the membrane.

ACKNOWLEDGEMENTS

We thank Professor Atsushi Nakazawa for helpful suggestions and stimulating discussions, and also Mr Mamoru Yamada for computer analyses. This work was supported by grants-in-aid for scientific research from the Ministry of Education, Science, and Culture of Japan.

REFERENCES

[1] Gill, D.M. (1976) Biochemistry 15, 1242–1248.
[2] Klapper, D.G., Finkelstein, R.A. and Capra, J.D. (1976) Immunochemistry 13, 605–611.
[3] Gill, D.M., Clements, J.D., Robertson, D.C. and Finkelstein, R.A. (1981) Infect. Immun. 33, 677–682.
[4] Holmgren, J. (1973) Infect. Immun. 8, 851–859.
[5] Heyningen, S.V. (1974) Science 183, 656–657.
[6] Donta, S.T. and Viner, J.P. (1975) Infect. Immun. 11, 982–985.
[7] Kurosky, A., Markel, D.E., Touchstone, B. and Peterson, J.W. (1976) J. Infect. Dis. 133 (suppl.) S14–S22.
[8] Lai, D.Y., Mendez, E. and Chang, D. (1976) J. Infect. Dis. 133 (suppl.) S23–S30.
[9] Gill, D.M. and Richardson, S.H. (1980) J. Infect. Dis. 141, 64–70.
[10] Evans, D.J., Chen, L.C., Curlin, G.T. and Evans, D.G. (1972) Nat. New Biol. 236, 137–138.
[11] Moss, J. and Richardson, S.H. (1978) J. Clin. Invest. 62, 281–285.
[12] Honda, T., Tsuji, T., Takeda, Y. and Miwatani, T. (1981) Infect. Immun. 34, 337–340.
[13] Geary, S.J., Marchlewicz, B.A. and Finkelstein, R.A. (1982) Infect. Immun. 36, 215–220.
[14] Takeda, Y., Honda, T., Sima, H., Tsuji, T. and Miwantani, T. (1983) Infect. Immun. 41, 50–53.
[15] Dallas, W.S. and Falkow, S. (1980) Nature 288, 499–501.
[16] Spicer, E.K. and Noble, J.A. (1982) J. Biol. Chem. 257, 5716–5721.
[17] Mekalanos, J.J., Swartz, D.J., Pearson, G.D.N., Harford, N., Groyne, F. and Wilde, M. (1983) Nature 306, 551–557.
[18] Lockman, H. and Kaper, J.B. (1983) J. Biol. Chem. 258, 13722–13726.
[19] Yamamoto, T., Tamura, T., Yokota, T. and Takano, T. (1982) Mol. Gen. Genet. 188, 356–359.
[20] Yamamoto, T. and Yokota, T. (1983) J. Bacteriol. 155, 728–733.
[21] Yamamoto, T., Tamura, T. and Yokota, T. (1984) J. Biol. Chem., in press.
[22] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828.
[23] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[24] Ikemura, T. (1981) J. Mol. Biol. 151, 389–409.
[25] Yamamoto, T., Yokota, T. and Kaji, A. (1981) J. Bacteriol. 148, 983–987.
[26] Pearson, G.D.N. and Mekalanos, J.J. (1982) Proc. Natl. Acad. Sci. USA 79, 2976–2980.
[27] Yamamoto, T., Tamura, T., Ryoji, M., Kaji, A., Yokota, T. and Takano, T. (1982) J. Bacteriol. 152, 506–509.
[28] Gennaro, M.L. and Greenaway, P.J. (1983) Nucleic Acids Res. 11, 3855–3861.
[29] Lai, C.Y., Xia, Q.C. and Salotra, P.T. (1983) Biochem. Biophys. Res. Commun. 116, 341–348.
[30] Wolf, M.J.S., Fridkin, M., Epstein, M. and Kohn, L.D. (1981) J. Biol. Chem. 256, 5481–5488.
[31] Tsuru, S., Matsuguchi, M., Ohtomo, N., Zinnaka, Y. and Takeya, K. (1982) J. Gen. Microbiol. 128, 497–502.