Vol. 201, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 30, 1994 Pages 466-471

IDENTIFICATION OF *dnaK* MULTIGENE FAMILY IN *SYNECHOCOCCUS* SP. PCC7942

Kaori Nimura, Hirofumi Yoshikawa and Hideo Takahashi*

Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Received April 10, 1994

Three *dnaK* gene homologs have been cloned and sequenced from cyanobacterium *Synechococcus* sp. PCC7942 using a set of primers designed from two conserved regions of known *dnaK* genes. This is the first example of triple genes for *dnaK* from prokaryotic cells. These three genes were derived from different loci of the chromosome.

The response to heat shock is a universal phenomenon. To date, all organisms investigated respond to elevated temperature by the induction of a set of proteins which have been specified as heat shock proteins [8]. One of these proteins has an apparent molecular mass of 70 kDa (HSP70; bacterial homolog known as the DnaK protein). Although the synthesis of HSP70 is enhanced under various stress conditions, it also constitutes a major protein under normal growth conditions and has been shown to be essential for cellular growth [5]. They function as molecular chaperons which assist in the correct folding of other polypeptides and/or their assembly into oligomeric structures but which are not components of the final functional structures [3, 4].

Cyanobacteria are prokaryotic cells which have developed oxygen-producing photosynthetic system similar to chloroplasts of higher plants. Under natural conditions they inhabit in the area with suitable amounts of sunlight where they are inevitably subject to a variety of stresses such as UV irradiation and high temperature. Here we report cloning and sequencing of *dnaK* homolog genes from a transformable cyanobacterium [11] *Synechococcus* sp. PCC7942. We found three *dnaK* homolog genes in this microorganism.

* Corresponding author. FAX: (813) 3813-0539.

0006-291X/94 $5.00

MATERIALS AND METHODS

Oligonucleotide primers

Oligodeoxyribonucleotides were synthesized using an ABI392 DNA/RNA synthesizer (Applied Biosystems Japan, Tokyo). Mixed oligonucleotides corresponding to two conserved regions of DnaK proteins were used as primers for polymerase chain reaction (PCR). One mixed primer (forward) is 5'GG(T/C)GG(T/C)AC(T/C)TT(T/C)GA(T/C)GT3' corresponding to the amino acid sequence GGTFDV and another (reverse) is 5'CCGTT(A/G)GC(A/G)TC(A/G)AT(A/G)TC(A/G)AA3' corresponding to FDIDANG. Primers used for inverse PCR are K1u: 5'CGTGGTTGTCGGCAGGGTT3' (corresponds to the nucleotide number 1521 to 1503 in Fig. 1), K1d: 5'CCAGTCGCACTCAGGTC3' (1955 to 1971 in Fig. 1), K2u: 5'TTCAGTTTGAGTGGCGCTG3' (945 to 927 in Fig. 2) and K2d: 5'CAAGAAGTCGGAAACCTTCT3' (1365 to 1384 in Fig. 2),

DNA sequencing

The PCR products to be sequenced were cloned into *Hinc*II site of pTZ18R (TOYOBO, Tokyo) DNA. For larger DNA fragments, DNAs were sonicated and the fragments 0.4 to 0.7 kb in size were recovered from an agarose gel by using GENECLEAN II (Funakoshi, Tokyo). After blunting ends, they were cloned into *Hinc*II site of pTZ18R. Sequencing was performed by the dideoxy chain termination method [12] using the Taq Dye Primer Cycle Sequencing Kit (Applied Biosystems Japan, Tokyo) with an ABI 373A DNA Sequencer (Applied Biosystems).

Southern hybridization

Genomic DNA (10μg) was completely digested with a restriction enzyme and electrophoresed on a 0.8% agarose gel, transfer to Hybond-N nylon membrane (Amersham), and hybridized with labeled probes specific for each *dnaK*. Plasmids containing each *dnaK* gene was labeled by using Photo ChemiProbe Kit (Takara Shuzo Co., LTD.,Kyoto) and bands were detected according to manufacturer's instruction.

RESULTS AND DISCUSSION

By using mixed oligonucleotide primers, about 850 bp fragments were amplified. The PCR products were cloned and nucleotide sequences were determined. We analyzed sequences of total 7 clones and found that there were at least three different sequences highly homologous to known *dnaK* genes. These three genes were tentatively named *orf1*, *orf2*, *orf3* respectively. Flanking regions of *orf1* and *orf2* genes were cloned by using the inverse PCR method [10]. Southern hybridization of the chromosomal DNA digested with several restriction enzymes indicated that a 2.7 kb *Bgl*II fragment was hybridized with *orf1* probe (data not shown). Therefore *Bgl*II fragments of about this size were isolated, self-ligated, then upstream and downstream regions of *orf1* was amplified by using primers K1u and K1d. Similarly 2.0 kb *Bgl*II fragment was used to amplify upstream region of *orf2* with primers K2u and K2d and for downstream region, 2.5 kb *Eco*RI fragment was used with same primers. Entire gene sequences of *orf1* and *orf2* and flanking regions were determined from these clones. The reading frame *orf1* encodes a polypeptide containing 655 amino acids whose molecular weight is 71 kDa, while *orf2* encodes for a

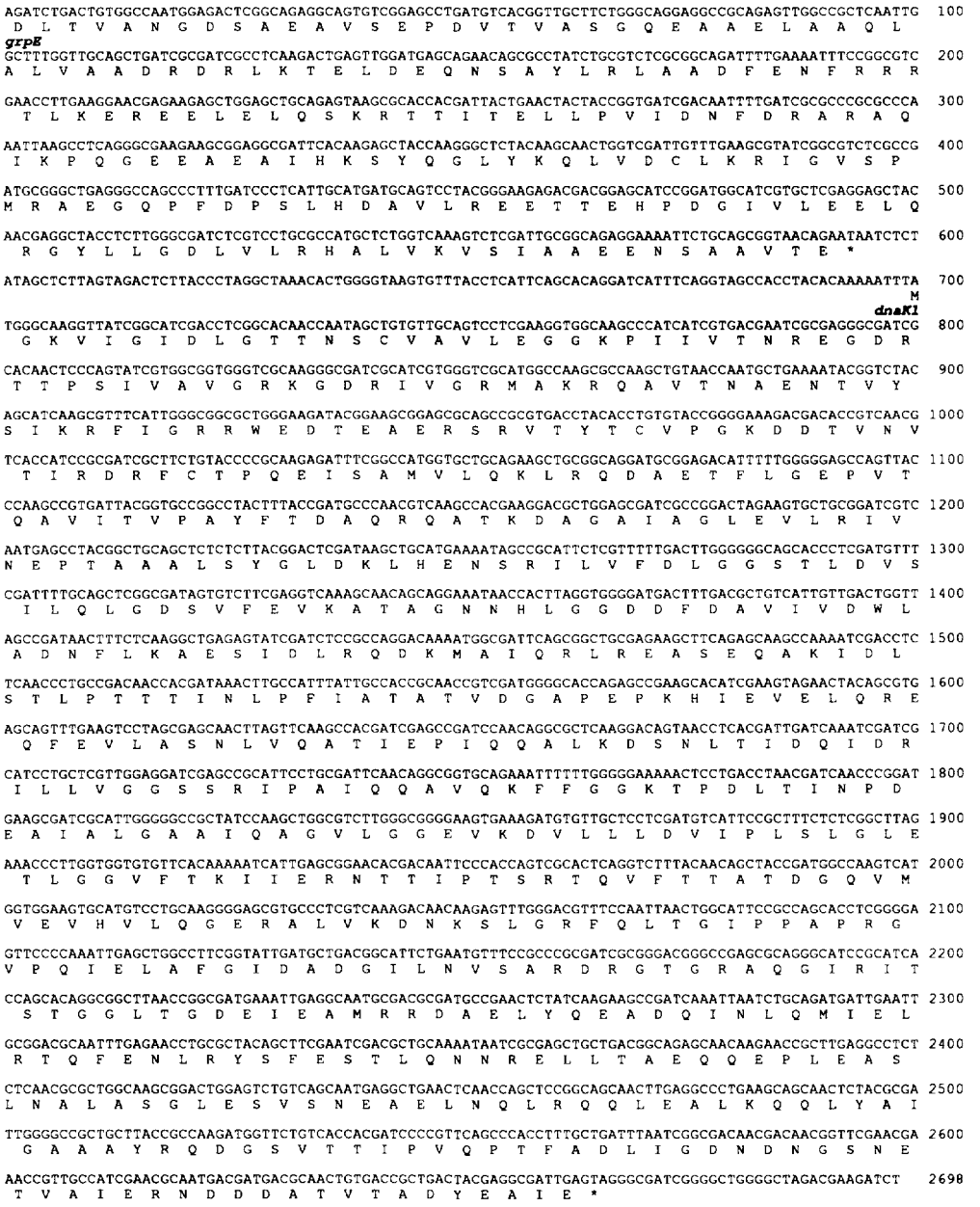

Fig. 1. Nucleotide sequence of the *Synechococcus* sp. strain PCC7942 *dnaK1* gene and of the flanking regions. The deduced amino acid sequences of the open reading frames corresponding to DnaK1 and GrpE are shown underneath. Sequence data of this region has been assigned in the GSDB, DDBJ, EMBL and NCBI data bases with accession no. D28550.

polypeptide of 68 kDa containing 634 amino acids. These amino acid sequences are remarkably conserved when compared with the HSP70 family from other bacteria or eukaryotes, therefore *orf1* and *orf2* were assigned to *dnaK1* and *dnaK2* respectively (Fig. 1 and 2). The highest score 54.7% was obtained

```
CGAGGTTCGGGAACCCCGCCTTGGAAGAGGGGGCGGCTCACCGGCAAGATAGTCGAAACTCAAAAGAGTTCACCCCTGTTGGCGGTTGGCCGTCAGCAGA  100

TGTAATCCCTCCCAAGCTGCACTATGGCCAAAGTTGTCGGAATCGACCTCGGAACCACCAACTCTTGCGTGGCTGTCATGGAGGGCGGCAAGCCCACTGT  200
                      M  A  K  V  V  G  I  D  L  G  T  T  N  S  C  V  A  V  M  E  G  G  K  P  T  V
                                dnaK2
GATCGCTAATGCGGAAGGTTTTCGCACCACTCCTTCAGTCGTTGCTTTTGCGAAAAACCAAGACCGCCTCGTGGGTCAAATCGCCAAACGCCAGGCGGTG  300
 I  A  N  A  E  G  F  R  T  T  P  S  V  V  A  F  A  K  N  Q  D  R  L  V  G  Q  I  A  K  R  Q  A  V

ATGAACCCCGAGAACACCTTCTACTCGGTTAAGCGCTTCATCGGCCGTCGTCCGGATGAAGTCACGAACGAACTGACCGAAGTGGCCTACAAAGTCGATA  400
M  N  P  E  N  T  F  Y  S  V  K  R  F  I  G  R  R  P  D  E  V  T  N  E  L  T  E  V  A  Y  K  V  D  T

CTTCGGGCAATGCCGTCAAGCTGGATAGCTCCAATGCTGGCAAGCAGTTCGCTCCTGAAGAAATTTCGGCGCAGGTGCTGCGCAAACTGGCCGAAGACGC  500
 S  G  N  A  V  K  L  D  S  S  N  A  G  K  Q  F  A  P  E  E  I  S  A  Q  V  L  R  K  L  A  E  D  A

CAGCAAATACCTGGGTGAAACCGTCACCCAAGCCGTGATCACGGTTCCGGCCTACTTCAATGACTCCCAGCGCCAAGCGACCAAAGACGCTGGCAAAATC  600
 S  K  Y  L  G  E  T  V  T  Q  A  V  I  T  V  P  A  Y  F  N  D  S  Q  R  Q  A  T  K  D  A  G  K  I

GCCGGCCTAGAAGTGCTGCGGATCATCAACGAGCCGACGGCAGCCGCGCTGGCCTACGGTCTTGATAAGAAGAGCAACGAACGCATCCTTGTCTTTAACT  700
A  G  L  E  V  L  R  I  I  N  E  P  T  A  A  A  L  A  Y  G  L  D  K  K  S  N  E  R  I  L  V  F  N  L

TGGGCGGCGGTACTTTCGACGTCTCGGTCTTGGAAGTGGGCGACGGCGTTTTTGAAGTGCTGGCGACCTCGGGTGATACCCACCTCGGTGGCGACGACTT  800
 G  G  G  T  F  D  V  S  V  L  E  V  G  D  G  V  F  E  V  L  A  T  S  G  D  T  H  L  G  G  D  D  F

CGACAAAAAAATCGTTGACTTCCTGGCTGGTGAATTCCAGAAGAACGAAGGCATCGATCTGCGCAAAGACAAGCAGGCTCTGCAGCGTCTGACGGAAGCC  900
 D  K  K  I  V  D  F  L  A  G  E  F  Q  K  N  E  G  I  D  L  R  K  D  K  Q  A  L  Q  R  L  T  E  A

GCTGAGAAAGCCAAAATCGAGCTGTCCAGCGCCACTCAAACTGAAATCAACCTGCCCTTCATCACGGCAACCCAAGACGGGCCGAAGCACCTCGACCTGA 1000
A  E  K  A  K  I  E  L  S  S  A  T  Q  T  E  I  N  L  P  F  I  T  A  T  Q  D  G  P  K  H  L  D  L  T

CCTTAACCCGCGCCAAGTTTGAAGAATTGGCTTCGGATCTGATCGATCGCTGCCGGATTCCGGTGGAGCAAGCGATCAAAGATGCCAAGTTGGCCCTGAG 1100
 L  T  R  A  K  F  E  E  L  A  S  D  L  I  D  R  C  R  I  P  V  E  Q  A  I  K  D  A  K  L  A  L  S

CGAAATTGACGAAATCGTCTTGGTCGGTGGTTCGACCCGGATTCCTGCGGTGCAGGCGATCGTCAAGCAAATGACGGGCAAAGAGCCCAACCAAAGTGTC 1200
 E  I  D  E  I  V  L  V  G  G  S  T  R  I  P  A  V  Q  A  I  V  K  Q  M  T  G  K  E  P  N  Q  S  V

AACCCCGATGAGGTGGTGGCGATCGGTGCGGCGATTCAAGGTGGCGTCTTGGCTGGGGAAGTCAAAGACATCCTGCTGCTCGACGTGACGCCACTATCCT 1300
N  P  D  E  V  V  A  I  G  A  A  I  Q  G  G  V  L  A  G  E  V  K  D  I  L  L  L  D  V  T  P  L  S  L

TGGGGGTAGAAACCCTTGGTGGCGTGATGACTAAGTTGATCCCACGCAACACCACTATCCCCACCAAGAAGTCGGAAACCTTCTCGACGGCGGCGGACGG 1400
 G  V  E  T  L  G  G  V  M  T  K  L  I  P  R  N  T  T  I  P  T  K  K  S  E  T  F  S  T  A  A  D  G

TCAAACCAACGTCGAAATCCACGTGCTCCAAGGCGAGCGCGAAATGGCCAGCGACAACAAGAGCTTGGGAACCTTCCGGCTGGATGGCATTCCGCCGGCT 1500
 Q  T  N  V  E  I  H  V  L  Q  G  E  R  E  M  A  S  D  N  K  S  L  G  T  F  R  L  D  G  I  P  P  A

CCCCGTGGCGTGCCCCAAATCGAAGTGATCTTCGACATCGACGCTAACGGCATCCTCAATGTCACGGCCAAAGACAAAGGGTCGGGCAAAGAGCAGTCGA 1600
P  R  G  V  P  Q  I  E  V  I  F  D  I  D  A  N  G  I  L  N  V  T  A  K  D  K  G  S  G  K  E  Q  S  I

TCAGCATCACCGGCGCTTCGACCTTGTCTGACAACGAAGTCGATCGCATGGTCAAAGACGCCGAAGCGAATGCAGCAGCGGACAAAGAACGGCGCGAACG 1700
 S  I  T  G  A  S  T  L  S  D  N  E  V  D  R  M  V  K  D  A  E  A  N  A  A  A  D  K  E  R  R  E  R

TATCGACCTGAAGAACCAAGCCGACACGCTGGTCTATCAGTCTGAGAAACAACTCAGCGAGCTGGGTGACAAGATCTCGGCTGATGAGAAAAGCAAAGTC 1800
 I  D  L  K  N  Q  A  D  T  L  V  Y  Q  S  E  K  Q  L  S  E  L  G  D  K  I  S  A  D  E  K  S  K  V

GAAGGCTTTATCCAAGAGCTGAAAGATGCCTTGGCTGCCGAAGACTACGACAAGATCAAGTCGATCATCGAGCAACTGCAGCAAGCTCTCTACGCCGCTG 1900
E  G  F  I  Q  E  L  K  D  A  L  A  A  E  D  Y  D  K  I  K  S  I  I  E  Q  L  Q  Q  A  L  Y  A  A  G

GCAGCAGCGTCTACCAGCAGGCTAGCGCTGAAGCTTCGGCCAACGCCCAAGCCGGTCCTTCCTCGTCCTCGAGCAGCAGCTCTGGCGATGATGATGTGAT 2000
 S  S  V  Y  Q  Q  A  S  A  E  A  S  A  N  A  Q  A  G  P  S  S  S  S  S  S  S  S  G  D  D  D  V  I

TGACGCAGAGTTCTCTGAGTCGAAGTAATCACCTAGGCAGTTTGGGAATGAATCGCCGGGCCGCAAGGTCCGGCTTTTTTATGCTTCAGCCTCGATTCAT 2100
 D  A  E  F  S  E  S  K  *

TTCTGCGATCGCCCGGCTGAAGCGCAGCTCACTTTCCCAACACGATCGCGGTTCGAGCGTCCCATCCAAGGGATACACTGAGCCTGTTCCTGTGAAAGAT 2200

CATCGATGGCTGCTGCACCGAGTTCCGTGGCACCGTTGGCTGCACGAGCCTTGCGGTTAGTAGGTCTCGTCCTTGTTCTCTCTGTGCTCGTTGACTATGT 2300
     M  A  A  A  P  S  S  V  A  P  L  A  A  R  A  L  R  L  V  G  L  V  L  V  L  S  V  L  V  D  Y  V
     orf2.3
TGCCAGTTTGGTCGGTGCCAACTGGAGCGAACCACGTTGGTACGTGGCCACCAGCAGCCAATGGATCGATCGCGGGGTTGTGCCGCTGGTGGGCCTGACG 2400
 A  S  L  V  G  A  N  W  S  E  P  R  W  Y  V  A  T  S  S  Q  W  I  D  R  G  V  V  P  L  V  G  L  T

TTCTGGTTTAGTGGCGATTGGCTCGCCCAACAACTCGATCCTTTGTCCGTCACGCAACAAAAACTGATCCGGGCGATCGCGCGGATTTTGCCCTTGGTGC 2500
F  W  F  S  G  D  W  L  A  Q  Q  L  D  P  L  S  V  T  Q  Q  K  L  I  R  A  I  A  R  I  L  P  L  V  L

TCACAGTCATTTTTATCCTGGCGGTCCCGATTCACCTCAGCAATGCCTTCACCAGCCAAGCCCAACGGCTGGAAGAAGTTCAGCAGCAGGTGCAGCAACA 2600
 T  V  I  F  I  L  A  V  P  I  H  L  S  N  A  F  T  S  Q  A  Q  R  L  E  E  V  Q  Q  Q  V  Q  Q  Q

GCGACAAGCCTCGATCTGCAGTTTCAAGGGGCGATCGCTCAGCAGAAACAGCAACTGAAGCAGCTATTTGGCACCCCCGGCCTGTTTGAGCAAGTCAAGG 2700
 R  Q  A  S  I  C  S  F  K  G  R  S  L  S  R  N  S  N  *

CCAGCGCCAAATTCCGGCCGATCAAGTTGGCGAGCTCGAAAAGCTAGTAGGCAAACCGGAAGCCGAAATCGATCGCTTTGTCGATCAGCGTGCGGCAGAA 2800

GCTAAGCAAACCCAAGAAGCTGAAATCAAAGCGAACCAAGAGAAGCAATCTAACCTCATTCGCCGCGAAGTCACCGGCAATACGATCCGAACTTCGCTGC 2900

TTGGGATCCTGCTCGTCATTGGCTATGGCGGCATCAGCTGGGGCATCCTGCGCCAAAGTCCGAACCACAACTGATCGCTGAGCTGCGCCTCCTCTAGCGG 3000

CGCTTTAGCCTCGCAGTTGTCGGATCTAACACTCGCTGCGATCGCCCTTGATCCCCTGCCCCCCCGCGGAGGAGTCGGATTGTGTCACAGTGAAGCAGGC 3100

CCTAGCTGCTAGCGATCGCTTGCTGAGACTGGAAACTTAGCGCTGAGCCAGCCCAAGGTGCTGCGCTGTGACCCTTTGCCTCACTCCCCGAAGACCATGT 3200

CTGACGCCCCCAATCACAGCTCTCTTAGTTTTTGGCAACGGCTGCGGCAGGATTTCAAAAATGATTTGATCGCCGGTTTGCTGGTGGTCATTCCCCTAGC 3300

AACCACAATCTGGCTGACGGCAGTGGTCTCCCGCTGGGTTGTCAACCTGCTGACAGGAATTC                                          3362
```

Fig. 2. Nucleotide sequence of the *Synechococcus* sp. strain PCC7942 *dnaK2* gene and of the flanking regions. The deduced amino acid sequences of DnaK2 and Orf2.3 are shown underneath. Sequence data of this region has been assigned in the GSDB, DDBJ, EMBL and NCBI data bases with accession no. D28551.

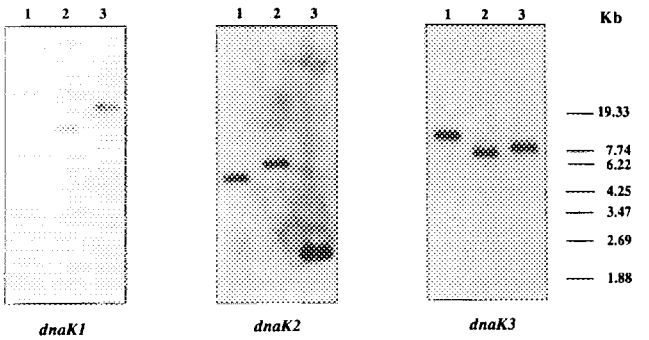

Fig. 3. Southern hybridizations of total cellular DNA with individual *dnaK* probes. Genomic DNA was completely digested with *Eco*RI (lane 1), *Bam*HI (lane 2), or *Hind*III (lane 3).

between DnaK1 and *Bacillus subtilis* DnaK protein [14]. DnaK2 shares 82.3% sequence identity with another cyanobacteria *Synechocystis* sp. PCC6803 protein [2] and 72.3% with *Cryptomonas* Φ chloroplast protein [13]. This surprisingly high homology between chloroplast and cyanobacterial protein is consistent with the postulated cyanobacterial origin of chloroplast [6].

A sequence analysis of the flanking regions of *orf3* is under way and we observed extensive homology between *orf3* and other *dnaK* genes, then we also designated *orf3 dnaK3*. Representative clones of each *dnaK* gene were used as probes for genomic southern hybridization. Results shown in Fig.3 indicate that these three genes are derived from different loci of the chromosome. This is the first example of a *dnaK* multigene family in prokaryotes except that a new suppressor gene of *hns-1* mutation in *Escherichia coli* has been found to encode a protein homologous to the HSP70 family[7].

Some microorganisms such as *B. subtilis* [15] and *Clostridium acetobutylicum* [9] have an operon consisting of the gene order of *grpE-dnaK-dnaJ* and the others such as *E. coli* [1] have the order of *dnaK-dnaJ*. On the other hand, a cyanobacterial species *Synechocystis* sp. PCC6803 is somehow unique at the organization of *dnaK* homolog gene which is not accompanied with *grpE* or *dnaJ* homolog gene [2]. It is interesting to know the gene organization of *dnaK* multigene family found in *Synechococcus* sp. PCC7942. At present, we identified a *grpE* gene homolog upstream of *dnaK1* (Fig. 1). The downstream region of *dnaK2* exhibits no similarity to *dnaJ* region even at the nucleotide level. It contains one major open reading frame named *orf2.3* which has no significant homology with known proteins.

ACKNOWLEDGMENTS

This research was supported by a grant-in-aid from the Ministry of Education, Science and Culture of Japan and partially supported by a grant from the 'Biodesign Research Program' from Riken.

REFERENCES

[1] Bardwell, J.C.A., Tilly, K., Craig, E., King, J., Zylicz, M., and Georgopoulos, C. (1986) J. Biol. Chem. 261, 1782-1785.
[2] Chitnis, P.R., and Nelson, N. (1991) J. Biol. Chem. 266, 58-65.
[3] Ellis, R.J. (1987) Nature 328, 378-379.
[4] Ellis, R.J., and van der Vies, S.M. (1991) Annu. Rev. Biochem. 60, 321-347.
[5] Gething, M-J., and Sambrook, J. (1992) Nature 355, 33-45.
[6] Gray, M.W., and Doolittle, W.F. (1982) Microbiol. Rev. 46, 1-42.
[7] Kawula, T.H., and Lelivelt, M.J. (1994) J. Bacteriol. 176, 610-619.
[8] Lindquist, S., and Craig, E.A. (1988) Annu. Rev. Genet. 22, 631-677.
[9] Narberhaus, F., Giebeler, K., and Bahl, H. (1992) J. Bacteriol. 174, 3290-3299.
[10] Ochman, H., Gerber, A.S., and Hartl, D.L. (1988) Genetics 120, 621-623.
[11] Porter, R.D. (1988): Meth. Enzymol. 167, 703-714.
[12] Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
[13] Wang, S., Liu, X-Q. (1991) Proc. Natl. Acad. Sci. USA 88, 10783-10787.
[14] Wetzstein, M., Dedio, J., and Schumann, W. (1990) Nucleic Acids Res. 18, 2172.
[15] Wetzstein, M., Völker, U., Dedio, J., Löbau, S., Zuber, U., Schiesswohl, M., Herget, C., Hecker, M., and Schumann, W.(1992) J. Bacteriol. 174, 3300-3310.