Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 848–852, February 1984
Genetics

Major heat shock gene of *Drosophila* and the *Escherichia coli* heat-inducible *dnaK* gene are homologous

(DNA sequence comparison/gene evolution/archaebacteria)

JAMES C. A. BARDWELL AND ELIZABETH A. CRAIG

Department of Physiological Chemistry, University of Wisconsin-Madison, Madison, WI 53706

Communicated by James F. Crow, October 24, 1983

ABSTRACT **The *Escherichia coli dnaK* gene is homologous to the major heat shock-induced gene in *Drosophila* (*Hsp70*). The primary DNA sequence of the entire protein-coding region of the *dnaK* gene was determined and compared with that of the *Hsp70* gene of *Drosophila*. The two sequences are homologous; the *dnaK* gene could encode a 69,121-Da polypeptide, 48% identical to the hsp70 protein of *Drosophila*. The homology between the *Hsp70* gene of *Drosophila* and the *E. coli dnaK* gene illustrates the remarkable conservation of the heat shock genes in evolution. In contrast to *Drosophila* and *Saccharomyces cerevisiae*, both of which contain multigene families related to the *Hsp70* gene, hybridization analyses indicate that *E. coli* contains only a single *Hsp70*-related gene, *dnaK*. Hybridization between the DNA of an archaebacterium *Methanosarcina barkeri* and the *Hsp70* genes of *Drosophila*, *Saccharomyces*, and *E. coli* has been detected, suggesting the existence of *Hsp70*-related genes in the three "primary kingdoms": eukaryotes, eubacteria, and archaebacteria.**

When cells from a variety of species are quickly heated to a few degrees above their normal growth temperature, the synthesis of a small set of proteins is strongly and rapidly induced. The heat shock response has been most extensively studied in *Drosophila*, but a similar response has been observed in cells of a broad spectrum of eukaryotes: *Saccharomyces cerevisiae* (1), *Dictyostelium* (2), tobacco, hamsters, chickens, and humans (3, 4).

The function of the heat shock proteins is not known. However, the synthesis of heat shock proteins has been correlated with the acquisition of resistance to heat in eukaryotic cells (5). Agents that interfere with oxidative phosphorylation, as well as various chemicals such as ethanol, elicit the heat shock response (5, 6). Thus, the induction is thought to be a general response to stress, not merely a response to an alteration in temperature.

Evidence is accumulating that indicates that these induced proteins have been conserved throughout eukaryotic evolution. In many organisms, heat induces the synthesis of a protein of ≈70,000 Da. Antibodies raised to a 70,000-Da heat-shock protein (hsp70) from chicken crossreact with heat-shock proteins of similar molecular size in eukaryotic species as divergent as yeast and man (7). The predicted amino acid sequences of hsp70 proteins of yeast are 72% identical to the *Drosophila* hsp70 protein (ref. 8; unpublished observations).

A heat shock response has also been observed in a prokaryote, *Escherichia coli*. On temperature up-shift, the rate of synthesis of at least 14 polypeptides is enhanced. This induction is under the control of a gene called htp^R (9). Four of these HTP (*h*igh *t*emperature *p*roduction) proteins have been identified: the *groEL*, *groES*, and *dnaK* gene products, and a lysine tRNA synthetase. The first three proteins are essential for the growth of bacteriophage λ (10, 11, 18). Furthermore, $groEL^-$, $groES^-$, and $dnaK^-$ mutants that are temperature sensitive for bacterial growth at 43°C have been isolated (11, 12), suggesting that these genes are essential for cell viability. In this paper, we report that the *dnaK* gene of *E. coli* is homologous to the *Hsp70* heat shock genes of higher organisms, thus establishing a relationship between the prokaryotic and eukaryotic heat shock systems.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Bacteriophage and Plasmids. λdnaK, a phage capable of tranducing *E. coli* $dnaK^-$ to $dnaK^+$, and λdnaJdnaK phage were obtained from M. Feiss (13). Deletion derivatives of λdnaJdnaK were obtained from H. Uchida (14). Three plasmids containing yeast genes related to the *Drosophila Hsp70* gene were used. *YG100* (8) and *YG102* (15) are 97% identical to each other and 67% identical to *Drosophila Hsp70*. Transcription of these three genes is enhanced by heat shock. *YG101* is 67% related to *YG100* and *YG102*, and 56% identical to *Drosophila Hsp70*. Another plasmid, B8, contains a single copy of a 87C *Drosophila Hsp70* gene (16). In all cases pBR322 is the vector.

General Methods. Gel electrophoresis, blotting of DNA to nitrocellulose, nick-translation, and DNA sequence analysis were carried out as described (16, 17). Labeling of DNA by polynucleotide kinase was as described (16) except that prior to labeling, blunt or recessed 5′ ends of DNA were converted to protruding 5′ ends using DNA polymerase Klenow fragment (New England BioLabs). Five units of enzyme were incubated with 20–30 μg of digested DNA in 50 mM NaCl/7 mM $MgCl_2$/7 mM Tris·HCl, pH 7.4, for 30 min at room temperature (Z. Burton, personal communication).

In calculating the percentage amino acid identity, perfect amino acid matches were counted and divided by the sum of the total number of residues and the number of gapped residues. To calculate percentage nucleotide identity, matching bases were counted and divided by the sum of the total number of base pairs and the number of gaps.

DNA·DNA Hybridizations. Hybridization of DNA labeled by nick-translation to DNA bound to nitrocellulose filters was carried out under nonstringent conditions. The hybridization solution was 30% formamide/0.75 M NaCl/0.075 M sodium citrate/0.1% $NaDodSO_4$/1.0 mM EDTA/10 mM Tris·HCl, pH 7.5/1× Denhart's solution (8). The filters were incubated with probe overnight at 37°C after a 4-hr incubation in hybridization buffer in the absence of probe. The filters were then washed in 0.75 M NaCl/0.075 M sodium citrate/0.2% $NaDodSO_4$ at 37°C for 4 hr with three changes of the wash solution.

RESULTS

Hybridization of Eukaryotic *Hsp70* Genes to the *E. coli dnaK* Gene. To determine whether *E. coli* contains se-

Abbreviation: kb, kilobase(s).

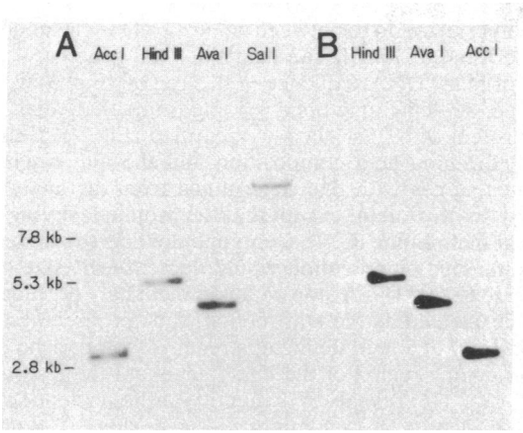

FIG. 1. Hybridization of a yeast heat shock gene and the *dnaK* gene to *E. coli* genomic DNA. Three-microgram portions of *E. coli* DNA were digested with restriction enzymes, electrophoresed on a 0.7% agarose gel, and transferred to nitrocellulose. (*A*) A fragment of a yeast *Hsp70* gene, *YG102*, encompassing the codons of amino acids 118–639 was labeled with ^{32}P by nick-translation and hybridized to the filter under low stringency. (*B*) A fragment of the *dnaK* gene encompassing the codons of amino acids 174–412 was labeled with ^{32}P by nick-translation and hybridized to the filter under conditions of low stringency.

quences related to eukaryotic 70,000-Da heat shock genes, hybridization experiments were carried out. *E. coli* genomic DNA was digested with restriction enzymes, subjected to gel electrophoresis, and transferred to nitrocellulose. The DNA was hybridized to a protein-coding portion of the yeast gene, *YG102*, which is related to the *Drosophila* 70,000-Da heat-shock gene. As shown in Fig. 1*A*, a single band of hybridization was observed after independent digestion with four different restriction enzymes. DNA of λ transducing phage containing the *dnaK* and *dnaJ* genes was fixed to nitrocellulose and probed with a portion of the protein-coding region of an *Hsp70*-related gene from yeast (*YG101*). *dnaK* was tested because it is a major heat shock gene in *E. coli* (9). An intense hybridization signal was observed, indicating probable similarity between the *dnaJ dnaK* region and the yeast gene. To map the region of hybridization on these phage, a yeast heat shock gene probe was hybridized to a series of λdnaJdnaK deletion derivatives. The *dnaK* gene had been previously mapped between the left end of the *Hin*dIII insert

```
          10         20         30         40         50
1    ATGGGTAAAA TAATTGGTAT CGACCTGGGT ACTACCAACT CTTGTGTAGC
51   GATTATGGAT GGCACCACTC CTCGCGTGCT GGAGAACGCC GAAGGCGATC
101  GCACCACGCC TTCTATCATT GCCTATACCC AGGATGGTGA AACTCTAGTT
151  GGTCAGCCGG CTAAACGTCA GGCAGTGACG AACCCGCAAA ACACTCTGTT
201  TGCGATTAAA CGCCTGATTG GTCGCCGCTT CCAGGACGAA GAAGTACAGC

251  GTGATGTTTC CATCATGCCG TTCAAAATTA TTGCTGCTGA TAACGGCGAC
301  GCATGGGTCG AAGTTAAAGG CCAGAAAATG GCACCGCCGC AGATTTCTGC
351  TGAAGTGCTG AAAAAAATGA AGAAAACCGC TGAAGATTAC CTGGGTGAAC
401  CGGTAACTGA AGCTGTTATC ACCGTACCGG CATACTTTAA CGATGCTCAG
451  CGTCAGGCAA CCAAAGACGC AGGCCGTATC GCTGGTCTGG AAGTAAAACG

501  TATCATCAAC GAACCGACCG CAGCTGCGCT GGCTTACGGT CTGGACAAAG
551  GCACTGGCAA CCGTACTATC GCGGTTTATG ACCTGGGTGG TGGTACTTTC
601  GATATTTCTA TTATCGAAAT CGACGAAGTT GACGGCGAAA AAACCTTCGA
651  AGTTCTGGCA ACCAACGGTG ATACCCACCT GGGGGGTGAA GACTTCGACA
701  GCCGTCTGAT CAACTATCTG GTTGAAGAAT TCAAGAAAGA TCAGGGCATT

751  GACCTGCGCA ACGATCCGCT GGCAATGCAG CGCCTGAAAG AAGCGGCAGA
801  AAAAGCGAAA ATCGAACTGT CTTCCGCTCA GCAGACCGAC GTTAACCTGC
851  CATACATCAC TGCAGACGCG ACCGGTCCGA AACACATGAA CATCAAAGTG
901  ACTCGTGCGA AACTGGAAAG CCTGGTTGAA GATCTGGTAA ACCGTTCCAT
951  TGAGCCGCTG AAAGTTGCAC TGCAGGACGC TGGCCTGTCC GTATCTGATA

1001 TCGACGACGT TATCCTCGTT GGTGGTCAGA CTCGTATGCC AATGGTTCAG
1051 AAGAAAGTTG CTGAGTTCTT TGGTAAAGAG CCGCGTAAAG ACGTTAACCC
1101 GCACGAAGCT GTAGCAATCG GTGCTGCTGT TCAGGGTGGT GTTCTGACTG
1151 GTGACGTAAA AGACGTACTG CTGCTGGACG TTACCCCGCT GTCTCTGGGT
1201 ATCGAAACCA TGGGCGGTGT GATGACGACG CTGATCGCGA AAAACACCAC

1251 TATCCCGACC AAGCACAGCC AGGTGTTCTC TACCGCTGAA GACAACCAGT
1301 CTGCGGTAAC CATCCATGTG CTGCAGGGTG AACGTAAACG TGCGGCTGAT
1351 AACAAATCTC TGGGTCAGTT CAACCTAGAT GGTATCAACC CGGCACCGCG
1401 CGGCATGCCG CAGATCGAAG TTACCTTCGA TATCGATGCT GACGGTATCC
1451 TGCACGTTTC CGCGAAAGAT AAAAACAGCG GTAAAGAGCA GAAGATCACC

1501 ATCAAGGCTT CTTCTGGTCT GAACGAAGAT GAAATCCAGA AAATGGTACG
1551 CGACGCAGAA GCTAACGCCG AAGCTGACCG TAAGTTTGAA GAGCTGGTAC
1601 AGACTCGCAA CCAGGGCGAC CATCTGCTGC ACAGCACCCG TAAGCAGGTT
1651 GAAGAAGCAG GCGACAAACT GCCGGCTGAC GACAAAACTG CTATCGAGTC
1701 TGCGCTGACT GCACTGGAAA CTGCTCTGAA AGGTGAAGAC AAAGCCGCTA

1751 TCGAAGCGAA AATGCAGGAA CTGGCACAGG TTTCCCAGAA ACTGATGGAA
1801 ATCGCCCAGC AGCAACATGC CCAGCAGCAG ACTGCCGGTG CTGATGCTTC
1851 TGCAAACAAC GCGAAAGATG ACGATGTTGT CGACGCTGAA TTTGAAGAAG
1901 TCAAAGACAA AAAATAA
```

FIG. 3. Nucleotide sequence of the protein-coding portion of the *dnaK* gene. All of the sequence was determined on both strands of the DNA with the exception of a 70-base pair stretch from nucleotide 370–440, which was determined twice on one strand. Two independent clones of *dnaK* were sequenced over the entire length to guard against errors due to cloning artifacts. We have sequenced across all restriction sites, so we are certain no gaps exist in our sequence.

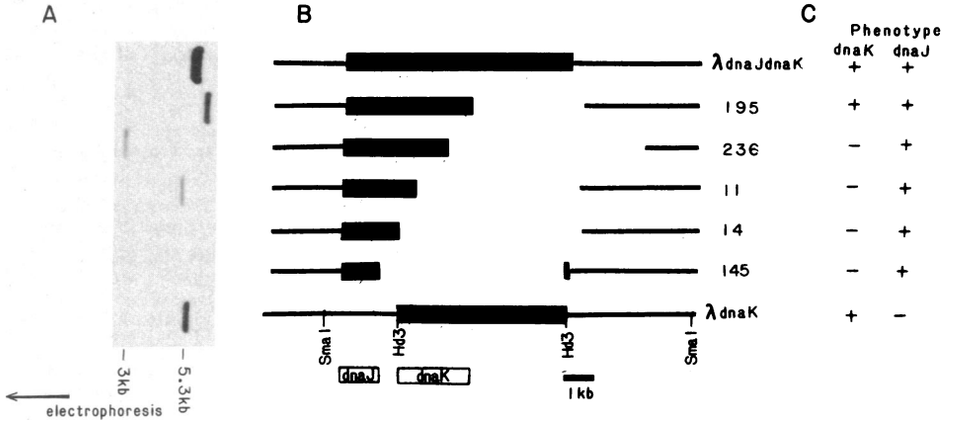

FIG. 2. Mapping of the hybridization between *E. coli* and eukaryotic heat shock genes to the *dnaK* gene. (*A*) DNA from λdnaK and deletion derivatives derived from λdnaJdnaK were cleaved with *Sma* I and *Hin*dIII (Hd3), electrophoresed through 0.7% agarose, and transferred to nitrocellulose. A fragment of the yeast heat shock gene *YG102* encompassing the codons of amino acids 118–639 was hybridized to the filter. (*B*) The physical structures of the λ deletions shown were determined by Saito (14) using heteroduplex analysis. *E. coli* DNA is represented by the thick lines. Thin lines represent flanking phage vector. (*C*) The dnaK and dnaJ phenotypes of these phages were determined by Saito (14) by testing the ability of each of these deletion mutants to plaque on bacterial strains containing mutations in the *dnaK* and *dnaJ* genes. The limits of *dnaJ* and *dnaK* designated at the bottom of the figure are based on such tests of these deletion derivatives.

in λ*dnaK* and the end point of the deletion in λdnaJ-dnaKΔ195 (ref. 14; Fig. 2*B*) by testing the ability of the deletions to complement *dnaK*$^-$ mutations. DNA from these phages were cleaved with restriction enzymes that excise the bacterial DNA from phage DNA and also separate the *dnaK* gene from the *dnaJ* gene. When the DNA was hybridized to a yeast heat shock gene probe, a band of hybridization was observed only with those deletion derivatives that contained DNA from the *dnaK* gene region (Fig. 2*A*). The only fragments that hybridized were those containing the *dnaK* sequences. A labeled plasmid (B8) containing a copy of the *Drosophila Hsp70* gene was also hybridized to a similar blot of the λdnaJdnaK deletions, and it showed a similar pattern of hybridization (results not shown).

DNA Sequence Analysis of *dnaK* and Comparison to *Hsp70* Gene. The 5.3-kilobase (kb) *Hin*dIII fragment containing sequences that hybridized to the *Drosophila* and yeast *Hsp70* genes was subcloned into pBR322. Restriction enzyme mapping and hybridization analyses localized the region of *Hsp70* hybridization to between an *Nru* I site in the middle of the insert and a *Hin*dIII site near the *dnaJ* gene.

The primary DNA sequence of this region (Fig. 3) was determined. A single long open-reading frame that could code for a protein of 69,121 Da was identified (Fig. 4). Both the predicted amino acid composition and the amino-terminal sequence agrees with that determined from the analysis of purified dnaK protein, except that the protein lacks an NH_2-terminal methionine (C. P. Georgopoulos, personal communication). This modification would decrease the size of the protein to 68,990 Da. A stop codon immediately precedes the AUG designated as the start codon, thus prohibiting initiation of *dnaK* at a methionine upstream. A comparison of the protein-coding region of the *dnaK* gene and the *Drosophila Hsp70* gene (16) reveals 57% identity at the nucleotide level and 48% identity at the amino acid level (Fig. 4). Some regions of the polypeptide show a greater similarity. Between amino acid 138 and amino acid 183, >90% of the amino acids are the same, and between amino acid 365 and 489, 64% are

```
E. coli dnaK  Met Gly Lys Ile Ile Gly Ile Asp Leu Gly Thr Thr Asn Ser Cys Val Ala Ile Met Asp Gly Thr Thr Pro Arg Val Leu Glu Asn Ala aa30
Dr. hsp70      *  Pro  Δ  Ala  *  *  *  *  *  *  *  *  Tyr  *  *  *  Gly Val Tyr Gln His Gly Lys Val Glu Ile Asn Ala Tyr Asp

E. coli dnaK  Glu Gly Asp Arg Thr Thr Pro Ser Ile Ile Ala Tyr Thr Gln Asp Gly Glu Thr Leu Val Gly Gln Pro Ala Lys Arg Gln Ala Val Thr aa60
Dr. hsp70      Gln  *  Asn  *  *  *  *  *  Tyr Val  *  Phe  *  Δ  *  Ser  *  Arg  *  Asn  *  Glu  *  *  *  Asn  *  Val Ala Met

E. coli dnaK  Asn Pro Gln Asn Thr Leu Phe Ala Ile Lys Arg Leu Ile Gly Arg Arg Phe Gln Asp Glu Glu Val Gln Arg Asp Val Ser Ile Met Pro aa90
Dr. hsp 70     *  *  Arg  *  *  Val  *  Asp Ala  *  *  *  *  *  *  Lys Tyr Asp  *  Pro Lys Ile Ala Glu  *  Met Lys His Trp  *

E. coli dnaK  Phe Lys Ile Ile Ala Ala Asp Asn Gly Asp Ala Trp  Δ  Val Glu Val Lys Gly Gln  Δ  Lys Met Ala Pro Pro  Δ  Gln Ile Ser Ala aa117
Dr. hsp70      *  *  Val Val Ser Asp Gly Gly Lys Pro Lys Ile Gly  *  *  Tyr  *  *  Glu Ser  *  Arg Phe Ala  *  Glu Glu  *  *  Ser

E. coli dnaK  Glu Val Leu Lys Lys Met Lys Lys Thr Ala Glu Asp Tyr Leu Gly Glu Pro Val Thr Glu Ala Val Ile Thr Val Pro Ala Tyr Phe Asn aa147
Dr. hsp70      Met  *  *  Thr  *  *  *  Glu  *  *  *  Ala  *  *  *  *  Ser Ile  *  Asp  *  *  *  *  *  *  *  *  *  *

E. coli dnaK  Asp Ala Gln Arg Gln Ala Thr Lys Asp Ala Gly Arg Ile Ala Gly Leu Glu Val Lys Arg Ile Ile Asn Glu Pro Thr Ala Ala Ala Leu aa177
Dr. hsp70      *  Ser  *  *  *  *  *  *  *  *  *  His  *  *  *  *  Asn  *  Leu  *  *  *  *  *  *  *  *  *  *  *

E. coli dnaK  Ala Tyr Gly Leu Asp Lys  Δ  Gly Thr Gly Asn Arg Thr Ile Ala Val Tyr Asp Leu Gly Gly Gly Thr Phe Asp Ile Ser Ile Ile Glu aa206
Dr. hsp70      *  *  *  *  *  *  Asn Leu Lys  *  Glu  *  Asn Val Leu Ile Phe  *  *  *  *  *  *  *  *  Val  *  *  Leu Thr

E. coli dnaK  Ile Asp Glu Val Asp Gly Glu Lys Thr Phe Glu Val Leu Ala Thr Asn Gly Asp Thr His Leu Gly Gly Glu Asp Phe Asp Ser Arg Leu aa236
Dr. hsp70      *  *  *  Gly  Δ  Δ  Δ  Ser Leu  *  *  *  Arg Ser  *  Ala  *  *  *  *  *  *  *  *  *  *  *  Asn  *  *

E. coli dnaK  Ile Asn Tyr Leu Val Glu Glu Phe Lys Lys Asp Gln Gly Ile Asp Leu Arg Asn Asp Pro Leu Ala Met Gln Arg Leu Lys Glu Ala Ala aa266
Dr. hsp70      Val Thr His  *  Ala  *  *  *  *  Arg Lys Tyr Lys Lys  *  *  *  Ser Asn  *  Arg  *  Leu Arg  *  *  Arg Thr  *  *

E. coli dnaK  Glu Lys Ala Lys Ile Glu Leu Ser Ser Ala Gln Gln Thr Asp Val Asn Leu Pro Tyr Ile Thr Ala Asp Ala Thr Gly Pro Lys His Met aa296
Dr. hsp70      *  Arg  *  *  Arg Thr  *  *  *  Ser Thr Glu Ala Thr Ile Glu  Δ  Δ  Δ  *  Asp  *  Leu Phe Glu  *  Δ  Gln Asp Phe

E. coli dnaK  Asn Ile Lys Val Thr Arg Ala Lys Leu Glu Ser Leu Val Glu Asp Leu Val Asn Arg Ser Ile Glu Pro Leu Lys Val Ala Leu Gln Asp aa326
Dr. hsp70      Tyr Thr  *  *  Ser  *  *  Arg Phe  *  Glu  *  Cys Ala Asn  *  Phe Arg Asn Thr Leu Gln  *  Val Glu Lys  *  *  Asn  *

E. coli dnaK  Ala Gly Leu Ser Val Ser Asp Ile Asp Asp Val Ile Leu Val Gly Gly Gln Thr Arg Met Pro Met Val Gln Lys Lys Val Ala Glu Phe aa356
Dr. hsp70      *  Lys Met Asp Lys Gly Gln  *  His  *  Ile Val  *  *  *  *  Ser  *  *  Ile  *  Lys  *  *  Ser Leu Leu Gln  *  *

E. coli dnaK  Phe  Δ  Gly Lys Glu Pro Arg Lys Asp Val Asn Pro Asp Glu Ala Val Ala Ile Gly Ala Ala Val Gln Gly Gly Val Leu Thr Gly Asp aa385
Dr. hsp70      *  His  *  *  Asn Leu Asn Leu Ser Ile  *  *  *  *  *  *  *  *  Tyr  *  *  *  *  *  Ala Ala Ile  *  Ser  *  *

E. coli dnaK   Δ  Δ  Δ  Δ  Val Lys Asp Val Leu Leu Leu Asp Val Thr Pro Leu Ser Leu Gly Ile Glu Thr Met Gly Gly Val Met Thr Thr Leu aa411
Dr. hsp70     Gln Ser Gly Lys Ile Gln  *  *  *  *  Val  *  *  Ala  *  *  *  *  *  *  *  *  *  Ala  *  *  *  *  *  Lys  *

E. coli dnaK  Ile Ala Lys Asn Thr Thr Ile Pro Thr Lys His Ser Gln Val Phe Ser Thr Ala Glu Asp Asn Gln Ser Ala Val Thr Ile His Val Leu aa441
Dr. hsp70      *  Glu Arg  *  Cys Arg  *  *  Cys  *  Gln Thr Lys Thr  *  *  *  Tyr Ser  *  *  *  Pro Gly  *  Ser  *  Gln  *  Tyr

E. coli dnaK  Gln Gly Glu Arg Lys Arg Ala Ala Asp Asn Lys Ser Leu Gly Gln Phe Asn Leu Asp Gly Ile Asn Pro Ala Pro Arg Gly Met Pro Gln aa471
Dr. hsp70      Glu  *  *  *  Ala Met Thr Lys  *  *  Asn Ala  *  *  Thr  *  Asp  *  Ser  *  *  Pro  *  *  *  *  *  Val  *  *

E. coli dnaK  Ile Glu Val Thr Phe Asp Ile Asp Ala Asp Gly Ile Leu His Val Ser Ala Lys Asp Lys Asn Ser Gly Lys Glu Gln Lys Ile Thr Ile aa501
Dr. hsp70      *  *  *  *  *  *  Leu  *  *  Asn  *  *  *  Asn  *  *  *  *  Glu Met Ser Thr  *  *  Ala Lys Asn  *  *  *

E. coli dnaK  Lys Ala Ser Ser Gly  Δ  Leu Asn Glu Asp Glu Ile Gln Lys Met Val Arg Asp Ala Glu Ala Asn Ala Glu Ala Asp Arg Lys Phe Glu aa530
Dr. hsp70      *  Asn Asp Lys  *  Arg  *  Ser Gln Ala  *  *  Asp Arg  *  *  Asn Glu  *  *  Lys Tyr  *  Asp Glu  *  Glu  *  His Arg

E. coli dnaK  Glu Leu Val Gln Thr Arg Asn Gln Gly Asp His Leu Leu His Ser Thr Arg Lys Gln Val Glu Glu Ala  Δ  Gly Asp Lys Leu Pro Ala aa559
Dr. hsp70      Gln Arg Ile Thr Ser  *  *  Ala Leu Glu Ser Tyr Val Phe Asn Val Lys Gln Ser  *  *  Gln  *  Pro Ala Gly  *  *  Asp Glu

E. coli dnaK  Asp Asp Lys Thr Ala Ile Glu Ser Ala Leu Thr Ala Leu Glu Thr Ala Leu Lys Gly Asp Lys Ala Ala Ile Glu  Δ  Δ  Δ  Ala aa586
Dr. hsp70      Ala  *  *  Asn Ser Val Leu Asp Lys Cys Asn Glu Thr Ile Arg Trp  *  Asp Ser Asn Thr Thr  *  Glu Lys  *  Glu Phe Asp His

E. coli dnaK  Lys Met Gln Glu Leu Ala Gln Val Ser Gln Lys Leu Met Glu Ile Ala Gln Gln Gln His Ala Gln Gln Gln Thr Ala  Δ  Gly Ala  Δ  aa614
Dr. hsp70      *  *  Glu  *  *  Δ  Δ  Δ  Thr Arg His Cys Ser Pro  *  Met Thr Lys Met  *  Gln  *  Gly Ala Gly  *  Ala  *  Gly Pro

E. coli dnaK   Δ  Asp Ala Ser Ala Asn Asn Ala Lys Asp Asp Asp Val Val Asp Ala Glu  Δ  Δ  Δ  Phe Glu Glu Val Lys Asp Lys Lys END
Dr. hsp70     Gly Ala Asn Cys Gly Gln Gln  *  Gly Gly  Δ  Δ  Phe Gly Gly Tyr Ser Gly Pro Thr Val  *  *  *  Asp  Δ  Δ  Δ  *
```

FIG. 4. Comparison of the predicted amino acid sequence of *E. coli* dnaK protein and *Drosophila* (Dr.) hsp70. Those amino acids identical in both sequences are shown by an asterisk. The first amino acid (aa) of the dnaK protein is designated 1. Those amino acids deleted in one sequence relative to the other are shown by a triangle.

the same. If one divided the dnaK protein into five equal segments, the second and fourth are more conserved (66% and 58%) than the first, third, and fifth (46, 42, and 23% respectively). *dnaK* is also homologous to the yeast *Hsp70*-related genes used as hybridization probes. At the nucleotide level, *dnaK* is 57% identical to two yeast genes, *YG101* and *YG102*. The amino acid residues conserved from *E. coli* to *Drosophila* are nearly all conserved to yeast as well. These regions of high identity may represent functionally conserved regions of the protein.

Hybridization of Isolated *dnaK* to *E. coli* Genomic DNA. Since it has been shown that the *S. cerevisiae* and *Drosophila melanogaster* genomes contain a family of *Hsp70*-related genes (8, 19), we wanted to determine whether *E. coli* also contained multiple *Hsp70*-related genes. The *dnaK* gene was hybridized under nonstringent conditions to fractionated *E. coli* genomic DNA. A single band was observed after hybridization to genomic DNA digested independently with three different restriction enzymes (Fig. 1*B*). Under similar hybridization conditions a yeast *Hsp70* fragment hybridized to yeast genomic DNA reveals 8–10 *Hsp70* homologous genes (8). The detection of only one band indicates that the *E. coli* genome contains a single *Hsp70* homologous sequence—*dnaK*.

Hybridization of 70,000-Da Heat-Shock Genes to an Archaebacterium, *Methanosarcina barkeri*. In an attempt to further establish the conservation of *Hsp70* genes in evolution, *M. barkeri* genomic DNA was digested with restriction enzymes, blotted to nitrocellulose, and hybridized to protein-coding portions of *Hsp70*-related genes. Three probes, one from *E. coli dnaK*, one from yeast, and one from *Drosophila Hsp70* were used in separate hybridizations. The same single band of hybridization was observed using either yeast and *Drosophila* or *E. coli* and *Drosophila* DNA as probe (Fig. 5). We suggest that *M. barkeri* contains *Hsp70*-related sequences.

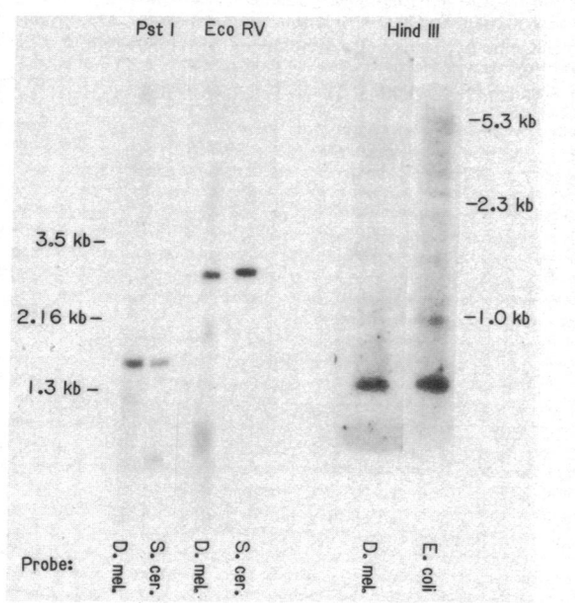

FIG. 5. Hybridization of *Hsp70* genes to DNA of an archaebacterium. *M. barkeri* genomic DNA was digested with restriction enzymes, electrophoresed, and transferred to nitrocellulose. Three protein-coding fragments of *Hsp70*-related genes were used as hybridization probes: a fragment of the *E. coli dnaK* gene encompassing amino acids 23–325, a fragment of the yeast (S. cer.) *YG102* gene encompassing amino acids 118–639, and a fragment of the *Drosophila* (D. mel.) *Hsp70* gene encompassing amino acids 18–337.

DISCUSSION

Because 70,000-Da heat shock-inducible genes exist in two distantly related species, *D. melanogaster* and *S. cerevisiae*, we wondered whether *E. coli* contained *Hsp70*-related sequences. A heat-inducible gene in *E. coli*, *dnaK*, is 57% identical to the major heat shock gene of *Drosophila* (*Hsp70*). This identity extends over the entire protein-coding region with some regions showing extreme conservation. A comparison of the predicted amino acid sequences of the *dnaK* and *Drosophila Hsp70* genes showed 48% overall identity. This identity increased to >90% in one 46-amino acid segment. The *dnaK* gene is also homologous to a yeast *Hsp70* gene. The sequence of the yeast gene is 57% identical to *dnaK* and 67% identical to the *Drosophila Hsp70*. Given the high degree of similarity observed between *dnaK* and other 70,000-Da heat shock genes, we assert that these genes have diverged from a common ancestor rather than converged from unrelated sequences. These results are a remarkable demonstration of prokaryotic–eukaryotic sequence homology.

Based on comparison of 5S RNA sequences, Hori and Osawa (20) have estimated that the prokaryotes and eukaryotes diverged about 1.8×10^9 years ago. Only a very few proteins have maintained recognizable sequence similarity between prokaryotic and eukaryotic species (21, 22). *c*-type cytochromes, serine proteases, and the β-subunit of the ATP synthetase complex are among those that have been well conserved. Cytochrome *c* shows 10–48% similarity between eukaryotic and prokaryotic species. Bacterial trypsin from *Streptomyces griseus* is 26–31% identical to eukaryotic serine proteases. One very highly conserved protein is the β-subunit of ATP synthetase (23). Sequences from maize chloroplasts and beef mitochondria are 64% and 69% identical to the *E. coli* β-subunit. The homology between the dnaK protein and *Drosophila* hsp70 is thus comparable with the homology seen between some of the most highly conserved proteins known.

The archaebacteria composed of the methanogens, extreme halophiles, and acidophiles form a unique biological grouping phylogenetically distant from typical bacteria (eubacteria) and eukaryotes. Three primary kingdoms have been proposed (24)—the eukaryotes, the eubacteria, and the archaebacteria. The archaebacteria differ significantly from eubacteria and eukaryotes. For example, all eubacteria contain muramic acid in their cell walls while archaebacteria do not. Membranes of archaebacteria contain unusual lipids not found in members of the two other groups and tRNA base modifications commonly found in eukaryotic and prokaryotic tRNA are absent in archaebacteria. The two bacterial lines of descent appear to be no more related to one another than either of them is to the eukaryotes (24). The detection of *Hsp70*-related sequences in an archaebacteria thus dramatically shows the extreme conservation of the *Hsp70* gene throughout evolution.

Both the *Drosophila* and yeast genomes contain a family of *Hsp70*-related genes (8, 19). In both cases, some members are heat shock inducible while others are transcribed under normal growth conditions and not inducible by heat treatment. Evidence presented here indicates that there is only one *Hsp70*-related gene, *dnaK*, in the *E. coli* genome. However, under normal growth conditions at 37°C the dnaK protein accounts for 1.4% of the weight fraction of cellular protein and is the seventh most abundant protein in the cell (25). Transcripts of one of the *Drosophila* Hsp70-related proteins (Hsc4) are very abundant in cells in all stages of development, comparable in abundance to some actin mRNAs (26). Therefore, in both eukaryotes and prokaryotes Hsp70-related proteins appear to be abundant under normal physiological conditions.

Although the function of the heat shock proteins is not known, the information from eukaryotic species including the abundance of these proteins suggests structural roles for the heat shock proteins either in the nucleo- or cytoskeleton. Studies of avian and mammalian heat shock proteins have shown that hsp70 is identical to a highly conserved polypeptide previously shown to copurify with intermediate filaments and microtubules (3). Antibodies to chicken hsp70 stains a pattern of stress fibers in the cytoplasm (27). Unfortunately, the information available concerning the dnaK protein does not obviously suggest its function in cellular metabolism. The dnaK protein is required for the replication of the phages λ and P22 but not T4, T7, φX174, or fd (28). Mutant λ phage able to grow on a *dnaK*$^-$ strain contain an altered *P* gene product (10). The *P* gene product is thought to interact with the λ *O* gene product and the host *dnaB* gene product to form a replisome structure at λ *ori* (29). The temperature-sensitive phenotype of some *dnaK* mutants is strong evidence that the *dnaK* gene is required for the normal growth of *E. coli*. It has been suggested that *dnaK* is involved in host DNA replication (30). However, the existence of mutants that prevent λ DNA replication but do not affect host growth (30) suggests that the functions involved in host and phage metabolism may be different. Further genetic and biochemical analysis of *dnaK* and its protein product should help elucidate the role of *hsp70* in both normal and stress situations.

We thank C. Gross and J. G. Zeikus for *E. coli* and *M. barkeri* DNA, respectively and H. Saito and M. Feiss for providing phages used in this study. We thank L. Stinson for excellent technical assistance in performing some of the experiments. This work was supported by a National Institutes of Health grant (E.A.C.). E.A.C. was supported by a U.S. Public Health Service research career development award; J.C.A.B. was supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada.

1. Miller, M. J., Xuong, N. & Geiduschek, E. P. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5222–5225.
2. Loomis, W. T. & Wheeler, S. (1980) *Dev. Biol.* **79,** 399–408.
3. Wang, C., Gomer, R. H. & Lazarides, E. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3531–3535.
4. Key, J. L., Lin, C. Y. & Chen, Y. M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3526–3530.
5. Li, G. C. & Werb, Z. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3218–3222.
6. Ashburner, M. & Bonner, J. J. (1979) *Cell* **17,** 241–254.
7. Kelley, P. M. & Schlesinger, M. J. (1982) *Mol. Cell. Biol.* **2,** 267–274.
8. Ingolia, T. D., Slater, M. R. & Craig, E. A. (1982) *Mol. Cell. Biol.* **2,** 1388–1398.
9. Neidhart, F. C. & Van Bogelen, R. (1981) *Biochem. Biophys. Res. Commun.* **100,** 894–900.
10. Georgopoulos, C. P. (1977) *Mol. Gen. Genet.* **151,** 35–39.
11. Tilly, K., Murialdo, H. & Georgopoulous, C. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1629–1633.
12. Itikawa, H. & Ryu, J. (1979) *J. Bacteriol.* **138,** 339–344.
13. Yochem, J., Uchida, H., Sunshine, M., Saito, H., Georgopoulos, C. P. & Feiss, M. (1978) *Mol. Gen. Genet.* **164,** 9–14.
14. Saito, H. & Uchida, H. (1978) *Mol. Gen. Genet.* 164, 1–8.
15. Craig, E. A., Ingolia, T. D., Slater, M., Manseau, L. & Bardwell, J. (1982) in *Heat Shock from Bacteria to Man*, eds. Schlesinger, M. J., Ashburner, M. & Tissieres, A. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 11–18.
16. Ingolia, T. D., Craig, E. A. & McCarthy, B. J. (1980) *Cell* **21,** 669–679.
17. Craig, E. A., McCarthy, B. J. & Wadsworth, S. C. (1979) *Cell* **16,** 575–588.
18. Georgopoulos, C. P. & Hohn B. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 131–135.
19. Ingolia, T. D. & Craig, E. A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 525–529.
20. Hori, H. & Osawa, S. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 381–385.
21. Doolittle, R. F. (1981) *Science* **214,** 149–159.
22. Dayhoff, M. O., ed. (1979) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, MD).
23. Runswick, M. J. & Walker, J. E. (1983) *J. Biol. Chem.* **257,** 3081–3089.
24. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5088–5090.
25. Herendeen, S. L., Van Bogelen, R. A. & Neidhardt, F. C. (1979) *J. Bacteriol.* **139,** 185–194.
26. Craig, E. A., Ingolia, T. D. & Manseau, L. (1983) *Dev. Biol.* **99,** 418–426.
27. Schlesinger, M. J., Kelley, P. M., Aliperti, G. & Malfer, C. (1982) in *Heat Shock from Bacteria to Man*, eds. Schlesinger, M. J., Ashburner, M. & Tissieres, A. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 243–250.
28. Wickner, S. H. (1978) *Annu. Rev. Biochem.* **47,** 1163–1191.
29. Klein, A., Lanka, E. & Schuster, H. (1980) *Eur. J. Biochem.* **105,** 1–6.
30. Saito, H. & Uchida, H. (1977) *J. Mol. Biol.* **113,** 1–25.