Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 5177–5181, August 1987
Biochemistry

Eukaryotic M_r 83,000 heat shock protein has a homologue in *Escherichia coli*

(gene evolution/human/yeast/*Drosophila*/protein sequence comparison)

JAMES C. A. BARDWELL* AND ELIZABETH A. CRAIG†

Department of Physiological Chemistry, University of Wisconsin Medical School, Madison, WI 53706

Communicated by Boris Magasanik, April 13, 1987

ABSTRACT **We have isolated a gene from *Escherichia coli* homologous to the gene encoding the M_r 83,000 *Drosophila* heat shock protein (hsp83). In *E. coli* the protein homologous to hsp83 is a heat shock protein called C62.5. The predicted amino acid sequence of C62.5 is 41% and 42% identical to the *Drosophila* and human hsp83 proteins, respectively. Selected regions of the protein have conservation as high as 90%. The gene encoding C62.5 (named *htpG*) is located between the *dnaZ* and *adk* genes at 11.1 minutes on the *E. coli* chromosome. The *htpG* gene appears to be a newly identified locus. The isolation of an *E. coli* homologue of hsp83 illustrates the remarkable conservation of heat shock proteins in evolution and will facilitate genetic and biochemical experiments aimed at determining the function of hsp83.**

When an organism is subjected to a sudden temperature increase of a few degrees, the synthesis of a small number of proteins is rapidly and strongly induced (1). This phenomenon is called the heat shock response. This response was first discovered in *Drosophila*, but a similar response has now been reported in an extraordinary variety of organisms, including the prokaryote *Escherichia coli* (2, 3). The universality of the heat shock response and the high degree of conservation of the induced proteins within eukaryotes suggest that these proteins are involved in important processes within the cell. Synthesis of the heat shock proteins in many organisms is also induced by viral infection, ethanol, azide, arsenite, and certain other toxic compounds; therefore, although the induction is more specific than a response to stress, it is also more general than just a response to heat (4). Although these proteins may provide protection from the toxic effects of these treatments, the protective mechanism is unknown.

We have been studying the evolutionary conservation of the heat shock proteins. We have shown (5) that the *dnaK* gene of *E. coli* is homologous to the *Drosophila* gene encoding a M_r 70,000 heat shock protein (hsp70). In this paper, we establish an additional similarity between the eukaryotic and prokaryotic heat shock systems. We have isolated a gene from *E. coli* called *htpG* that is homologous to the M_r 83,000 heat shock protein gene (*Hsp83*) of eukaryotes.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. The plasmid p82R, containing the *Drosophila Hsp83* gene cloned into pUC13, was obtained from R. Blackman (Harvard University). The plasmid pLC30-3 from the Clarke–Carbon library, known to complement mutations in the *dnaZX* gene, was obtained from F. Neidhardt (University of Michigan) (6). Four bacterial strains were used. Both χ2844 [F⁻ *tsx-462*::Tn*10*, λ⁻] and N43 [F⁻ *acrA1*, *ara-14*, Δ*lac-85*, *galK2 rpsL197*, *malA1 xyl-5*, *mtl-1*] were obtained from the *E. coli* Genetic Stock Center; CAG552 [zbc21::Tn*10* (95% linked to *purE*) *cycA2*, *cycA1*, *ilv277*, *metB5 his53*, *lacY29*, *pdxC3*, *proC24*, *pyrF30*, *rpsL97*, *tonA32*, *tsx63*, *zyl14*] was obtained from C. Gross (University of Wisconsin at Madison); and JC7623, which is AB1157 *recB21*, *recC22 sbcB15*, *sbcC201* (7), was obtained from M. Nomura (University of California at Irvine). The strains containing Tn*10* insertions were used in cotransduction experiments, the *acr*A mutant strain was used for complementation tests, and the *recB*⁻, *recC*⁻, *sbcB*⁻ strain was used for linear transformations.

General Methods. Gel electrophoresis, nick-translation, and Maxam–Gilbert DNA sequencing were done as described (5). Southern hybridization was carried out under nonstringent conditions [30% (vol/vol) formamide, 0.75 M NaCl at 37°C] as described (5). The filters were washed in 0.75 M NaCl/0.075 M sodium citrate/0.2% $NaDodSO_4$ at 37°C for 4 hr and then 50°C for 1 hr with changes of the wash solution every hour. Radioactive labeling of bacterial heat shock proteins with [^{3}H]leucine and of plasmid proteins in minicells with [^{35}S]methionine were as described in ref. 8. Plasmid-encoded proteins were identified by comigration with heat shock proteins on two-dimensional O'Farrell gels (9). An extract containing ^{35}S-labeled plasmid-produced protein was mixed with a ^{3}H-labeled extract from a heat-shocked *E. coli* culture. The mixed extract was resolved on a two-dimensional gel and analyzed by differential autoradiography as described by Neidhardt *et al.* (6).

Cloning of the *htpG* Gene. When *E. coli* genomic DNA was transferred to nitrocellulose and hybridized at low stringency with *Drosophila* Hsp83 probes, distinct bands of hybridization were seen. We cloned the hybridizing sequences into pBR322 by the following procedure. *E. coli* genomic DNA was digested with *Eco*RI and size-fractionated by agarose gel electrophoresis, and a 6-kilobase (kb) hybridizing fraction was then electroeluted from the agarose. After ligation with *Eco*RI-digested pBR322 and transformation into *E. coli*, the resulting partial library was screened by probing Southern blots of digests of mini-plasmid preparations with a *Drosophila* Hsp83 probe.

Mapping. To map the *htpG* gene on the *E. coli* chromosome, a kanamycin resistance cassette from pUC-4K (10) was inserted into the *Sal* I site in cloned *E. coli* DNA flanking the *htpG* gene. This kanamycin resistance marker was introduced into the chromosome by linear transformation (11) and mapped to within a 10-minute interval with a set of

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.

Abbreviations: hsp70, M_r 70,000 heat shock protein; hsp83, M_r 83,000 heat shock protein; *Hsp83*, gene for M_r 83,000 heat shock protein.

*Present address: Laboratory of Molecular Oncology, National Cancer Institute, Frederick Cancer Research Facility, Frederick, MD 21701-1013.

†To whom reprint requests should be addressed.

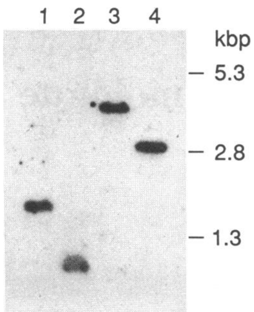

FIG. 1. Southern hybridization of the *Drosophila* heat shock gene *Hsp83* to *E. coli* genomic DNA. A fragment of the *Drosophila Hsp83* gene encompassing the codons for amino acids 17–319 was labeled with ^{32}P by nick-translation and hybridized at low stringency to *E. coli* genomic DNA blotted to nitrocellulose. Lanes: 1, *E. coli* DNA digested with *Hin*cII; 2, *Hae* III digest; 3, *Eco*RV digest; 4, *Dde* I digest. An *Eco*RI digest (data not shown) yielded a 6-kb band of hybridization. kbp, Kilobase pairs.

Hfr strains whose origins are distributed around the chromosome (obtained from C. Gross). The kanamycin resistance marker was then tested for P1 cotransduction with mapped Tn*10* insertions in the region as described in the results section. Genetic complementation by cross-streaking the *acrA* mutant strain N43 with strains containing Clarke–Carbon plasmid pLC30-3 was as described (12).

Sequence Comparison. The amino acid sequences were aligned using the program GAP (13). GAP uses the algorithm of Needleman and Wunsch modified to allow the imposition of a gap length penalty. Identities were given a value of 1, mismatches were given a value of 0, gaps were given a value of −3, and the gap length penalty was 0.2. In this study the percentage identity is defined as the number of identical amino acids observed after alignment of the two sequences divided by the number of residues in the shorter sequence.

RESULTS

Hybridization of the Eukaryotic *Hsp83* Gene to *E. coli* Genomic DNA. To determine whether *E. coli* contains sequences related to the eukaryotic *Hsp83* heat shock gene, total *E. coli* DNA was digested with various restriction enzymes, transferred to nitrocellulose, and hybridized at low stringency with a probe made from a protein-coding portion of the *Drosophila Hsp83* gene. As shown in Fig. 1, a single band of hybridization was observed. After *Eco*RI-digestion, DNA was isolated from the region of the gel that had been shown to hybridize to the probe and was used to construct a plasmid library. Two out of the 144 plasmid clones screened hybridized to the *Drosophila* probe. One recombinant plasmid, pBJ1, containing a 6-kb *Eco*RI fragment was chosen for analysis.

DNA Sequence of the *E. coli* Gene and Comparison to Eukaryotic *Hsp83*. Restriction enzyme mapping and hybridization analysis of pBJ1 localized the region that hybridized to *Hsp83* to a 2.5-kb *Sal* I-*Eco*RI fragment. When the DNA

```
                                                                                                       -35 HS             -10 HS
-165  TCGACGGCACCACGTTGTCACACTCCCTTG CCGGGCGTCATAAGATTCGTTTTTAAGCAA ACGAGAGCAGGATCACCTGCTCTCGCTTGA AATTATTCTCCCTTGTCCCCATCTCTCCCA  -46
                                           htpG start
   1                                            MetLysGlyGlnGlu ThrArgGlyPheGlnSerGluValLysGln LeuLeuHisLeuMetIleHisSerLeuTyr   25
 -45  CATCCTGTTTTTAACCTTAAAATGGCATTA TTGAGGTAGACCTACATGAAAGGACAAGAA ACTCGTGGTTTTCAGTCAGAAGTGAAACAG CTTCTGCACCTGATGATCCATTCTCTCTAT   75

  26  SerAsnLysGluIlePheLeuArgGluLeu IleSerAsnAlaSerAspAlaAlaAspLys LeuArgPheArgAlaLeuSerAsnProAsp LeuTyrGluGlyAspGlyGluLeuArgVal   65
  76  TCCAATAAAGAAATCTTCCTGCGTGAGCTT ATCTCTAACGCCTCCGATGCGGCGGACAAG CTGCGTTTCCGTGCGCTCTCTAACCCGGAC CTGTACGAAGGTGATGGCGAACTACGCGTT  195

  66  ArgValSerPheAspLysAspLysArgThr LeuThrIleSerAspAsnGlyValGlyMet ThrArgAspGluValIleAspHisLeuGly ThrIleAlaLysSerGlyThrLysSerPhe  105
 196  CGTGTCTCTTTCGATAAAGACAAGCGTACG CTGACCATCTCCGATAACGGCGTGGGGATG ACCCGCGACGAAGTGATTGACCATCTGGGG ACTATCGCTAAATCCGGTACCAAATCATTC  315

 106  LeuGluSerLeuGlySerAspGlnAlaLys AspSerGlnLeuIleGlyGlnPheGlyVal GlyPheTyrSerAlaPheIleValAlaAsp LysValThrValArgThrArgAlaAlaGly  145
 316  CTCGAATCCCTGGGTTCTGACCAGGCGAAA GACAGCCAGCTGATCGGTCAGTTTGGTGTT GGTTTCTACTCTGCGTTTATCGTGGCCGAC AAAGTGACCGTGCGTACTCGCGCGGCAGGC  435

 146  GluLysProGluAsnGlyValPheTrpGlu SerAlaGlyGluGlyGluTyrThrValAla AspIleThrLysGluAspArgGlyThrGlu IleThrLeuHisLeuArgGluGlyGluAsp  185
 436  GAAAAACCAGAAAATGGCGTCTTCTGGGAA TCGGCTGGCGAAGGTGAATACACCGTTGCC GACATCACCAAAGAAGATCGTGGTACTGAA ATCACCCTGCATCTGCGTGAAGGCGAAGAC  555

 186  GluPheLeuAspAspTrpArgValArgSer IleIleSerLysTyrSerAspHisIleAla LeuProValGluIleGluLysArgGluGlu LysAspGlyGluThrValIleSerTrpGlu  225
 556  GAGTTCCTCGATGACTGGCGCGTGCGTTCC ATCATCAGCAAATACTCCGACCATATCGCG CTGCCGGTAGAGATCGAAAAACGCGAAGAG AAAGACGGCGAAACCGTTATCTCCTGGGAG  675

 226  LysIleAsnLysAlaGlnAlaLeuTrpThr ArgAsnLysSerGluIleThrAspGluGlu TyrLysGluPheTyrLysHisIleAlaHis AspPheAsnAspProLeuThrTrpSerHis  265
 676  AAAATCAACAAAGCGCAGGCGCTGTGGACT CGTAACAAGTCGGAAATCACCGATGAAGAG TACAAAGAGTTCTACAAACACATCGCCCAC GACTTTAATGATCCGCTGACCTGGAGCCAC  795

 266  AsnArgValGluGlyLysGlnGluTyrThr SerLeuLeuTyrIleProSerGlnAlaPro TrpAspMetTrpAsnArgAspHisLysHis GlyLeuLysLeuTyrValGlnArgValPhe  305
 796  AACCGTGTTGAAGGTAAGCAGGAGTACACC AGCCTGCTGTACATCCCGTCCCAGGCTCCG TGGGATATGTGGAACCGCGATCATAAACAC GGCCTGAAACTGTATGTTCAGCGTGTGTTC  915

 306  IleMetAspAspAlaGluGlnPheMetPro AsnTyrLeuArgPheValArgGlyLeuIle AspSerSerAspLeuProLeuAsnValSer ArgGluIleLeuGlnAspSerThrValThr  345
 916  ATCATGGACGACGCAGAACAGTTCATGCCG AACTATCTGCGCTTCGTGCGTGGTCTGATT GACTCCAGCGATCTGCCGCTGAACGTTTCC CGTGAAATCCTCCAGGACAGCACGGTAACG 1035

 346  ArgAsnLeuArgAsnAlaLeuThrLysArg ValLeuGlnMetLeuGluLysLeuAlaLys AspAspAlaGluLysTyrGlnThrPheTrp GlnGlnPheGlyLeuValLeuLysGluGly  385
1036  CGTAACCTGCGCAATGCGCTGACCAAGCGT GTGCTGCAAATGCTGGAAAAACTGGCGAAA GACGACGCGGAAAAATACCAGACCTTCTGG CAACAGTTTGGCCTGGTACTGAAAGAAGGT 1155

 386  ProAlaGluAspPheAlaAsnGlnGluAla IleAlaLysLeuLeuArgPheAlaSerThr HisThrAspSerSerAlaGlnThrValSer LeuGluAspTyrValSerArgMetLysGlu  425
1156  CCGGCGGAAGATTTCGCTAACCAGGAAGCG ATCGCCAAACTGCTGCGTTTTGCTTCTACC CATACCGATTCTTCTGCGCAGACCGTATCT CTGGAAGACTACGTTTCCCGCATGAAAGAA 1275

 426  GlyGlnGluLysIleTyrTyrIleThrAla AspSerTyrAlaAlaAlaLysSerSerPro HisLeuGluLeuLeuArgLysLysGlyIle GluValLeuLeuLeuSerAspArgIleAsp  465
1276  GGGCAGGAGAAAATCTACTACATCACCGCA GACAGCTATGCGGCAGCGAAGAGCAGCCCG CACCTGGAACTGCTGCGTAAGAAAGGCATC GAAGTTCTGCTGCTTTCCGACCGCATCGAT 1395

 466  GluTrpMetMetAsnTyrLeuThrGluPhe AspGlyLysProPheGlnSerValSerLys ValAspGluSerLeuGluLysLeuAlaAsp GluValAspGluSerAlaLysGluAlaGlu  505
1396  GAGTGGATGATGAACTATCTGACTGAGTTC GACGGTAAACCGTTCCAGTCGGTGTCTAAA GTTGACGAGTCGCTTGAAAAACTGGCTGAC GAAGTTGATGAGAGCGCGAAAGAAGCGGAG 1515

 506  LysAlaLeuThrProPheIleAspArgVal LysAlaLeuLeuGlyGluArgValLysAsp ValArgLeuThrHisArgLeuThrAspThr ProAlaIleValSerThrAspAlaAspGlu  545
1516  AAAGCACTGACTCCGTTCATCGACCGTGTG AAAGCCCTGCTCGGCGAGCGCGTGAAAGAT GTCCGTCTGACTCACCGTCTGACCGATACG CCAGCGATCGTTTCGACCGACGCGGACGAA 1635

 546  MetSerThrGlnMetAlaLysLeuPheAla AlaAlaGlyGlnLysValProGluValLys TyrIlePheGluLeuAsnProAspHisVal LeuValLysArgAlaAlaAspThrGluAsp  585
1636  ATGAGCACTCAGATGGCGAAACTGTTCGCT GCGGCGGGCCAGAAAGTGCCAGAAGTGAAA TACATCTTCGAACTGAACCCGGATCACGTA CTGGTGAAACGTGCGGCAGATACTGAAGAT 1755

 586  GluAlaLysPheSerGluTrpValGluLeu LeuLeuAspGlnAlaLeuLeuAlaGluArg GlyThrLeuGluAspProAsnLeuPheIle ArgArgMetAsnGlnLeuLeuValSerEnd  624
1756  GAAGCGAAGTTCAGCGAGTGGGTAGAACTG CTGCTGGATCAGGCGCTGCTGGCAGAACGC GGCACGCTGGAAGATCCGAACCTGTTTATT CGTCGTATGAACCAGCTGCTGGTTTCCTGA 1875
          ------------>      <------- -----                                                                           "-35"
1876  TGTAATGCCGGATGACCTTCGTGTCATCCG GCATTTTTCTTTTCATCATCTGCACTTTCC GCAAATTATCTCGCCATTAACCGTTTCAGC CCCAGGTGCCTTTCTTGAGGCAATCGCCTG 1995

            "-10"                                                  MetArgIleIleLeu
1996  TTGGTGGTATCGTTTATCGCTTTTTCAAAA AATTCGACACATTTTAAGGGGATTTTCGCA ATGCGTATCATTCTG  2068
                                                                   adk start
```

FIG. 2. DNA sequence of the *htpG* heat shock gene. The sequence and characterization of the heat shock promoter of the *htpG* gene has been reported (14). HS, heat shock promoter sequence. The 624-amino acid protein sequence resulting from translation of the 1875-base-pair open reading frame is shown above the DNA sequence. All of the sequence was determined on both strands of the DNA with the exception of base pairs 1390–1459 and 2030–2068, which were determined on one strand only. We sequenced across all restriction sites. The arrows indicate the inverted repeat structure of a potential transcription terminator. The promoter sequences upstream from the *adk* start were found by inspection so are enclosed in quotation marks. The sequence of *adk* and flanking DNA, base pairs 1858–2068, has been reported (15). Our sequence differs from the published *adk* sequence at three positions: base pairs 1867, 1904–1906, and 1926.

sequence of this region was determined, an open reading frame that could encode a protein of M_r 71,429 was found (Fig. 2). A comparison of the predicted amino acid sequence of this open reading frame with the *Drosophila* M_r 83,000 heat shock protein (hsp83) revealed 41% identity (Fig. 3). The human and yeast hsp83 proteins are 42% and 40% identical with the *E. coli* protein, respectively (17, 18). If conservative amino acid substitutions are allowed in the comparison, the percent similarity between the *E. coli* and eukaryotic proteins is 57%. The homology extends for the entire length of the *E. coli* protein. Some segments show extreme conservation. Two 20-amino acid stretches near the amino terminus are ≥90% identical between the *E. coli* and eukaryotic sequences, and the region from amino acid 298 to amino acid 340 of the *E. coli* protein has 65% identity or 81% similarity. Two major segments of the eukaryotic hsp83 protein are absent in the *E. coli* protein homologous to hsp83. One deletion removes a highly hydrophilic stretch of 50 amino acids; the other deletion shortens the carboxyl terminus by 35 residues. As a result of these differences, the *E. coli* gene homologous to *Hsp83* encodes a smaller protein of M_r 71,429. The amino acids in the regions missing in the *E. coli* protein are poorly conserved among the eukaryotic hsp83 proteins. Although the hsp83 homologous protein in *E. coli* has a M_r of ≈70,000, it shows no homology to the hsp70 protein of eukaryotes. This is expected because the hsp83 protein of eukaryotes is not homologous to hsp70. The protein in *E. coli* that is homologous to hsp70 is a heat shock protein called DnaK (5). It has a M_r of 69,121.

The *E. coli* Protein Homologous to the Eukaryotic hsp83 Protein is the C62.5 Heat Shock Protein. Transcripts of the *E. coli* gene homologous to *Hsp83* increase in abundance after a temperature upshift (14). To identify which of the *E. coli* heat shock proteins is the hsp83 homologue, minicells containing a *Pst* I–*Eco*RI subclone of the *E. coli* gene on a plasmid were labeled with [^{35}S]methionine. The *E. coli* protein synthesized in minicells comigrates during two-dimensional gel electrophoresis with the C62.5 heat shock protein of *E. coli* (Fig. 4). Since the gene encoding C62.5 was unknown, the C62.5 protein had been named by F. Neidhardt based on its coordinates on two-dimensional gels (8). Although the structural gene had not been identified it was provisionally called *htpG* (for *h*igh *t*emperature *p*rotein *G*) (3). To confirm that the C62.5 protein is encoded by the open reading frame observed, an in-frame deletion derivative of the C62.5 clone was constructed *in vitro*. We fused the *Alu* I site at the codon for amino acid 15 to the *Pvu* II site at the codon for amino acid 620, removing all but 18 codons of the putative *htpG* gene. Minicells containing this plasmid synthesized only pBR322-encoded proteins. Furthermore, the amino acid composition and amino-terminal protein sequence predicted from the DNA sequence of the hsp83 homologue agree with those of the purified C62.5 protein (J. Spence and C. Georgopoulos, personal communication). We conclude that the *E. coli* hsp83 homologue is the C62.5 heat shock protein.

The Gene for C62.5 Is Located Between the *dnaZ* and *adk* Genes. We reasoned that if *htpG* were located at a previously identified gene locus that information would help us understand the function of C62.5. Inserting a kanamycin resistance gene into the chromosome near the *htpG* gene allowed mapping of the *htpG* gene. Hfr mapping placed this kan-

```
  1  MKGQETRGFQSEVKQLLHLMIHSLYSNKEIFLRELISNASDAADKLRFRALSNPDLYEGDGELRVRVSFDKDKRTLTIS E.coli C62.5
  1  MPEEAETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNASDALDKIRYESLTDPSKLDSGKELYIKLIPNKTAGTLTII Dm Hsp83

 80  DNGVGMTRDEVIDHLGTIAKSGTKSFLESLGSDQAKDSQLIGQFGVGFYSAFIVADKVTVRTRAAGEKPENGVFWESAGE E.coli C62.5
 81  DTGIGMTKSDLVNNLGTIAKSGTKAFMEAL--QAGADISMIGQFGVGFYSAYLVADKVTVTSKNNDDEQYV---WESSAG Dm Hsp83

161  GEYTV-ADITKE-DRGTEITLHLREGEDEFLDDWRVRSIISKYSDHIALPVEIEKREEKDGETV---------------- E.coli C62.5
156  GSFTVRADNSEPLGRGTKIVLYIKEDQTDYLEESKIKEIVNKHSQFIGYPIKLLVEKEREKEVSDDEADDEKKEGDEKKE Dm Hsp83

222  ---------------------------------ISWEKINKAQALWTRNKSEITDEEYKEFYKHIAHDFNDPLTWSHNR E.coli C62.5
236  METDEPKIEDVGEDEDADKKDKDAKKKKTIKEKYTEDEELNKTKPIWTRNPDDISQEEYGEFYKSLTNDWEDHLAVKHFS Dm Hsp83

268  VEGKQEYTSLLYIPSQAPWDMW-NRDHKHGLKLYVQRVFIMDDAEQFMPNYLRFVRGLIDSSDLPLNVSREILQDSTVTR E.coli C62.5
316  VEGQLEFRALLFIPRRTPFDLFENQKKRNNIKLYVRRVFIMDNCEDLIPEYLNFMKGVVDSEDLPLNISREMLQQNKVLK Dm Hsp83

347  NLRNALTKRVLQMLEKLAKDDAEKYQTFWQQFGLVLKEGPAEDFANQEAIAKLLRFASTHTDSSAQT-VSLEDYVSRMKE E.coli C62.5
396  VIRKNLVKKTMELIEELT-EDKENYKKFYDQFSKNLKLGVHEDSNNRAKLADFLRF---HTSASGDDFCSLADYVSRMKD Dm Hsp83

426  GQEKIYYITADSYAAAKSSPHLELLRKKGIEVLLLSDRIDEWMMNYLTEFDGKPFQSVSKVDESLEKLADEVDESAKEAE E.coli C62.5
472  NQKHVYFITGESKDQVSNSAFVERVKARGFEVVYMTEPIDEYVIQHLKEYKGKQLVSVTK--EGLELPEDESEKKKREED Dm Hsp83

506  KALTPFIDR-VKALLGERVKDVRLTHRLTDTPAIVSTDADEMSTQMAKLFAAAGQKVPEV------KYIFELNPDHVLV- E.coli C62.5
550  KAKFESLCKLMKSILDNKVEKVVVSNRLVDSPCCIVTSQFGWSANMERIMKAQALRDTATMGYMAGKKQLEINPDHPIVE Dm Hsp83

578  ----KRAADTEDEAKFSEWVELLLDQALLAERGTLEDPNLF---IRRMNQLLVS*                          E.coli C62.5
630  TLRQKADADKNDKAV-KDLVILLFETSLLSSGFSLDSPQVHASRIYRMIKLGLGIDEDEPMTTDDAQSAGDAPSLVEDTE Dm Hsp83

709  DASHMEEVD*                                                                        Dm Hsp83
```

FIG. 3. Comparison of the predicted amino acid sequence (in the single-letter code) of the *E. coli* C62.5 heat shock protein and *Drosophila melanogaster* (Dm) hsp83 sequence from Blackman and Meselson (16). Those amino acids identical in both sequences are connected by a solid oval. Open ovals indicate replacements by highly conserved amino acids. Conservative replacements are defined as being within the following groups: (aspartic acid and glutamic acid), (lysine and arginine), (serine and threonine), (phenylalanine and tyrosine), and (isoleucine, leucine, valine, and methionine). Gaps have been inserted to improve the alignment and are marked by hyphens within the sequences.

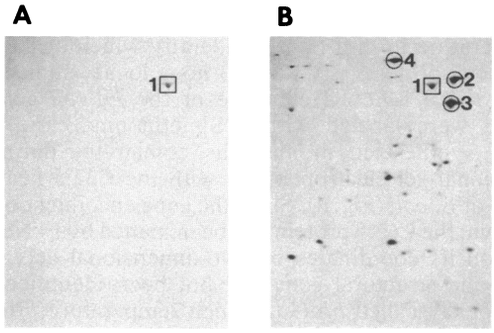

FIG. 4. Comigration of the protein produced by the hsp83 homologous clone and the C62.5 heat shock protein of *E. coli*. Minicells containing the C62.5 clone pBJ1 were labeled with [^{35}S]-methionine and mixed with [^{3}H]leucine-labeled heat shocked *E. coli* cells. The cells were lysed, and their proteins were resolved by two-dimensional gel electrophoresis (9). (*A*) Autoradiogram showing plasmid produced ^{35}S-labeled protein spot in minicells. (*B*) Fluorogram of same gel showing both the ^{35}S-labeled plasmid protein and ^{3}H-labeled *E. coli* heat shock proteins. Heat shock proteins were identified according to Neidhardt *et al.* (8): protein 1, C62.5; protein 2, DnaK; protein 3, GroEL; protein 4, F84.1.

amycin resistance marker between 10 and 20 minutes on the *E. coli* map. Kanamycin resistance was found to cotransduce at approximately equal frequencies with *tsx*::Tn*10* at 9.4 minutes and zbc21::Tn*10* at 12.2 minutes, placing *htpG* between the two markers, in the vicinity of 11 minutes. Subsequently, the *htpG* clone was found to overlap with previously isolated clones from this region that contain the *dnaZX*, *apt*, and *adk* genes. Restriction mapping placed *htpG* between the *dnaZX* and the *adk* genes, both of which had been sequenced (refs. 15 and 19 and Fig. 5).

***htpG* Appears to Be a Newly Identified Gene.** The goal of mapping *htpG* was to determine if it was identical to any previously identified gene. Since the gene encoding C62.5 is located in the same region as *ushA*, *acrA*, *hemH*, and *ras*, we wished to determine whether *htpG* was identical to any of these loci. The *ushA* gene has been cloned and sequenced and is not *htpG* (22). The Clarke–Carbon plasmid pLC30-3 complements the *acrA*1 and *dnaZ* mutations. This plasmid overlaps with the left end of the C62.5 clone pBJ1 but not with the part containing the *htpG* gene. Restriction mapping of these plasmids indicates that the gene order is *acrA-apt-dnaZX-htpG-adk*. Unfortunately, it appears that strains containing characterized mutations in the *ras* (*ra*diation *s*ensitivity) and *hemH* (*hem*in biosynthesis) genes no longer exist.

DISCUSSION

Both eukaryotes and prokaryotes, when subjected to a sudden temperature increase, rapidly accelerate the synthesis rate of a set of proteins. The question has arisen as to whether these responses are only superficially similar, or whether the heat shock response has been conserved as a unit. We have shown (5) that the M_r 70,000 heat shock protein (hsp70) of eukaryotes is homologous to the *E. coli* heat shock

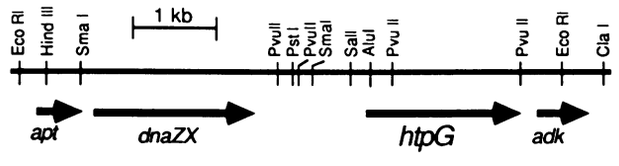

FIG. 5. Restriction map of the *htpG* gene and flanking regions. The thick black arrows indicate the position and orientation of the known genes in the region (see refs. 15 and 19–21). Not all *Alu* I sites are shown. kb, Kilobase pairs.

protein DnaK. We have found another parallel between the two heat shock systems; the C62.5 heat shock protein in *E. coli* is 41% and 42% identical to the hsp83 proteins of *Drosophila* and humans. Ninety-four percent of the amino acid residues that are identical between the *E. coli* and *Drosophila* proteins are conserved in human hsp83 as well (17). Approximately 1.8×10^9 years have passed since the ancestors of humans and *E. coli* diverged (23), and only a very small number of proteins are known to be as highly conserved as hsp83. Some of the proteins that are the most highly conserved between eukaryotes and prokaryotes are chloroplast and mitochondrial proteins. The β subunit of ATP synthetase from barley chloroplasts is 68% identical with the *E. coli* β subunit (24). Elongation factor Tu from yeast mitochondria is 64% identical to the *E. coli* elongation factor Tu (25). The cytosolic form of this elongation factor in yeast is 32% identical to the *E. coli* protein (26). Other very highly conserved cytosolic proteins include the main-stream enzymes glyceraldehyde-3-phosphate dehydrogenase at 51%–53% identity between prokaryotic and eukaryotic species (27), triose phosphate isomerase at 45% (27), and the heat shock proteins hsp70 at 49%–50% (5, 28) and hsp83 at 40%–42%. Therefore, the heat shock proteins are among the most conserved proteins known. Some regions of hsp83 are ≥90% conserved in all species for which the sequence is known. These regions where the sequence is highly conserved may represent functionally conserved regions of the protein. The extreme conservation of the heat shock proteins implies that these proteins may be performing the same or similar roles in all species.

The function of hsp83 is not known. In eukaryotes, cell fractionation studies indicate that it is an abundant, cytoplasmic protein and immunocytological localizations show diffuse cytoplasmic staining (29). hsp83 has been reported to associate transiently with retroviral transforming proteins (30), steroid hormone receptor complexes (31, 32), and to bind to actin (33). Only a small portion of hsp83 is found in such complexes, while the bulk of the protein sediments as a monomer (34). It is not clear what physiological roles these associations may play. Although the *E. coli* hsp83 homologous protein, C62.5, has been identified as a protein whose synthesis rate is accelerated after heat shock, nothing is known about its functon. Identification of the gene for C62.5 allows genetic and biochemical analysis that should help elucidate the role of hsp83.

We thank D. Stafford and C. Georgopoulos for communicating results prior to publication; the members of the C. Gross laboratory for advice on genetic mapping; L. Manseau and W. Boorstein for advice on computer analysis; and K. Tilly, M. Werner-Washburne, J. Doctor, M. Cockrem, J. Erickson, and D. Cowing for the critical reading of this manuscript. This work was supported by Grant GM72870 from the National Institutes of Health.

1. Craig, E. A. (1985) *CRC Crit. Rev. Biochem.* **18,** 239–280.
2. Lindquist, S. (1986) *Annu. Rev. Biochem.* **55,** 1151–1191.
3. Neidhardt, F. C., VanBogelen, R. A. & Vaughn, V. (1984) *Annu. Rev. Genet.* **18,** 295–329.
4. Nover, L. (1984) *Heat Shock Response of Eukaryotic Cells* (Springer, New York), pp. 7–12.
5. Bardwell, J. C. A. & Craig, E. A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 848–852.
6. Neidhardt, F. C., Vaughn, V., Phillips, T. A. & Bloch, P. L. (1983) *Microbiol. Rev.* **47,** 231–284.
7. Lloyd, R. G. & Buckman, C. (1985) *J. Bacteriol.* **164,** 836–844.
8. Neidhardt, F. C., VanBogelen, R. A. & Lau, E. T. (1983) *J. Bacteriol.* **153,** 597–603.
9. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 4007–4021.
10. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
11. Winans, S. C., Elledge, S. J., Krueger, J. H. & Walker, G. C. (1985) *J. Bacteriol.* **161,** 1219–1221.
12. Hübscher, U. & Kornberg, A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6284–6288.

13. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
14. Cowing, D. W., Bardwell, J. C. A., Craig, E. A., Woolford, C., Hendrix, R. W. & Gross, C. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2679–2683.
15. Brune, M., Schumann, R. & Wittinghofer, F. (1985) *Nucleic Acids Res.* **13,** 7139–7151.
16. Blackman, R. K. & Meselson, M. (1986) *J. Mol. Biol.* **188,** 499–515.
17. Rebbe, N. F., Ware, J., Bertina, R. M., Modrich, P. & Stafford, D. W. (1987) *Gene*, in press.
18. Farrelly, F. W. & Finkelstein, D. B. (1984) *J. Biol. Chem.* **259,** 5745–5751.
19. Yin, K.-C., Blinkowa, A. & Walker, J. R. (1986) *Nucleic Acids Res.* **14,** 6541–6549.
20. Kodaira, M., Biswas, S. B. & Kornberg, A. (1983) *Mol. Gen. Genet.* **192,** 80–86.
21. Hershey, H. V. & Taylor, M. W. (1986) *Gene* **43,** 287–297.
22. Burns, D. M. & Beacham, I. R. (1986) *Nucleic Acids Res.* **14,** 4325–4342.
23. Hori, H. & Osawa, S. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 381–385.
24. Zurawski, G. & Clegg, M. T. (1984) *Nucleic Acids Res.* **12,** 2549–2559.
25. Nagata, S., Tsunetsugu-Yokota, Y., Naito, A. & Kaziro, Y. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6192–6196.
26. Nagata, S., Nagashima, K., Tsunetsugu-Yokota, Y., Fujimura, K., Miyazaki, M. & Kaziro, Y. (1984) *EMBO J.* **3,** 1825–1830.
27. Doolittle, R. F., Feng, D. F., Johnson, M. S. & McClure, M. A. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51,** 447–455.
28. Hunt, C. & Morimoto, R. I. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6455–6459.
29. Lai, B. T., Chin, N. W., Stanek, A. E., Keh, W. & Lanks, K. W. (1984) *Mol. Cell. Biol.* **4,** 2802–2810.
30. Brugge, J., Yonemoto, W. & Darrow, D. (1983) *Mol. Cell. Biol.* **3,** 9–19.
31. Schuh, S., Yonemoto, W., Brugge, J., Bauer, V. J., Riehl, R. M., Sullivan, W. P. & Toft, D. O. (1985) *J. Biol. Chem.* **260,** 14292–14296.
32. Sanchez, E. R., Toft, D. O., Schlesinger, M. J. & Pratt, W. B. (1985) *J. Biol. Chem.* **260,** 12398–12401.
33. Koyasu, S., Nishida, E., Kadowaki, T., Matsuzaki, F., Iida, K., Harada, F., Kasuga, M., Sakai, H. & Yahara, I. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8054–8058.
34. Lanks, K. W. & Kasambalides, E. J. (1979) *Biochim. Biophys. Acta* **578,** 1–12.