

A gene encoding a DnaK/hsp70 homolog in *Escherichia coli*

(chaperone/heat shock cognate protein/ferredoxin operon)

BRENT L. SEATON AND LARRY E. VICKERY*

Department of Physiology and Biophysics, University of California, Irvine, CA 92717

Communicated by Helmut Beinert, November 29, 1993 (received for review May 6, 1993)

ABSTRACT Eukaryotic organisms have been shown to have multiple forms of hsp70-class stress-related proteins, but only a single family member, DnaK, has been found in prokaryotes. We report here the identification of a heat shock cognate gene, designated *hsc*, in *Escherichia coli*. The amino acid sequence deduced from *hsc* predicts a 65,647-Da polypeptide having 41% sequence identity with DnaK from *E. coli*, and overexpression produces a protein (Hsc66) with properties similar to DnaK. In contrast to *dnaK*, however, the *hsc* gene lacks a consensus heat shock promoter sequence, and expression is not induced by elevated temperature. The *hsc* gene is located near 54 min on the physical map, immediately upstream of the *fdx* gene, which encodes a [2Fe–2S] ferredoxin; evidence is presented that the *hsc* and *fdx* genes make up a bicistronic operon in which expression of the ferredoxin is coupled to that of Hsc66. The function of Hsc66 is not known, but the coregulation of its expression with that of ferredoxin suggests the possibility of a specific role in association with the ferredoxin protein.

The 70-kDa heat shock proteins (hsp70) and their cognates (hsc70) make up a ubiquitous, multigene family of highly conserved proteins, which are involved in diverse protein-protein interactions (reviewed in ref. 1). They are important under normal conditions as well as during stress and have been implicated in a variety of processes including stabilization of protein-folding intermediates (2), protein assembly and disassembly (3), protein secretion (4, 5), and protein degradation (6). Eukaryotic organisms have been found to contain multiple hsp70 family members; for example, nine distinct proteins are produced in *Saccharomyces cerevisiae* (3, 7), six have been identified in *Drosophila* (8), and at least eight have been described in mammals (9). In contrast, only a single hsp70-class protein, DnaK, has been reported in prokaryotes. The most extensively characterized of these hsp70 proteins is the DnaK protein from *Escherichia coli*. DnaK plays a role in the heat shock response (10, 11), but it is expressed at levels of $\approx 1\%$ of the cell protein and performs important cellular functions under nonstress conditions (12–14). Hybridization analyses in *E. coli* have revealed only one gene, *dnaK*, located near 0.3 min on the genetic linkage map (15, 16). The presence of a single gene encoding a “stress 70-type” protein in prokaryotes would seem to suggest that all members of the multigene eukaryote hsp70 family evolved from a single DnaK-like ancestral protein (cf. refs. 17 and 18).

We report here the identification of a second hsp70-related gene in *E. coli*. The gene,¹ designated *hsc*, encodes a protein of ≈ 66 kDa (Hsc66), which shows $\approx 40\%$ sequence identity with DnaK and other hsp70-class proteins. The *hsc* gene is found near 54 min on the *E. coli* chromosome and is located immediately upstream of the *fdx* gene, which encodes a [2Fe–2S] ferredoxin (19, 20). These genes appear to make up

a bicistronic operon in which expression of Hsc66 and ferredoxin is coregulated.

EXPERIMENTAL PROCEDURES

General Methods. Expression of Hsc66 was carried out in *E. coli* strain MZ-1 (21). Sequencing, bacterial transformation, and oligonucleotide purification were carried out as described by Sambrook *et al.* (22), and β -galactosidase activities were determined as described by Miller (23). SDS/PAGE was carried out according to Laemmli (24). Western immunoblotting was carried out by the method of Towbin *et al.* (25) using enhanced chemiluminescence detection (Amersham).

Plasmids. The plasmid p66-Fdx, used to overexpress Hsc66, contained the *hsc* and *fdx* genes and flanking regions derived from clone DT10 originally isolated from an *E. coli* B genomic library (19, 20).² An *EcoRI* fragment containing the insert was cloned into pBS(+/-) (Stratagene) to yield the plasmid pDT10 and into pAblue (27) to yield the plasmid pADT10. Five hundred and four base pairs of 5' flanking DNA were deleted from pADT10 by digesting with *Nco* I and *Hind*III, and the overhangs were filled in with Klenow DNA polymerase and ligated. The resulting plasmid, p66-Fdx, contained the *hsc* gene, including 188 bp of 5' flanking DNA, under control of the λ p_L promoter.

Plasmid p66-Lac, used for analyses of Hsc66 expression, was constructed by amplifying the region of pDT10 containing the N-terminal nine amino acids of Hsc66 and 690 bp of upstream sequence. PCR primers were constructed such that the 5' end of the PCR product would contain an *EcoRI* site and the 3' end would contain an *Sma* I site to allow fusion of the *hsc* coding sequence in-frame with a *lacZ* coding sequence in pMLB1034 (28). The PCR product was digested with *EcoRI* and *Sma* I, and the 730-bp fragment was ligated to pMLB1034 that had been digested with *EcoRI* and *Sma* I.

Plasmid pFdx-Lac, used for promoter analyses, was constructed using PCR to amplify a region from pDT10 including ≈ 900 bp upstream of the *fdx* coding region and the bases encoding the first 11 amino acids of ferredoxin. PCR primers were designed such that the 3' end of the amplified sequence would contain a *Bam*HI site. The PCR product was cleaved with *Acc* I, at a site located 91 bp upstream of the *fdx* coding sequence, and *Bam*HI, and the resulting 124-bp fragment was ligated to the 2.4-kb *EcoRI*–*Acc* I fragment of pDT10 containing the remaining upstream sequences of DT10. This fragment was then ligated to the 3.1-kb *EcoRI*–*Bam*HI fragment of pMLB1034. The resulting pFdx-Lac contains 2.5 kb of sequence upstream of the *fdx* gene followed by the sequence encoding the first 11 amino acids of ferredoxin fused in-frame to β -galactosidase beginning at codon 8. Upstream deletion derivatives of pFdx-Lac shown in Fig. 5 were prepared by

*To whom reprint requests should be addressed.

¹The sequence reported in this paper has been deposited in the GenBank data base (accession no. U05338).

²The *hsc* and *fdx* genes are also present in *E. coli* K12 and are located in λ clones 7F8 and 5E10 (20) in the miniset library isolated by Kohara *et al.* (26).

digesting pFdx-Lac with the appropriate restriction enzymes, isolating the large vector fragment by agarose electrophoresis, and ligating the purified deletion construct.

Protein Expression and Purification. MZ-1 cells transformed with plasmid p66-Fdx were grown in Terrific broth (22) to $OD_{600} \approx 0.5$, induced by heating to 42°C for 2 hr, and subsequently grown overnight at 37°C. Cells were harvested by centrifugation and disrupted by French press. Protein extracts were fractionated by anion-exchange chromatography and molecular sieving chromatography on Sephacryl S-300. The description of a more complete purification procedure will be published (L. W. Goodman and L.E.V.).

Primer Extension. Primer extension analysis was carried out using a synthetic oligonucleotide complementary to nucleotides 22–41 of the *hsc* coding sequence. The oligonucleotide was labeled at the 5' end using [γ -³²P]ATP and hybridized to 50 μ g of total RNA isolated from *E. coli* strain DH5 α previously transformed with pDT10. RNA was isolated from late-logarithmic phase cells by breaking with glass beads in the presence of hot (65°C), water-saturated phenol. The mixture was vortexed for 30 s followed by incubation at 65°C for 30 s; this cycle was repeated twice. The aqueous phase was extracted with chloroform and precipitated with ethanol. RNA samples were resuspended in RNase-free water and treated with RNase-free DNase. Hybridizations and primer extensions were performed using the Promega primer extension kit following the included protocol. After extension with reverse transcriptase, samples were digested with RNase A for 30 min at 37°C and extracted with phenol. RNA-DNA hybrids were precipitated with ethanol, denatured, and subjected to electrophoresis on a 6% polyacrylamide sequencing gel. Unlabeled oligonucleotide served as a primer for the sequencing ladder comparison.

RESULTS AND DISCUSSION

Identification of the *hsc* Gene. An open reading frame of 1848 bp encoding a possible hsp70-class protein was detected during sequencing of DNA in the 5' flanking region of the *fdx* gene of *E. coli* (Fig. 1). Analysis of codon usage with the *E. coli* codon bias (29) shows a clear preference for this reading frame, and a sequence resembling a Shine-Dalgarno sequence for ribosome binding and translation (30) is found immediately upstream of the initiation AUG. Sequence showing similarity to a -10 consensus promoter sequence is also observed from -63 to -68 bp, but no -35 region consensus sequence (31) is apparent.

The translated DNA sequence of this reading frame predicts a polypeptide of 616 amino acids with a molecular mass of 65,647 Da. A search of the GenBank data base using the deduced amino acid sequence indicated that the predicted protein showed homology to prokaryotic and eukaryotic hsp70-class stress proteins, and among the 100 proteins exhibiting the highest similarity scores, all were either heat shock or heat shock cognate proteins. Because of the apparent lack of a heat shock promoter consensus sequence in the 5' flanking region of the gene (cf. ref. 7; see also below), we considered the gene product to be a heat shock cognate protein and designated the gene *hsc* and the predicted protein Hsc66.

Comparison of Hsc66 with DnaK. The protein exhibiting the highest degree of sequence similarity to Hsc66 is DnaK of *E. coli*, and a comparison of the amino acid sequences predicted for the two proteins is presented in Fig. 2. The alignment shown yields a sequence identity of 41% and a similarity of 60% over the region of residues 17–616 of Hsc66. Similarities to other hsp70 proteins are also notable: Hsc66 has 36% identity with bovine hsc70 (32) and 39% identity with yeast Ssc1p over the same region (33). The similarities observed are especially notable in the N-terminal two-thirds of the proteins; this is the most highly conserved region in hsp70 proteins and

has been identified as an ATPase domain in other forms of hsp70 (34). A number of the conserved residues have been shown to be involved in ATP binding in bovine Hsc70 (35, 36), and Hsc66 residues 208–230 show homology to the ATP binding sites of protein kinases (37). These similarities suggest that, like other hsp70 proteins, Hsc66 may possess ATPase activity. In addition, Thr-212 of Hsc66 aligns with Thr-199 of DnaK, a site of autophosphorylation (38), raising the possibility that Hsc66 may also be subject to regulation by phosphorylation at this position.

Significant differences between the sequences predicted for Hsc66 and DnaK, however, are apparent. It was necessary to introduce several gaps in the sequence of Hsc66 to optimize the alignment. Alignment of the sequences of Hsc66 and DnaK with the structure of the ATPase fragment of bovine Hsc70 (35, 36) suggests that the regions in which gaps were introduced may correspond to residues near the surface of the folded protein; thus, these differences may reflect different surface structural features of Hsc66 compared to DnaK and bovine Hsc70 in those regions. In addition, Hsc66 is predicted to have an N-terminal extension not present in DnaK and lacks 17 C-terminal residues found in DnaK. The 16-residue N-terminal extension of Hsc66, which is not present in *E. coli* DnaK, is unusual because a similar extension is absent from the predicted sequences of DnaK proteins found in other prokaryotes. Some eukaryotic forms of hsp70 contain N-terminal extensions, which function in targeting to, or retention in, the endoplasmic reticulum or mitochondria (for review, see ref. 38), but these do not show sequence similarity to the N terminus of Hsc66. Moreover, the N-terminal sequence of Hsc66 does not show homology to known signal sequences of membrane-bound or periplasmic proteins of *E. coli*, and its role remains to be determined. The divergence between the amino acid sequences of Hsc66 and DnaK in the C-terminal region is similar to the variability observed in other hsp70 proteins. The C-terminal domain is believed to be involved in protein recognition (37–39), and the differences observed suggest that Hsc66 is likely to interact with different target protein(s) within the cell.

Expression of Hsc66. To establish the identity of the *hsc* gene product, the plasmid p66-Fdx, containing 188 bp of 5' flanking DNA and the putative coding region, was constructed to overexpress Hsc66 in MZ-1 cells. Fractionation of extracts from induced cells revealed a major band of ≈ 66 kDa, which was partially purified by anion-exchange and gel-filtration chromatography. Fig. 3 *Left* shows the preparation following SDS/PAGE and blotting to a poly(vinylidene difluoride) membrane. The major band of ≈ 66 kDa was subjected to N-terminal amino acid sequencing for nine cycles and was identified as Hsc66; the first residue of the mature protein was found to be alanine, indicating that the formylmethionine was removed as found for other proteins in *E. coli* which have alanine as the penultimate N-terminal residue (40). DnaK, which has properties similar to Hsc66, copurified in the preparation and is visible as a minor band that migrates with an apparent molecular mass of ≈ 75 kDa, an anomaly previously reported (41).

Western immunoblot analyses were carried out on the partially purified preparation to test the relatedness of Hsc66 and DnaK. Antisera to DnaK from three rabbits were separately tested, and in no case was cross-reactivity with Hsc66 observed. As shown in Fig. 3 *Right*, DnaK is readily detected in the preparation by each antiserum, whereas Hsc66, although present in larger amounts, is not detected, suggesting that the major epitopes present in DnaK are not conserved in Hsc66.

Characterization of the *hsc*-*fdx* Operon. The 5' flanking region of the *hsc* gene does not contain sequences resembling the consensus sequences found in heat shock promoters (-CnCcTTGAA- in the -35 region and -CCCCATnT- in the -10 region), which are recognized by the heat shock σ^{32} factor (7).

```

-180          -160          -140          -120          -100
TCAAACGTGT GAAAAAGATG TTTGATACCC GCCATCAGTT GATGGTTGAA CAGTTAGACA ACGAGACGTG GGACGCGGCG GCGGATACCG
          -80          (-10)          -60          -40          -20          S-D
TGCGTAAAGCT CCGTTTTCTC GATAAACTGC GAAGCAGTGC CGAACAACCT GAAGAAAAAC TGCTCGATT TTAATTTCTG GAAGCTAAAC
1
ATG GCC TTA TTA CAA ATT AGT GAA CCT GGT TTG AGT GCC GCG CCG CAT CAG CGT CGT CTG GCG GCC GGT ATT GAC
Met Ala Leu Leu Gln Ile Ser Glu Pro Gly Leu Ser Ala Ala Pro His Gln Arg Arg Leu Ala Ala Gly Ile Asp
          90          120          150
CTG GGC ACA ACC AAC TCG CTG GTG GCG ACA GTG CGC AGC GGT CAG GCC GAA ACG TTA GCC GAT CAT GAA GGC CGT
Leu Gly Thr Thr Asn Ser Leu Val Ala Thr Val Arg Ser Gly Gln Ala Glu Thr Leu Ala Asp His Glu Gly Arg
          180          210          240          270          300
CAC CTG CTG CCA TCT GTT GTT CAC TAT CAA CAG CAA GGG CAT TCG GTG GGT TAT GAC GCG CGT ACT AAT GCA GCG
His Leu Leu Pro Ser Val Val His Tyr Gln Gln Gln Gly His Ser Val Gly Tyr Asp Ala Arg Thr Asn Ala Ala
          330          360          390          420          450
CTC GAT ACC GCC AAC ACA ATT AGT TCT GTT AAA CGC CTG ATG GGA CGC TCG CTG GCT GAT ATC CAG CAA CGC TAT
Leu Asp Thr Ala Asn Thr Ile Ser Ser Val Lys Arg Leu Met Gly Arg Ser Leu Ala Asp Ile Gln Gln Arg Tyr
          480          510          540          570          600
CCG CAT CTG CCT TAT CAA TTC CAG GCC AGC GAA AAC GGC CTG CCG ATG ATT GAA ACG GCG GCG GGG CTG CTG AAC
Pro His Leu Pro Tyr Gln Phe Gln Ala Ser Glu Asn Gly Leu Pro Met Ile Glu Thr Ala Ala Gly Leu Leu Asn
          630          660          690          720          750
CCG GTG CGC GTT TCT CCG GAC ATC CTC AAA GCA CTG GCG GCG CGG GCA ACT GAA GCC CTG GCA GGC GAG CTG GAT
Pro Val Arg Val Ser Ala Asp Ile Leu Lys Ala Leu Ala Arg Ala Thr Glu Ala Leu Ala Gly Glu Leu Asp
          780          810          840          870          900
GGT GTA GTT ATC ACC GTT CCG GCG TAC TTT GAC GAT GCC CAG CGT CAG GGC ACC AAA GAC GCG GCG CGT CTG GCG
Gly Val Val Ile Thr Val Pro Ala Tyr Phe Asp Asp Ala Gln Arg Gln Gly Thr Lys Asp Ala Ala Arg Leu Ala
          930          960          990          1020          1050
GGC CTT CAC GTC CTG CGC TTA CTT AAC GAA CCG ACC GCT GCG GCT ATC GCC TAC GGG CTG GAT TCC GGT CAG GAA
Gly Leu His Val Leu Arg Leu Leu Asn Glu Pro Thr Ala Ala Ala Ile Ala Tyr Gly Leu Asp Ser Gly Gln Glu
          1080          1110          1140          1170          1200
GGC GTG ATC GCC GTT TAT GAC CTC GGT GGC GGG ACG TTT GAT ATT TCC ATT CTG CGC TTA AGT CGC GGC GTG TTT
Gly Val Ile Ala Val Tyr Asp Leu Gly Gly Gly Thr Phe Asp Ile Ser Ile Leu Arg Leu Ser Arg Gly Val Phe
          1230          1260          1290          1320          1350
GAA GTG CTG GCA ACC GGC GGT GAT TCC GCG CTC GGC GGC GAT GAT TTC GAC CAT CTG CTG GCG GAT TAC ATT CGC
Glu Val Leu Ala Thr Gly Gly Asp Ser Ala Leu Gly Gly Asp Asp Phe Asp His Leu Leu Ala Asp Tyr Ile Arg
          1380          1410          1440          1470          1500
GAG CAG GCG GGC ATT CCT GAT CGT AGC GAT AAC CGC GGT CAG CGT GAA CTG CTG GAT GCC GCC ATT GCA GCC AAA
GAG Gln Ala Gly Ile Pro Asp Arg Ser Asp Asn Arg Val Gln Leu Leu Asp Ala Ala Ile Ala Lys Lys
          1530          1560          1590          1620          1650
ATC GCG CTG AGC GAT GCG GAC TCC GTG ACC GTT AAC GTT GCG TGG CAG GGC GAA ATC AGC CGT GAA CAA TTC
Ile Ala Leu Ser Asp Ala Asp Ser Val Thr Val Asn Val Ala Gly Trp Gln Gly Glu Ile Ser Arg Glu Gln Phe
          1680          1710          1740          1770          1800
AAT GAA CTG ATC GCG CCA CTG GTA AAA CGA ACC TTA CTG GCT TGT CGT CGC GCG CTG AAA GAC GCG GGT GTA GAA
Asn Glu Leu Ile Ala Pro Leu Val Lys Arg Thr Leu Leu Ala Cys Arg Arg Ala Leu Lys Asp Ala Gly Val Glu
          1830          1848          Fdx →
GCT GAT GAA GTG CTG GAA GTG GTG ATG GTG GGC GGT TCT ACT CGC GTG CCG CTG GTG CGT GAA CGG GTA GGC GAA
Ala Asp Glu Val Leu Glu Val Val Met Val Gly Gly Ser Thr Arg Val Pro Leu Val Arg Glu Arg Val Gly Glu
          1860          1890          1920          1950          1980
TTT TTC GGT CGT CCA CCG CTG ACT TCC ATC GAC CCG GAT AAA GTC GTC GCT ATT GGC GCG GCG ATT CAG GCG GAT
Phe Phe Gly Arg Pro Pro Leu Thr Ser Ile Asp Pro Asp Lys Val Val Ala Ile Gly Ala Ala Ile Gln Ala Asp
          2010          2040          2070          2100          2130
ATT CTG GTG GGT AAC AAG CCA GAC AGC GAA ATG CTG CTG CTT GAT GTG ATC CCA CTG TCG CTG GGC CTC GAA ACG
Ile Leu Val Gly Asn Lys Pro Asp Ser Glu Met Leu Leu Leu Asp Val Ile Pro Leu Ser Leu Gly Leu Glu Thr
          2160          2190          2220          2250          2280
ATG GGC GGT CTG GTG GAG AAA GTG ATT CCG CGT AAT ACC ACT ATT CCG GTG GCC CGC GCT CAG GAT TTC ACC ACC
Met Gly Gly Leu Val Glu Lys Val Ile Pro Arg Asn Thr Thr Ile Pro Val Ala Arg Ala Gln Asp Phe Thr Thr
          2310          2340          2370          2400          2430
TTT AAA GAT GGT CAG ACG GCG ATG TCT ATC CAT GTA ATG CAG GGT GAG CGC GAA CTG GTG CAG GAC TGC CGC TCA
Phe Lys Asp Gly Gln Thr Ala Met Ser Ile His Val Met Gln Gly Glu Arg Glu Leu Val Gln Asp Cys Arg Ser
          2460          2490          2520          2550          2580
CTG GCG CGT TTT GCG CTG CGT GGT ATT CCG GCG CTA CCG GCT GGC GGT GCG CAT ATT CGC GTG ACG TTC CAG GTC
Leu Ala Arg Phe Ala Leu Arg Gly Ile Pro Ala Leu Pro Ala Gly Gly Ala His Ile Arg Val Thr Phe Gln Val
          2610          2640          2670          2700          2730
GAT GCC GAC GGT CTT TTG AGC GTG ACG GCG ATG GAG AAA TCC ACC GGC GTT GAG GCG TCT ATT CAG GTC AAA CCG
Asp Ala Asp Gly Leu Leu Ser Val Thr Ala Met Lys Ser Thr Gly Val Glu Ala Ser Ile Gln Val Lys Pro
          2760          2790          2820          2850          2880
TCT TAC GGT CTG ACT GAC AGC GAA ATC GCT TCG ATG ATC AAA GAC TCA ATG AGC TAT GCC GAG CAG GAC GTA AAA
Ser Tyr Gly Leu Thr Asp Ser Glu Ile Ala Ser Met Ile Lys Asp Ser Met Ser Tyr Val Glu Gln Asp Val Lys
          2910          2940          2970          3000          3030
GCC CGA ATG CTG GCA GAA CAA AAA GTA GAA GCG GCG CGT GTG CTG GAA AGT CTG CAC GGC GCG CTG GCT GCT GAT
Ala Arg Met Leu Ala Glu Gln Lys Val Glu Ala Ala Arg Val Leu Glu Ser Leu His Gly Ala Leu Ala Ala Asp
          3060          3090          3120          3150          3180
GCC GCG CTG TTA AGC GCC GCA GAA CGT CAG GTC ATT GAC GAT GCT GCC GCT CAC CTG AGT GAA GTG GCG CAG GGC
Ala Ala Leu Leu Ser Ala Ala Glu Arg Gln Val Ile Asp Asp Ala Ala Ala His Leu Ser Glu Val Ala Gln Gly
          3210          3240          3270          3300          3330
GAT GAT GTT GAC GCC ATC GAA AAA GCG ATT AAA AAC GTA GAC AAA CAA ACC CAG GAT TTC GCC GCT CGC CGC ATG
Asp Asp Val Asp Ala Ile Glu Lys Ala Ile Lys Asn Val Asp Lys Gln Thr Gln Asp Phe Ala Ala Arg Arg Met
          3360          3390          3420          3450          3480
GAC CAG TCG GTT CGT CGT GCG CTG AAA GGC CAT TCC GTG GAC GAG GTT TAA T ATG CCA AAG ATT GTT ATT TTG
Asp Gln Ser Val Arg Arg Ala Leu Lys Gly His Ser Val Asp Glu Val * Met Pro Lys Ile Val Ile Leu

```

FIG. 1. Nucleotide sequence and deduced amino acid sequence of the *hsc* gene. DNA sequence numbering begins with the predicted initiator methionine of Hsc66; the initiator methionine for the *fdx* gene is indicated by Fdx. Possible regulatory sequences are underlined (−10, promoter sequence; S-D, Shine-Dalgarno sequence), and the proposed site of transcriptional initiation (see Fig. 3) is indicated by an arrow (→).

To test whether other sequences present might function in heat shock induction of the gene, we used the vector p66-Lac, which contains the *lacZ* gene fused in-frame with bases encoding the first nine amino acids of Hsc66 together with 690 bp of 5' flanking DNA; control of expression of the chimeric gene is thus under control of *hsc* promoter sequences. This plasmid was introduced into *E. coli* JM109 cells, and β -galactosidase activity of cell extracts was determined before and after subjecting

cultures to heat shock. No increase in β -galactosidase activity was observed following a shift to 46°C or 51°C for up to 30 min (data not shown), suggesting that Hsc66 is not induced by heat shock and is subject to other control mechanisms.

The observation that only a single base separates the termination codon for *hsc* and the initiation codon for *fdx* suggests that the two genes might function as a bicistronic operon. To determine whether the genes are cotranscribed, the plasmid

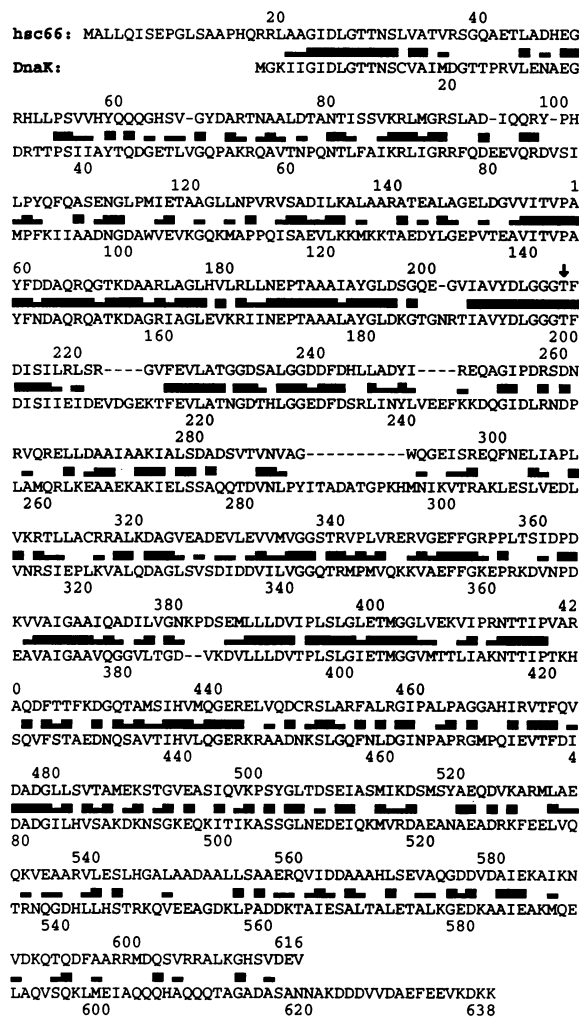


FIG. 2. Comparison of the deduced amino acid sequences of *E. coli* Hsc66 and DnaK. Amino acid identities are denoted by a thick line, and similarities are denoted by a thinner line.

pFdx-Lac was constructed (Fig. 4). This plasmid encodes a ferredoxin- β -galactosidase fusion protein under the control of promoter elements immediately upstream of the *fdx* gene as well as those in the 5' flanking region of the *hsc* gene. Deletion derivatives of pFdx-Lac were made using unique restriction sites at varying distances from the 5' end of the insert. Plasmids were introduced into *E. coli* JM109 cells, and the cells were grown to midlogarithmic phase for determination of β -galactosidase activity levels. Deletion of the *EcoRI*-*HindIII* region of the upstream sequence reduced β -galactosidase activity by only 7%, whereas deletion of the *EcoRI*-*Nru I* region reduced activity by 94%; no further reduction was observed upon deletion of bases to within 90 bp of the *fdx* initiation codon. These findings suggest that under the growth conditions used expression of the *fdx* gene is primarily regulated by promoter sequences between the *HindIII* and *Nru I* sites.

A derivative of plasmid pFdx-Lac was also constructed in which an 8-bp linker was inserted into the *Nru I* site found at position 750 of the *hsc* coding sequence. This insertion causes a shift of the reading frame of *hsc* and introduces a termination codon 28 codons after the site of linker insertion. As shown in Fig. 4, this frameshift reduced the β -galactosidase activity of the *fdx-lacZ* fusion \approx 7-fold. This finding suggests the possibility that *hsc* and *fdx* are translationally coupled, with translation of the ferredoxin mRNA dependent on translation of the Hsc66 mRNA.

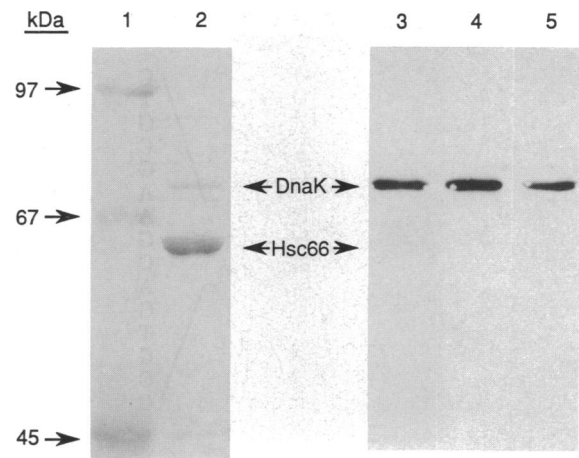


FIG. 3. SDS/PAGE and immunoblot analysis of Hsc66. A partially purified preparation of Hsc66 was subjected to SDS/PAGE in a 10% gel, and the gel was blotted to a poly(vinylidene difluoride) membrane and stained with Coomassie blue to visualize proteins. Lane 1, molecular mass markers; lanes 2-5, \approx 3 μ g of total protein. The membrane containing lanes 1 and 2 shows Coomassie blue-stained protein bands. The membranes containing lanes 3-5 were individually probed with antisera to *E. coli* DnaK obtained from three rabbits (114, 115B, and 116C, respectively) provided by Graham Walker (Massachusetts Institute of Technology); cross-reacting proteins were detected using a peroxidase-conjugated goat anti-rabbit second antibody and luminol chemiluminescence exposure to autoradiography film.

To determine the 5' end of the *hsc-fdx* transcript(s), primer extension reactions were performed using total RNA isolated from *E. coli* cells transformed with plasmid pDT10. Reactions were primed with an oligonucleotide complementary to nucleotides 22-41 within the coding sequence of *hsc* or to nucleotides 43-60 within the coding sequence of *fdx*. The results using the primer within the *hsc* gene showed a single major transcript starting 57 bp upstream of the initiation ATG (Fig. 5). The DNA sequence upstream at positions -68 to -63 (TAAACT) shows similarity and spacing to that of -10 promoter sequences; no sequence showing similarity to the -35 consensus promoter sequence (TTGACG; ref. 31), however, is apparent. Using the primer within the *fdx* gene no single, major transcriptional start site was observed within the resolution of the sequencing gel. Instead, a band of moderate intensity (\approx 1/10th as intense as the *hsc* transcript) was seen, which corresponded to a transcriptional start site 21 bp upstream of the *fdx* initiation ATG; multiple sites of weak intensity were also observed along the full

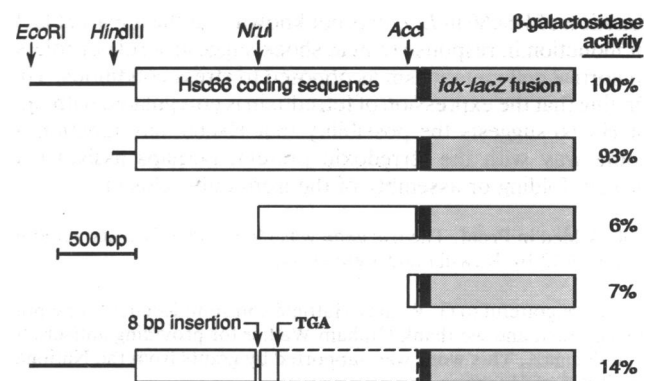


FIG. 4. Promoter analysis of the *hsc-fdx* operon. Deletions and insertions within the plasmid pFdx-Lac were made at the indicated restriction sites. β -Galactosidase activities were measured as described in *Experimental Procedures* and are reported as percent activity of the parent plasmid (100% = 4227 Miller units/mg of protein).

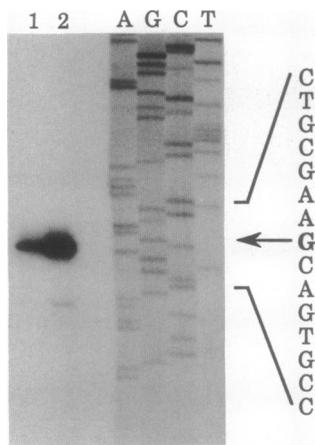


FIG. 5. Primer extension analysis of the transcriptional start site of the *hsc* gene. A [γ - 32 P]ATP-labeled primer complementary to the 5' end of the *hsc* gene was hybridized to 50 μ g of total RNA isolated from *E. coli* strain DH5 α transformed with pDT10 and was extended with reverse transcriptase. Products were analyzed by electrophoresis on a 6% polyacrylamide sequencing gel. Samples in lanes 1 and 2 represent reactions in which the hybridization products were digested with RNase A (lane 1) or were left undigested (lane 2). Lanes A, G, C, and T are products of sequencing reactions using the same oligonucleotide as primer. The sequence of the antisense strand is shown on the right, and the putative start site, G at position -57, is indicated by the arrow.

length of the resolving gel (data not shown). These results, in accordance with the deletion studies described above, suggest that under the conditions used the primary transcript is initiated upstream of the *hsc* gene.

The finding that the *hsc* and *fdx* genes are coregulated suggests that they make up a bicistronic operon. While no complete consensus promoter sequence can be identified within the 5' flanking region of the *hsc* gene, expression levels observed with the Hsc66- and ferredoxin- β -galactosidase fusions and the primer extension results are consistent with a primary site of transcriptional initiation 57 bp upstream of the Hsc66 coding sequence. In addition, the presence of a sequence resembling a ρ -independent termination signal 120 bp downstream of the *fdx* gene and the absence of open reading frames in this region (19) suggest that no additional genes are encoded in the operon. Additional studies are needed, however, to define the transcript and its regulation.

The Function of Hsc66. Hsp70 proteins participate in a variety of processes involving protein folding, protein assembly and disassembly, protein secretion, and protein degradation. The function of Hsc66 in *E. coli* is not known, but the apparent lack of induction in response to heat shock suggests a role or role(s) in normal cell metabolism as opposed to stress conditions. The finding that the expression of ferredoxin is coregulated with that of Hsc66 suggests the possibility that Hsc66 may function in some way with the ferredoxin protein, perhaps assisting in protein folding or assembly of the iron-sulfur cluster.

Note Added in Proof. The *hsc* gene was independently discovered in *E. coli* K12 by Kawula and Lelivelt (42).

We are grateful to G. Wesley Hatfield and John Keener for helpful discussions, and we thank Graham Walker for providing antisera to *E. coli* DnaK. This work was supported by grants from the National Institutes of Health.

1. McKay, D. B. (1993) *Adv. Prot. Chem.* **44**, 67-98.
2. Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K. & Hartl, F. U. (1992) *Nature (London)* **356**, 683-689.

3. Lindquist, S. & Craig, E. A. (1988) *Annu. Rev. Genet.* **22**, 631-677.
4. Chirico, W. J., Waters, M. G. & Blobel, G. (1988) *Nature (London)* **332**, 805-810.
5. Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A. & Schekman, R. (1988) *Nature (London)* **332**, 800-805.
6. Chiang, H.-L., Terlecky, S. R., Plant, C. P. & Dice, J. F. (1989) *Science* **246**, 382-385.
7. Nover, L. (1991) *Heat Shock Response* (CRC, Boca Raton, FL).
8. Craig, E. A., Ingolia, T. D. & Manseau, L. J. (1983) *Dev. Biol.* **99**, 418-426.
9. Arrigo, A.-P. & Welch, W. J. (1987) *J. Biol. Chem.* **262**, 15359-15369.
10. Straus, D. B., Walter, W. A. & Gross, C. A. (1989) *Genes Dev.* **3**, 2003-2010.
11. Straus, D., Walter, W. & Gross, C. (1990) *Genes Dev.* **4**, 2202-2209.
12. Bukau, B. & Walker, G. C. (1989) *J. Bacteriol.* **171**, 2337-2346.
13. Bukau, B. & Walker, G. C. (1989) *J. Bacteriol.* **171**, 6030-6038.
14. Herendeen, S. L., VanBogelen, R. A. & Neidhardt, F. C. (1979) *J. Bacteriol.* **139**, 185-194.
15. Bardwell, J. C. A. & Craig, E. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 848-852.
16. Neidhardt, F. C., VanBogelen, R. A. & Vaughn, V. (1984) *Annu. Rev. Genet.* **18**, 295-329.
17. Gupta, R. S. & Singh, B. (1992) *J. Bacteriol.* **174**, 4594-4605.
18. Hughes, A. L. (1993) *Mol. Biol. Evol.* **10**, 243-255.
19. Ta, D. T. & Vickery, L. E. (1992) *J. Biol. Chem.* **267**, 11120-11125.
20. Ta, D. T., Seaton, B. L. & Vickery, L. E. (1992) *J. Bacteriol.* **174**, 5760-5761.
21. Nagai, K. & Thogersen, H. C. (1984) *Nature (London)* **309**, 810-812.
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
23. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab. Press, Plainview, NY).
24. Laemmli, U. K. (1970) *Nature (London)* **227**, 680-685.
25. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354.
26. Kohara, Y., Akiyama, K. & Isono, K. (1987) *Cell* **50**, 495-508.
27. Brandt, M. E. & Vickery, L. E. (1992) *Arch. Biochem. Biophys.* **294**, 735-740.
28. Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984) *Experiments with Gene Fusions* (Cold Spring Harbor Lab. Press, Plainview, NY).
29. Gribskov, M., Devreux, J. & Burgess, R. R. (1984) *Nucleic Acids Res.* **12**, 539-549.
30. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342-1346.
31. Hawley, D. K. & McClure, W. R. (1983) *Nucleic Acids Res.* **11**, 2237-2255.
32. DeLuca-Flaherty, C. & McKay, D. B. (1990) *Nucleic Acids Res.* **18**, 5569.
33. Craig, E. A., Kramer, J., Shilling, J., Werner-Washburne, M., Holmes, S., Kosic-Smiths, J. & Nicolet, C. M. (1989) *Mol. Cell. Biol.* **9**, 3000-3008.
34. Chappell, T. G., Konforti, B. B., Schmid, S. L. & Rothman, J. E. (1987) *J. Biol. Chem.* **262**, 746-751.
35. Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. (1990) *Nature (London)* **346**, 623-628.
36. Flaherty, K. M., McKay, D. B., Kabsch, W. & Holmes, K. C. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5041-5045.
37. Hannink, M. & Donoghue, D. J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7894-7898.
38. McCarty, J. S. & Walker, G. C. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9513-9517.
39. Gething, M.-J. & Sambrook, J. (1992) *Nature (London)* **355**, 33-45.
40. Flinta, C., Persson, B., Jornvall, H. & von Heijne, G. (1986) *Eur. J. Biochem.* **154**, 193-196.
41. Zylicz, M. & Georgopoulos, C. (1984) *J. Biol. Chem.* **259**, 8820-8825.
42. Kawula, T. H. & Lelivelt, M. J. (1994) *J. Bacteriol.* **176**, 610-619.