JOURNAL OF BACTERIOLOGY, July 1992, p. 4594–4605
0021-9193/92/144594-12$02.00/0

Vol. 174, No. 14

Cloning of the HSP70 Gene from *Halobacterium marismortui*: Relatedness of Archaebacterial HSP70 to Its Eubacterial Homologs and a Model for the Evolution of the HSP70 Gene

RADHEY S. GUPTA* AND BHAG SINGH

Department of Biochemistry, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

Received 11 November 1991/Accepted 5 May 1992

Heat shock induces the synthesis of a set of proteins in *Halobacterium marismortui* whose molecular sizes correspond to the known major heat shock proteins. By using the polymerase chain reaction and degenerate oligonucleotide primers for conserved regions of the 70-kDa heat shock protein (HSP70) family, we have successfully cloned and sequenced a gene fragment containing the entire coding sequence for HSP70 from *H. marismortui*. HSP70 from *H. marismortui* shows between 44 and 47% amino acid identity with various eukaryotic HSP70s and between 51 and 58% identity with its eubacterial and archaebacterial homologs. On the basis of a comparison of all available HSP70 sequences, we have identified a number of unique sequence signatures in this protein family that provide a clear distinction between eukaryotic organisms and prokaryotic organisms (archaebacteria and eubacteria). The archaebacterial (viz., *H. marismortui* and *Methanosarcina mazei*) HSP70s have been found to contain all of the signature sequences characteristic of eubacteria (particularly the gram-positive bacteria), which suggests a close evolutionary relationship between these groups. In addition, detailed analyses of HSP70 sequences that we have carried out have revealed a number of additional novel features of the HSP70 protein family. These include (i) the presence of an insertion of about 25 to 27 amino acids in the N-terminal quadrants of all known eukaryotic and prokaryotic HSP70s except those from archaebacteria and the gram-positive group of bacteria, (ii) significant sequence similarity in HSP70 regions comprising its first and second quadrants from organisms lacking the above insertion, (iii) highly significant similarity between a protein, MreB, of *Escherichia coli* and the N-terminal half of HSP70s, (iv) significant sequence similarity between the N-terminal quadrant of HSP70 (from gram-positive bacteria and archaebacteria) and the m-type thioredoxin of plant chloroplasts. To account for these and other observations, a model for the evolution of HSP70 proteins involving gene duplication is proposed. The model proposes that HSP70 from archaebacteria (*H. marismortui* and *M. mazei*) and the gram-positive group of bacteria constitutes the ancestral form of the protein and that all other HSP70s (viz., other eubacteria as well as eukaryotes) containing the insert have evolved from this ancient protein.

All bacterial and eukaryotic species studied to date exhibit increased synthesis of a set of proteins referred to as stress or heat shock proteins (HSPs) in response to a sudden increase in physiological temperature as well as exposure to other stressors (e.g., hypoxia, amino acid analogs, ethanol, etc.) (27, 34). One of the proteins whose synthesis is greatly induced under these conditions has an apparent molecular mass of 70 kDa (HSP70; bacterial homolog known as the DnaK protein). Although the synthesis of HSP70 is greatly enhanced by various physiological stressors, it also constitutes a major protein under normal growth conditions and has been shown to be essential for cellular growth. Gene cloning and sequencing studies on HSP70 show that the primary structure of this protein has been highly conserved during evolution in species ranging from bacteria to plants to humans (27, 34).

In recent years, although HSP70 homologs have been cloned from numerous bacterial and eukaryotic species (1, 2, 4, 6, 12, 14, 18, 21, 32, 37, 38, 41, 42), none of the gene or protein structures of HSP70s from any archaebacteria, which have been proposed to constitute the third primary lineage distinct from eubacteria and eukaryotes (44–47), has yet been determined. In the present paper, we describe the cloning and complete nucleotide sequence of the HSP70 gene from *Halobacterium marismortui*, which belongs to the extreme halophile group of archaebacteria (44, 47). The cloning strategy is based on the polymerase chain reaction (PCR) employing degenerate oligonucleotide primers for conserved regions of HSP70 (12). We have carried out detailed analyses of HSP70 sequences from different species which provide evidence of gene duplication in the evolution of the HSP70 family of proteins. These studies also reveal that HSP70 from *H. marismortui* contains many structural features in common with the gram-positive group of bacteria, indicating its close evolutionary relationship to this group.

MATERIALS AND METHODS

Bacterial strains. *H. marismortui* ATCC 43049 was purchased from the American Type Culture Collection, Rockville, Md. A culture of this strain was also kindly provided by P. P. Dennis, University of British Columbia, Vancouver, Canada. The cells were grown at 37°C in halobacterium high-salt starch medium (medium 1218; American Type Culture Collection), as recommended by the supplier. High-molecular-weight DNA from *H. marismortui* cells was prepared by the method of Mevarech et al. (33).

Heat shock response. To determine the effect of heat shock on protein synthesis, an exponentially growing culture of *H. marismortui* was divided into several portions. Individual

* Corresponding author.

portions were transferred to different temperatures (50 or 55°C), while a control culture was maintained at 37°C. After 30 min at elevated temperatures, [^{35}S]methionine (50 μCi/ml; specific activity, 1,200 Ci/mmol) was added to different cultures. After 30 min of labeling, the cells were centrifuged and washed with cold, unlabeled medium, and the cell pellets were dissolved in sodium dodecyl sulfate (SDS)-lysis buffer (62.5 mM Tris-HCl [pH 7.5], 1% SDS, 1% β-mercaptoethanol, 10% glycerol). The proteins in different samples were electrophoresed on SDS–10% polyacrylamide gels, which were stained and fluorographed as described previously (1).

PCR. Oligonucleotide primers with opposite orientations were custom synthesized for two conserved regions of the HSP70 family of proteins (12). The primers were synthesized at the Central Facility of the Institute of Molecular Biology, McMaster University, Hamilton, Ontario, Canada. The forward (5′-CARGCNACNAARGAYGCNGG-3′) and the reverse (5′-GCNACNGCYTCRTCNGGRTT-3′) primers (where N = A, C, G, or T; Y = C or T; and R = A or G) were made for the sequences QATKDAG and NPDEAVA, respectively. The primers were degenerate to allow for all possible codon usages (12). Both of these primers have similar degrees of degeneracy (512-fold).

PCR amplification using *H. marismortui* DNA and the above sets of primers was carried out in a similar manner to that described elsewhere (12). After 30 cycles of amplification, an aliquot of the reaction mixture was analyzed on a 1% agarose gel to visualize and photograph the amplified product(s). The DNA from the amplified fragment was eluted by using a GENECLEAN kit (Bio 101, Inc., La Jolla, Calif.), and after filling in the ends of the fragment with Klenow, it was ligated to *Sma*I-digested and dephosphorylated pGEM-7zf(+) vector (Promega). After transformation of *Escherichia coli* JM109 cells with the plasmid, DNAs from a number of the clones containing the expected size of insert were sequenced by using forward- and reverse-sequencing primers. The inserts whose sequences resembled a consensus HSP70 gene sequence in the amplified region were used as probes in DNA hybridization and colony screening studies.

Screening for genomic clones. *H. marismortui* DNA was digested with appropriate restriction enzymes and run on 1% agarose gels. The gel region corresponding to the size range of interest (based on Southern blot analysis) was excised, gene cleaned, and then ligated in pGEM-7zf(+) vector digested with the same enzyme and dephosphorylated. The ligated vector was used to transform *E. coli* JM109 cells, and the colonies obtained were screened with the *H. marismortui* HSP70 probe. The DNA was sequenced by the dideoxy chain termination method with a Sequenase kit (United States Biologicals, Inc.).

Data base searches and sequence comparison analyses. The computer searches of various protein data bases were performed by using the FASTA program of the National Biomedical Research Foundation Protein Identification Resource in conjunction with the University of Wisconsin Genetics Computer Group (GCG) program package (28, 35). These programs were accessed on the CAN/SND Molecular Biology Database System of the National Research Council of Canada. The FASTA analysis was carried out with HSP70 sequences, and the top 100 scores were examined for the length and quality of sequence overlap. The statistical significance of similarity between any two sequences was evaluated by using the RDF2 program in the FASTA package. This program evaluates the sequences in pairs to determine whether the observed sequence similarity is due to common ancestory or simply to the locally biased amino acid composition (35). This program compares two sequences, calculating the initial and optimized score, and then shuffles the second sequence a specified number of times (keeping the amino acid composition of the shuffled sequence identical to the unshuffled sequence) and again calculates the initial and optimized scores. The statistical significance of the observed similarity could be assessed from two different perspectives (35). One of these is provided by the z value, which is calculated by subtracting the mean score of the randomly shuffled sequences from the score of the unshuffled sequence and then dividing it by the standard deviation of the distribution of the shuffled sequence. Pearson and coworkers (28, 35) have suggested that when z values are <3, one should be skeptical of a conclusion based on sequence similarity. However, z values >6 generally indicate highly significant similarity, pointing to common ancestory (28, 35). A second perspective to evaluate the significance of observed similarity is based on the highest score of the shuffled sequences. If the highest score of the shuffled sequences (about 100 shuffles) is lower than that of the original unshuffled sequence, then it is again strongly indicative of the significance of the observed sequence similarity (35). The pairwise sequence alignment between various proteins was carried out by using the BESTFIT program of the GCG6 software package. The phylogenetic tree for HSP70 sequences was constructed by using the CLUSTAL program of the PC Gene Software package (Intelligenetics). The program initially calculates pairwise similarity scores by the method of Wilbur and Lipman (43). In the next step, these scores are used to construct a phylogenetic tree by using UPGMA (unweighted pair group maximum averages) or the average linkage cluster analysis method (39). In the final stage, sequences are aligned by using the program PALIGN, minimizing the distances (i.e., gaps) between various sequences.

Nucleotide sequence accession number. Sequence data for *H. marismortui* HSP70 have been deposited in the GenBank data base under accession no. M84006.

RESULTS

Heat shock response of *H. marismortui*. We initially examined the nature of proteins synthesized in *H. marismortui* upon heat shock. In these experiments, *H. marismortui* cells growing at 37°C were shifted to elevated temperatures (50 and 55°C), and the nature of the synthesized proteins was determined by labeling the cells with [^{35}S]methionine. As seen in Fig. 1, while synthesis of the majority of the proteins was only slightly increased upon shifting the cells to either 50 or 55°C (because of the temperature effect), a few proteins with approximate masses of 100, 90, 70, and 60 kDa were synthesized in much larger amounts. In addition to these proteins, the synthesis of a few other proteins of ~55 and 45 kDa also appears to be enhanced upon heat shock. The observed response is very similar to that reported previously by Daniels et al. (7) for *H. marismortui* and *H. trapanicum*, but it differs somewhat from the response of *Methanococcus voltae*, where, upon heat shock, a number of low molecular weight HSPs were also induced (20).

Cloning of the *H. marismortui* HSP70 gene. To clone the HSP70 gene, initially the PCR reaction was carried out by using *H. marismortui* DNA and a set of degenerate oligonucleotide primers made for HSP70 sequences which are conserved in all eukaryotic and prokaryotic species (12). In

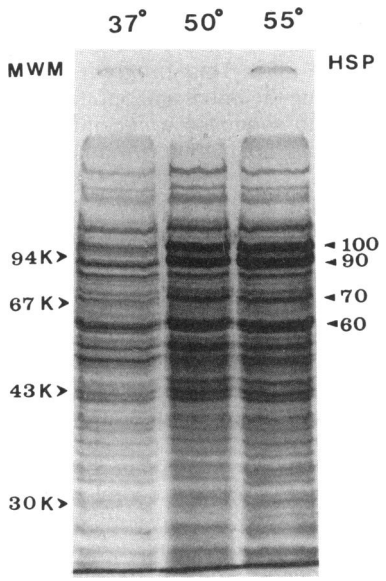

FIG. 1. Effect of elevated temperature on protein synthesis in *H. marismortui* culture. Portions of the culture growing at 37°C were shifted to either 50 or 55°C, and after 30 min the cells at different temperatures were labeled with [^{35}S]methionine for 30 min. Total cellular proteins from different cultures were analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography. Arrowheads on the right indicate the approximate M_r (×1,000) of the main proteins whose synthesis was induced upon heat shock. The positions of molecular weight markers (MW (*M*; ×1,000) are indicated on the left.

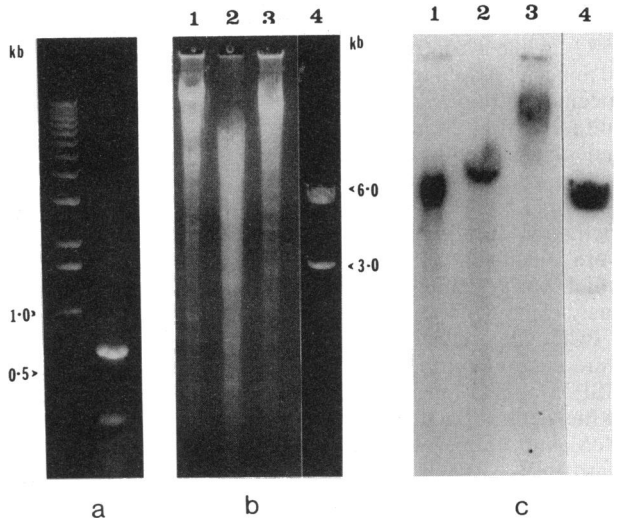

FIG. 2. Cloning of the *H. marismortui* HSP70 gene. (a) Agarose gel electrophoresis of the PCR reaction product of *H. marismortui* DNA and degenerate oligonucleotide primers for conserved regions of HSP70. The left lane contains molecular size markers. (b and c) Southern blot analyses of *H. marismortui* genomic DNA and that of the positive clone. Lanes 1 to 3, *H. marismortui* DNA digested to completion with the enzymes *Hin*dIII, *Cla*I, and *Bam*HI, respectively; lane 4, *Hin*dIII digestion pattern of the positive clone Archm P-2. The lower band in lane 4 corresponds to the plasmid. (b) Ethidium bromide staining of the gel. (c) Hybridization of the blot to the 0.65-kb PCR probe.

these experiments, strong amplification of a 0.65-kb fragment, which is the expected size based on the positions of the primers in the HSP70 sequences, was observed (Fig. 2a). The amplified PCR fragment was subcloned in a plasmid vector, and its complete nucleotide sequence was determined. The deduced amino acid sequence of the cloned fragment was unique, but it showed a high degree of homology to the HSP70 family of proteins (results not shown), indicating that it corresponded to HSP70. To isolate a genomic clone from *H. marismortui*, the DNA was digested with several restriction enzymes (*Hin*dIII, *Cla*I, and *Bam*HI), blotted, and probed with the 0.65-kb cloned PCR fragment. In Southern blots of *H. marismortui* DNA digested with different restriction enzymes, the cloned fragment showed specific hybridization to fragments in the range of about 6 to 12 kb (Fig. 2b and c). For *Hin*dIII-digested DNA, strong and specific hybridization to a band of about 6 kb was observed. To clone the *H. marismortui* HSP70 gene, the 5.5- to 6.5-kb region from *Hin*dIII-digested DNA was excised and subcloned in the plasmid pGEM-7zf(+). Screening of the resulting library with the cloned probe identified several positive clones, each containing an approximately 6-kb insert (lane 4 of Fig. 2b and c) and showing a similar enzyme digestion pattern. Restriction enzyme digestion and Southern blot analysis of the insert indicated that the sequence hybridizing to the PCR probe was contained within an approximately 2.8-kb *Xho*I-*Nsi*I fragment (not shown). To sequence the insert, nested sets of deletions using exonuclease III in both orientations were made and the nucleotide sequences of both DNA strands were determined. The nucleotide sequence of a portion of this fragment is shown in Fig. 3. It contained an open reading frame of 1,996 bp encoding a protein of 68,891 Da (calculated). The probe sequence matched exactly with nucleotides 388 to 1,029 in the above sequence. Codon usage for the encoded protein showed a strong bias towards codons with either G or C in the third position, reflecting the high G + C content (~62%) of the sequenced fragment and of the halobacterial genome (40). In addition, similar to other halobacterial proteins (26), many additional acidic residues (at the C-terminal end) were present in the *H. marismortui* HSP70 sequence (Fig. 3).

Sequence comparisons and analyses. Alignment of the deduced amino acid sequence of the *H. marismortui* protein with known prokaryotic and representative eukaryotic HSP70 sequences (Fig. 4) showed extensive similarity throughout its length, confirming its identity as an HSP70 homolog. Recently (after original submission of this paper), the sequence of HSP70 from the archaebacterium *Methanosarcina mazei* was also reported (30), and it is included in the comparison shown in Fig. 4. Pairwise alignment of the HSP70 sequences using the BESTFIT program of the GCG package revealed that *H. marismortui* HSP70 showed between 51 and 58% amino acid identity over its entire length with the archaebacterial and eubacterial homologs and between 44 and 47% identity with the eukaryotic counterparts (Table 1). Additionally, between 15 and 20% amino acid residues were found to be conservative replacements in various HSP70 pairs (Table 1). The observed high degree of sequence similarity in HSP70 sequences over their entire lengths, in species covering all three primary kingdoms (or domains), indicates that HSP70 constitutes one of the most conserved proteins known to date.

From the sequence alignment shown in Fig. 4, as well as an examination of other HSP70 sequences available in various nucleic acid and protein sequence data bases (results not shown), a number of unique and distinguishing features

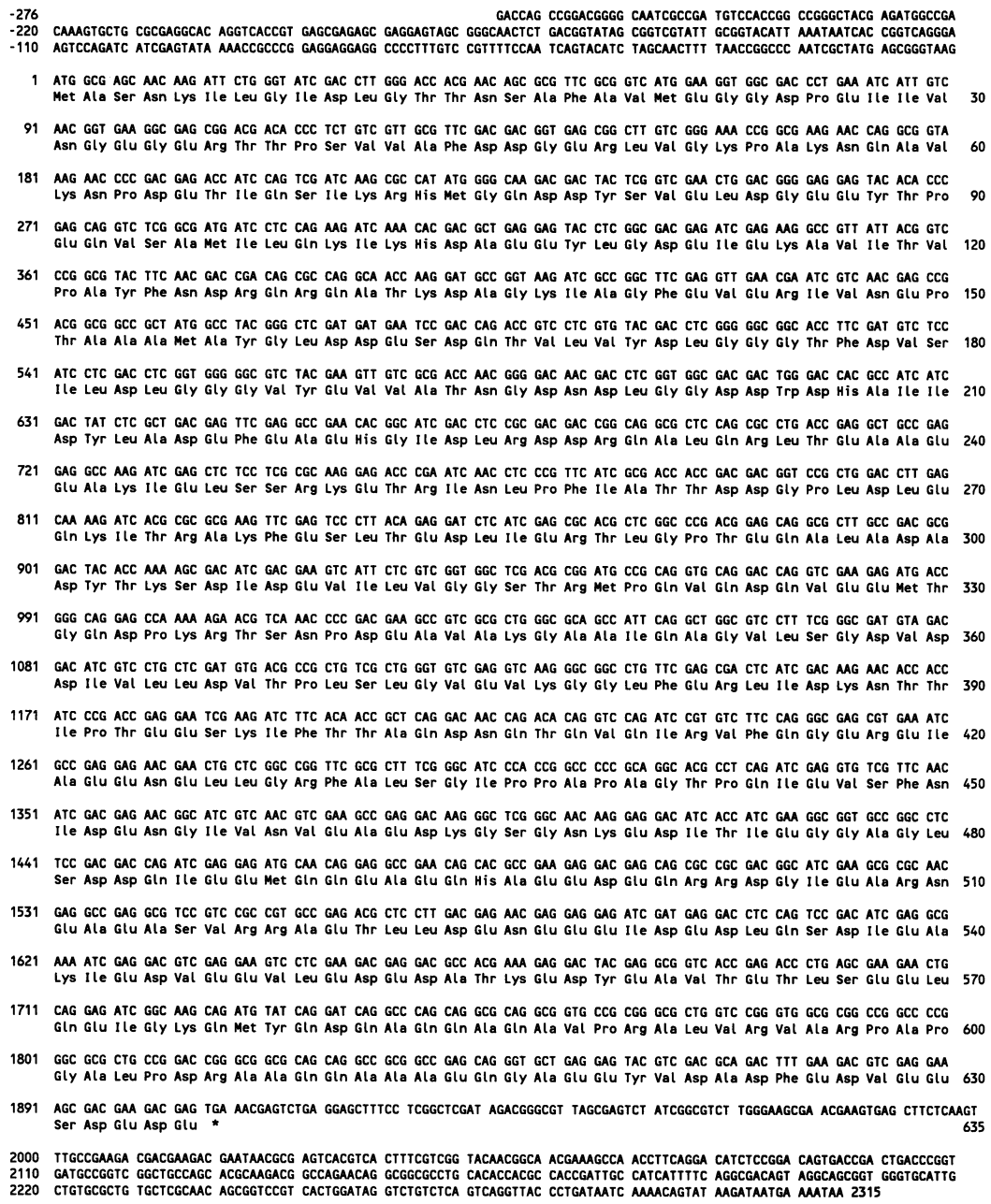

FIG. 3. Nucleotide sequence of the *H. marismortui* HSP70 gene and of the flanking region. The deduced amino acid sequence of the open reading frame corresponding to HSP70 is shown underneath. The sequence of the 0.65-kb PCR probe matches exactly with nucleotides 388 to 1,029 in this sequence.

of the HSP70 family could be identified. These include (i) a large gap or deletion of about 25 to 27 amino acids in the N-terminal quadrant of HSP70 from the two archaebacteria as well as from all gram-positive bacteria but not from any of the other species, (ii) a deletion of five amino acids near the middle, which is present in all eukaryotic HSP70 sequences but is not seen in any of the eubacterial homologs, (iii) an insertion of three to five amino acids near the middle, which is specific for the eukaryotic sequences, and (iv) the insertion of an arginine (R) in the C-terminal quadrant of all eukaryotic HSP70 sequences. In view of their specificity, the last three sequence signatures (Fig. 4, stars) could be used to distinguish between various eukaryotes and eubacteria. As can be seen in Fig. 4, HSP70 from *H. marismortui* and *M. mazei* contained all of the signature sequence characteristics of eubacteria, particularly the gram-positive bacteria, indicating their close relationship to this group. It is also noteworthy that, in contrast to the cytoplasmic HSP70s, the mitochondrial HSP70s from *Saccharomyces cerevisiae* (Fig. 4, row i) and other organisms (results not shown) showed the characteristics of eubacterial HSP70s, supporting the endosymbiotic origin of mitochondria from eubacteria (17).

We also examined HSP70 sequences for possible internal repeats. The results of these studies indicated that the amino

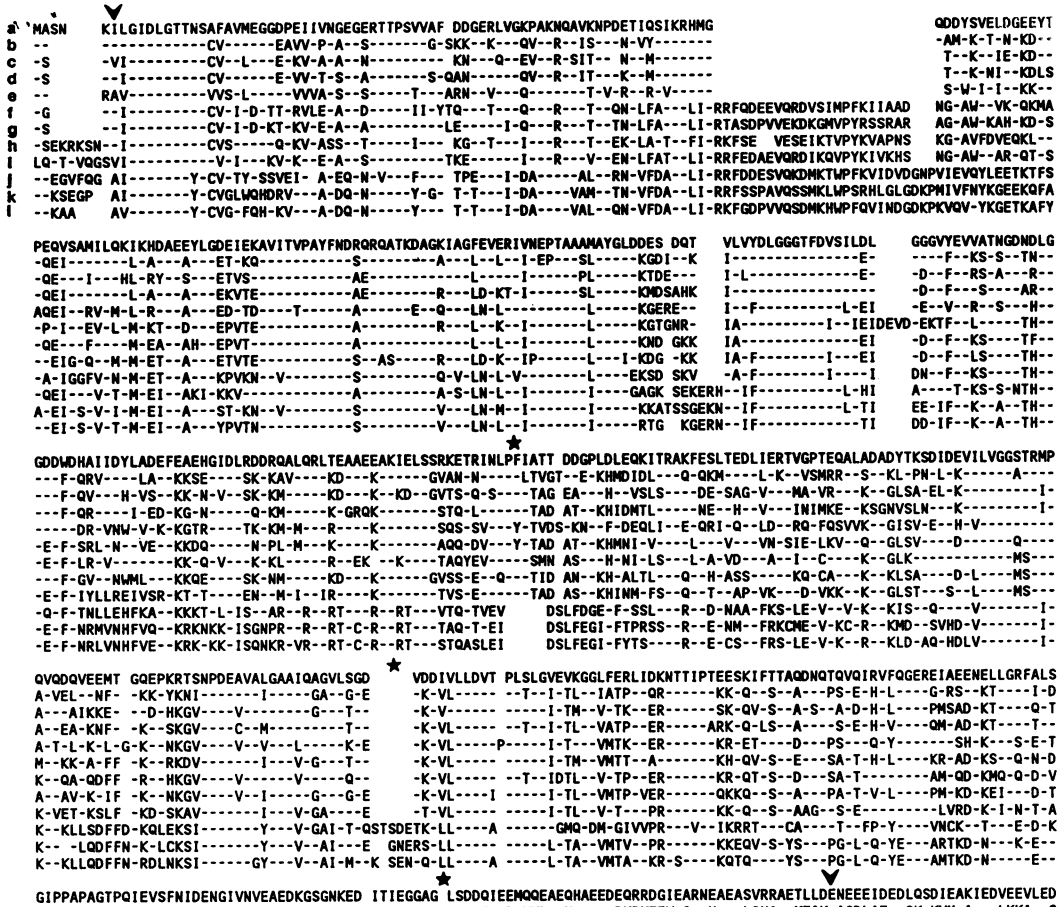

FIG. 4. Alignment of HSP70 sequences from different species. Rows: a, *H. marismortui*; b, *M. mazei* S6 (30); c, *B. subtilis* (18); d, *C. perfringens* (12); e, *M. leprae* (32); f, *E. coli* (2); g, *C. crescentus* (14); h, *Chlamydia trachomatis* (4); i, yeast mitochondrial SSC1 (6); j, yeast cytoplasmic SSB1 (38); k, maize (37); l, human (21). Sequence alignment is based on PC Gene Software alignment and BESTFIT analyses of the sequences. Residues identical to the *H. marismortui* sequence are denoted by dashes. Stars show sequence features which distinguish various eukaryotes from prokaryotes. The sequence in between the arrowheads marks the conserved region which could be aligned for various HSP70 sequences.

acids in the first quadrant of HSP70 (amino acids 1 to 160) showed significant similarity to those in the next quadrant (amino acids 161 to 320) (Table 2). The similarity between these two segments was seen most clearly in the case of gram-positive bacteria and archaebacteria which contained the large deletions in the N-terminal quadrant. An alignment of the amino acids in the first and second quadrant of HSP from *Bacillus subtilis* is shown in Fig. 5a; the two segments could be aligned with only a few gaps. Of the 139-amino-acid overlap, 33 residues (23.7%) were found to be identical and an additional 33 residues (23.7%) were found to be conservative replacements. In one stretch of eight amino acids, seven were found to be identical. The similarity between the two segments could be even higher if additional gaps in these sequences are introduced. The significance of the observed similarity between the first and second quadrants was eval-

TABLE 1. Similarity matrix of HSP70 sequences

Source of sequence	% Identity or similarity with HSP70 sequence from[a]:											
	A	B	C	D	E	F	G	H	I	J	K	L
(A) *H. marismortui*		58.0	57.7	54.1	55.9	52.3	56.9	51.5	50.5	47.6	44.3	46.3
(B) *M. mazei*	73.0		65.6	59.3	65.4	58.3	61.5	60.4	55.2	48.5	49.3	49.0
(C) *B. subtilis*	74.9	79.8		58.3	67.1	58.1	63.3	60.0	56.2	48.6	48.2	48.7
(D) *M. leprae*	72.3	76.5	75.2		58.4	56.6	57.6	56.8	52.5	47.7	48.3	49.3
(E) *C. perfringens*	73.0	78.0	79.6	74.4		59.7	64.5	59.4	56.4	46.0	46.3	47.2
(F) *E. coli*	71.7	74.0	75.5	72.1	76.0		66.5	59.8	59.6	48.0	47.7	49.4
(G) *C. crescentus*	72.9	77.2	77.7	73.7	76.7	79.7		59.7	60.5	48.7	50.7	50.8
(H) *Chlamydia trachomatis*	68.2	75.9	75.0	74.9	74.1	74.6	74.9		57.1	47.0	47.6	48.8
(I) *S. cerevisiae* (mitochondria)	66.6	72.3	74.0	69.7	72.2	75.0	76.4	72.1		49.2	48.3	46.9
(J) *S. cerevisiae* (SSB1)	65.9	66.1	67.3	66.1	66.0	65.9	65.9	66.1	67.1		57.9	59.8
(K) Maize	65.7	68.1	68.5	67.6	66.3	66.1	68.9	66.0	67.2	72.0		75.2
(L) Human	66.8	67.3	68.8	69.2	67.9	66.6	68.9	66.9	65.7	74.4	85.0	

[a] Sequence alignment was carried out by using the BESTFIT program of the GCG6 package. Upper and lower triangles indicate the percent identity and percent similarity, respectively, between pairs of sequences.

uated by RDF2 analysis, and it was found to be significant ($P < 0.001$) (Table 2).

Sequence similarity between HSP70 and other proteins. We have also examined whether the HSP70 family of proteins shows significant similarity to any other proteins in the data base. To do this, initially a FASTP analysis of proteins in various data bases (National Biomedical Research Foundation, EMBL, and SWISS-Protein) using HSP70 sequences was carried out. The statistical significance of the similarity between HSP70 and the top 100 sequences thus identified (excluding the known HSP70 sequences) was evaluated individually by RDF2 analysis (Table 2). Such analysis revealed that a protein, MreB of E. coli (M_r 36,958; 347 amino acids), which is involved in cell division and the formation of the rod-shaped structure of the cells (9), showed highly significant similarity to the HSP70 family of proteins. The optimized alignment score of various HSP70s was between 6.7 and 14.2 standard deviations away from the mean alignment score of randomly shuffled MreB sequences. This large difference between the unshuffled and randomly shuffled sequences strongly suggests that the observed homology between MreB and the HSP70 family of proteins is highly significant. The significance of the observed homology between HSP70 and MreB is also evident from the fact that the alignment score of the unshuffled sequence in all cases was much higher than the maximum score of the shuffled MreB sequence.

An alignment of the HSP70 sequence from *H. marismortui* with the MreB protein of *E. coli* is shown in Fig. 5b; the two proteins show considerable homology throughout their length, with several long stretches of complete identity. The homology of MreB to HSP70 was observed from the beginning of the HSP70 sequence to nearly half its length, with only a small number of gaps. In an overlap of 328 amino acids, 91 identical and 77 conservative replacements were observed, thereby giving an overall similarity of about 51.2%. The sequence similarity between MreB and other HSP70s was comparable to the above percentage (Table 2). However, when alignment was carried out with other HSP70 sequences which do not contain the deletion in the N-terminal quadrant, then a large gap in the MreB sequence in the

TABLE 2. Statistical significance of sequence similarities

Protein sequences	Optimized score[a]			
	Unshuffled	Mean ± SD	Maximum	*z*
HSP70-I (*B. subtilis*) × HSP70-II (*B. subtilis*)	54	30.6 ± 7.0	52	3.3
HSP70-I × thioredoxin (spinach chloroplast)	65	27.5 ± 5.6	45	6.7
HSP70 (*B. subtilis*) × thioredoxin	64	30.6 ± 6.4	48	5.2
HSP70 (*H. marismortui*) × MreB (*E. coli*)	154	44.4 ± 9.6	73	11.4
HSP70 (*B. megaterium*) × MreB	173	44.9 ± 10.3	74	12.4
HSP70 (*C. perfringens*) × MreB	149	44.8 ± 9.8	79	10.6
HSP70 (*E. coli*) × MreB	109	43.5 ± 9.7	71	6.7
HSC70 (Chinese hamster) × MreB	147	39.9 ± 7.5	60	14.2
HSP70 (human) × MreB	158	41.6 ± 10.2	85	11.4
GRP78 (human) × MreB	155	44.3 ± 10.3	74	10.8
SSC1P (*S. cerevisiae* mitochondria) × MreB	133	46.0 ± 11.5	84	7.6
HSP70 (*Trypanosoma bruceii* mitochondria) × MreB	126	40.6 ± 9.4	75	9.1
HSP70 (*B. subtilis*) × actin (bovine)	32	38.5 ± 8.9	73	−0.74
HSP70 (Chinese hamster) × actin	32	39.5 ± 10.0	71	−0.75

[a] The statistical significance of similarity between any two sequences was evaluated by using the RDF2 program (see Materials and Methods) (35). This program compares two sequences and calculates an unshuffled optimized similarity score. The second sequences in each case were shuffled 100 times, and for each shuffled sequence a similarity score was calculated with the sequence. The maximum and average scores of these randomly shuffled sequences are indicated. The *z* values were calculated by subtracting the mean score of randomly shuffled sequences from the score of the unshuffled sequence and then dividing by the standard deviation of the distribution of shuffled scores. HSP70-I and HSP70-II refer to the first (amino acids 1 to 160) and second (amino acids 161 to 320) quadrants of HSP70 sequences.

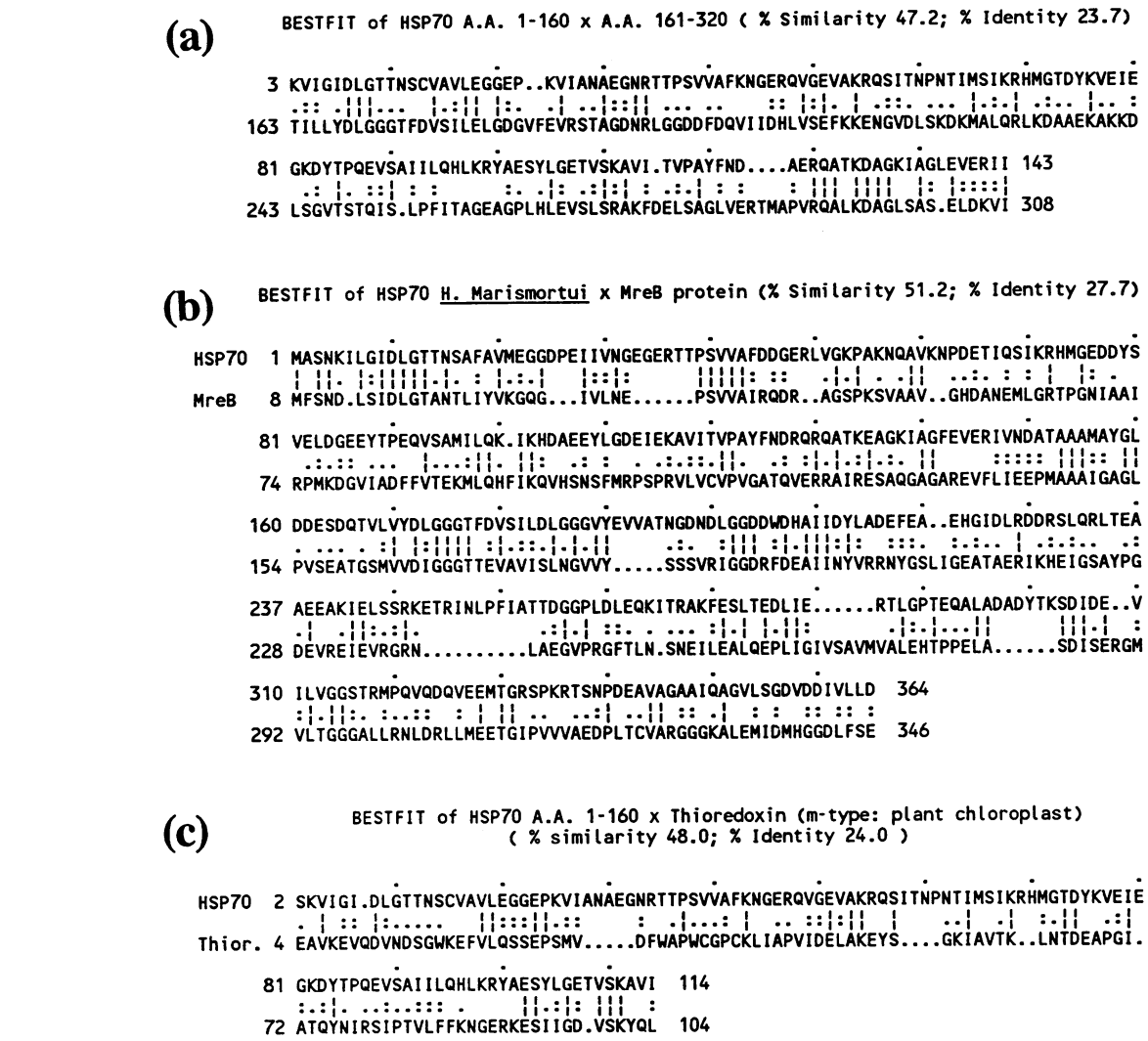

FIG. 5. (a) Sequence similarity between the first and second quadrants of the HSP70 sequence of *B. subtilis*; (b) amino acid sequence alignment of HSP70 from *H. marismortui* (top lines) and the MreB protein (bottom lines) of *E. coli*; (c) alignment of the N-terminal sequence of HSP70 from *B. subtilis* (top line) with the sequence of m-type thioredoxin from spinach chloroplast (bottom lines). Sequence similarities were computed by the BESTFIT program of the GCG6 software package. Symbols: vertical line, identical residues; colon, conservative amino acid replacements; dot, evolutionarily related substitutions (8). Dots within sequences denote gaps.

corresponding position was required for proper alignment of the sequences (results not shown).

In addition to MreB, a somewhat lower but significant homology of HSP70 with the m-type thioredoxin from spinach chloroplast was also noted (31). Interestingly, this latter protein, which has only 104 amino acid residues, showed maximum similarity to the first quadrant of the HSP70 sequence from gram-positive bacteria. An alignment of these two sequences is shown in Fig. 5c; the N-terminal quadrant of HSP70 could be aligned with the thioredoxin sequence with only a small number of gaps. In an overlap of 104 amino acids, 24% identical and an additional 24% conservative amino acid substitutions were observed. In contrast to the MreB protein and thioredoxin, no significant similarity was observed between the HSP70s and bovine actin sequences (Table 2), for which similarity in the three-dimensional structures has been reported (11).

DISCUSSION

This paper reports the cloning and sequencing of the HSP70 gene from *H. marismortui*, an extremely halophilic group of bacteria constituting one of the subdivisions of archaebacteria (44, 46, 47). After submission of this manuscript, the sequence of HSP70 from a methanogenic archaebacterium, *M. mazei*, was also reported (30). These reports provide definitive identification and characterization of HSP70 homologs in species belonging to the archaebacterium domain, and they establish the ubiquity of both HSP70 as well as the heat shock response in all forms of life.

Our analysis of various HSP70 sequences that are currently available has revealed several novel features of this protein family that are of much interest. We have observed that HSP70s from *H. marismortui*, *M. mazei*, and various gram-positive groups of bacteria (*B. subtilis*, *Bacillus mega-*

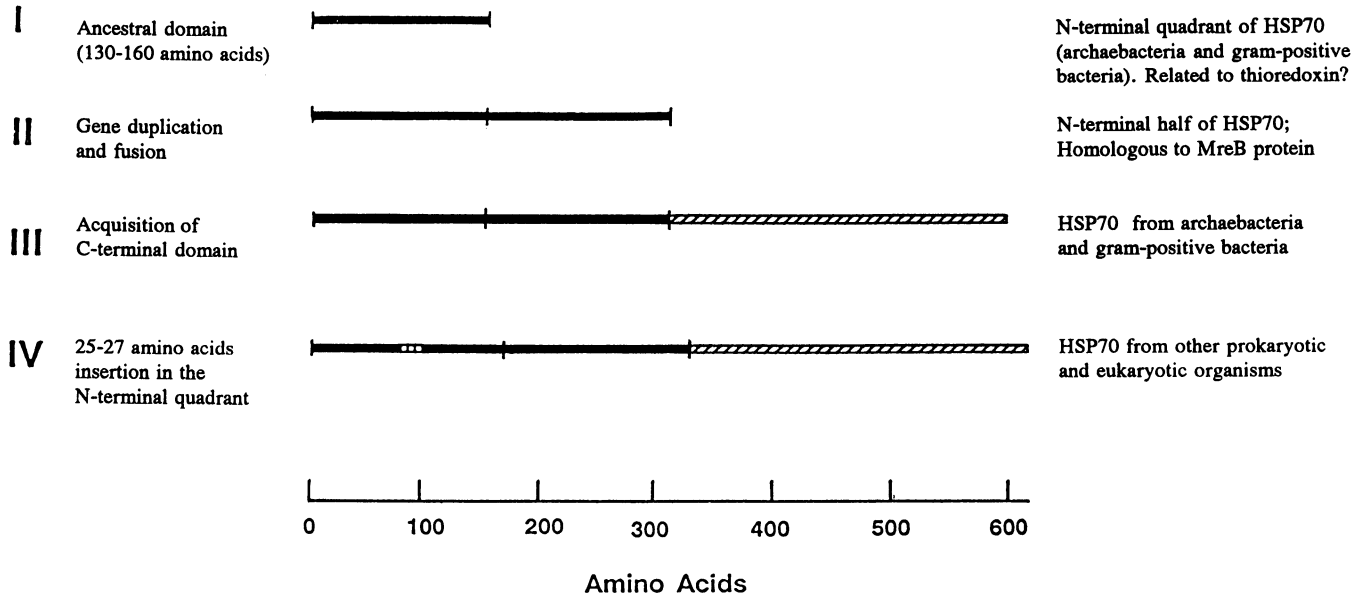

FIG. 6. Hypothetical model for the evolution of HSP70. The proposed stages in the evolution are indicated on the left, and some of the proteins showing sequence similarity to the various stages are noted on the right.

terium, *Clostridium perfringens*, *Mycobacterium leprae*, and *Mycobacterium paratuberculosis*) lack a stretch of 25 to 27 amino acids in the N-terminal quadrant which is present in HSP70s from all other prokaryotic and eukaryotic species. The significance of this observation is discussed below. In addition, significant sequence similarity between the first and second quadrants of HSP70s (i.e., amino acids 1 to 160 and 161 to 320) from two archaebacteria and various gram-positive bacteria has been observed. This similarity between the first and second quadrants of HSP70 was not readily apparent in other HSP70s, which do not contain the 25 to 27 amino acid gap in the N-terminal quadrant. These observations fit well with the recently reported three-dimensional crystal structure of the N-terminal fragment of bovine HSP70 (386 amino acids). The crystal structure of this fragment shows the presence of two distinct lobes of approximately equal sizes and similar tertiary structures with a deep cleft in between (10). The 25 amino acids (from 82 to 106) corresponding to the gap are located on the outside of lobe I and appear as an appendage (10). Most importantly, the boundaries of the two lobes as determined from X-ray crystal structure data (lobe I, amino acids 1 to 188 [minus a 25-amino-acid deletion which falls in this region]; lobe II, amino acids 189 to the end [386] [minus a 25-amino-acid deletion]) coincides very well with the boundaries of the two segments as deduced from sequence alignment and overlapping. The observed sequence and structural similarities between the first and second quadrants of HSP70 indicate that these two segments arose by duplication of an ancestral domain and suggests that the 25 to 27 amino acids corresponding to the gap were probably inserted at a later stage. The observed significant sequence similarity between the first quadrant of HSP70 and the m-type thioredoxin from spinach chloroplast (Table 2; Fig. 5c) suggests that these proteins probably evolved from a common ancient gene or domain.

Sequence comparison studies have identified yet another very compelling and highly significant similarity between the HSP70 family of proteins and a protein, MreB of *E. coli* (9) (Table 2). The latter protein, which consists of 347 amino acids, shows >50% similarity (identical plus conserved residues) with approximately the first half of the HSP70 protein. Statistical analysis of the HSP70 and MreB sequences shows that the observed similarity between these proteins is highly significant and that they almost certainly are derived from a common ancestral protein (Table 1; see Materials and Methods). The homology of MreB to only the N-terminal half of HSP70 further suggests that this protein evolved from a predecessor of HSP70 before the C-terminal fragment was acquired. One observation that is of considerable interest is that, while the MreB protein could be readily aligned with HSP70s from the two archaebacteria as well as various gram-positive groups of bacteria, its alignment with other HSP70s (from other eubacteria and eukaryotes) requires the introduction of a gap of 25 to 27 amino acids in the N-terminal region, corresponding to the position of the deletion or insertion noted above. Since MreB is postulated to have diverged from an ancient progenitor of the HSP70 family of proteins, the absence of the 25- to 27-amino-acid insertion strongly suggests that this insertion was not present in the ancient HSP70 protein. This observation, in conjunction with the alignment results for the first and second quadrants of HSP70 (discussed above), strongly suggests that the HSP70s from archaebacteria and gram-positive groups of bacteria which lack the N-terminal insertion represent a more ancient version of this protein.

On the basis of the above analyses and the structural features of HSP70, a tentative model for the evolution of the HSP70 family of proteins can be proposed (Fig. 6). It is suggested that the evolution of the ancestral HSP70 began with an ancient 130- to 160-amino-acid domain corresponding to its N-terminal quadrant and probably related to the m-type thioredoxin (31) (Fig. 6, stage I). Duplication and fusion of the gene for this domain, followed by divergent evolution of the two domains, resulted in another ancestral gene from which the MreB gene or protein evolved (stage II). It is proposed that at a later time, a C-terminal domain was acquired by this ancient gene or protein to give rise to

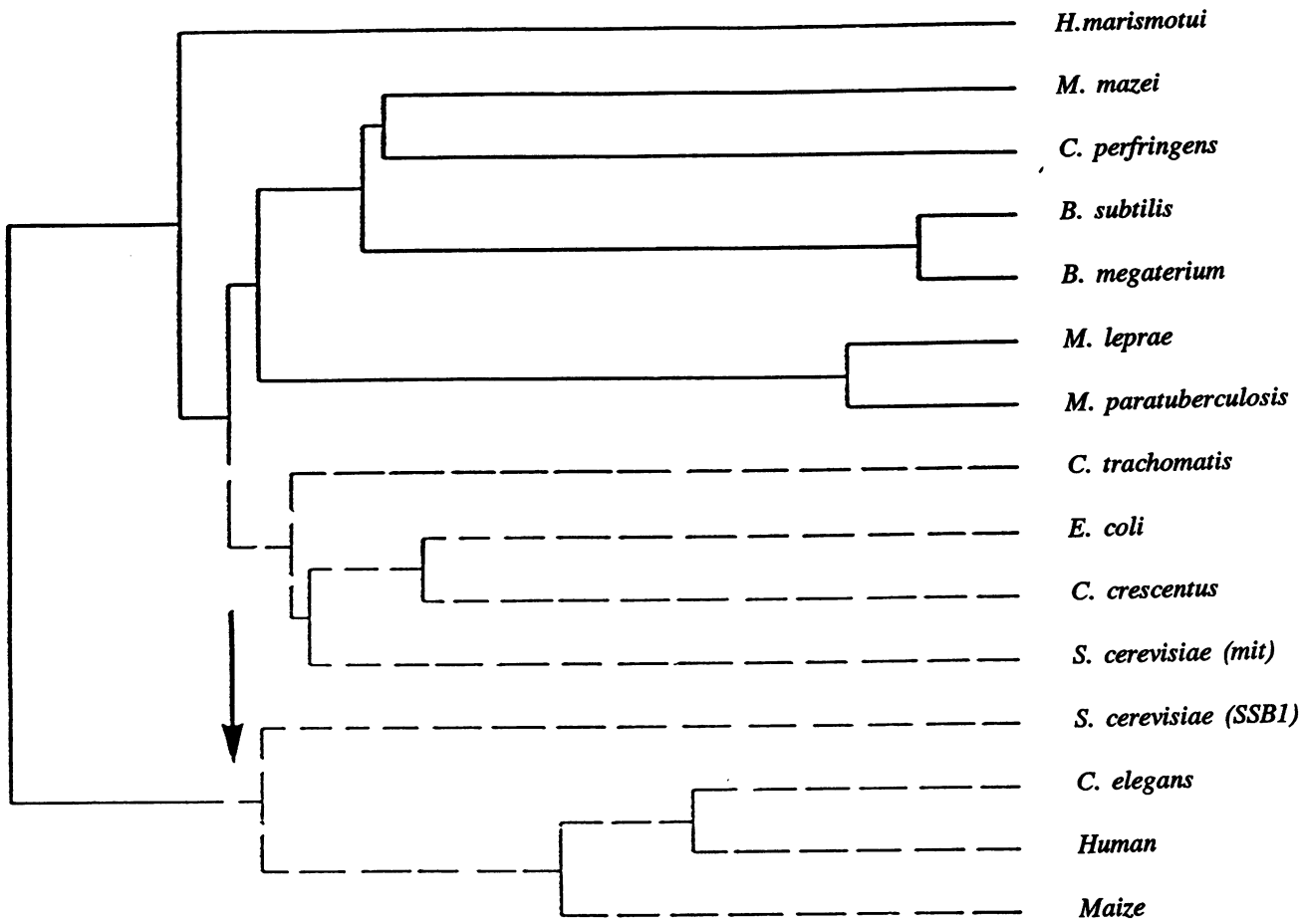

FIG. 7. Dendogram based on HSP70 sequence data showing the relationship of the archaebacteria *H. marismortui* and *M. mazei* to other eukaryotes and prokaryotes. The tree was constructed by the CLUSTAL program of PC Gene Software. Dotted lines represent organisms which contain the insert in the N-terminal quadrant. The arrow indicates the interspecies transfer of the insert (or gene) from a prokaryotic ancestral cell to eukaryotes. *C. trachomatis*, *Chlamydia trachomatis*; *C. elegans*, *Caenorhabditis elegans*.

the ancestral HSP70 version, represented by those present in archaebacteria and gram-positive bacteria (stage III). Since the above sequence features of HSP70 are highly conserved and present in all organisms examined (representing the three primary domains), it is expected that these evolutionary events took place in the universal ancestor before the divergence of eubacteria, archaebacteria, and eukaryotes. At a later time (stage IV), it is postulated that a gene fragment corresponding to the insertion in the N-terminal quadrant got inserted in this ancient gene to give rise to other HSP70s found in other eubacteria and eukaryotes.

The sequence data on the HSP70 family of proteins, because of their ubiquitous presence and very high degree of sequence conservation, also provide a useful resource and model system for investigating the deep phylogenetic relationships, such as those among archaebacteria, eubacteria, and eukaryotes. Figure 7 shows a dendogram based on all known prokaryotic (archaebacterial and eubacterial) HSP70 sequences and a few representative eukaryotic sequences. There are several points of interest in this dendogram which are discussed below.

(i) As seen, the dendogram consists of an unrooted tree apparently with two main branches, one consisting of all eukaryotes and the second including archaebacteria as well as various eubacteria. The division of organisms into these two main groups is in accordance with the much higher degree of sequence similarity in species within one group compared with that between the two groups. The observed division of organisms into eukaryotic and prokaryotic clusters is also supported by our observation that all of the species within these two groups contain specific signature sequences that are unique to each group and which distinguish them from the other.

(ii) Of the two archaebacterial species examined, *H. marismortui* forms the deepest branch of the prokaryotic cluster, whereas *M. mazei* clusters with the gram-positive group of bacteria. It should be noted that, in contrast to the prokaryotic and eukaryotic species which could be readily distinguished from each other based upon the extent of sequence homology, no such distinction could be made between archaebacterial and eubacterial sequences. The two archaebacterial sequences examined also do not contain any unique signature sequence that distinguishes them from the gram-positive eubacterial group. These results are of considerable interest because previously at least three alternate phylogenics for the evolution of archaebacteria have been proposed. On the basis of rRNA sequence data, Woese and others have suggested that archaebacteria are monophyletic and constitute one of the three primary urkingdoms (or domains), which is quite distinct from both eubacterial and

eukaryotic domains (16, 43–47). In a second model proposed by Cavalier-Smith (5), a gram-positive ancestor of both eukaryotes and archaebacteria is suggested. Cavalier-Smith has noted several similarities between gram-positive bacteria and archaebacteria and presented arguments that archaebacteria are fundamentally prokaryotes and not a third type of organism (5), a view consistent with our results. However, the model of Cavalier-Smith proposes that gram-positive bacteria are derived from gram-negative bacteria, an inference not supported by the present results. In the third model, proposed by Lake (23–25), archaebacteria are considered to be polyphyletic. In this model, the halobacteria and methanogens are grouped along with eubacteria, whereas the extreme thermophiles (referred to as eocytes) are grouped with eukaryotes (24, 25). The HSP70 sequence data on the two archaebacterial species examined, a halobacterium and a methanogen, seem in accordance with this model. However, additional sequence data on other groups of archaebacteria (particularly the extreme thermophiles) are needed to confirm this inference.

(iii) One surprising observation from the sequence data as well as the dendogram is the similarity between the archaebacteria and gram-positive bacteria and the very deep branching of members of the latter group. It should be mentioned that the ancient nature and deep branching of certain species of gram-positive bacteria have been previously noted by Woese and coworkers (44, 45) and advocated by Cavalier-Smith (5). However, this fact has not received much attention in the past. The specific relationship of archaebacteria to gram-positive bacteria is also indicated by the fact that, similar to gram-positive bacteria, several halpohiles and methanogens show a gram-positive reaction and contain a thick and homogeneous cell wall characteristic of the gram-positive group of bacteria (3, 40). Additionally, several enzymes from archaebacteria (e.g., citrate synthase, malate dehydrogenase, pyruvate dehydrogenase, succinate thiokinase, etc.) show similar structures and biochemical properties, as seen for the corresponding enzymes from gram-positive bacteria (and often eukaryotes) but not from other eubacteria (5, 15, 29). In view of these observations, the relationship of the gram-positive bacteria to the archaebacteria needs to be further investigated with different model systems.

(iv) Within the archaebacterial and eubacterial cluster, the deepest branching divides the various organisms into two groups. One group consists of the two archaebacteria as well as various gram-positive bacteria, i.e., all the organisms which lack the 25- to 27-amino-acid insert in their N-terminal quadrants. The second group consists of all other eubacterial species containing the insertion, including chlamydiae (which form the deepest branch of this group), *E. coli*, *Caulobacter crescentus*, and the endosymbiont that gave rise to mitochondria. As seen in Fig. 7, on the basis of the presence or absence of this insert, the various prokaryotic organisms could be divided into two halophyletic branches (Fig. 7). Thus, the time at which this insertion took place seems to mark an important event in the evolution and divergence of prokaryotic organisms. It is of much interest to examine HSP70 sequences from other deep-rooted prokaryotic organisms such as other groups of archaebacteria, thermotogales, and green nonsulfur bacteria to see whether they conform to or support the observed division.

(v) As discussed previously and as shown in Fig. 6, our analyses of HSP70 sequences suggest that the HSP70s which lack the insert in the N-terminal quadrant constitute the ancient form of the protein and other HSP70 sequences containing the insert are derived from it. It should be noted that this insert is present at the same position in all eukaryotic organisms examined (including various animals, plants, yeast cells, drosophila, leishmania, trypanosomes, etc. [results not shown]) as well as all prokaryotes except for archaebacteria and gram-positive bacteria (Fig. 4). In addition, the length of this insert as well as its sequence is highly conserved in various species belonging to the two urkingdoms (Fig. 4). Thus, it seems unlikely that this insertion event took place completely independently in the two cases. To account for this observation, three possibilities could be considered. First, it is possible that a common or related event such as infection with a plasmid, transposable element, or virus containing the insert took place at the same time in both a progenitor of all eukaryotic cells as well as a member of the eubacterial lineage. Second, a related possibility is that the above event took place initially in one urkingdon (viz., the eubacterial) and then got transferred laterally to the other; in fact, the transfer of genetic information from eubacteria to eukaryotes is a known phenomenon (19). Third, since this insert is present in all eukaryotic species, it is possible that the ancestral eukaryotic cell evolved or originated from an ancient prokaryotic cell (eubacteria or archaebacteria) which contained this insertion. In view of the reported greater similarity of certain groups of archaebacteria to eukaryotic organisms (e.g., thermoacidophiles) (5, 13, 15, 22–27), it is of much interest to determine whether the above insert is present in their HSP70 sequences.

(vi) Lastly, the HSP70 sequence data also enable us to draw some tentative inferences regarding the root of the universal tree of life. As discussed above, the archaebacterial and eubacterial HSP70 sequences are much more closely related to each other than to the eukaryotic lineage. Our analyses also suggest that the HSP70 from archaebacteria and the gram-positive group of bacteria lacking the insertion constitutes a more ancient form of the protein. Accordingly, an HSP70 sequence lacking the insert (from *H. marismortui*) constitutes the deepest branch of the archaebacterial and eubacterial cluster, from which the derivation of other eubacterial sequences could be readily explained. Our results thus suggest that the archaebacteria and eubacteria are sister or related urkingdoms. If the above arguments and assumptions are correct, then the root of the universal tree should lie either within this group or in between the deepest branching member of the eukaryotic group and that of the archaebacterial and eubacterial cluster. This view of the universal tree of life is in accordance with that based on rRNA sequence data (44–46), but it is at variance with the phylogenetic relationship deduced from protein sequence data on duplicated gene families (13, 22, 36). We do not have any satisfactory explanation for the observed discrepancies at present. However, possible explanations for these could include differences in the evolutionary rates for specific proteins as well as among various species, examination or consideration of different species in various studies, the probable polyphyletic nature of the archaebacterial urkingdom, or an indeterminate quality of sequence alignments and extent of homologies between protein or nucleic acid sequences that are used to deduce the phylogenetic relationships. It should be mentioned that, in comparison to other protein sequences that have been previously examined (viz., RNA polymerases, H^+-ATPase, EF-Tu, and EF-G, F_1-ATPase, lactate dehydrogenase, and malate dehydrogenase) (13, 22, 36), both the extent of homology as well as the quality of sequence alignment are far superior in the case of

the HSP70 family of proteins. In fact, as indicated earlier, HSP70 is the most conserved protein that is known to date. Lastly, the changes in nucleotide sequences that result from large differences in the base composition of DNAs between various species (e.g., low-G-C-content and high-G-C-content gram-positive bacteria) could also distort the phylogenetic relationships. As acknowledged by Woese (45), this problem is quite serious when comparisons are made among noncoding nucleic acid sequences (e.g., rRNA), since there is no rational way to correct for such changes. However, in comparing protein sequences, the base composition-induced changes have a minimal effect because of codon degeneracy and the selective use of codons rich in specific bases. In view of the above considerations, further investigations with the highly conserved protein models such as those of HSP70 should prove particularly useful in clarifying the deep evolutionary relationships among various urkingdoms.

ACKNOWLEDGMENTS

This work was supported by a research grant from the Medical Research Council of Canada to R.S.G.

We thank B. Sweet for secretarial assistance and for typing the manuscript.

REFERENCES

1. **Ahmad, S., R. Ahuja, T. J. Venner, and R. S. Gupta.** 1990. Identification of a protein altered in mutants resistant to microtubule inhibitors as a member of the major heat shock protein (hsp70) family. Mol. Cell. Biol. **10:**5160–5165.
2. **Bardwell, J., and E. A. Craig.** 1989. Major heat shock gene of *Drosophila* and the *Escherichia coli* heat inducible *dnaK* gene are homologous. Proc. Natl. Acad. Sci. USA **81:**848–852.
3. **Beveridge, T. J., G. D. Sprott, and P. Whippey.** 1991. Ultrastructure, inferred porosity, and Gram-staining character of *Methanospirillum hungatei* filament termini describe a unique cell permeability for this archaebacterium. J. Bacteriol. **173:** 130–140.
4. **Birkelund, S., A. G. Lundemose, and G. Christiansen.** 1990. The 75-kilodalton cytoplasmic *Chlamydia trachomatis* L2 polypeptide is a DnaK-like protein. Infect. Immun. **58:**2098–2104.
5. **Cavalier-Smith, T.** 1987. The origin of eukaryote and archaebacterial cells. Ann. N.Y. Acad. Sci. **503:**17–71.
6. **Craig, E. A., J. Kramer, J. Shilling, M. Werner-Washburne, S. Holmes, J. Kosic-Smithers, and C. M. Nicolet.** 1989. SSC1, an essential member of the yeast HSP70 multigene family, encodes a mitochondrial protein. Mol. Cell. Biol. **9:**3000–3008.
7. **Daniels, C. J., A. H. Z. McKee, and W. F. Doolittle.** 1984. Archaebacterial heat shock proteins. EMBO J. **3:**745–749.
8. **Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt.** 1979. A model of evolutionary change in proteins, p. 345–362. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5. National Biomedical Research Foundation, Washington, D.C.
9. **Doi, M., M. Wachi, F. Ishino, S. Tomioka, M. Ito, Y. Sakagami, A. Suzuki, and M. Matsuhashi.** 1988. Determinations of the DNA sequence of the *mre*B gene and of the gene products of the *mre* region that function in the formation of the rod shape of *Escherichia coli* cells. J. Bacteriol. **170:**4619–4624.
10. **Flaherty, K. M., C. DeLuca-Flaherty, and D. B. McKay.** 1990. Three-dimensional structure of the ATPase fragment of a 70 K heat shock cognate protein. Nature (London) **346:**623–628.
11. **Flaherty, K. M., D. B. McKay, W. Kabsch, and K. C. Holmes.** 1991. Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. Proc. Natl. Acad. Sci. USA **88:**5041–5045.
12. **Galley, K. A., B. Singh, and R. S. Gupta.** 1992. Cloning of HSP70 gene from *Clostridium perfringens* using a general polymerase chain reaction based approach. Biochim. Biophys. Acta **1130:**203–208.
13. **Gogarten, J. P., H. Kibak, P. Dittrich, L. Taiz, E. J. Bowman, B. J. Bowman, M. F. Monolson, R. J. Poole, T. Date, T. Osima, J. Konishi, K. Denda, and M. Yoshida.** 1989. Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA **86:**6661–6665.
14. **Gomes, S. L., J. W. Gober, and L. Shapiro.** 1990. Expression of the *Caulobacter* heat shock gene *dnaK* is developmentally controlled during growth at normal temperatures. J. Bacteriol. **172:**3051–3059.
15. **Gorisch, H., W. Grossebuter, and T. Hartl.** 1986. Archaebacterial malate dehydrogenase and citrate synthases: the enzymes from *thermoplasma* and *sulfolobus*. Syst. Appl. Microbiol. **7:**421.
16. **Gouy, M., and W.-H. Li.** 1989. Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. Nature (London) **339:**145–147.
17. **Gray, M. W.** 1989. The evolutionary origin of organelles. Trends Genet. **5:**294–299.
18. **Hearne, C. M., and D. J. Ellar.** 1989. Nucleotide sequence of a *Bacillus subtilis* gene homologous to the *dnaK* gene of *Escherichia coli*. Nucleic Acids Res. **17:**8373.
19. **Heinemann, J. A.** 1991. Genetics of gene transfer between species. Trends Genet. **7:**181–185.
20. **Herbert, A. M., A. M. Kropinski, and K. F. Jarrell.** 1991. Heat shock response of the archaebacterium *Methanococcus voltae*. J. Bacteriol. **173:**3224–3227.
21. **Hunt, C., and R. I. Morimoto.** 1985. Conserved features of eukaryotic hsp70 genes revealed by comparison with the nucleotide sequence of human hsp70. Proc. Natl. Acad. Sci. USA **82:**6455–6459.
22. **Iwabe, M., K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata.** 1989. Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA **86:**9355–9359.
23. **Lake, J. A.** 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. Nature (London) **331:**184–186.
24. **Lake, J. A.** 1989. Origin of the eukaryotic nucleus: eukaryotes and eocytes are genotypically related. Can. J. Microbiol. **35:** 109–118.
25. **Lake, J. A.** 1991. Tracing origins with molecular sequences: metazoan and eukaryotic beginnings. Trends Biochem. Sci. **16:**46–50.
26. **Lam, W. N., A. Cohen, D. Tsoulunas, and W. F. Doolittle.** 1990. Genes for tryptophan biosynthesis in the archaebacterium *Haloferax volcanii*. Proc. Natl. Acad. Sci. USA **87:**6614–6618.
27. **Lindquist, S., and E. A. Craig.** 1988. The heat shock proteins. Annu. Rev. Genet. **22:**631–677.
28. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227:**1435–1441.
29. **Lohlen-Werhahn, G., P. Golpfert, and H. Eggerer.** 1988. Purification and properties of an archaebacterial enzyme: citrate synthase from *Sulfolobus solfataricus*. Biol. Chem. Hoppe-Seyler **369:**109–113.
30. **Macario, A. J. L., C. B. Dugan, and E. C. D. Macario.** 1991. A *dnaK* homolog in the archaebacterium *Methanosarcina mazei* S6. Gene **108:**133–137.
31. **Maeda, K., A. Tsugita, D. Dalzoppo, F. Vilbois, and P. Schurmann.** 1986. Further characterization and amino acid sequence of m-type thioredoxins from spinach chloroplasts. Eur. J. Biochem. **154:**197–203.
32. **McKenzie, K. R., E. Adams, W. J. Britton, R. J. Garsia, and A. Basten.** 1991. Sequence and immunogenicity of the 70-kDa heat shock protein of *Mycobacterium leprae*. J. Immunol. **147:**312–319.
33. **Mevarech, M., S. Hirsch-Twizer, S. Goldman, E. Yakobson, H. Eisenberg, and P. P. Dennis.** 1989. Isolation and characterization of the rRNA gene clusters of *Halobacterium marismortui*. J. Bacteriol. **171:**3479–3485.
34. **Morimoto, R. I., A. Tissieres, and C. Georgopoulos.** 1990. The stress response function of the proteins, and perspectives, p. 1–15. *In* R. I. Morimoto, A. Tissieres, and C. Georgopoulos (ed.), Stress proteins in biology and medicine. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
35. **Pearson, W. R.** 1990. Rapid and sensitive sequence comparison

with FASTP and FASTA. Methods Enzymol. **183:**62–98.
36. **Puhler, G., H. Leffers, F. Gropp, P. Palm, H.-P. Klenk, F. Lottspeich, R. A. Garrett, and W. Zillig.** 1989. Archaebacterial DNA-dependent RNA polymerases testify to the evolution of eukaryotic nuclear genome. Proc. Natl. Acad. Sci. USA **86:** 4569–4573.
37. **Rochester, D. E., J. A. Winer, and D. M. Shah.** 1986. The structure and expression of maize genes encoding the major heat shock protein, hsp70. EMBO J. **5:**451–458.
38. **Slater, M. R., and E. A. Craig.** 1989. The SSB1 heat shock cognate gene of the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res. **17:**4891.
39. **Sneath, P. H. A., and R. R. Sokal.** 1973. Numerical taxonomy. W. H. Freeman & Co., New York.
40. **Stanier, R. Y., J. L. Ingraham, M. L. Wheelis, and P. R. Painter.** 1987. The archaebacteria, p. 330–343. General microbiology, 5th ed. Macmillan Education Ltd., London.
41. **Stevenson, K., N. F. Inglis, B. Rae, W. Donachie, and J. M. Sharp.** 1991. Complete nucleotide sequence of a gene encoding the 70 Kd heat shock protein of *Mycobacterium paratuberculosis*. Nucleic Acids Res. **19:**4552.
42. **Sussman, M. D., and P. Setlow.** 1987. Nucleotide sequence of a *Bacillus megaterium* gene homologous to the *dnaK* gene of *Escherichia coli*. Nucleic Acids Res. **15:**3923.
43. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80:**726–730.
44. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
45. **Woese, C. R.** 1991. The use of ribosomal RNA in reconstructing evolutionary relationships among bacteria, p. 1–24. *In* R. K. Selander, A. G. Clark, and T. S. Whittmay (ed.), Evolution at the molecular level Sinauer Associates, Inc., Publishers, Sunderland, Mass.
46. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains archala, bacteria, and eucarya. Proc. Natl. Acad. Sci. USA **87:**4576–4579.
47. **Woese, C. R., and G. J. Olsen.** 1986. Archaebacterial phylogeny: perspectives on the urkingdoms. Syst. Appl. Microbiol. **7:**161–177.