Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 1159–1163, February 1990
Biochemistry

An essential member of the HSP70 gene family of *Saccharomyces cerevisiae* is homologous to immunoglobulin heavy chain binding protein

(yeast/endoplasmic reticulum/evolution)

RICHARD C. NICHOLSON*, DAVID B. WILLIAMS, AND LAURENCE A. MORAN†

Department of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada

Communicated by Bruce M. Alberts, July 28, 1989

ABSTRACT **Immunoglobulin heavy chain binding protein (BiP) is present in the lumen of the mammalian endoplasmic reticulum, where it associates transiently with a variety of newly synthesized secretory and membrane proteins or permanently with mutant proteins that are incorrectly folded. We describe a unique member of the *Saccharomyces cerevisiae* 70-kDa heat shock protein gene family (HSP70) that encodes a protein homologous to mammalian BiP. The DNA sequence contains a 2046-nucleotide open reading frame devoid of introns, and examination of the predicted amino acid sequence reveals features not found in most other yeast HSP70 proteins but which are present in BiP. Most notable are a 42-residue sequence at the N terminus that exhibits characteristics of a cleavable signal sequence and a C-terminal sequence, -His-Asp-Glu-Leu, that is involved in determining endoplasmic reticulum localization in yeast. The 5′ flanking region of this gene contains two overlapping sequences between nucleotides −146 and −169 that closely resemble consensus heat shock elements. The yeast BiP gene is strongly heat shock-inducible, whereas the BiP genes in various other species are either weakly or non-heat-inducible. We demonstrate that a functional BiP gene is essential for vegetative growth. An evolutionary comparison of amino acid sequences of 34 HSP70 proteins from 17 species suggests that BiP genes share a common ancestor, which diverged from other HSP70 genes near the time when eukaryotes first appeared.**

Members of the 70-kDa heat shock protein multigene family are among the most highly conserved genes known. Eukaryotic genomes contain several related members of this family. All of the proteins encoded by these genes probably interact with other cellular proteins and catalyze inter- and intramolecular rearrangements (folding) in a reaction that is probably coupled to hydrolysis of ATP (1, 2). Such enzymes have been referred to as molecular chaperones (3).

The most prominent members of the gene family are the genes encoding the so-called heat shock cognate proteins of 70 kDa (hsc70). Most eukaryotes have several hsc70 genes; and one or more of these genes is transcribed in all cells producing an abundant cytoplasmic protein. hsc70 catalyzes the uncoating of clathrin-coated vesicles by disrupting clathrin–clathrin interactions (4, 5), but this is undoubtedly only one of many similar activities that it carries out (6, 7).

The second class of HSP70 genes includes the classic heat-inducible genes after which the family is named. Most eukaryotic genomes contain several closely related members of this class of genes (8–10). Upon induction by stress, the ≈70-kDa heat shock proteins (hsp70) concentrate in the nucleus and nucleolus, where they are likely to mediate protein repair or reassembly after stress damage (11).

The HSP70 gene family includes a gene that encodes a 78-kDa glucose-regulated protein (GRP78) also known as immunoglobulin heavy chain binding protein (BiP; refs. 1, 12, and 13). The product of the BiP gene is localized to the lumen of the mammalian ER, where it has been shown to associate transiently during the normal biogenesis of a variety of newly synthesized membrane and secretory proteins and permanently with underglycosylated or mutant proteins that fail to leave the ER (13–16). Mammalian BiP is synthesized constitutively, but its rate of synthesis can be enhanced by a variety of stress conditions including glucose starvation, inhibition of N-linked glycosylation, calcium ionophores, amino acid analogues, and overproduction of improperly folded proteins that accumulate in the ER, but not substantially by heat shock (12, 17–19). Based on these findings it has been proposed that BiP recognizes incorrectly folded proteins in the ER, preventing their aggregation and possibly promoting proper folding (1, 2).

We have sought to identify the gene‡ encoding a protein homologous to BiP in *Saccharomyces cerevisiae* with the intention of generating mutants that may be useful in elucidating the function of this member of the HSP70 gene family.

MATERIALS AND METHODS

Strains. The *S. cerevisiae* strains used were: LL20 (*MATα, leu2-3, leu2-112, his3-11, his3-15, can1*), W3031A-H (*MAT*a, *can1-100, leu2-3, leu2-112, trp1-1, ura3-1, ade2-1*), W3031B-T (*MATα, can1-100, leu2-3, leu2-112, his3-11, ura3-1, ade2-1*), and LP112-HT produced by mating W3031A-H and W3031B-T (20).

Isolation of the Yeast BiP Gene. A yeast (strain LL20) genomic library was provided by A. Percival-Smith and J. Segall (21). Six different recombinants were isolated from this library by screening for homology to a *Drosophila* HSP70 gene (22). Plasmid YG2C2 contained a single yeast HSP70 gene within a 18-kilobase (kb) insert of genomic DNA. Most of the sequence was determined by using a modification of the supercoiled template technique (23).

Gene Disruption. Plasmid YG2C2 was digested with *Xho I* and *Xba I* and the fragment, containing part of the 5′ flanking region and most of the coding region of the BiP gene

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: BiP, immunoglobulin heavy chain binding protein; ER, endoplasmic reticulum; GRP78, glucose-regulated protein of 78 kDa (=BiP); hsp70, heat shock-induced protein of ≈70-kDa; hsc70, heat shock cognate protein of 70-kDa (the major constitutively expressed protein); HSP70, the gene family to which hsp70, hsc70, and BiP belong.

*Present address: Laboratoire de Genetique Moleculaire des Eucaryotes, Institut de Chimie Biologique, 11 rue Humann, 67085 Strasbourg Cedex, France.
†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M31006).

(nucleotides −243 to 1783 in Fig. 1), was subcloned into pUC13 and designated YG2C2SX. Subsequently, the 0.68-kb *Bgl* II fragment within the coding region was removed and replaced with a 3.0-kb *Bgl* II fragment carrying the *LEU2* gene. The disrupted BiP gene was excised from the plasmid and used for integrative transformations of both diploid (LP112-HT) and haploid (W3031A-H and W3031B-T) strains by the spheroplast method (24).

RESULTS AND DISCUSSION

Identification and Characterization of the Yeast BiP Gene. In the yeast *S. cerevisiae*, the HSP70 gene family has been extensively characterized by E. Craig and her co-workers; they have isolated eight members of the family (25–27). We have independently detected and cloned five members of the yeast HSP70 gene family by screening a genomic library with a *Drosophila* hsp70 gene probe. One of these genes (contained on plasmid YG2C2) was shown to be a single-copy gene not closely related to the other members of the family (22). Transcripts of this gene are present at a low level during optimal growth temperature (28°C) but increase dramatically (5- to 10-fold) after transfer of the cells to the stress temperature (42°C) (data not shown). The entire sequence of this gene is shown in Fig. 1.

A long open reading frame begins at the indicated ATG (underlined, nucleotide +40) and continues, without introns, for 2046 nucleotides. The hypothetical protein encoded by this gene has a predicted molecular mass of 74,400 daltons. The 5′ flanking sequence of this HSP70 gene contains up-

```
-591                                                              ACTTCA
-585 ATGTCTAATGCTAGAGCAGATGCACACATTATACACGGTTCTACAGTAACATATAGAGTAATGTC
-520 TTTGAACACGTCAACAACTCCTCGGGAGCCCAACATCGCCTTGATCTGATCGATCCCCATGAACT
-455 CAGCATGTGCTACTCCAGTTAATGACTTGTTCGTATCGTTCATGCCATAAGCCATCACCTGGCCA
-390 GTTGGCGTATGTACAAAGATGCAAGCTACCGGTGTCTCATCGTGGTCAAGAGCGTATCTAGCCAA
-325 ACGGACAGCTGTCCTCATATGTTTAATATGCTGCATAGTGTGAGAGTCCTCTAAGAAAAATGGCG
-260 TCGGTGGTCGGCGAACTCGAGCAAAGTGTAGATCCCATTAGGACTCATCATTCATCTAATTTTGC
-195 TATGTTAGCTGCAACTTTCTATTTTAATAGAACCTTCTGGAAATTTCACCCGGCGCGGCACCCGA
-130 GGAACTGGACAGCGTGTCGAAAAAGTTGCTTTTTTATATAAAGGACACGAAAAGGGTTCTCTGGA
 -65 AGATATAAATATGGCTATGTAATTCTAAAGATTAACGTGTTACTGTTTTACTTTTTTAAAGTCCC
  +1  CAAGAGTAGTCTCAAGGGAAAAAGCGTATCAAACATACC ATG TTT TTC AAC AGA CTA
  58  AGC GCT GGC AAG CTG CTG GTA CCA CTC TCC GTG GTC CTG TAC GCC CTT
 106  TTC GTG GTA ATA TTA CCT TTA CAG AAT TCT TTC CAC TCC TCC AAT GTT
 154  TTA GTT AGA GGT GCC GAT GAT GTA GAA AAC TAC GGA ACT GTT ATC GGT
 202  ATT GAC TTA GGT ACT ACT TAT TCC TGT GTT GCT GTG ATG AAA AAT GGT
 250  AAG ACT GAA ATT CTT GCT AAT GAG CAA GGT AAC AGA ATC ACC CCA TCT
 298  TAC GTG GCA TTC ACC GAT GAT GAA AGA TTG ATT GGT GAT GCT GCA AAG
 346  AAC CAA GTT GCT GCC AAT CCT CAA AAC ACC ATC TTC GAC ATT AAG AGA
 394  TTG ATC GGT TTG AAA TAT AAC GAC AGA TCT GTT CAG AAG GAT ATC AAG
 442  CAC TTG CCA TTT AAT GTG GTT AAT AAA GAT GGG AAG CCC GCT GTA GAA
 490  GTA AGT GTC AAA GGA GAA AAG AAG GTT TTT ACT CCA GAA GAA ATT TCT
 538  GGT ATG ATC TTG GGT AAG ATG AAA CAA ATT GCC GAA GAT TAT TTA GGC
 586  ACT AAG GTT ACC CAT GCT GTC GTT ACT GTT CCT GCT TAT TTC AAT GAC
 634  GCG CAA AGA CAA GCC ACC AAG GAT GCT GGT ACC ATC GCT GGT TTG AAC
 682  GTT TTG AGA ATT GTT AAT GAA CCA ACC GCA GCC GCC ATT GCC TAC GGT
 730  TTG GAT AAA TCT GAT AAG GAA CAT CAA ATT ATT GTT TAT GAT TTG GGT
 778  GGT GGT ACT TTC GAT GTC TCT CTA TTG TCT ATT GAA AAC GGT GTT TTC
 826  GAA GTC CAA GCC ACT TCT GGT GAT ACT CAT TTA GGT GGT GAA GAT TTT
 874  GAC TAT AAG ATC GTT CGT CAA TTG ATA AAA GCT TTC AAG AAG AAG CAT
 922  GGT ATT GAT GTG TCT GAC AAC AAC AAG GCC CTA GCT AAA TTG AAG AGA
 970  GAA GCT GAA AAG GCT AAA CGT GCC TTG TCC AGC CAA ATG TCC ACC CGT
1018  ATT GAA ATT GAC TCC TTC GTT GAT GGT ATC GAC TTA AGT GAA ACC TTG
1066  ACC AGA GCT AAG TTT GAG GAA TTA AAC CTA GAT CTA TTC AAG AAG ACC
1114  TTG AAG CCT GTC GAG AAG GTT TTG CAA GAT TCT GGT TTG GAA AAG AAG
1162  GAT GTT GAT GAT ATC GTT TTG GTT GGT GGT TCT ACT AGA ATT CCA AAG
1210  GTC CAA CAA TTG TTA GAA TCA TAC TTT GAT GGT AAG AAG GCC TCC AAG
1258  GGT ATT AAC CCA GAT GAA GCT GTT GCA TAC GGT GCA GCC GTT CAA GCT
1306  GGT GTC TTA TCC GGT GAA GAA GGT GTC GAA GAT ATT GTT TTA TTG GAT
1354  GTC AAC GCT TTG ACT CTT GGT ATT GAA ACC ACT GGT GGT GTC ATG ACT
1402  CCA TTA ATT AAG AGA AAT ACT GCT ATT CCT ACA AAG AAA TCC CAA ATT
1450  TTC TCT ACT GCC GTT GAC AAC CAA CCA ACC GTT ATG ATC AAG GTA TAC
1498  GAG GGT GAA AGA GCC ATG TCT AAG GAC AAC AAT CTA TTA GGT AAG TTT
1546  GAA TTA ACC GGC ATT CCA CCA GCA CCA AGA GGT GTA CCT CAA ATT GAA
1594  GTC ACA TTT GCA CTT GAC GCT AAT GGT ATT CTG AAG GTG TCT GCC ACA
1642  GAT AAG GGA ACT GGT AAA TCC GAA TCT ATC ACC ATC ACT AAC GAT AAA
1690  GGT AGA TTA ACC CAA GAA GAG ATT GAT AGA ATG GTT GAA GAG GCT GAA
1738  AAA TTC GCT TCT GAA GAC GCT TCT ATC AAG GCC AAG GTT GAA TCT AGA
1786  AAC AAA TTA GAA AAC TAC GCT CAC TCT TTG AAA AAC CAA GTT AAT GGT
1834  GAC CTA GGT GAA AAA TTG GAA GAA GAA GAC AAG GAA ACC TTA TTA GAT
1882  GCT GCT AAC GAT GTT TTA GAA TGG TTA GAT GAT AAC TTT GAA ACC GCC
1930  ATT GCT GAA GAC TTT GAT GAA AAG TTC GAA TCT TTG TCC AAG GTC GCT
1978  TAT CCA ATT ACT TCT AAG TTG TAC GGA GGT GCT GAT GGT TCT GGT GCC
2026  GCT GAT TAT GAC GAC GAA GAT GAA GAT GAC GAT GGT GAT TAT TTC GAA
2074  CAC GAC GAA TTG TAG ATAAAATAGTTAAAAATTTTTGCTGCTGGAAGCTT
```

FIG. 1. Sequence of the yeast BiP gene. The two overlapping heat shock elements beginning at −169 are underlined. There are two TATA homologies at −96 and −62 (boldface). The unique transcription start site (+1) occurs at CAAG. The open reading frame begins at the first ATG (+40). A potential transcription termination signal at the 3′ end of the gene is indicated. This signal includes the termination codon (TAG) of the long open reading frame.

stream regulatory sequences known as heat shock elements (28, 29). Between nucleotides −146 and −169 there are two overlapping regions that match the heat shock element consensus sequence (C--GAA--TTC--G) in seven of eight nucleotides.

The DNA sequence shows two TATA homologies at positions −56 to −62 and −89 to −96. However, only a single transcriptional start site occurs *in vivo* (nucleotide +1) as determined by S1 nuclease digestion of DNA·RNA hybrids (not shown). In yeast cells, Zaret and Sherman (30) have proposed a consensus sequence of TAG. . .TAGT. . . (AT-rich)TTT as a transcription termination signal in some genes. This gene contains an exact match to the proposed consensus sequence beginning at the stop codon TAG (nucleotide 2086).

The predicted amino acid sequence of the protein encoded by the yeast HSP70 gene that we have cloned is compared directly to the amino acid sequence of rat BiP, the amino acid sequence of yeast hsc70A1 (product of the *SSA1* gene), and the consensus HSP70 sequence (Fig. 2). Small deletions (indicated by dashed lines) have been inserted in these sequences to maximize homology. The boxed regions of homology indicate that the gene we have isolated encodes a protein that is closely related to both the yeast hsc70A1 sequence and that of rat BiP. The homology to rat BiP is greater than that to the other yeast gene, although the differences are not striking. From amino acid 4 to the C terminus, the amino acid sequence is identical to rat BiP in 421 positions, compared with only 394 amino acids in common with the other yeast gene. The rat BiP and yeast hsc70A1 sequences, on the other hand, are identical at 404 positions.

A survey of 33 HSP70-related protein sequences from 17 species shows that the gene we have isolated and the rat and hamster BiPs share characteristics not common in other HSP70s. For example, most reported HSP70 protein sequences show a common, highly conserved, stretch of amino acids, Gly-Ile-Asp-Leu-Gly-Thr-Thr-Tyr-Ser-Cys-Val, which begins a few residues from the N terminus, but the predicted protein sequence of the gene that we have cloned contains an unusually long amino acid sequence (53 amino acids) preceding this highly conserved sequence. We suspect that a portion of this is a hydrophobic leader sequence found in many examples of proteins that are translocated across the membrane of the ER (32), including mammalian BiP (12). The N-terminal extensions of the yeast gene product and rat BiP, although quite different in length, share several features (Fig. 2). They both contain one or more basic amino acid residues near the N terminus, they both have a stretch of hydrophobic amino acids near the middle of the extension followed by basic residue(s) and then a group of acidic amino acids. Finally, both extensions contain the triplet Gly-Thr-Val immediately prior to the highly conserved sequence noted above. This triplet is not observed in this location in any other members of the gene family. The predicted cleavage sites for the leader sequence of rat BiP (residue 18; ref. 12) and for the yeast gene product (residue 42; see ref. 32) are shown in Fig. 2.

We have noted other differences between BiP sequences and those of the other members of the HSP70 family. For example, the highly conserved tryptophan-87 in most HSP70s is replaced by leucine in BiPs (the only other protein that does not have tryptophan at this position is the yeast hsc70A1 shown in Fig. 2). In the region from amino acid 232 to amino acid 298, the BiP sequence diverges from that of other HSP70s, but the putative yeast BiP and the rat BiP are more similar to each other than to the yeast hsc70A1 or the consensus sequence in this region. Secondary structure predictions suggest that in most HSP70s, this region is an extended α-helix with nine heptad repeats where the first and fourth amino acids are hydrophobic. Such sequences are likely to form coiled coils and are found in many multimeric structural proteins, including clathrin light and heavy chains and inter-

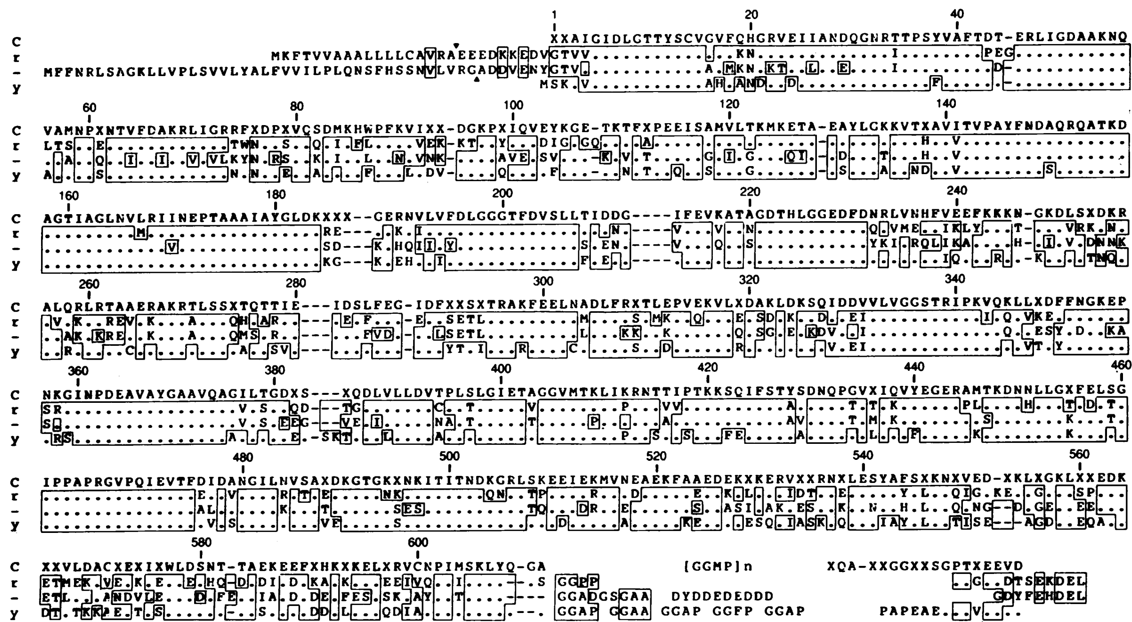

FIG. 2. Hypothetical amino acid sequences of several HSP70 genes. The consensus sequence in lines C was determined by comparing all known HSP70 sequences and choosing the amino acid that was most commonly found when different groups of proteins were examined (i.e., all mammalian hsp70 and hsc70 sequences count as a single group, thus avoiding the usual nonscientific bias towards the numerous mammalian sequences that have been determined). The predicted sequence from the longest open reading frame in Fig. 1 is shown in lines (-). The rat BiP gene sequence (12) is shown in lines r; the yeast gene *SSA1* sequence (9, 10, 31) is shown in lines y. The sequences have been aligned for maximal homology based on the analysis of all available sequences. The boxed areas indicate that identical amino acids are present in at least two of the sequences. Dots designate residues that are identical to the consensus sequence. Amino acid positions are numbered with reference to the *Drosophila* HSP70 sequence of the gene at locus *87C* until amino acid 550, where the numbering system corresponds to the consensus sequence. Gaps in all four sequences represent regions where insertions are present in other HSP70 genes. The standard single letter amino acid abbreviations are used except that "X" in the consensus sequence stands for any amino acid. Arrowheads indicate the predicted cleavage sites of the N-terminal leader sequence.

mediate filaments (33). The short sequence Ser-Glu-Thr-Leu (amino acids 291–294) appears to be a good indicator of BiP homology because it is not found in any other HSP70 protein and is present in the putative *Plasmodium* BiP (34).

The C-terminal amino acid sequence of the putative yeast BiP protein shows no similarity with the analogous sequence in yeast hsc70A1 or the consensus sequence (Fig. 2), but a comparison of this region with rat BiP shows remarkable homology. Five of the last eight amino acids in each of these proteins are identical. The three nonidentical amino acids represent relatively conservative changes. The rat BiP C-terminal sequence Lys-Asp-Glu-Leu has been implicated in the ER localization of the protein (35), and the analogous sequence His-Asp-Glu-Leu present in the putative yeast BiP has recently been shown to have the same effect when linked to the C terminus of invertase (36). This feature, in combination with the N-terminal extension, most clearly distinguishes BiPs from the other HSP70 family members and suggests that yeast BiP may also be localized to the lumen of the endoplasmic reticulum. These unique amino acids and the overall homology of our yeast gene to BiP suggest that the gene we have cloned corresponds to yeast BiP and that the BiP genes have evolved independently of the other members of the HSP70 gene family.

Disruption of the Yeast BiP Gene. To determine whether the yeast BiP gene is essential, we constructed a plasmid in which the *Bgl* II fragment within the coding region of the gene was replaced by the yeast *LEU2* gene (Fig. 3). A restriction fragment containing the disrupted gene from this plasmid was used in integrative transformations to replace the wild-type gene. The diploid strain LP112-HT as well as the haploid strains, W3031A-H (*MAT***a**) and W3031B-T (*MAT*α), were used as transformation recipients with selection for leucine prototrophy. No transformants were obtained with the haploid strains, suggesting that a complete copy of the gene is essential

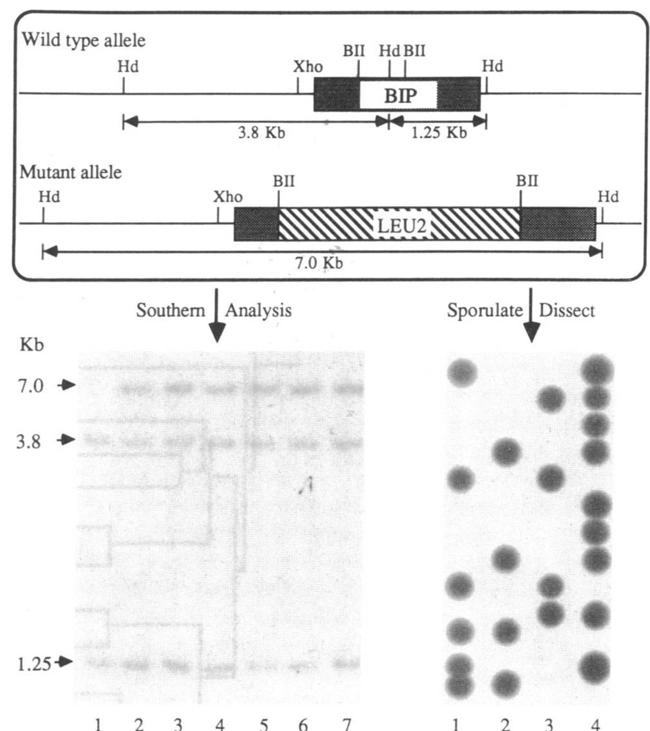

FIG. 3. Disruption of the yeast BiP gene. (*Lower Left*) DNA from untransformed diploid cells (lane 1) and diploid cells transformed with the disrupted BIP gene (lanes 2–7) were probed with nick-translated plasmid YG2C2SX. (*Lower Right*) Transformants were sporulated and 19 asci were dissected. Shown are results of tetrad analysis of 13 of the asci (spores numbered 1–4; colonies were grown on nutrient-rich plates).

for viability following transformation. The genomes of several independently isolated diploid transformants were analyzed for the presence of a new 7-kb *Hin*dIII fragment that would indicate that one of the endogenous genes had been disrupted. The results, shown in Fig. 3, revealed that the probe hybridized to three fragments, a 3.8-kb fragment and a 1.25-kb fragment originating from the wild-type gene, and the expected 7-kb fragment derived from the disrupted gene. Thus, insertion occurred at one of the two allelic wild-type loci.

Several transformed diploid strains were induced to sporulate to assess the effects of a disrupted BiP gene in haploid cells. Upon dissection of 19 tetrads, only one or two viable spores were recovered per tetrad (Fig. 3). Furthermore, all viable spores gave rise to colonies that were leucine auxotrophs, demonstrating that the disrupted gene carrying the *LEU2* marker cosegregated with the nonviable phenotype. Microscopic examination of the nonviable spores showed that no cell divisions had occurred. This is unlike the result obtained with the only other yeast HSP70 gene shown to be essential (*SSC1*), where residual protein in the haploid spore may be sufficient for several cell divisions (26). The viability of haploid cells containing the disrupted BiP gene has been restored by rescue with a centromere-containing plasmid possessing a complete copy of the intact gene (not shown). This rescue confirms that the disruption of the BiP gene is responsible for nonviability as opposed to some other alteration outside of the gene that may have occurred during cloning or transformation. We conclude that the yeast BiP gene is essential for growth under the conditions used in the experiment. The yeast BiP gene also has been isolated by M. Rose and by M.-J. Gething and J. Sambrook (see ref. 37); it is identical to the *KAR2* gene.

There are at least nine members of the yeast HSP70 gene family (see ref. 37). Three of these, including the gene we have sequenced, are unique members of the family. One of these unique members, *SSC1*, is essential for normal growth, whereas another, *SSD1*, is not (26). The closely related pair of HSP70 genes, *SSB1* and *SSB2*, encode hsc70-like proteins, and deletion of either one of them has no effect on cell growth. Disruption of both genes on the other hand renders yeast cells cold sensitive (25). The remaining genes, *SSA1–SSA4*, are a closely related group of genes whose products are functionally similar, and disruption of three of these genes is lethal (37). These results in combination with our current findings suggest that the *SSC1* gene, the *SSB1/SSB2* pair, the *SSA* group, and the gene for BiP each encode proteins with distinct functions and/or cellular locations in spite of their close relatedness.

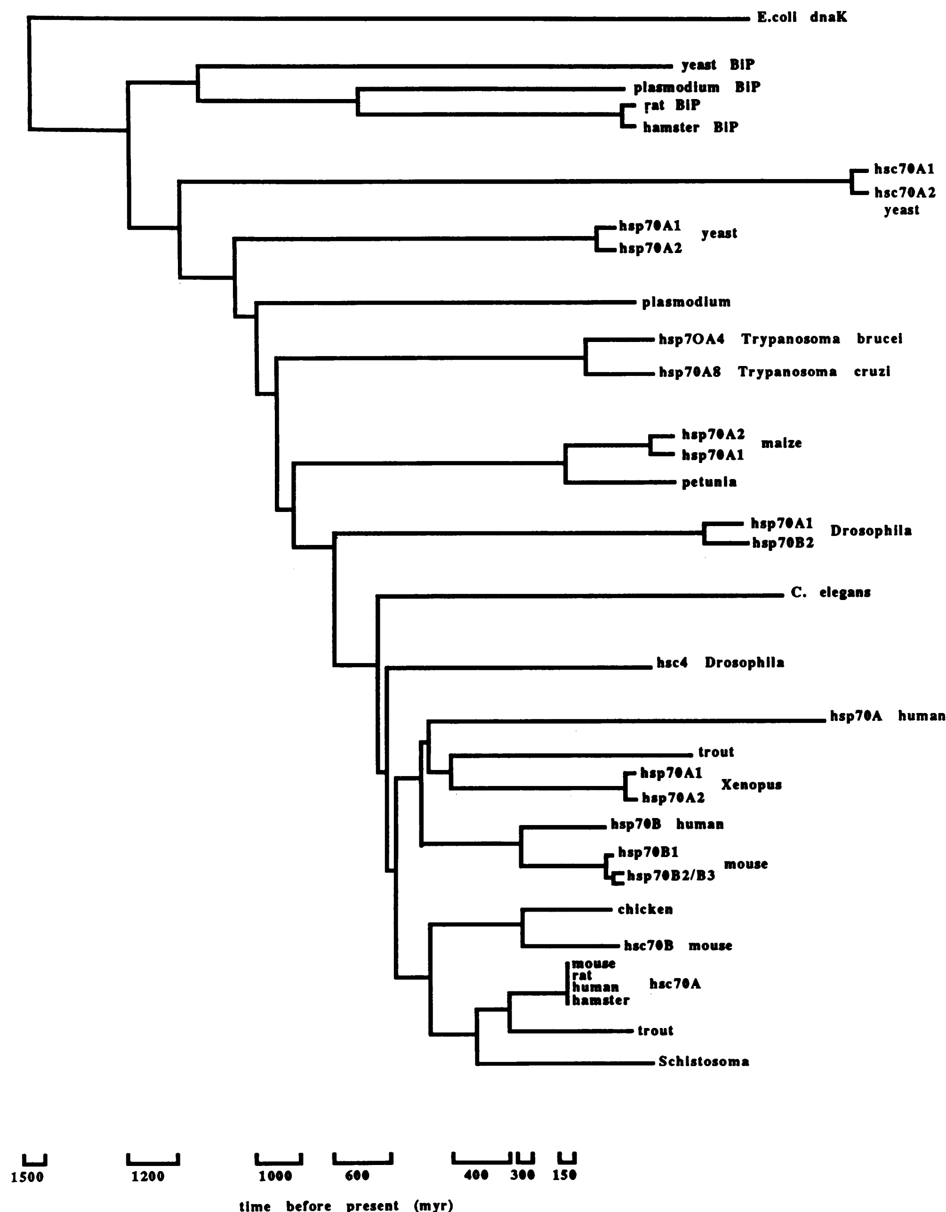

FIG. 4. Comparison of HSP70-related proteins. The diagram was derived from pairwise comparisons of all available sequences by using the minimum distances that were internally consistent. The length of each horizontal line corresponds to the number of amino acid differences between proteins from amino acids 3–610. The vertical lines are arbitrary. In many cases published sequences have been corrected and updated but the relationships shown here should be considered tentative because of the high probability of sequencing errors and cloning artifacts. The approximate time of divergence of some species is shown at the bottom of the dendrogram. Contact one of the authors (L.A.M.) for an annotated version of the data base with complete references.

Pelham (1) has suggested that heat shock proteins in general bind to proteins with exposed hydrophobic surfaces to limit their aggregation. The energy of ATP hydrolysis is then used to alter the conformation of the protein, releasing it from its substrate and simultaneously distorting the substrate (1). In yeast, hsc70 proteins have a specialized function in the cytosol; they promote the translocation of proteins through the ER and mitochondrial membranes, possibly by maintaining newly translated proteins in a relaxed conformation (6, 7, 38). They also play a role in uncoating clathrin-coated vesicles (4, 5). The *SSC1* gene product is located in mitochondria, where it may promote disaggregation and proper folding of mitochondrial proteins subsequent to translocation (37). Similarly, yeast BiP probably catalyzes analogous reactions in the ER.

Species Variation of hsp70 Primary Structure. We have compared the amino acid sequence of yeast BiP with that of 33 other complete or partial sequences of related proteins. The relationships shown in Fig. 4 resemble a phylogenetic tree, but we prefer to call this diagram a sequence relationship dendrogram, since we have not included time (of species divergence) as a parameter. We have not attempted to correct for multiple substitutions at the same site, and we have not taken into account synonymous nucleotide substitutions, since only the amino acid sequence was analyzed. True phylogenetic relationships and mutation rates require complex manipulation of nucleotide sequence data, and the theoretical justification of these manipulations is not strong when inter-kingdom comparisons are made. In particular, extreme differences in codon bias ("A/T pressure") between different kingdoms complicate such comparisons.

The sequence relationship dendrogram shown in Fig. 4 reveals several interesting sequence relationships. As expected, yeast BiP protein is most closely related to the rat, hamster, and *Plasmodium* BiPs (12, 33, 39). Furthermore, the amino acid sequences of these four proteins are quite distantly related to all other sequences. Thus, the yeast, rat, hamster, and *Plasmodium* genes are likely to be orthologous, and presumably the first eukaryotes contained a gene from which BiP evolved independently of the other HSP70 genes. This hypothesis is consistent with the concept that selection in favor of organisms with internal membrane compartments required concomitant duplication of an ancestral HSP70 gene and subsequent divergence when the products of each gene became localized to either cytoplasmic (hsc70/hsp70) or ER (BiP) compartments.

The data in Fig. 4 also show that the other sequenced yeast HSP70 genes can be grouped into two pairs that are distantly related to each other. This observation suggests that early yeast ancestors contained at least three different members of the HSP70 multigene family. Furthermore, the dendrogram suggests that all of the known vertebrate proteins, with the exception of BiP, are closely related to each other; apparently they arose from a common ancestor after the evolution of the first chordates (although gene conversion between parologous genes could account for the observed similarities). These vertebrate proteins can be roughly divided into two groups; those that are related to the rat hsc70 protein and those more closely related to the mouse and human hsp70/hsp68 proteins. The hsc70-like group contains proteins that are major components of normal cells (rat hsc70, mouse hsc70), others that are developmentally regulated (mouse hsc70B, testis specific), and proteins that are only synthesized in response to stress (chicken and trout 70.14). Thus, closely related HSP70 genes can be regulated in different ways. The fact that the yeast BiP gene is strongly heat inducible whereas the rat, hamster, and human BiP genes are either weakly inducible or unaffected by heat shock shows that the sequences that regulate expression are not as conserved as the coding region.

Note Added in Proof. The yeast BIP gene has been characterized independently by M. D. Rose *et al.* (40) and K. Normington *et al.* (41).

We thank Tony Percival-Smith and Jacqueline Segall for the yeast genomic library and David Law for his advice on constructing the strain containing the disrupted gene. This work was supported by grants from the Medical Research Council of Canada. D.B.W. is a Scholar of the Medical Research Council of Canada.

1. Pelham, H. R. B. (1986) *Cell* **46**, 959–961.
2. Pelham, H. (1988) *Nature (London)* **332**, 776–777.
3. Ellis, R. J. (1987) *Nature (London)* **328**, 378–379.
4. Chappell, T. G., Welch, W. J., Schlossman, D. M., Palter, K. B., Schlesinger, M. J. & Rothman, J. C. (1986) *Cell* **45**, 3–13.
5. Ungewickell, E. (1985) *EMBO J.* **4**, 3385–3391.
6. Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A. & Shekman, R. (1988) *Nature (London)* **332**, 800–805.
7. Chirico, W. J., Waters, M. G. & Blobel, G. (1988) *Nature (London)* **332**, 805–810.
8. Neidhardt, F. C., Van Bogelen, R. A. & Vaughn, V. (1984) *Annu. Rev. Genet.* **18**, 295–329.
9. Craig, E. A. (1985) *CRC Crit. Rev. Biochem.* **18**, 239–280.
10. Lindquist, S. (1986) *Annu. Rev. Biochem.* **55**, 1151–1191.
11. Craig, E. A. & Lindquist, S. (1988) *Annu. Rev. Genet.* **22**, 631–677.
12. Munro, S. & Pelham, H. R. B. (1986) *Cell* **46**, 291–300.
13. Bole, D. G., Hendershot, L. M. & Kearney, J. F. (1986) *J. Cell Biol.* **102**, 1558–1566.
14. Gething, M.-J., McCammon, K. & Sambrook, J. (1986) *Cell* **46**, 939–950.
15. Dorner, A. J., Bole, D. G. & Kaufman, R. J. (1987) *J. Cell Biol.* **105**, 2665–2674.
16. Kassenbrock, C. K., Garcia, P. D., Walter, P. & Kelly, R. B. (1988) *Nature (London)* **333**, 90–93.
17. Lee, A. S. (1987) *Trends Biochem. Sci.* **12**, 20–23.
18. Kelley, P. M. & Schlesinger, M. J. (1978) *Cell* **15**, 1277–1286.
19. Kozutsumi, Y., Segal, M., Normington, K., Gething, M.-J. & Sambrook, J. (1988) *Nature (London)* **332**, 462–464.
20. Law, D. T. S. & Segall, J. (1988) *Mol. Cell. Biol.* **8**, 912–922.
21. Percival-Smith, A. & Segall, J. (1984) *Mol. Cell. Biol.* **4**, 142–150.
22. Nicholson, R. C. (1986) Ph.D. Thesis (Univ. of Toronto, Toronto, ON).
23. Haltiner, M., Kempe, T. & Tijian, R. (1985) *Nucleic Acids Res.* **13**, 1015–1025.
24. Hinnen, A., Hicks, J. B. & Fink, G. B. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1929–1933.
25. Craig, E. A. & Jocobson, K. (1985) *Cell* **38**, 841–849.
26. Craig, E. A., Kramer, J. & Kosic-Smithers, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4156–4160.
27. Craig, E. A., Slater, M. R., Stone, D. E., Park, H. O. & Boorstein, W. R. (1987) in *RNA Polymerase and the Regulation of Transcription*, 16th Steenbock Symposium, eds. Reznikoff, W. S., Burgess, R. R., Dahlberg, J. E., Gross, C. A., Record, M. T., Jr., & Wickens, M. R. (Elsevier, New York), pp. 267–278.
28. Pelham, H. R. B. (1982) *Cell* **30**, 517–528.
29. Pelham, H. R. B. (1985) *Trends Genet.* **1**, 31–35.
30. Zaret, K. S. & Sherman, F. (1982) *Cell* **28**, 563–573.
31. Ingolia, T. D., Slater, M. R. & Craig, E. A. (1982) *Mol. Cell. Biol.* **2**, 1388–1398.
32. Von Heijne, G. (1985) *J. Mol. Biol.* **183**, 99–105.
33. Jackson, A. P., Seow, H.-F., Homles, N., Drickamer, K. & Parnham, P. (1987) *Nature (London)* **326**, 154–159.
34. Peterson, M. G., Crether, P. E., Thompson, J. K., Corcoran, L. M., Coppel, R. L., Brown, G. V., Anders, R. F. & Kemp, D. J. (1988) *DNA* **7**, 71–78.
35. Munro, S. & Pelham, H. R. B. (1987) *Cell* **48**, 899–907.
36. Pelham, H. R. B., Hardwick, K. G. & Lewis, M. J. (1988) *EMBO J.* **7**, 1757–1762.
37. Deshaies, R. J., Koch, B. D. & Scheckman, R. (1988) *Trends Biochem. Sci.* **13**, 384–388.
38. Murakami, H., Pain, D. & Blobel, G. (1988) *J. Cell Biol.* **107**, 2051–2057.
39. Ting, J., Wooden, S. K., Kriz, R., Kelleher, K., Kaufman, R. J. & Lee, A. S. (1987) *Gene* **5**, 147–152.
40. Rose, M. D., Misra, L. M. & Vogel, J. P. (1989) *Cell* **57**, 1211–1221.
41. Normington, K., Kohno, K., Yasunori, K., Gething, M.-J. & Sambrook, J. (1989) *Cell* **57**, 1223–1236.