Identification, Characterization, and Expression of the BiP Endoplasmic Reticulum Resident Chaperonins in *Pneumocystis carinii*

TIMOTHY T. STEDMAN† AND GREGORY A. BUCK*

Department of Microbiology and Immunology, Medical College of Virginia Campus, Virginia Commonwealth University, Richmond, Virginia 23298-0678

Received 17 April 1996/Returned for modification 18 June 1996/Accepted 2 August 1996

We have isolated, characterized, and examined the expression of the genes encoding BiP endoplasmic reticulum (ER) resident chaperonin responsible for transport, maturation, and proper folding of membrane and secreted proteins from two divergent strains of *Pneumocystis carinii*. The BiP genes, *Pcbip* and *Prbip*, from the *P. c. carinii* (prototype) strain and the *P. c. rattus* (variant) strain, respectively, are single-copy genes that reside on chromosomes of ~330 and ~350 kbp. Both genes encode ~72.5-kDa proteins that are most homologous to BiP genes from other organisms and exhibit the amino-terminal signal peptides and carboxy-terminal ER retention sequences that are hallmarks of BiP proteins. We established short-term *P. carinii* cultures to examine expression and induction of *Pcbip* in response to heat shock, glucose starvation, inhibition of protein transport or N-linked glycosylation, and other conditions known to affect proper transport, glycosylation, and maturation of membrane and secreted proteins. These studies indicated that *Pcbip* mRNA is constitutively expressed but induced under conditions known to induce BiP expression in other organisms. In contrast to mammalian BiP genes but like other fungal BiP genes, *P. carinii* BiP mRNA levels are induced by heat shock. Finally, the *Prbip* and *Pcbip* coding sequences surprisingly exhibit only ~83% DNA and ~90% amino acid sequence identity and show only limited conservation in noncoding flanking and intron sequences. Analyses of the *P. carinii* BiP gene sequences support inclusion of *P. carinii* among the fungi but suggest a large divergence and possible speciation among *P. carinii* strains infecting a given host.

*Pneumocystis carinii* is an important opportunistic pathogen of immunocompromised patients and infects up to 80% of patients with AIDS (2). Relatively little is known of the basic biology of this organism, largely because of its relatively recent ascendance to the status of an important opportunistic pathogen but also because of the lack of long-term continuous culture systems. The cell surface composition of this lung pathogen has been examined in some detail (11, 12, 40, 62), major antigens have been characterized (25, 29, 58), and ultrastructural analyses have identified most of the prototype eukaryotic organelles in each of the two basic life cycle stages of *P. carinii*. Although these surface antigens are clearly critical in all interactions between the pathogen and the host, the mechanisms by which they are transported, processed, and matured have not been studied. The phylogenetic placement of *P. carinii* among the fungi was recently confirmed by analyses of several rRNA, mitochondrial, and structural protein coding genes, although this classification remains somewhat controversial (14, 15, 19, 54, 60). Finally, recent reports have suggested that *P. carinii* strains have restricted host ranges but that multiple strains may infect a given host (3, 50, 52, 55). Two distinct isoforms, prototype and variant, have been identified in rats (10), and multiple karyotype variants of the prototype isoform have been reported (9). A proposed revision in the taxonomic nomenclature designates the prototype and variant isoforms as *P. c. carinii* and *P. c. rattus*, respectively (53).

Members of the ubiquitous HSP70 family of proteins play crucial roles in controlling protein folding and trafficking (reviewed in reference 21) and are among the most highly conserved proteins known (23, 26). The HSP70 gene family diverged into different functional classes of molecular chaperones early in the evolution of eukaryotes (4). The HSP70 homolog BiP is an essential resident protein of the endoplasmic reticulum (ER) and plays an important role in the folding and assembly of immature proteins within the ER lumen (6). BiP, also known as grp78, is a 72- to 78-kDa glucose-regulated protein that constitutes 5 to 10% of the luminal protein (21). BiP binds transiently to nascent, incompletely folded proteins destined for secretion, proteins involved in the secretory pathway, and transmembrane proteins prior to N-linked glycosylation or folding in the ER. The principal role of BiP appears to be the stabilization of nascent chains in a folding-competent state and preventing formation of protein aggregates (reviewed in reference 24), although BiP also appears to aid translocation of nascent chains into the ER.

Herein, we characterize BiP genes from the prototype and variant strains of rat *P. carinii*. We report the genomic and cDNA sequences for *P. c. carinii* BiP (*Pcbip*) and the genomic sequence for *P. c. rattus* BiP (*Prbip*). The hallmark leader peptide and ER retention signal sequences of BiP proteins are identified in the inferred protein sequence. We use short-term *P. carinii* cultures to show that, as in other organisms, *Pcbip* expression is probably constitutively expressed but induced in response to heat shock, glucose starvation, and other conditions known to affect protein transport and maturation. DNA elements that may be responsible for this regulation were identified in sequences upstream from the *Pcbip* coding sequence.
Finally, phylogenetic analysis by parsimony- and distance matrix-based methods supported classification of *P. carinii* among the fungi, although the BiP genes of the prototype and variant isoforms show a surprising divergence, suggesting possible speciation among *P. carinii* isolates from a given host.

**MATERIALS AND METHODS**

*P. carinii* isolation. *P. carinii* was isolated from the lungs of infected Sprague-Dawley rats (Charles River, Raleigh, N.C.) immunosuppressed by twice-weekly subcutaneous injection of 25 mg of cortisone acetate (Cortone Acetate; Merck, Dohme) essentially as described previously (64). Samples showing bacterial contamination, as determined by Gram staining, were discarded.

**PFGE.** One-tenth (10 μl) of the intact cell pellet suspension from each *P. carinii* preparation was reserved for pulsed-field gel electrophoresis (PFGE) performed essentially as described previously (64). The chromosome-bearing agarose blocks were electrophoresed in a 1% SeaKem GTG LE agarose gel (FMC) by using a contour-clamped homogeneous electric field (CHEF) DR-II apparatus (Bio-Rad) at 14°C and 6 V/cm for 40 h with a 20- to 60-s ramp.

**Nucleic acid isolation and library construction.** Genomic DNA and total RNA were isolated from fresh *P. carinii* cell pellets as described previously (64) by standard protocols (48). For library construction, DNA was extracted from *P. carinii* isolate 319931, derived from a single rat naturally coinfected with both the prototype and variant strains, as determined by PFGE karyotype analysis. The library was screened with probes end-labeled as described in Materials and Methods. The probe hybridized strongly to a ~330-kbp chromosome (arrowhead) in the *P. carinii* prototype isolate (lanes 3 to 5) but only weakly to the ~330-kbp chromosome in the *P. rattus* variant isolates coinfecting with a minority of *P. carinii* (lanes 6 and 7). (C) Southern blot analysis using a *Prbip* gene-specific oligonucleotide probe p380Br5′ end-labeled as described in Materials and Methods. The probe hybridized to a ~355-kbp *P. rattus* variant strain chromosome (arrowhead) but not to the ~330-kbp *P. carinii* prototype strain chromosome (lanes 6 and 7).

**A. PFGE**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

**B. *Prbip* probe**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

**C. *Prbip* probe**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 1.** Karyotype analysis of *Pc* and *Pr* genes from *P. carinii* prototype and *P. rattus* variant strains. (A) CHEF gel of *P. carinii* chromosomes. Lanes: 1, bacteriophage λ concatemer molecular size standards; 2, *S. cerevisiae* chromosome standards; 3 to 5, *P. carinii* prototype strain, isolates 101393, 102093, and 101193, respectively; 6 and 7, natural coinfections of *P. rattus* variant strain (predominant) and *P. carinii* prototype strain, isolates 319931 and 319932, respectively. The approximate sizes of the λ markers are indicated. (B) Southern blot analysis of gel shown in panel A using *Prbip* gene-specific oligonucleotide probe p380Ar 5′ end-labeled as described in Materials and Methods. The probe hybridized strongly to a ~330-kbp chromosome (arrowhead) in the *P. carinii* prototype isolate (lanes 3 to 5) but only weakly to the ~330-kbp chromosome in the *P. rattus* variant isolates coinfecting with a minority of *P. carinii* (lanes 6 and 7). (C) Southern blot analysis using a *Prbip* gene-specific oligonucleotide probe p380Br5′ end-labeled as described in Materials and Methods. The probe hybridized to a ~355-kbp *P. rattus* variant strain chromosome (arrowhead) but not to the ~330-kbp *P. carinii* prototype strain chromosome (lanes 6 and 7).
After treatment, the cells were pelleted at 13,000 rpm prior to heat shock or treatment with reagents as described in Results and Discussion. The cellsuspensions were diluted to an optical density at 600 nm of 0.5. The cell homogenates were strained through a sterile 60-mesh wire tissue sieve (VWR). Lung tissue homogenates were prepared in low-glucose Dulbecco modified Eagle medium (Life Technologies) supplemented with 2% fetal calf serum. Lung DNAs were isolated from infected lungs or control lungs, and stored at −70°C. RNA was isolated from infected lungs or control lungs; RNA was treated with RNase-free DNase (Promega) and stored at −70°C.

The RNA samples were heat shocked in the presence of heat shock protein 70 (HSP70) antibodies at 37°C for 30 min. The RNA samples were treated with RNase-free DNase (Promega) and stored at −70°C.

The cDNA template was synthesized from total P. carinii total RNA with reverse transcriptase (Life Technologies) with 400 μM dNTPs, 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl2, 200 units of RNase inhibitor, and 200 units of reverse transcriptase (Life Technologies). The cDNA products were purified with Glass-Max columns (Life Technologies) and poly(C)/tailed by using terminal deoxynucleotidyl transferase (New England Biolabs) and 0.5 mM dUTP. PCR primers were designed using the Primer3 program (15) with the OligoLabeling Kit (Pharmacia) or oligonucleotides 5′-GAGCTCTCGAGGTAATTCGCTGCTTTCGC-3′. The PCR products were purified with Glass-Max columns (Life Technologies) and amplified with 5.5 mM dNTPs, 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 200 units of Taq polymerase (Roche) with the RACE anchor primer and the gene-specific primer RACE154. Where not indicated, the buffers were those recommended by the enzyme manufacturers.

Cloning, cloning, and sequencing of PCR products. Specific PCR products were electrophoresed in 2% NuSieve GTG agarose (FMC) and excised. The purified DNA was digested with glass-mesh sieves and ligated into EcoRI-BamHI-linearized pUC19. All clones were verified by sequence analysis.

Southern and Northern blot analysis. Southern blots of CHEF gels were hybridized to the cloned HSP70-specific probes described above. Hybridization signals were quantified on a Betascope radioanalytic imager (Betascope). The sequences of Pcbip and Prbip have been deposited in GenBank under the accession numbers L46790 and U40994, respectively.

RESULTS AND DISCUSSION

Cloning of HSP70 homolog genes from prototype and variant rat P. carinii. Degenerate oligonucleotide primer pairs (HSP9-HSP6r, HSP8-HSP4, HSP6f-HSP4), targeted to distinct conserved domains of HSP70 genes, were used to amplify these domains from a prototype strain of rat-derived P. carinii as described in Materials and Methods. PCR amplification of P. carinii DNA with these primer pairs yielded products of the predicted size, whereas control PCR amplifications of rat liver DNA with the same primers produced no visible products (data not shown). Three specific PCR products were ligated into pUC19, sequenced, and found to encode 1,173 bp of overlapping DNA sequence with two discontinuous open reading frames exhibiting homology to eukaryotic HSP70 genes.

To verify the P. carinii origin of the PCR products, the cloned inserts from these plasmids were used as probes on Southern blots of CHEF gels containing P. carinii chromosomes (Fig. 1). P. carinii strains used for PCRs contained 17 chromosomes in a pattern similar to that described by Cushion et al. (9) for P. c. rattus (prototype strain) form 1, except that we observed an additional chromosome of ~555 kbp and consider chromosomes at ~450 and 430 kbp to be doublets (Fig. 1A, lanes 3 to 5). All three probes hybridized to a single 330-kbp P. c. carinii chromosome (data not shown, but see Fig. 1B, lanes 3 to 5). No hybridization to yeast or rat DNA controls was observed (data not shown), confirming the origin of the amplified DNA to be P. carinii, and not host, DNA.

A genomic library was generated in λ-DASH and screened with the cloned HSP70-specific PCR fragments as hybridization probes. The P. carinii used for generation of the genomic library was from isolate 319931 (Fig. 1A, lane 6), which, along with isolate 319932 (lane 7), displayed a complex CHEF karyotype indicating that these isolates were from natural coinfections of a single rat host with predominantly the P. c. rattus variant strain and a minor component of the P. c. carinii prototype strain as defined by Cushion et al. (9, 10). Two sets of

FIG. 2. Genomic context of Pcbip and Prbip. The genomic contexts of the BIP genes in P. c. carinii (A) and P. c. rattus (B) were characterized by Southern analysis using various oligonucleotides and PCR products as probes (data not shown). Boxed regions indicate the extent of the known DNA sequence. Shaded regions represent exons, with the relative positions of the transcription start point (+1), translation initiation (AUG), and termination (TAA) codons indicated. Unshaded boxed regions represent introns and the 5′ or 3′ untranscribed regions of the genes. Selected restriction sites are indicated, as are the approximate sizes of the restriction fragments. The exact localizations of the PstI restriction sites flanking Pcbip are not known. The regions between the HindIII sites flanking Pcbip and the HindIII and EcoRI sites flanking Prbip are drawn to scale.
FIG. 3. BiP DNA and protein sequences in prototype and variant strain rat P. carinii. (A) Alignment of BiP homolog genes from the P. c. carinii prototype strain (Pcbip) and the P. c. rattus variant strain (Prbip). Numbering is relative to the 111 transcription startpoint of Pcbip. Sequence identities between Prbip and Pcbip are shown as periods and gaps are indicated with hyphens in the Prbip sequence. As a result of poor homology, alignment of the 5' untranslated regions of these sequences was not possible and identities in this region are not indicated. Translation start (ATG) and stop (TAA) codons and putative polyadenylation signal sequences (AATAAA) are indicated in bold. Introns are shown in lowercase letters, and putative branchpoint consensus sequences of YTRAT (Y = pyrimidine and R = purine) are indicated. Potential promoter elements (GC-rich sequences, HSEs, and UPRs) in Pcbip are underlined, with conserved residues indicated by the following symbols:

(A)

---

(B)

Downloaded from http://iai.asm.org/ on January 6, 2021 by guest
clones were isolated and shown to bear HSP70-related genes that displayed similar but distinct restriction patterns (Fig. 2) and DNA sequences (Fig. 3). These two sets of genomic clones may have been derived from prototype and variant strains of P. carinii. This hypothesis was confirmed by hybridization of gene-specific probes to PFGE blots of chromosomes from the P. carinii isolate used for generation of the library and from known P. c. carinii prototype strain isolates (Fig. 1B and C).

Oligonucleotide probes specific for one of the cloned HSP70 homologs hybridized strongly to a 330-kbp chromosome in the prototype isolates (Fig. 1B, lanes 3 to 5) and weakly to a ~330-kbp chromosome present in the variant strain (Fig. 1C) but absent in the prototype strain. Finally, hybridization profiles of these gene-specific probes on Southern blots of restriction enzyme-digested genomic DNA from either the prototype or variant P. carinii were consistent with single-copy genes (data not shown). Thus, we concluded that we had cloned related HSP70 homologs from both prototype and variant strain P. carinii.

The cloned P. carinii HSP70 homologs are BiP genes. The sequences of the genes from the two P. carinii strains both exhibited four discontinuous open reading frames with homology to HSP70 and were interrupted by at least three introns (Fig. 2 and 3). The three introns and corresponding splice sites were confirmed by RT-PCR of total P. carinii RNA by using gene-specific primers and sequencing of the products (data not shown), and the 5′ end of the mRNA from the prototype P. c. carinii strain was cloned by RACE-PCR as described in Materials and Methods. The sequence of this clone identified the transcription start point (+1) and a fourth intron in the 5′ untranslated leader of the gene from this strain. We have been unable to identify the transcription start point (+1) of the mRNA of the gene from the variant P. c. rattus strain because of a lack of RNA. However, both of the P. carinii HSP70 homologs apparently consist of five exons and four introns and encode proteins of ~72.5 kDa from a translation initiation codon in exon 2. Northern analysis of P. c. carinii RNA showed a major transcript of ~2.5 kb (see below), consistent with the use of any of three potential poly(A) addition sequences 78, 156, and 252 bp downstream from the stop codon (Fig. 3), and a poly(A) tail length of 100 to 300 residues.

The HSP70 family includes molecular chaperones which function in protein folding and trafficking in various cellular compartments. Orthologous functional classes of these proteins evolve at similar rates in different organisms and are distinguishable from their paralogous counterparts, which arose by gene duplication and evolve independently by a low but unequal rate of evolutionary divergence in individual domains throughout the highly conserved sequences in each HSP70 class (26). Thus, to identify the function of the cloned P. carinii HSP70 homologs, we performed pairwise comparisons of their protein sequences with the sequences of HSP70 homologs with known function in other organisms. This analysis showed that the P. carinii homologs were more similar to known BiP ER resident HSP70s than to any other HSP70 family members (data not shown). Moreover, two sequence motifs exclusive to BiP HSP70 members are prominent in the proteins encoded by the HSP70-related genes isolated from the two P. carinii strains (Fig. 3B), namely, (i) a hydrophobic amino-terminal signal peptide and cleavage site required for targeting and translocation of BiP into the ER (6, 24) and (ii) a carboxyl-terminal signal sequence, XDEL, required for recycling BiP to the ER lumen (32, 37). The ER retention signal, HDEL, encoded by the prototype P. c. carinii clone is identical to that found in the yeasts Saccharomyces cerevisiae (39, 46) and Kluyveromyces lactis (49). The SDEL ER retention signal of the P. c. rattus HSP70-related protein is identical to that observed in Plasmodium falciparum BiP. Thus, we conclude that the cloned HSP70-related genes, PcbiP and PcbiP, encode BiP homologs PcBiP and PrBiP from the prototype and variant strains P. c. carinii and P. c. rattus, respectively.

Expression and stress induction of PcbiP. BiP genes in other organisms are expressed constitutively, but BiP mRNA levels are induced by conditions that cause an accumulation of unfolded or immature proteins in the ER, e.g., heat shock, glucose starvation mediated by α-deoxyglucose treatment, inhibition of N-linked glycosylation with tunicamycin, inhibition of export from the ER with brefeldin A, or perturbation of intracellular calcium levels by agents including the ER-specific Ca2+ ATPase inhibitor thapsigargin or the calcium ionophore A23187 (30, 31, 34, 43, 57). These treatments induce BiP mRNA levels in S. cerevisiae (39, 46, 57), Schizosaccharomyces pombe (42), and mammalian cells (31, 34, 43, 45), although mammalian BiP genes are not significantly induced by heat shock (61). We examined the P. carinii BiP mRNA levels in response to treatment with these agents in short-term in vitro cultures as described in Materials and Methods. Preincubations were either at 29°C to parallel analyses of BiP expression in S. cerevisiae and other fungi (42) or at 35°C to approximate temperatures within a host lung. PcbiP expression was measured by Northern hybridization of RNA isolated from short-term cultures of prototype P. carinii after treatment with the various agents (Fig. 4). Specific mRNA levels were compared with BiP mRNA levels in untreated cultures normalized to the ~1.4-kb P. c. carinii actin mRNA control detected with the P. carinii actin-specific probe PcaCT (Table 1).

PcbiP mRNA was detected under all conditions examined, suggesting that as for other organisms, expression of BiP in P. carinii is constitutive. However, the levels of PcbiP mRNA were clearly increased two- to threefold by a 39 or 42°C heat shock or by treatment with α-deoxyglucose, tunicamycin, or brefeldin A (Fig. 4; Table 1). Thus, the presence of malformed or improperly matured proteins apparently induces BiP expression in P. carinii much as it does in other organisms. In contrast, neither the calcium ionophore A23187 nor thapsigargin enhanced PcbiP mRNA levels under the conditions of these experiments (Fig. 4, lane 7; Table 1). Together, these results showed that PcbiP is regulated in patterns that are similar but not identical to those of previously characterized BiP genes in other organisms.

These analyses were limited by the lack of a robust continuous culture system for P. carinii, and therefore, we believe that the relative PcbiP stress response observed may have been underestimated. It is likely, if not probable, that the physical...
insult associated with isolation of *P. carinii* from the host lung induces a stress response in the remaining viable organisms. Such a stress response was previously reported during isolation of rabbit type II pneumocytes (5). Thus, these experiments demonstrated, but may underestimate, the induction of *Pcbip* mRNA levels in response to stress.

**cis-acting elements associated with *Pcbip* expression.** Promoter elements responsible for basal expression and induction of BiP genes have been identified in *S. cerevisiae* (28, 36) and some mammals (7, 45). Promoter elements responsible for constitutive BiP expression include GC-rich elements resembling Sp1 transcription factor bindingsites localized withina few hundred basepairs of the transcription start point of the gene. Heat-inducible BiP genes in *S. cerevisiae*, *S. pombe*, and *K. lactis* also contain multiple heat shock elements (HSEs), which bind specific heat shock transcription factors and act as controlling elements for transcriptional regulation in response to heat or other forms of stress. These HSEs are loosely conserved, with a consensus sequence of 5’-CNGAANNTTCT NNG-3’ or contiguous arrays of the 5-bp unit NGAA and NTTCN arranged in alternating orientations (reviewed in reference 33). Mammalian BiP genes, which are not heat inducible, lack functional HSEs (61). Finally, conditions leading to glucose starvation or to disruption of ER-golgi body transport, glycosylation, or intracellular Ca^{2+} levels mediate induction of BiP through a specific cis-acting DNA element referred to as an unfolded protein response (UPR) element in *S. cerevisiae* (28, 36). This element exhibits a loose consensus sequence with an analogous domain in vertebrate and *Caenorhabditis elegans* BiP promoters (45).

Since regulation of *Pcbip* follows patterns observed for BiP genes in other organisms, we examined the 5’ flanking region of *Pcbip* for sequences resembling conserved regulatory elements associated with BiP gene expression. This examination revealed multiple TATA-like elements, including one ~25 bp upstream from the *Pcbip* transcription start point, and two GC-rich elements (Fig. 3), which together could form a basal transcription promoter. Candidate HSEs and a UPR-like element can also be identified in the 5’ flanking region of *Pcbip* (Fig. 3). The putative UPR-like element overlaps one of the potential GC-rich elements, similar to that described for the UPR and GC-rich elements in the *S. cerevisiae* BiP gene. The presence of these elements in the *Pcbip* 5’ flanking region is consistent with our empirical observations demonstrating stress response regulation of this gene, although verification of the functions of these putative regulatory elements is currently not possible because of the lack of a genetic system for *P. carinii*. Similar elements can also be identified in the *Pribp* gene sequence, but we have less confidence in their localization because we have been unable to map the transcription start point of this gene and the 5’ untranslated regions of *Pcbip* and *Pribp* are markedly divergent.

All introns in both *Pcbip* and *Pribp* conform to the consensus 5’ GYA splice donor sites and 3’ YAG splice acceptor sites of fungi (Fig. 3). The GC dinucleotide in the 5’ splice donor site of *Pribp* intron 2 is an unusual variant of the nearly ubiquitous GT (GU in RNA) dinucleotide but has been observed previously in fungal introns and introns in the *P. carinii* dihydrofolate reductase and thymidylate synthase genes (13, 16). Each BiP intron also displays a possible branch point consensus of T-YTRAT, similar to the *S. pombe* consensus CTRAY (35), within 20 bp of the 3’ splice acceptor sites (Fig. 3).

**Molecular divergence between *Pribp* and *Pcbip*.** Between the translation initiation and stop codons, *Pcbip* is a total of 29 bp (20 intronic bp and 9 coding bp) larger than *Pribp*. Pairwise alignments showed ~83% identity between the coding sequences of *Pcbip* and *Pribp* but only ~58% identity within the introns (Fig. 3). Thus, the introns in these two BiP genes are conserved in size and position but are not highly conserved in sequence. In contrast, the *P. carinii* BiP gene introns show no conservation in size or position when compared with introns in

### Table 1. Induction of *Pcbip* expression in transient in vitro *P. carinii* culture

<table>
<thead>
<tr>
<th>Isolates (no. of rats)</th>
<th>Control temp (°C)</th>
<th>Heat shock temp (°C)</th>
<th>Relative fold induction of <em>Pcbip</em> mRNA* treated with:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HS 10’</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>39</td>
<td>2.0 ± 0.2</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>42</td>
<td>1.3 ± 0.1</td>
</tr>
</tbody>
</table>

* Fold induction of *Pcbip* relative to untreated *P. carinii* (data not shown) and normalized to actin expression by using *P. carinii*-specific oligonucleotide probes p380Ar and PcAct, as described in Materials and Methods. HS 10’ and 30’, heat shock for 10 and 30 min, respectively. Tunc, 1 mM tunicamycin; α-d-glc, 10 mM α-deoxyglucose; Bref-A, 5 μM brefeldin A; Thaps, 2 μM thapsigargin; A23187, 10 μM A23187 (calcium ionophore).

* ND, not determined.
the BiP genes of other organisms (data not shown). A convincing alignment of the 5′ flanking noncoding domains of Prbip and Pcbip was not possible, and we were therefore unable to verify, in Pcbip, the presence of an intron corresponding to that in the 5′ untranslated region of Pcbip. The AT compositions of Pcbip and Prbip were ~61 and 63% for coding sequence, respectively, and ~62 and ~66% including intron and 5′ flanking sequence, respectively, thus conforming to the low-GC and high-AT codon bias previously reported for rat P. carinii genes (18). However, codon usage varied significantly between these genes since ~44% of the codons differ.

Apparent amino acid sequence identity between the PcBiP and PrBiP proteins is ~90% (Fig. 3B), which is greater than that between the P. carinii BiP genes and other P. carinii HSP70 homologs or HSP70 homologs of other organisms (51). However, the BiP homologs are more divergent than other proteins and RNAs that have been examined in both the prototype and variant strains of P. carinii; e.g., thymidylate synthase, β-tubulin, arom, 23S mitochondrial rRNA, and 18S nuclear rRNA show ~93 to 96% identity (10, 53). The homology observed between the P. carinii BiPs is considerably lower than the ~98% observed between rat and human BiPs and is comparable to the ~90% observed between the BiPs of the distantly related trypanosomes Trypanosoma cruzi and Trypanosoma brucei (Table 2). Together, these data show a surprisingly high degree of divergence between the normally highly conserved BiP gene homologs of these two P. carinii strains and are consistent with previous suggestions (9, 10, 52) that the prototype and variant rat P. carinii strains could be considered different species.

**Phylogenetic analysis of PcBiP and PrBiP.** BiP genes and proteins are ideal for phylogenetic analysis because of their high degree of primary sequence conservation. Thus, the PcBiP and PrBiP protein sequences were aligned with BiP sequences from representative members of each of the major eukaryotic kingdoms by use of the Pileup and Lineup programs in the GCG Sequence Analysis Package (data not shown). The aligned sequences extended from Gly-36 to Ala-603 of PcBiP (Gly-35 to Ala-602 of PrBiP), excluding the amino and carboxyl termini which could not be aligned unambiguously. An alignment of the cDNAs of each BiP homolog was also constructed on the basis of the protein alignment and corrected for codon placement in gaps (data not shown). The sequence of the prokaryotic HSP70 homolog Escherichia coli DnaK was included in the analysis to root the BiP phylogenetic trees.

Two methods of assessing phylogenetic distance were employed for tree construction. First, a protein parsimony approach using the PHYLIP (version 3.5) program PROTPARS with the global search option (17) generated a tree with the topology shown in Fig. 5A. More-favorable phylogenetic trees were sought by swapping and rearranging branches across small branch lengths. The initial alignment was bootstrapped 250 times with the program SEQBOOT, and a consensus tree was derived with the program CONSENSE to obtain bootstrap confidence values for clades found in the parsimony tree. Second, a distance matrix approach was performed with the DNA alignment by use of the maximum likelihood distance algorithm in DNADIST. The alignment was weighted to consider only the first two bases of each codon to reduce homoplastic noise associated with the third base position, attributable to large evolutionary distance and the strong codon bias of several of the organisms, including Giardia lamblia, P. falciparum, and P. carinii. The alignment was bootstrapped 250 times, the maximum likelihood distance matrix data were collected, and a consensus neighbor-joining (47) tree was generated (Fig. 5B).

The trees generated by the protein parsimony and DNA distance approaches are similar. Both methods group species belonging to the metazoa and higher fungi as monophyletic taxa, with bootstrap confidence supporting these groups at or near 100%. In addition, P. c. carinii and P. c. rattus grouped as terminal ancestors and with the ascomycetes S. pombe, K. lactis, and S. cerevisiae, as a monophyletic group distinct from other groups of eukaryotes. The first genetic evidence for classification of P. carinii with the fungi was based on the sequence of the 18S rRNA (14, 56). Wakefield et al. (60) analyzed mitochondrial rRNA to attempt a more definitive placement of P. carinii within the fungi and proposed similarity with the ustomycetous red yeasts, although a sequence-based phylogenetic analysis was not shown. More recently, analysis of the α-actin gene also supported classification of P. carinii with the fungi (19). Our analysis of the P. carinii BiP homologs, which are large, highly conserved proteins with many phylogenetically informative sites, strongly supports inclusion of P. c. carinii and P. c. rattus among the higher fungi, as represented here by three ascomycetes. However, relative placement of the P. carinii strains within this group is currently precluded by the paucity of available fungal BiP sequences.

Members of the Proctoctista displayed a polyphyletic origin in these analyses. The oomycete Phytophthora cinnamomi branched well before the fungi, supporting early divergence of oomycetes prior to speciation of the fungi proper (1, 27). Not surprisingly, a unique rooting point could not be determined for the apicomplexan malarial parasite P. falciparum, which branched inconsistently at the base of fungal and plant clades in trees generated by both methods. Apicomplexa share apparent ancestry with both dinoflagellates and ciliates (20), are proposed to have arisen via an endosymbiotic pathway involving an ancestral eukaryote and a rhodophytic alga (63), and
contain residual plastid genomes. Like BiP, the nuclear-encoded *P. falciparum* enolase also resembles plant enolases more closely than those of other eukaryotes (44), perhaps suggesting a major influence of the ancestral plastid genome on the nuclear genome. However, a large mutational distance separates *Plasmodium* spp. from other eukaryotes, and these organisms have been subjected to accelerated rates of evolutionary change (20), a property known to confound the construction of evolutionary trees. Of the lower protists, the trypanosomes were monophyletic, as was the primitive mitochondrion-lacking eukaryote *G. lamblia*.

Previous phylogenetic analyses with members of the HSP70 gene family have included a few BiP sequences (4, 22, 39). Despite small branch lengths and although not supported at a high level of confidence by bootstrap methods, we infer from the BiP trees presented here a more recent common ancestry for animals and fungi than for either one with plants, consistent with proposals of others (38, 59) in this long-standing controversy. Our results are in contrast to a previous parsimony analysis using members of the HSP70 family of proteins (22), which concluded that the BiP homolog of *S. cerevisiae* is deep branching among the eukaryotes. The consensus tree in the latter analysis did not support this topology, however, and more importantly, *S. cerevisiae* was the only fungus examined. In contrast, several members of the ascomycetes and related fungi have been used herein, permitting a clearer resolution of the evolutionary lineage of higher fungi.

ACKNOWLEDGMENTS

We thank Tom Reynolds and the MCV-VCU Nucleic Acids Core Facility for oligonucleotide synthesis and automated sequencing. This work was supported by grants from the NIH, the Jeffress Memorial Trust, and the A. D. Williams Foundation.

REFERENCES


Editor: D. H. Howard