

Post-transcriptional Repair of a Split Heat Shock Protein 90 Gene by mRNA *trans*-Splicing*[§]♦

Received for publication, December 3, 2010, and in revised form, December 31, 2010. Published, JBC Papers in Press, January 5, 2011, DOI 10.1074/jbc.C110.208389

Rishi Kumar Nageshan[‡], Nainita Roy[‡], Adrian B. Hehl^{§1}, and Utpal Tatu^{‡2}

From the [‡]Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India and the [§]Institute of Parasitology, University of Zürich, Winterthurerstrasse 266a, CH-8046 Zürich, Switzerland

Heat shock protein 90 participates in diverse biological processes ranging from protein folding, cell cycle, signal transduction and development to evolution in all eukaryotes. It is also critically involved in regulating growth of protozoa such as *Dictyostelium discoideum*, *Leishmania donovani*, *Plasmodium falciparum*, *Trypanosoma cruzi*, and *Trypanosoma evansi*. Selective inhibition of Hsp90 has also been explored as an intervention strategy against important human diseases such as cancer, malaria, or trypanosomiasis. *Giardia lamblia*, a simple protozoan parasite of humans and animals, is an important cause of diarrheal disease with significant morbidity and some mortality in tropical countries. Here we show that the *G. lamblia* cytosolic *hsp90* (*glhsp90*) is split in two similar sized fragments located 777 kb apart on the same scaffold. Intrigued by this unique arrangement, which appears to be specific for the Giardiinae, we have investigated the biosynthesis of GlHsp90. We used genome sequencing to confirm the split nature of the giardial *hsp90*. However, a specific antibody raised against the peptide detected a product with a mass of about 80 kDa, suggesting a post-transcriptional rescue of the genomic defect. We show evidence for the joining of the two independent Hsp90 transcripts *in-trans* to one long mature mRNA presumably by RNA splicing. The splicing junction carries hallmarks of classical *cis*-spliced introns, suggesting that the regular *cis*-splicing machinery may be sufficient for repair of the open reading frame. A complementary 26-nt sequence in the “intron” regions adjacent to the splice sites may assist in positioning the two pre-mRNAs for processing. This is the first example of post-transcriptional rescue of a split gene by *trans*-splicing.

Whole genome sequencing has galvanized investigation of biological processes and host-pathogen interactions. Protozoan parasites cause life-threatening diseases in humans and

animals accounting for hundreds of millions of deaths annually worldwide. In the last 10 years, the genome sequences of many important pathogens such as *Giardia lamblia*, *Leishmania donovani*, *Plasmodium falciparum*, *Trichomonas vaginalis*, and *Trypanosoma cruzi* have been completed (1–5) and have helped biological research toward development of new drug candidates. The first partial genome sequence of *G. lamblia* was reported in 2000 (4), and more recently, in 2007, the full genome sequence was published and is available on GiardiaDB (5). *G. lamblia* (syn. *Giardia intestinalis*, *Giardia duodenalis*) is an extracellular enteroparasite causing giardiasis, a diarrheal disease, in humans and animals. Initially thought to be a primitive eukaryote, *G. lamblia* turned out to be a highly simplified extracellular parasite that has undergone significant reductive evolution (6). The rapidly accumulating genomic and post-genomic data on *Giardia* and related genera now provide the necessary background to investigate complex biological processes in a comparative manner across species boundaries.

Heat shock protein 90 (Hsp90)³ is an essential molecular chaperone that is implicated in growth and development of many protozoan species such as *Dictyostelium*, *Leishmania*, *Plasmodium*, and *Trypanosoma* (7–10). Hsp90 has also been explored as a drug target in *Plasmodium* and *Trypanosoma* (10). Recent years have seen many new roles being ascribed to Hsp90. In addition to its ability to fold proteins and regulate cell cycle and development, recent studies also suggest an ability to guide evolution in eukaryotes (11–13). The function of Hsp90 as a sensor of environmental cues is especially crucial in protozoan parasites, which often need to respond to radical changes of milieu within and outside their hosts (7, 10).

In all organisms investigated to date, Hsp90 proteins are encoded by a single open reading frame (ORF), which usually contains multiple introns. In the genome sequence of three *Giardia* isolates, no contiguous *hsp90* ORF was predicted, but two fragments separated by a large stretch of sequence on the same scaffold were detected and annotated as *hsp90*. Here we investigated the split nature of *glhsp90* and the consequences of this unique genetic rearrangement. We report a post-transcriptional repair mechanism that generates a *trans*-spliced, mature mRNA coding for a *bona fide* Hsp90 protein from the two Hsp90 pre-mRNAs.

* This work was supported by grants from the Department of Science and Technology, New Delhi, the Department of Biotechnology, New Delhi, India, and MALSIG, by a Junior Research Fellowship from the Council of Scientific and Industrial Research and United Grants Commission, New Delhi, India (to R. K. N. and N. R.), and by Swiss National Science Foundation Grant 31003A125389 (to A. B. H.).

♦ This article was selected as a Paper of the Week.

§ The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Fig. S1 and Table S1.

¹ To whom correspondence may be addressed. Tel.: 41-44-635-8526/30 and 41-44-635-8501; E-mail: adrian.hehl@access.uzh.ch.

² To whom correspondence may be addressed. Tel.: 011-91-80-22932823; E-mail: tatu@biochem.iisc.ernet.in.

³ The abbreviations used are: Hsp90, heat shock protein 90; ACN, acetonitrile; 17AAG, 17-(allylamino)-17-demethoxygeldanamycin; nt, nucleotide(s).

EXPERIMENTAL PROCEDURES

Cultivation of Parasites—*G. lamblia* Portland P1 or WB-C6 (assemblage A) parasites were cultured in TYI-S33 (14) supplemented with 12% fetal bovine serum and subcultured with 5×10^4 cells/tube from log phase parasites. The parasites were harvested by chilling on ice for 20 min followed by repeatedly inverting the tubes to dislodge the parasites and finally pelleted down at $700 \times g$ for 5 min.

PCR and Whole Genome Sequencing—Genomic DNA was isolated as described previously (15) with minor modifications. To confirm the position of *hspN*, the following primers against the flanking (200 bp upstream and downstream) region were used: sense, 5'-GGGCATGAGCTGCCGG-3', and antisense, 5'-GTCTGCACGCCATAGACGC-3'. Similarly, the *hspC* (200 bp upstream and downstream) gene position was also confirmed from sense primer 5'-CCGCATGCTGAGGGTGC-3' and antisense primer 5'-CCGTGCAGCTCTAGCACAATTAC-3'.

Total RNA was prepared by TRI Reagent (Ambion) according to the manufacturer's protocol. Five μ g of total RNA was used for cDNA preparation using oligo(dT) primers (Fermentas cDNA kit). A full-length *hsp90* ORF was amplified with specific primers overlapping the *hsp90* start codon of *hspN* (ORF 98054 in GiardiaDB) (sense, 5'-ATGCCCGCTGAAGTCTTCGAGTTCCAG-3') and *hspC* (ORF 13864 in GiardiaDB) (antisense, 5'-TCAGTCAACTTCGTCAACGTCCTCCTC-3'). As an independent determination of the exact site of the transition from the transcript derived from ORF 98054 into that derived from ORF 13864, a PCR fragment from a cDNA template was generated with internal primers HspC-internal (sense, 5'-GCGAATTCAGGTCCACGAGCACGTGAAC-3') and HspN-internal (antisense 5'-GCGAATTCCTGTGATGTAGTAGATCGAC-3'). The resulting 640-bp product was cloned into the EcoRI restriction sites of pBluescript (Stratagene) and sequenced.

To rule out the possibility of *cis*-splicing, PCR was carried out using sense and antisense primers, 5'-ATGCTCCAGAA-GAATCGC-3' and 5'-CCGTGCAGCTCTAGCACAATTAC-3', respectively, as test primers. Amplification of the middle domain served as a positive control using 5'-CAGGGACCCGAAGGACGTCACCGAG-3' as sense and 5'-CATATACTCGTCGATGGCCTCGTCC-3' as antisense primers. RNA was extracted using TRI Reagent as described previously. RT-PCR was carried out using the above primers targeted toward amplifying a specific product to look for the presence of any putative long mRNA transcript.

Whole genome sequencing was carried out with the paired end genomic sequencing methodology with 72-bp read lengths in the Illumina GA II X sequencer. More than 17 million reads were obtained, and 15.7 million high quality reads were passed through the pipeline for alignment with the reference genome sequence (*G. lamblia* WB, ATCC 50803). The genome coverage was calculated at $165\times$, and SNPs and Indels were tabulated.

Western Blot Analysis—Parasites were lysed with 20 mM Tris HCl, pH 6.8, with 1% Triton X-100, and protease inhibitor mixture (G-Biosciences). A high speed supernatant was

resolved on a reducing 10% SDS-PAGE gel and blotted to nitrocellulose filters. A rabbit anti-GlHsp90 antibody was raised against a peptide, NKQPALWTRDPKDVTEDE, specific to HspN (Custom Synthesis, Mumbai, India) and was used to probe the filters.

In-gel Digestion—A narrow slice corresponding to a Gl-Hsp90 band was cut from the stained SDS-PAGE gel and further sliced into smaller gel plugs. After several washes with 100 mM ammonium bicarbonate (NH_4HCO_3) (Sigma-Aldrich) buffer in 50% acetonitrile (ACN), the gel plugs were subjected to a reduction step using 10 mM dithiothreitol (DTT) (Sigma-Aldrich) in 100 mM NH_4HCO_3 buffer (45 min at 56 °C). Alkylation was performed with a solution of 55 mM iodoacetamide (Sigma-Aldrich) in 100 mM NH_4HCO_3 (30 min at room temperature in the dark) followed by in-gel digestion with 20 μ l of trypsin (10 ng/ μ l) (Promega) in 50 mM NH_4HCO_3 (overnight at 37 °C). The reaction was stopped by storing at -20°C , and peptides were extracted in 5% formic acid. Samples were vacuum-dried and reconstituted in 5% formic acid.

Mass Spectrometry and Database Searching—The protein digest was analyzed by automated nanoflow LC-MS/MS. The sample was loaded onto PepMap C18 reverse phase column connected to a Tempo nano-HPLC system. The peptides were eluted from the analytical column by a linear gradient of 95% Solvent A (98% H_2O , 2% ACN, and 0.1% formic acid) to 10% Solvent A and Solvent B (2% H_2O , 98% ACN, 0.1% formic acid). The spectra were acquired on a Q-STAR Elite mass spectrometer equipped with Applied Biosystems NanoSpray II ion source. The data were acquired in a data-dependent mode, one MS spectrum followed by three MS/MS spectra. Data analysis was performed with the Analyst QS 2.0 software. For identification of proteins, the processed data were searched against NCBI nr database using the Mascot protein identification tool version 2.0 with precursor and fragment mass tolerances of 0.15 Da, cysteine carbamidomethylation as fixed modification, and methionine oxidation, lysine acetylation, glutamine, and asparagine deamidation as variable modifications. The resulting MS/MS-based peptide identifications were manually verified in the ProteinPilot 2.0 software.

Determination of Effect of Hsp90 Inhibition—Parasites from log phase were harvested, and 1.5×10^4 cells were treated with 0–10 mM concentration of 17-(allylamino)-17-demethoxygeldanamycin (17AAG) with 10-fold dilutions. 0.5% dimethyl sulfoxide (DMSO) was used as control in 300 μ l of TYI-S33 culture medium. The assay was performed in 96-well plates that were sealed with Parafilm and incubated in a humidified 5% CO_2 incubator. The parasite viability was scored by trypan blue exclusion and by assessing the motility of the parasites as described previously (16).

RESULTS

A Genome Search Reveals Two Partial ORFs Homologous to Hsp90—The *Giardia* genome (isolate WB, assemblage A) codes for two ORFs coding for predicted proteins with strong homologies to *hsp90*. The N-terminal domain with a part of a middle domain, annotated in GiardiaDB and designated

trans-Spliced Heat Shock Protein 90 Gene Product

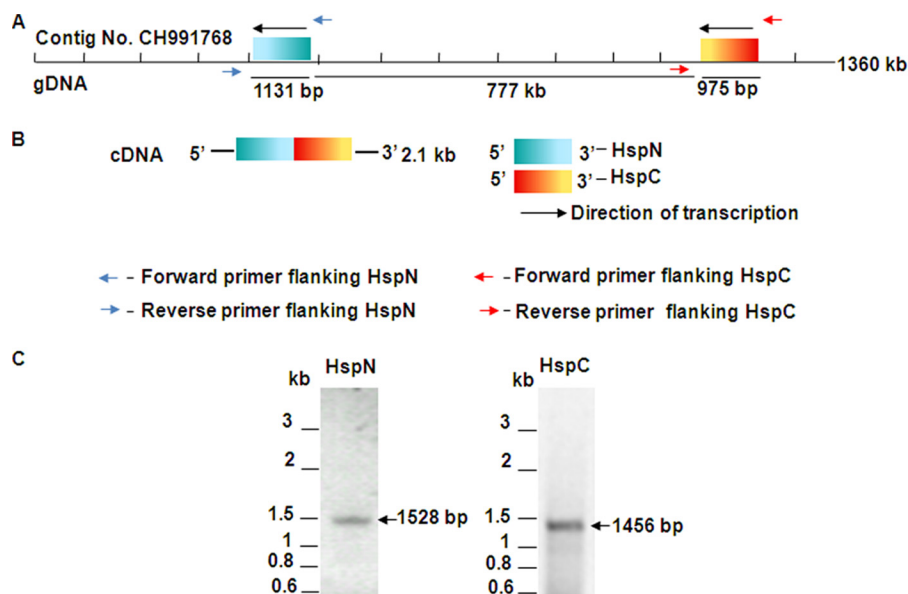


FIGURE 1. *hsp90* is arranged as a split gene in *G. lamblia* genome. A, representation of gene position of *glhsp90* as reported previously by Morrison *et al.* (5). Blue and red arrows represent primer sets for *hspN* and *hspC*, respectively. Contig No., number of group of overlapping clones; gDNA, genomic DNA. B, graphical representation of fused cDNA of *GIHsp90*, with *hspN* and *hspC* fused in head-to-tail pattern. C, PCR amplification using specific primers against the flanking region of *hspN* and *hspC* (200 bp upstream sense primer and downstream antisense primer) gives a specific 1.5-kb-long (*hspN*) and 1.4-kb-long (*hspC*) product, respectively, confirming the genomic position and organization of the *hsp90* split gene.

here as *hspN* (GL5083_98054, 376 amino acids), and the remaining middle domain with the C-terminal domain ending with DEVD, represented here as *hspC* (ORF GL50803_13864, 324 amino acids) (5). The predicted ORFs as well as the presence of a large intervening genomic region of 777 kb suggested that the *glhsp90* fragments were transcribed as separate mRNAs (Fig. 1A). We also noticed that a catalytic Arg residue, which is a hallmark of all Hsp90s reported so far, was not encoded in either of the predicted ORFs in GiardiaDB.

Genome resequencing using a Portland isolate of *G. lamblia* ruled out the possibility that the split gene was erroneously annotated. Paired end whole genome sequencing was carried out as described under "Experimental Procedures." The tagged genomic library was sequenced at 72-bp read lengths. About 15.7 million high quality reads were obtained, out of which 15.3 million reads aligned to the previously published reference genome of *G. lamblia*, which covered about 97% of the whole genome. We obtained significant coverage of about 168× of the reference sequence. As shown in Fig. 1A, we were able to confirm *hsp90*, represented as two fragments. We did, however, observe several single nucleotide polymorphisms and other minor InDels (supplemental Table S1). Recently, the sequences of two additional assemblages (B and E) were integrated into GiardiaDB (29). The gene pages for the two predicted *hsp90* ORFs show a high degree of synteny in the surrounding 10 kbp. Further, the genomic organization was confirmed by PCR using primers specific to flanking regions of both *hspN* and *hspC* ORFs. We used primers 200 bp upstream and downstream of these ORFs. PCR was carried out using genomic DNA isolated from *G. lamblia* culture as described under "Experimental Procedures." As shown in the Fig. 1C, specific amplicons of the expected size of 1.5 and 1.4 kb were obtained on using spe-

cific primers for *hspN* and *hspC*, respectively. These results confirmed the split gene organization of *hsp90* in the genome of *G. lamblia*. Altogether, the whole genome sequencing as well as PCR approach confirmed the presence of one *hspN* and one *hspC* fragment each, separated by a stretch of 777 kb of genomic sequence.

Identification of a Large Hsp90 Protein in Giardia—In contradiction to the genomic organization of *hsp90* as two fragments, we identified a large protein using a polyclonal anti-*GIHsp90* antibody raised against a peptide in the HspN sequence. Western blot analysis of separated *G. lamblia* lysates incubated with the antibodies revealed a specific band at ~80 kDa (Fig. 2A). To confirm the presence of an Hsp90 protein of this mass, we excised a narrow Coomassie Brilliant Blue-stained gel strip for in-gel digestion with trypsin and analysis by mass spectrometry. The analysis showed the significant presence of tryptic peptides corresponding to both HspN-derived and HspC-derived proteins (Fig. 2, B and E). Indirect immunofluorescence analysis using the anti-Hsp90 serum showed a cytoplasmic localization of the detected protein, which appeared to be excluded from the nuclei (Fig. 2D). This is consistent with the subcellular distribution of Hsp90 proteins in other eukaryotes.

Cytoplasmic Hsp90 chaperones are essential proteins required for the growth and development of many protozoa (8–10, 17, 18). Lethality of *hsp90* gene knock-outs in many eukaryotes (19, 20) and sensitivity to its inhibitor Geldanamycin in several protozoa including *Eimeria sp.*, *L. donovani*, *P. falciparum*, *Toxoplasma gondii*, as well as *Trypanosoma sp.* (9, 10, 21, 22) have been demonstrated. We treated *G. lamblia* cells with the Geldanamycin derivative 17AAG *in vitro* and found a dose-dependent inhibition of growth. The IC₅₀ value

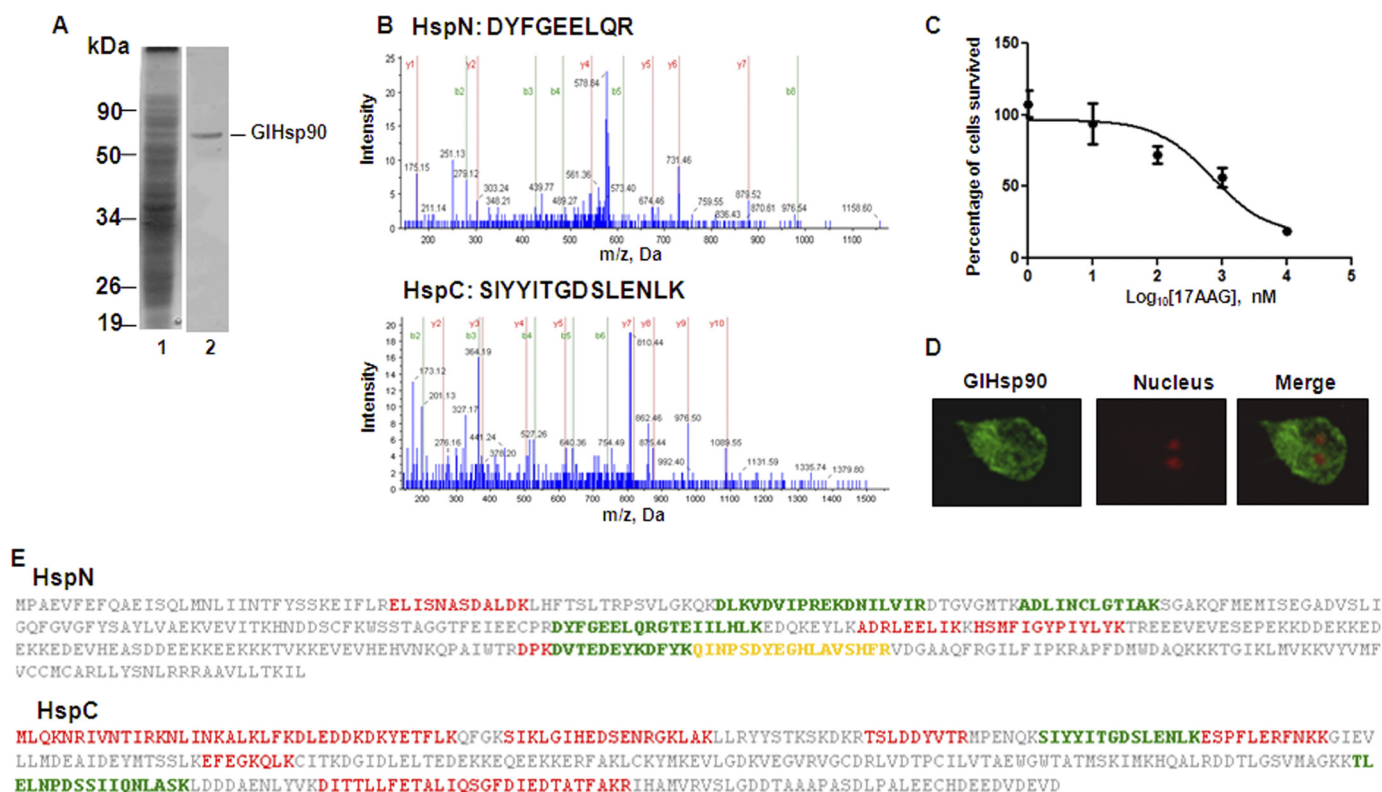


FIGURE 2. **A** full-length cytosolic Hsp90 in *G. lamblia*. **A**, GIHsp90 is expressed as a full-length, 80-kDa protein. Lane 1, total protein profile from the trophozoite stage of *G. lamblia* cells fractionated on 10% SDS-PAGE stained with Coomassie Brilliant Blue. Lane 2, Western blot analysis of total protein probed with anti-GIHsp90 antibody showing a specific band corresponding to 80 kDa. **B**, MS/MS spectrum of signature peptides from HspN and HspC from the 80-kDa band. **C**, GIHsp90 is essential for growth of parasites *in vitro*. The graph represents the percentage of survival of parasites against increasing concentrations of 17AAG. The IC₅₀ value of 17AAG was determined to be 711 nM *in vitro*. Error bars indicate S.D. **D**, indirect immunofluorescence analysis using anti-GIHsp90 antibody shows a diffused cytosolic localization of GIHsp90 in the trophozoite stage of the parasite. **E**, HspN and HspC peptides identified by MS/MS analysis are color-coded where green, yellow, and red indicate peptides identified with 95–100%, 59–95%, and <50% confidence, respectively.

was calculated at 711 nM, which is in accordance with that in other organisms.

A Long mRNA Spanning Both the HspN and the HspC Sequences Is Present in the Transcriptome—Identification of a large protein with the anti-Hsp90 antiserum indicated the presence of a full-sized Hsp90 in *G. lamblia*. Based on the available data, we predicted the presence of an mRNA combining the HspN and HspC fragments in the *G. lamblia* transcriptome. To test this hypothesis, we used RT-PCR to amplify a predicted full Hsp90 mRNA as well as delineate a possible junction between the two Hsp90 fragments from total RNA. The external primers covering the start of the *hspN* and the end of the *hspC* coding regions as shown in GiardiaDB yielded a PCR product of ~2.1 kb (Fig. 3A). A smaller, 635-nt fragment was amplified with internal primers *hspN*-internal and *hspC*-internal. This product was not obtained if genomic DNA was used as a template in the reaction as a control (Fig. 3B). Sequencing of ~2.1 kb product revealed a contiguous coding region of 2112 nucleotides starting with the predicted ATG codon of ORF 98054 (*hspN*) and ending with the predicted stop codon of ORF 13864 (*hspC*) (supplemental Fig. S1). We reasoned that the sequenced mRNA was a splicing product combining the *hspN* fragment somewhere upstream of the predicted stop codon with the *hspC* fragment. Indeed, we could determine a transition from one into the other at 1037 nt downstream of the predicted start codon

inside the ORF 98054 coding region (Fig. 3C). The opposite splicing site in the *hspC* fragment localized 99 nt upstream of the predicted ORF 13864, however. Thus, although the splicing reaction discarded 30 codons of ORF 98054, 33 codons were added from the region upstream of the predicted ORF 13864. The combined mRNA codes for a giardial Hsp90 of 704 amino acids whose closest homologs are the Hsp90 proteins in the related species *Spironucleus barkhanus* (696 amino acids, 72% identity) and *Retortamonas* sp. (602 amino acids, 68% identity). We ruled out the possibility of a putative large pre-mRNA, which on *cis*-splicing could result in mature Hsp90 mRNA. We have attempted amplification of the region encompassing the coding region of *hspC* and 197 bp downstream of *hspC* in the non-coding region (Fig. 3D, *discontinuous arrows*) using a PCR-based approach. Using the primer set described under “Experimental Procedures,” we used either genomic DNA or cDNA as template. As shown in Fig. 3D, on using cDNA as template, there was no amplicon obtained on RT-PCR using the primer set corresponding to *hspC* ORF and downstream of it (lane 1). On the other hand, specific amplification was observed when genomic DNA was used as a template with the same set of primers (Fig. 3D, lane 2). In the case of a *cis*-splicing event, we would have observed a 1172-bp-long amplicon on RT-PCR. The absence of a specific product on RT-PCR confirmed the absence of the puta-

trans-Spliced Heat Shock Protein 90 Gene Product

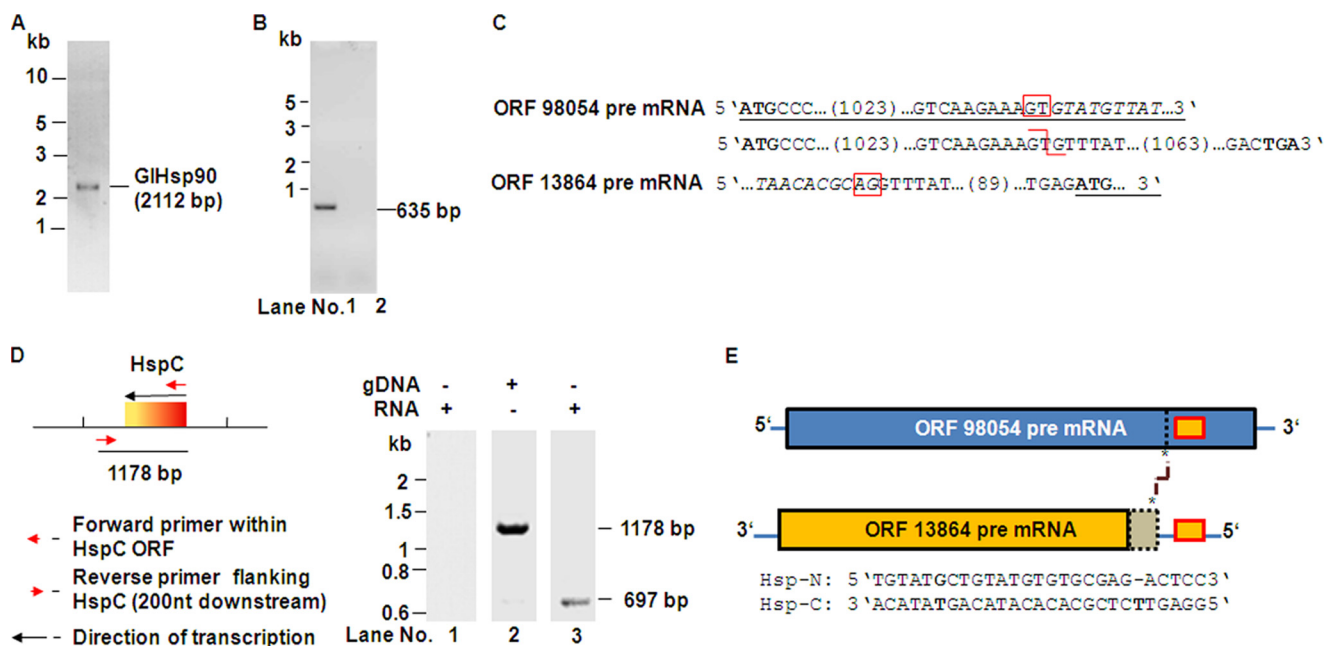


FIGURE 3. Splicing junction of GIHsp90. A, full-length GIHsp90 amplified from cDNA, using sense primer from *hspN* and antisense primer against *hspC*. B, a PCR product spanning the splice site of the mature Hsp90 mRNA could be amplified with internal primers from cDNA (lane 1) but not from genomic DNA (lane 2). C, alignment showing the “exon” borders (top and bottom lines) and the transition sequence in the mature GIHsp90 mRNA (middle line) determined by sequencing of cDNAs. The sequence segments coding for predicted *hsp90* fragments in GiardiaDB are underlined. The GU and AG dinucleotides at the intron borders are boxed. Discarded sequences are shown in *italic letters*. The transition point in the sequence of the spliced mRNA is depicted as an angled red line. The numbers in brackets indicate the number of nucleotides. Predicted start and stop codons are in **boldfaced letters**. D, representation of primers on the 777-kb fragment showing the annealing sites at *hspC* ORF and downstream flanking region. Red arrows represent primer annealing sites at *hspC* ORF and downstream flanking region. Arrows in black represent the direction of transcription of *hspC* ORF. Amplification using forward primer annealing within *hspC* ORF and reverse primer against flanking region of *hspC* (197 bp downstream) could not amplify gene product from RNA (lane 1), whereas specific product was obtained on using genomic DNA as template (lane 2). The middle domain served as a positive control (lane 3) for RNA. E, graphical depiction of the complementary 26-mer motifs (red boxes) on the two Hsp90 pre-mRNAs. Base pairing would bring the cleavage sites (asterisks) into close approximation. The respective sequences and orientations are shown below. Dotted lines indicate ORF boundaries that do not correspond to those predicted in GiardiaDB.

tive 779-kb-long transcript in the RNA pool, thus ruling out *cis*-splicing as a mechanism of RNA processing.

Based on these findings, we postulate that *G. lamblia* uses splicing of two separate mRNAs *in-trans* as a means to correct the genetic split of *hsp90* and to generate a mature mRNA coding for the ~80-kDa Hsp90 protein identified by Western blot. This is supported by the presence of canonical GU and AG dinucleotides at the “intron” borders (Fig. 3C), suggesting that the two pre-mRNAs are substrates for the conventional *cis*-splicing machinery. This leaves the question of how they could be spatially linked to allow a splicing reaction. Interestingly, we identified an almost perfectly matching, complementary region 6 nt downstream of the splicing junction in the *hspN* sequence and 35 nt upstream of the splice site in the *hspC* sequence. As shown in Fig. 3E, base pairing of the 26-nt regions of the pre-mRNAs would bring the splicing sites into close proximity.

DISCUSSION

Hsp90 is a highly conserved molecular chaperone that assists protein folding and participates in the regulation of cell cycle as well as in signal transduction pathways in eukaryotes. The list of clients regulated by Hsp90 is growing, as are its roles in different biological processes. With the exception of the *Giardia* homolog, *hsp90* is coded by a single gene in all biological systems examined so far and has been shown to possess variable number of introns (23). It is constitutively

expressed at high levels under all stages of cell cycle and development and is capable of further induction in response to stress (24, 25).

Examination of the *G. lamblia* genomic sequence revealed *glhsp90* to be represented as two separate genes, interrupted by a 777-kb stretch. On multiple sequence alignment of GIHsp90 with canonical Hsp90s, we observed the two genes, namely *hspN* and *hspC*, to align with the N termini and C termini of canonical *hsp90* gene sequences. To rule out the possibility that the split *hsp90* gene was a result of erroneously aligned genome fragments, we resequenced the whole genome of *G. lamblia* and confirmed that indeed the N and C termini of GIHsp90 are coded by two independent genes separated by a 777-kb intervening sequence. This sole example of a split *hsp90* gene intrigued us to further study Hsp90 from *G. lamblia*.

Are these fragments independently expressed, or is there a post-transcriptional event that would result in the generation of a full-length *hsp90* gene product? Using primers specific to *hspN* and *hspC* and an RT-PCR approach, we were able to detect an amplicon corresponding to the full-length GIHsp90 message. Sequencing of this amplicon confirmed the presence of a fused message arising from genes corresponding to the N- and C-terminal fragments of GHsp90. The fused message was capable of giving rise to a full-length GIHsp90 protein bearing all the hallmarks of a canonical Hsp90. We investi-

gated the presence of a full-length GlHsp90 using Western blotting as well as a mass spectrometry approach. Indeed Western blotting analysis using antibodies specific to GlHsp90 revealed the presence of a full-length *glhsp90* gene product migrating at the expected size of about 80 kDa. Partial sequencing of the corresponding protein band from total lysate using mass spectrometry confirmed the presence of peptides corresponding to GlHsp90. Our results proved beyond doubt the presence of a full-length GlHsp90 protein arising out of *hspN* and *hspC* genes interspersed by a 777-kb sequence.

G. lamblia is tetraploid with two equivalent diploid nuclei and proliferates asexually in vertebrate hosts. The 12-Mbp genome is very compact with only four genes containing a single intron. A genome rearrangement placing >770 kb of sequence inside the *hsp90* ORF would have been compensated by the remaining intact alleles. However, the current split gene structure in *Giardia*, which apparently occurred post-speciation, shows that this disruption has been genetically fixed. Because Hsp90 function is essential, as shown by inhibition of *Giardia* growth after treatment with Geldanamycin derivative, this required an alternative mechanism to produce a full-length protein. The absence of the putative 779-kb-long pre-mRNA transcript formally ruled out the possibility of *cis*-splicing as a potential mechanism to generate a full-length mature mRNA. Based on examination of mRNA and genomic sequences, a *trans*-splicing reaction directed by canonical dinucleotides GU-AG is the only plausible explanation for the production of a mature mRNA containing both segments of Hsp90. *Giardia* possesses the necessary machinery to remove introns in four genes (26), but RNA *trans*-splicing has not been documented. In trypanosomatids and nematodes, this post-transcriptional reaction adds capped spliced leader sequences to the 5'-terminal regions of pre-RNAs to generate mature mRNAs from polycistronic precursors (27). There is no example in the literature for the joining of two separate coding RNAs by *trans*-splicing, although synthetic adenovirus encoded mRNAs are spliced *in-trans* to allow production of large proteins for gene therapy (28). Although the exact mechanism of the splicing reaction in *Giardia* remains to be determined, the presence of complementary sequence adjacent to the splice sites suggests a positioning function for the assembly of a splicing complex.

Hsp90 is one of the most highly conserved proteins in the biological kingdom with limited gene variation within and across species. In addition to its conserved primary structure, its oligomeric structure as well as its higher order structure as a multichaperone complex is well conserved. *G. lamblia* genome also shows the presence of all the members of Hsp90 co-chaperone team described in other systems. It is unclear what advantage has led to the selection of a split *hsp90* gene. The occurrence of a post-transcriptional repair suggests that one or both fragments may lead to production of partial proteins that exert distinct functions. Alternatively, repair of split genes may be a means to increase complexity in a virtually intron-less eukaryote. Indeed, the presence of fragmented as well as full-length dynein heavy chain genes in the *Giardia* genome may support this scenario. The documented post-

transcriptional repair of giardial Hsp90 may be the first example of an ancient mechanism for rescuing disruption of essential genes.

Acknowledgments—We thank Genotypic Technology, Bangalore, India for extended support in genome sequencing. We also thank Prof. R Sehgal, Department of Parasitology, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India for providing *Giardia* cultures. We thank Shatakshi Ranade for valuable inputs in the Hsp90 inhibitor sensitivity experiment.

REFERENCES

1. Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Müller, S., Dessi, D., Fiori, P. L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuer, F. D., Simoes-Barbosa, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Pertea, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osogawa, K., de Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M., Jr., Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y. S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M., and Johnson, P. J. (2007) *Science* **315**, 207–212
2. El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, C. G., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osogawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D., and Andersson, B. (2005) *Science* **309**, 409–415
3. Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., Bothe, G., Brusch, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Müller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schäfer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B., and Myler, P. J. (2005) *Science* **309**, 436–442
4. Adam, R. D. (2000) *Int. J. Parasitol.* **30**, 475–484
5. Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse,

- S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J. E., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner, D. S., Samuelson, J., Svard, S. G., and Sogin, M. L. (2007) *Science* **317**, 1921–1926
6. Simpson, A. G., and Roger, A. J. (2002) *Curr. Biol.* **12**, R691–693
7. Neckers, L., and Tatu, U. (2008) *Cell Host Microbe* **4**, 519–527
8. Sawarkar, R., Roy, N., Rao, S., Raman, S., Venketesh, S., Suguna, K., and Tatu, U. (2008) *J. Mol. Biol.* **383**, 24–35
9. Wiesgigl, M., and Clos, J. (2001) *Mol. Biol. Cell* **12**, 3307–3316
10. Pallavi, R., Roy, N., Nageshan, R. K., Talukdar, P., Pavithra, S. R., Reddy, R., Venketesh, S., Kumar, R., Gupta, A. K., Singh, R. K., Yadav, S. C., and Tatu, U. (2010) *J. Biol. Chem.* **285**, 37964–37975
11. Rutherford, S. L., and Lindquist, S. (1998) *Nature* **396**, 336–342
12. Queitsch, C., Sangster, T. A., and Lindquist, S. (2002) *Nature* **417**, 618–624
13. Cowen, L. E., and Lindquist, S. (2005) *Science* **309**, 2185–2189
14. Diamond, L. S., Harlow, D. R., and Cunnick, C. C. (1978) *Trans. R. Soc. Trop. Med. Hyg.* **72**, 431–432
15. Yee, J., and Dennis, P. P. (1992) *J. Biol. Chem.* **267**, 7539–7544
16. Harris, J. C., Plummer, S., Turner, M. P., and Lloyd, D. (2000) *Microbiology* **146**, 3119–3127
17. Péroval, M., Péry, P., and Labbé, M. (2006) *Int. J. Parasitol.* **36**, 1205–1215
18. Ahn, H. J., Kim, S., and Nam, H. W. (2003) *Biochem. Biophys. Res. Commun.* **311**, 654–659
19. Devaney, E., O'Neill, K., Harnett, W., Whitesell, L., and Kinnaird, J. H. (2005) *Int. J. Parasitol.* **35**, 627–636
20. Borkovich, K. A., Farrelly, F. W., Finkelstein, D. B., Taulien, J., and Lindquist, S. (1989) *Mol. Cell. Biol.* **9**, 3919–3930
21. Echeverria, P. C., Matrajt, M., Harb, O. S., Zappia, M. P., Costas, M. A., Roos, D. S., Dubremetz, J. F., and Angel, S. O. (2005) *J. Mol. Biol.* **350**, 723–734
22. Pavithra, S. R., Banumathy, G., Joy, O., Singh, V., and Tatu, U. (2004) *J. Biol. Chem.* **279**, 46692–46699
23. Chen, B., Piel, W. H., Gui, L., Bruford, E., and Monteiro, A. (2005) *Genomics* **86**, 627–637
24. Pan, F., Zarate, J. M., Tremblay, G. C., and Bradley, T. M. (2000) *J. Exp. Zool.* **287**, 199–212
25. Chen, B., Kayukawa, T., Monteiro, A., and Ishikawa, Y. (2005) *Insect. Mol. Biol.* **14**, 697–702
26. Nixon, J. E., Wang, A., Morrison, H. G., McArthur, A. G., Sogin, M. L., Loftus, B. J., and Samuelson, J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3701–3705
27. Hannon, G. J., Maroney, P. A., Yu, Y. T., Hannon, G. E., and Nilsen, T. W. (1992) *Science* **258**, 1775–1780
28. Nakai, H., Storm, T. A., and Kay, M. A. (2000) *Nat. Biotechnol.* **18**, 527–532
29. Franzén, O., Jerlström-Hultqvist, J., Castro, E., Sherwood, E., Ankarklev, J., Reiner, D. S., Palm, D., Andersson, J. O., Andersson, B., and Svärd, S. G. (2009) *PLoS Pathog.* **5**, e1000560