Genetic complexity of the human hsp 60 gene

Nicole Anne-Marie Pochon and **Bernard Mach**

Jeantet Laboratory of Molecular Genetics, Department of Genetics and Microbiology, University of Geneva Medical School, 1211 Geneva 4, Switzerland

Abstract

Hsp 60 is a chaperonin protein, homologous to GroEL of *Escherichia coli* and highly conserved across species. Immune response induced by the hsp 60 equivalent of numerous microorganisms elicits in animals and man a dominant cross-reactive T lymphocyte response. Hsp 60 has been strongly implicated as an example of molecular mimicry in the pathogenicity of autoimmune diseases and, more recently, in T cell-mediated protection. Curiously, in spite of this interest, the gene encoding HSP 60 has not yet been cloned. Sequencing of numerous PCR-derived HSP 60 clones, obtained following amplification of genomic DNA revealed multiple distinct but highly related sequences. These were all different from the sequence encoding the expressed protein and all had interrupted reading frames. PCR amplification from mRNA, however, yielded only the sequence expected for the expressed hsp 60 protein. This apparent paradox was resolved by cloning and sequencing HSP 60-specific genomic clones: the majority of these clones corresponded to intronless genes having the characteristics of retro-pseudogenes and were flanked by unrelated DNA sequences. In addition, several genomic clones were isolated that corresponded to a unique functional HSP 60 gene. This gene is composed of multiple exons, some very short. The transcription start site was identified and 750 bp of 5' flanking sequence were determined. The human HSP 60 gene is induced by heat. We conclude that hsp 60 is encoded by a single highly fragmented gene, that co-exists with multiple HSP 60 retro-pseudogenes, normally not expressed.

Introduction

Hsp 60 is a mammalian heat shock protein known to function as an essential chaperonin (1,2). It has an important role in protein folding and in transport within mitochondria (3). Human hsp 60 is homologous to the *Escherichia coli* protein GroEL (4–8), also an important chaperonin, and to the yeast mif4 protein (9). The mechanism of action of GroEL has been studied in detail (10–13). It is one of the most highly conserved protein in both pro- and eukaryotes Recently, a cytosolic GroEL homologue, distinct from mitochondrial hsp 60, has been identified (6,14,15). It is quite different in structure from the HSP 60 gene discussed in this paper.

In addition to its essential role as a chaperonin, hsp 60 has received much attention as a dominant cross-reactive antigen and has been implicated in the pathogenesis of autoimmune diseases (reviewed in 16,17). The hsp 60 homologous proteins from several microorganisms, including frequent human pathogens, such as *Mycobacterium tuberculosis*, elicit a strong and dominant immune response in their mammalian

hosts (18–20). Because of the extensive sequence conservation of the hsp 60 protein in pro- and eukaryotes, hsp 60-specific cross-reactive T cells are generated in several animal species, capable of recognizing distinct epitopes of the human hsp 60 protein (21) Furthermore, in the case of several animal experimental autoimmune models, such hsp 60-specific T cells were found able to transfer the disease upon injection into healthy animals (22–24). The possible implication of hsp 60 in the pathogenicity of autoimmune diabetes and arthritis was further strengthened by claims of induction of the diseases by specific hsp 60 peptides (25) and of vaccination against diabetes by such hsp 60 peptides (26–28). Very recently, it was reported that activation of T cells that recognize a specific MHC class II restricted epitope of hsp 60 can protect against adjuvant arthritis (29). This important observation implies that cross-reactivity to self hsp 60 epitopes might be involved in T cell-mediated protection against autoimmune diseases, a concept that implies a re-evaluation of the role of

hsp 60 in immunopathology It is not known to what extent the hsp 60 chaperonin may be implicated in the processing and/or the transport of antigens towards MHC-mediated antigen presentation.

In spite of this interest in hsp 60, both as a chaperonin and as a cross-reactive antigen possibly involved in autoimmune diseases, the gene encoding HSP 60 has not been cloned. This situation is probably a consequence of the genetic complexity of HSP 60 genes presented in this paper. Indeed, we have identified and cloned multiple non-expressed HSP 60 pseudogenes, without introns, probably representing retroposons, as well as a single expressed HSP 60 gene, with multiple introns. It has been previously assumed that, by analogy with other HSP genes such as HSP 70, the HSP 60 gene did not contain introns (30,31). These data now open the way to functional studies on the regulation of HSP 60 gene expression and to molecular studies exploring a possible genetic linkage between HSP 60 locus and disease susceptibility.

Methods

Preparation of DNA

DNA was extracted from peripheral blood lymphocytes or culture cell lines according to the procedure of Gros-Bellard *et al* (32) with minor modifications described by Tiercy *et al.* (33) without DNA dialysis.

Amplification of genomic DNA

DNA samples (1 μg) were amplified by PCR, in a 50 μl reaction using Taq polymerase (1 unit per reaction, Boehringer Mannheim, Rotkreuz, Switzerland). The PCR reactions were carried out with 50 pmol of primers, 200 μM dNTP, 2 mM $MgCl_2$ and 1×PCR buffer containing 50 mM KCl, 10 mM Tris, pH 8 4, 100 μg/ml BSA. Twenty five cycles of amplifications were performed on the programmable thermocycler (Techne PHC-1) in which each cycle consists of 1 min at 94°C, 30 s at 55°C and 30 s at 72°C

Subcloning and sequencing of PCR products

The PCR products were extracted with phenol–chloroform, precipitated with ethanol and dissolved in water. *Pst*I and *Hind*III restriction sites, incorporated into the 5' and 3' primers respectively, enable the amplified products to be subcloned into the Bluescript vectors (Stratagene, Heidelberg, Germany) One quarter of the amplified product was transformed into competent JM103 cells. The DNA sequence of the transformed clones was determined directly on the double-stranded clones by the dideoxy chain termination method (34) using the T7 enzyme system (Pharmacia).

Probes used for filters hybridizations

A fragment of 1.3 kb, obtained by genomic amplification with the primer Hsp5' and Hsp6 2, subcloned and entirely sequenced, was used as a probe (F1) and ^{32}P-labeled by random primed synthesis. This fragment corresponds to an HSP 60 pseudogene. The oligonucleotides probes were 5' end labeled with [32P]ATP and polynucleotide kinase. Oligonucleotides sequences:

Hsp5'	5'-GTCACTGCAGGTGTAGACCTTTTAGCCG-'3
Hsp5 1	5'-GTCACTGCAGGAAGAAAGGGTGTCATCAC-3'
Hsp5.2	3'-GACTTCTACAACTACCTCTTCGAATGAC-5'
Hsp6 2	3'-GGTCGTTACTGGTAACGATTCGAATGAC-5'
Hsp254A	3'-AAGATCATAGGTCAGGTAAC-5'
Hsp65+80	3'-GAGTGAGCCCGAATACG-5'
Hsp65+9F	3'-GTTACCCACAGTCTTTCGC-5'
Hsp65+9	3'-CAATGGGTGTCAGAAAGCG-5'
Hsp254G	5'-GTTGCCTGACCTGTGATCTT-3'

Southern analysis

Genomic DNA was isolated from cultured HeLa cells and 15 μg sample were digested to completion by one or several restriction enzymes. After electrophoresis on 0 7% agarose gels, the DNA was transferred to Gene Screen Plus membranes according to the suggestions of the manufacturer (NEN, Regensdorf, Switzerland). The blots were hybridized to the F1 probe with a sp. act of $1 5×10^6$ c.p.m /ml The hybridization conditions were 50% formamide, 10×Denhardt's, 1 M NaCl, 1% SDS, 50 mM Tris, pH 7.5, 200 μg/ml herring sperm DNA at 42°C The filters were washed in 1×SSC at 65°C.

Preparation of RNA

Total RNA was extracted from frozen cell pellets by using guanidium isothiocyanate (34) and poly(A)$^+$ RNA was purified by oligo(dT)–cellulose chromatography. The nuclear and cytoplasmic RNA were extracted from Raji and THP1 cell lines by the method described by M. Wilkinson (35)

PCR on total RNA

Total RNA was treated with DNase I, 30 min at 37°C. The first strand of the cDNA was obtained by reverse transcriptase reaction on 50 μg of total RNA in a volume of 50 μl with 100 pmol of random hexanucleotides, 500 μM of each dNTP, 32 units of RNase Inh. (Biofinex, Praroman, Switzerland) and 200 units of MMLV reverse transcriptase (BRL, Basel, Switzerland) in 1×reverse transcriptase buffer (BRL) The reaction was carried out 1 h at 42°C, heated at 95°C and twice precipitated to eliminate the non-incorporated primers. Half of the reverse transcript product was amplified with the primers Hsp5.1/Hsp5.2 as described for the genomic PCR. The amplified products were subcloned and sequenced.

PCR on nuclear and cytoplasmic RNA

The first strands of the cDNAs were synthesized as mentioned before on 5 μg of RNA pretreated by DNase I. The PCR were carried out with the primers Hsp5' and Hsp6.2 under the same conditions used for genomic DNA.

Contamination of RNA samples with DNA was tested by direct amplification on 5 μg of the same RNA preparation without reverse transcription

The amplified products were separated on 1% agarose gels and hybridized directly in the gel, as described by Ugozzoli *et al.* (36), with the 5'-end-labeled oligo Hsp254A.

Screening of a human genomic library

A human partial *Sau*3A genomic library cloned in EMBL3 (generously provided by Catherine Nguyen, INSERM) was screened with the F1 probe and the specific oligonucleotide Hsp254A.

Filters were hybridized to the ^{32}P-labeled F1 probe The hybridization conditions were 4×SSC, 5×Denhardt's, 0.1% SDS, 5 mM EDTA, 10 mM NaP and 100 µg/ml of herring sperm DNA at 68°C overnight. Filters were washed at final stringency in 0 1×SSC at 68°C.

The 5'-end-labeled Hsp254A was added to 5×SSC, 20 mM NaP, 10×Denhardt's, 5% SDS and 200 µg/ml of herring sperm DNA at 50°C

Positively hybridizing recombinant phages were plaque purified, and phage DNA was prepared. A restriction map was determined with the enzymes *Sal*I, *Eco*RI and *Bam*HI. The genomic inserts were subcloned in the Bluescript vector and their sequences determined with the specific oligonucleotides derived by the cDNA sequence

Primer extension

For primer extension analysis, the Hsp65+9F and Hsp65+80 oligonucleotide primers complementary to a sequence within

exon 2 and 3 respectively of the HSP 60 gene were 5'-end-labeled with [32P]ATP.

Samples of 20 µg of total RNA and 5 µg of poly(A)+ RNA were co-precipitated with 5×10[6] or 5×10[7] c.p.m. of each of the primers, then resuspended in 10 µl of H_2O and heated for 5 min at 65°C with 4 µl of 5×reverse transcriptase buffer (BRL). Annealing was carried out until the temperature has reached 42°C. The extension reaction was performed in presence of 150 µM deaza-dGTP, 2 mM each dNTP, 32 units RNAsin, 200 mM DTT and 200 units reverse transcriptase (BRL) for 1 h at 42°C. The products were analyzed on 6% acrylamide (38:2) gels.

RNase protection

For RNase protection, the 634 bp *Nae*I–*Bam*HI fragment of the 0.91 kb subcloned containing the HSP 60 promoter, was cloned in the Bluescript vector. The plasmid was digested with *Nae*I and transcribed with the T3 RNA polymerase in the
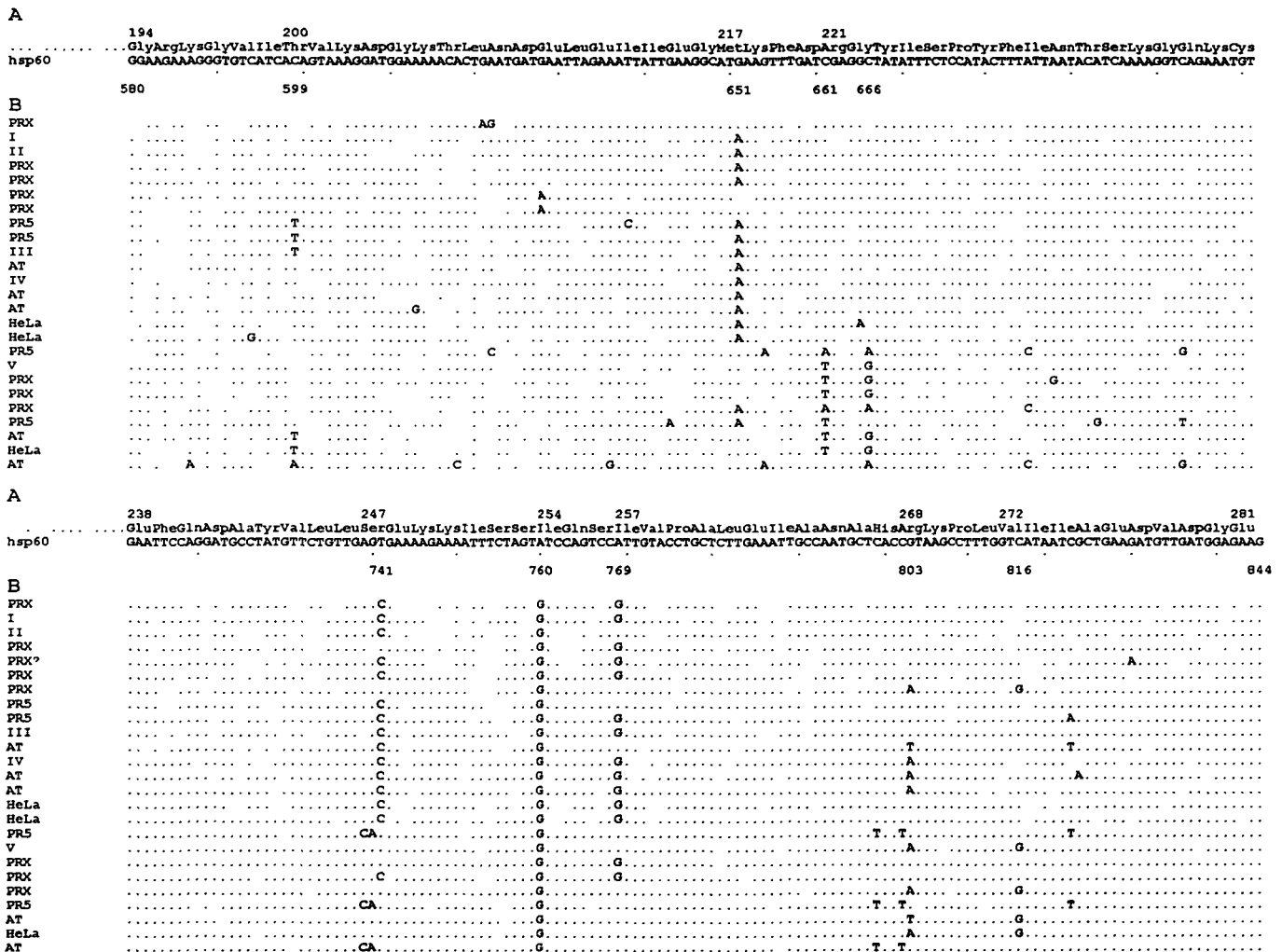
Fig. 1. Nucleotide sequence comparison of HSP 60 cDNA and HSP 60 amplified genomic sequences. (A) HSP 60 cDNA sequence from nucleotides 580 to 844. (B) Twenty five distinct homologous sequences of amplified genomic products obtained with primers Hsp5.1 and Hsp5 2 Five different human genomes were used for PCR amplification (PRX, PR5, AT, Raji and HeLa). Nucleotides identical to the published cDNA sequence are represented by a dot. The sequences I, II, III, IV and V were amplified in the five different genomes. Specific sequences are identified by their genomic name. The most frequent mutated positions are indicated at the top of the alignment

presence of [^{32}P]UTP This generates a 600 nucleotide probe homologous to the exon E1. Hybridization of the probe with poly(A)$^+$ RNA and total RNA, digestion with RNase A and T1, and PAGE were done as described previously (37).

Results

Evidences for a family of HSP 60 genes

The DNA sequence analysis of the clones obtained after PCR amplification of human DNA with several sets of HSP 60-specific primers gave unexpected results. Primers Hsp5' and Hsp6 2 delimited a cDNA fragment of 1.3 kb and primers Hsp5.1 and Hsp5.2 a fragment of 260 bp, coding for a portion of the hsp 60 protein that contains a well-studied immunoreactive T cell epitope (38,39). Two separate PCR amplifications were performed with each primer pair, on DNA prepared from nine different unrelated individuals. PCR amplification was performed under conditions tested to minimize polymerase-induced nucleotide substitution. All PCR products, when analyzed by agarose gel electrophoresis, corresponded to the expected cDNA size (data not shown).

Amplified DNA was subcloned and sequenced. From the product of the Hsp5.1 and Hsp5.2 primers, 64 different clones were sequenced, some several times, and this revealed an extensive sequence diversity and allowed the identification of 25 distinct highly homologous sequences (Fig 1B). A

striking feature of all these sequences was that none of them corresponded to the HSP 60 cDNA sequence! All the different sequences obtained from genomic DNA differed from cDNA by one or multiple nucleotide substitutions. Five of these different sequences (referred to as I, II, III, IV and V in Fig. 1B) were observed in the DNA from each of the nine individuals studied Nine positions (at nucleotides 599, 651, 661, 666, 741, 760, 769, 803, and 816 according to the cDNA sequence) were preferentially mutated. The mutation at position 661 (C to T) creates a stop codon. Other stop codons were observed in the PCR products of 1.3 kb, resulting from the deletion of one nucleotide at position 1266 or 1405 (data not shown). As a result of one these interruptions in the reading frame, it appears that none of these multiple HSP 60 genomic sequences can encode for a full length protein. Sequencing of multiple clones obtained from several independent PCR amplifications of the same genomic DNA confirmed that the observed diversity in HSP 60 DNA sequences was not the result of PCR artifacts (data not shown). Surprisingly, all these different clones are characterized by the presence, at position 760 of the cDNA sequence, of a G instead of an A (Fig. 1). All sequences were performed on both DNA strands and confirmed following a second PCR amplification.

These data suggest the existence, in human DNA, of a family of HSP 60 pseudogenes. This was confirmed by Southern blot hybridization (Fig. 2). DNA was digested with the enzymes *Bam*HI, *Eco*RI and *Pst*I. Hybridization reveals multiple HSP 60-specific fragments of various sizes. It is likely that the variation in the signal intensity of the different bands results from sequence divergence and/or from the existence of pseudogene clusters.
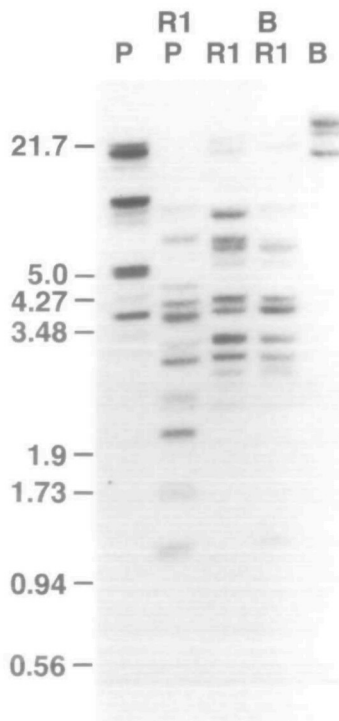
Fig. 2. Southern blot analysis of the HSP 60 gene. HeLa genomic DNA was digested with *Bam*HI (B), *Eco*RI (R1) and *Pst*I (P), and the F1 probe was used to hybridize the membrane. The position of the molecular weight markers (lambda *Hind*III–*Eco*RI) is indicated on the left.
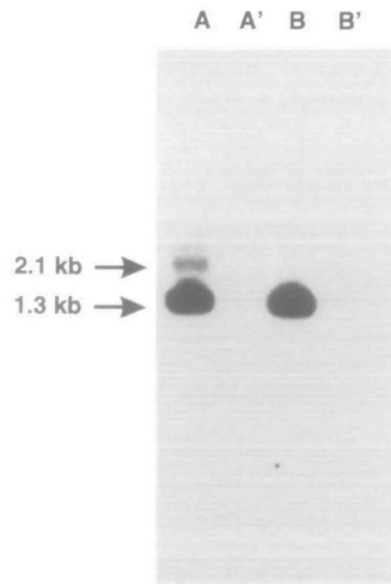
Fig. 3. Detection of an unspliced HSP 60 nuclear form RT-PCR on 5 µg of nuclear (A) and cytoplasmic (B) total RNA was performed with primers Hsp5' and Hsp6.2 The absence of genomic DNA was checked by direct amplification on 5 µg of nuclear (A') and cytoplasmic (B') total RNA without reverse transcriptase The PCR products were separated on an agarose gel and hybridized directly in the dried gel with oligonucleotide HSP 254A.

Expression of a single HSP 60 RNA sequence

The analysis of expressed HSP 60 transcripts by Northern blots was performed either with a 1.3 kb probe (obtained by PCR from genomic DNA, see above) or with oligonucleotides specific of the cDNA sequence around position 760, and containing either A (Hsp254A) or G, like all pseudogenes (Hsp254G). Northern blots showed a single HSP 60 mRNA band of 2.3 kDa, hybridizing with both the cDNA probe and the Hsp254A oligonucleotide, while no signal was obtained with the Hsp254G oligonucleotide (data not shown) This indicates that only the HSP 60 gene with A at position 760 is expressed and none of the pseudogenes.

PCR amplification was then performed on cDNA prepared from cytoplasmic RNA from different individuals using primers Hsp5.1 and Hsp5 2. The amplified fragments were subcloned and 16 such subclones were sequenced. In contrast to our finding with genomic DNA, all HSP 60 sequences derived from mRNA were identical to the published cDNA sequence, including the A at position 760. None of the multiple genomic HSP 60 sequences identified earlier corresponded to that cDNA sequence (Fig. 1). In addition, the RT-PCR product was hybridized with the Hsp254A and the Hsp254G oligonucleotide probes Again, hybridization was positive with the probe specific for the expressed gene, while no signal was obtained with the pseudogene probe Hsp254G (data not shown)

### Evidence for an unspliced nuclear form of HSP 60 RNA

To resolve the apparent paradox of the multiple mutated HSP 60 genes, we looked for the presence of unspliced nuclear HSP 60 transcripts. PCR was performed on cDNA prepared from either nuclear or cytoplasmic RNA and the resulting amplification product analyzed by Southern blot, using an oligonucleotide probe (Hsp254A). As shown in Fig. 3, a 1.3 kb PCR product, corresponding in length to the HSP 60 cDNA, was observed in both the cytoplasmic and nuclear RNA. In nuclear RNA, however, a larger RNA (2 1 kb) was present, indicative of an incompletely spliced HSP 60 transcript.

Isolation and characterization of the expressed HSP 60 gene

Considering the previous data, it could be argued that the expressed HSP 60 gene should have introns and be characterized by the presence of an A at position 760. In order to isolate the expressed HSP 60 gene, >1×10⁶ phages from a non-amplified human genomic library were screened in duplicate with the 1.3 kb HSP 60 probe (obtained by amplification of genomic DNA). After three rounds of hybridization, 45 phages were isolated. Three of these (28C1, 28C2 and 28C3) hybridized specifically with oligonucleotide Hsp254A, suggesting that these clones corresponded to the expressed HSP 60 gene The other 42 phages all hybridized to the Hsp254G oligonucleotide, as expected from signals derived from the HSP 60 pseudogenes Sequence analysis of selected clones among these 42 phages indicated that they indeed contain the various HSP 60 sequences previously identified in amplified genomic DNA (Fig. 1), including the characteristic stop codons. Importantly the DNA sequences flanking these intronless pseudogenes are all different (see below).

Restriction analysis of clones 28C1, 28C2 and 28C3 and hybridization with different oligonucleotides, specific for the coding region of the cDNA, indicates that these three clones correspond to the same Sau3A genomic fragment of 17 kb,
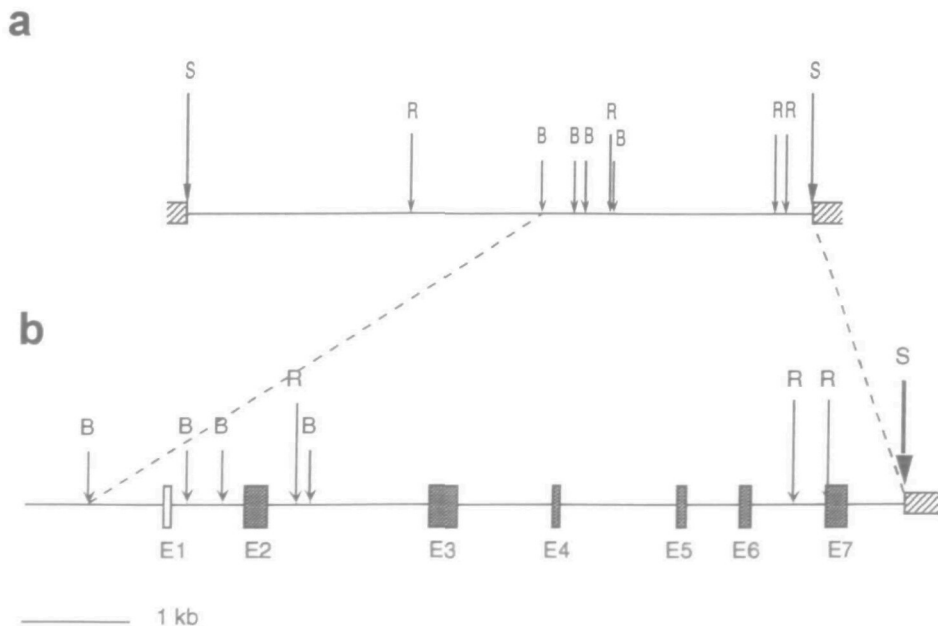
Fig. 4. Restriction map of phage 28C2 encoding the 5' part of the HSP 60 gene. (a) Restriction map of the 17 kb *Sau3A* genomic fragment inserted in the Sall site of the EMBL3 phage (hatched boxes). Restriction enzymes *EcoRI* (R), *BamHI* (B) and *Sall* (S) were used. (b) Map of the HSP 60 seven exons (boxes) encoded by clone 28C2. E1 (white box) is not translated, E2–E7 (gray boxes) are translated. The exons and the exon–intron boundaries were sequenced but the size of the introns was not precisely defined.

5'
CTTGCCGCCGCCCCGC<u>AG</u> <u>gta</u>cgcggcc...I₋₂ ...ccatccc<u>ag</u> AA <u>ATG</u> CTT CGG TTA
      Exon1   *Exon 2*

      ATG GGG CCA A<u>AG</u> <u>gta</u>ccagtat...I₁₇₄...tttcttc<u>tag</u> GGA AGA ACA GTG ATT
          Exon 2   *Exon 3*

      GAA ATC AGG AG<u>AG</u> <u>gta</u>ggaatgt...I₄₂₇...gttgaac<u>tag</u> GT GTG ATG TTA GCT
          Exon 3   *Exon 4*

      GAA ATT GCA C<u>AG</u> <u>gta</u>aggactt...I₅₁₀...attatt<u>tcag</u> GTT GCT ACG ATT
          Exon 4   *Exon 5*

      ATC ACA GTA A<u>AG</u> <u>g</u>caagtgtgt...I₆₀₆...cttttca<u>tag</u> GAT GGA AAA ACA
          Exon 5   *Exon 6*

      AAT ACA TCA AA<u>AG</u> <u>gta</u>agaacaa...I₇₀₀...atttttac<u>ag</u> GT CAG AAA TGT GAA
          Exon 6   *Exon 7*

      CTC GTC TTG AAT <u>AG</u> <u>gta</u>ataagca...I₈₇₀
   *Exon 7*

Fig. 5. HSP 60 intron–exon junctions. The nucleotide sequence of the exons and introns are represented with capital and small letters, respectively. The exon–intron consensus sequences are underlined. The intron positions (I_x) are indicated relative to the double underlined initiation codon.

that encodes the 5' part of the HSP 60 gene. A restriction map of phage 28C2 is shown in Fig. 4. The genomic fragments were subcloned and all exons, as well as the intron–exon boundaries, were sequenced. The HSP 60 open reading frame begins at the first ATG in exon 2 (E2) and is interrupted by five introns within clone 28C2. The nucleotide sequence of the six small exons (E2–E7) is identical to the cDNA sequence (including A at position 760). The 5' untranslated region of the HSP 60 transcript is mostly contained in exon 1 (E1), with the last two nucleotides in exon 2 (Fig. 5). The 3' part of the clone ends in an intronic *Bam*HI site ligated in the *Sal*I site of the EMBL3 long arm. The sequences of all intron–exon junctions are shown in the Fig. 5. They all correspond to the eukaryotic consensus sequence of splicing.

Promoter of the HSP 60 gene and initiation site

Because of the potential importance of the regulation of HSP 60 gene expression in immunopathology, the exact nature of the initiation site and the structure of the HSP 60 promoter were investigated. The transcription initiation site of the expressed HSP 60 gene was identified by primer extension and RNase protection experiments. For primer extension, we used a synthetic 19-mer (Hsp65+9), nine nucleotides downstream of the ATG. As shown in Fig 6, oligo Hsp65+9 primed the synthesis of a major product extending 62 nucleotides within exon 1. Other weaker products corresponded to extension of 56, 68, 70 and 88 nucleotides into exon 1 (Fig. 6), and reflect additional minor start sites.

To confirm this result, RNase protection assays were performed using the same RNAs. A genomic fragment of 443 bp, containing the promoter region, the first exon (E1) and 167 bp of the first intron, was used to generate a labeled RNA probe *in vitro*. Figure 7 shows one major and four weaker protected fragments, whose lengths are consistent with the results of the primer extension analysis. The position of the
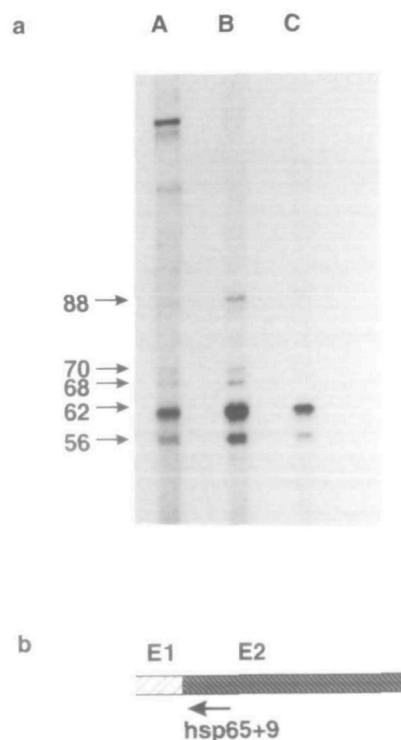
Fig. 6. Identification of initiation sites by primer extension. (a) The HSP 60 mRNA 5' end was identified by reverse transcription using primer Hsp 65+9. The molecular weights are deduced from a sequence reaction using the same primer, the numbers refer to the ATG position. Lanes A and B contain 20 µg total RNA with 10^6 and 10^7 c.p.m of probe respectively. Lane C contains 5 µg poly(A)$^+$ RNA and 10^6 c.p.m. of probe. Five initiation sites are detected, the major site is localized 62 nucleotides upstream of the ATG, as determined by the sequencing ladder. (b) Map showing the position of the primer relative to the exons 1 and 2 (E1 and E2 respectively)
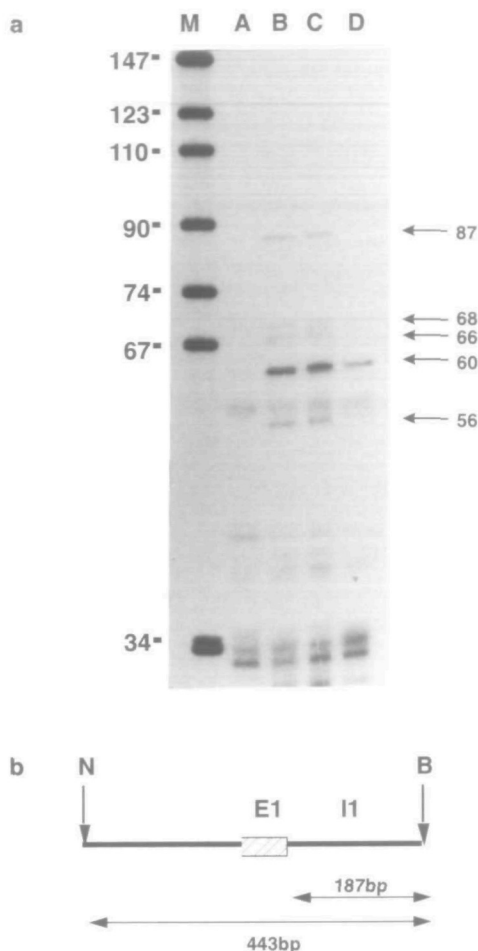
Fig. 7. Identification of initiation sites by RNase protection (a) RNase protection of the HSP 60 mRNA Five initiation sites are detected with a major site estimated at 60 nucleotides upstream of the ATG, according to the molecular weight marker M: MspI-digested pBR322 marker. (A) yeast tRNA, 50 μg; (B), (C) and (D) 50, 30 and 10 μg of total RNA respectively. (b) Map of the probe used in (a). RNase protection using a NaeI (N)–BamHI (B) fragment containing the promoter, the first exon (E1) and 187 bp of the first intron (I1)

major initiation site (+1) is indicated on the map as well as on the sequence of the HSP 60 promoter in Fig. 8.

About 750 nucleotides upstream of the HSP 60 transcription initiation sites were sequenced. This region was compared with the 5' flanking sequence of several cloned pseudogenes The entire transcribed region, including the 5' untranslated region, shows extensive homology to the HSP 60 pseudogenes examined (98%). For each pseudogene the sequence homology extends up to one of the initiation site predicted by the previous experiment. Upstream of this transcription start site, the sequences are totally divergent, indicating that all these different HSP 60 genes reside indeed in different DNA contexts (see example in Fig. 8A).

The HSP 60 gene promoter is characterized by the absence of a classical TATAA box and by the presence of a single heat shock response element (HSE), nGAAn, located 233 bp

upstream of the major initiation site (Fig. 8B). Promoters of other heat shock protein genes frequently contain several such elements (40–42). Two potential Sp1 sites are also present 64 and 150 bp upstream from the major initiation site.
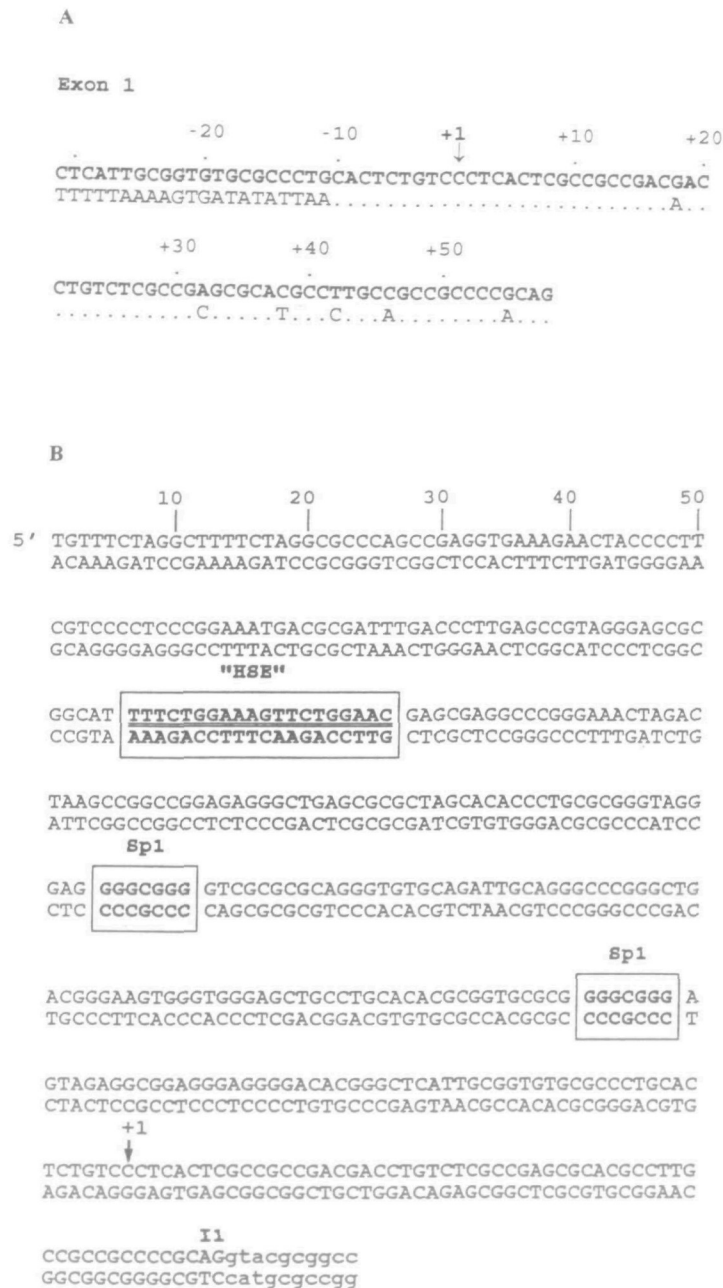
In agreement with earlier studies on the inducibility of the HSP 60 gene by heat (43), we confirmed by Northern blot experiment that HSP 60 mRNA is indeed induced in several cell types following exposure to 42°C (data not shown).

Discussion

The chaperonin hsp 60, the mammalian homologue of the GroEL of E. coli, is of interest to cell biologists because of its essential role as a catalyst of protein folding At the same time, the well-documented dominant cross-reactive immune response to bacterial hsp 60 in various animal models has made hsp 60 an interesting candidate as an auto-antigen in the pathogenicity of autoimmune diseases (39,44) In fact, it has been demonstrated that hsp 60-specific CD4 T cells, elicited in different animal species by bacterial hsp 60, react to specific epitopes of mammalian hsp 60, including human hsp 60 (25,45–47). Furthermore, the fact that such hsp 60-specific T cell lines or clones can trigger autoimmune diabetes or arthritis in healthy animals by passive transfer (23) has strengthened the models implicating hsp 60 in these diseases Finally, it has been claimed that specific hsp 60 peptides, corresponding to the epitope recognized by T cells, could induce autoimmune arthritis or diabetes in rat and mouse (25), whereas the same peptides, when administered as tolerogens, could protect, or vaccinate, against the same autoimmune diseases (26). Although these last experiments have not been generally reproduced, the role of hsp 60 epitopes as dominant cross-reactive T cell antigens and the functional role of such hsp 60-specific T cells in transferring autoimmune diseases is generally accepted. The recent demonstration of protection against adjuvant arthritis by T lymphocytes that specifically recognize an hsp 60 epitope (29) suggests yet a different, or additional, role for hsp 60 cross-reactivity and may lead us to reconsider the role of hsp 60 in the pathogenesis of autoimmune diseases.

It is therefore quite surprising that, in spite of such a general interest for hsp 60 in cell biology and in immunopathology, the HSP 60 gene had not been cloned. The extensive genetic complexity of HSP 60 genes reported here, as well as the finding that, contrary to earlier assumptions and to the case of other HSP genes, the HSP 60 gene contains multiple introns, probably explain earlier failure to identify this gene.

Our initial studies confronted us with a paradox: (i) multiple different HSP 60 sequences were identified following amplification of genomic DNA, none of which corresponded to the expressed HSP 60 sequence, and (ii) from mRNA (and cDNA), however, amplification lead exclusively to the expressed HSP 60 sequence. The data reported here explain these findings on the basis of multiple intronless pseudogenes, scattered in the human genome, and of a single expressed HSP 60 containing multiple introns. From genomic DNA, the numerous intronless pseudogenes are preferentially amplified by PCR, whereas expressed mRNA contains exclusively the processed product of the unique expressed, intron-containing, HSP 60

A

Exon 1

```
              -20         -10         +1          +10         +20
               .           .          ↓           .           .
CTCATTGCGGTGTGCGCCCTGCACTCTGTCCCTCACTCGCCGCCGACGAC
TTTTTAAAAGTGATATATTAA..........................A..

              +30         +40         +50
               .           .           .
CTGTCTCGCCGAGCGCACGCCTTGCCGCCGCCCCGCAG
...........C.....T...C...A........A...
```

B

```
              10          20          30          40          50
               |           |           |           |           |
5' TGTTTCTAGGCTTTTCTAGGCGCCCAGCCGAGGTGAAAGAACTACCCCTT
   ACAAAGATCCGAAAAGATCCGCGGGTCGGCTCCACTTTCTTGATGGGGAA


   CGTCCCCTCCCGGAAATGACGCGATTTGACCCTTGAGCCGTAGGGAGCGC
   GCAGGGGAGGGCCTTTACTGCGCTAAACTGGGAACTCGGCATCCCTCGGC
                        "HSE"
   GGCAT TTTCTGGAAAGTTCTGGAAC GAGCGAGGCCCGGGAAACTAGAC
   CCGTA AAAGACCTTTCAAGACCTTG CTCGCTCCGGGCCCTTTGATCTG

   TAAGCCGGCCGGAGAGGGCTGAGCGCGCTAGCACACCCTGCGCGGGTAGG
   ATTCGGCCGGCCTCTCCCGACTCGCGCGATCGTGTGGGACGCGCCCATCC
                 Sp1
   GAG GGGCGGG GTCGCGCGCAGGGTGTGCAGATTGCAGGGCCCGGGCTG
   CTC CCCGCCC CAGCGCGCGTCCCACACGTCTAACGTCCCGGGCCCGAC

                                                 Sp1
   ACGGGAAGTGGGTGGGAGCTGCCTGCACACGCGGTGCGCG GGGCGGG A
   TGCCCTTCACCCACCCTCGACGGACGTGTGCGCCACGCGC CCCGCCC T

   GTAGAGGCGGAGGGAGGGGACACGGGCTCATTGCGGTGTGCGCCCTGCAC
   CTACTCCGCCTCCCTCCCCTGTGCCCGAGTAACGCCACACGCGGGACGTG
                +1
                 ↓
   TCTGTCCCTCACTCGCCGCCGACGACCTGTCTCGCCGAGCGCACGCCTTG
   AGACAGGGAGTGAGCGGCGGCTGCTGGACAGAGCGGCTCGCGTGCGGAAC

                I1
   CCGCCGCCCCGCAGgtacgcggcc
   GGCGGCGGGGCGTCcatgcgccgg
```

Fig. 8. (A) Sequence comparison of the 5′ region of the expressed HSP 60 gene (top) and the pseudogene 9A (bottom). Sequence homology (~98%) begins exactly at one of the HSP 60 transcription initiation sites (10 bp upstream of the main initiation site identified as +1) and extends beyond exon 1, to the entire transcribed region (data not shown). (B) Structure of the HSP 60 promoter. The putative Sp1 sites and the palindromic sequence 'nGAAn' of the HSE are shown in boxes. The major initiation site (+1) is indicated by an arrow, the first intron (I1) by small caps.

gene. There is no evidence for expression of any of the HSP 60 pseudogenes, characterized by a G at position 760. Earlier attempts at cloning the HSP 60 gene had lead to pseudogenes and the report suggesting expression of hsp 60 from an intronless gene (48–50) should probably be re-evaluated on the basis of the findings reported here We have not attempted to quantify the number of HSP 60 pseudogenes, but the screening of the genomic library suggests the presence of at

least 15 such genes. The existence of numerous pseudogenes in the mammalian genome has been observed in the case of several other genes (51–53), generally housekeeping genes, but also in the case of the HSP 70 gene (54,55) It is generally believed that such pseudogenes result from multiple retroposition events, involving reverse transcription of cellular mRNA and probably mediated by a retrovirus (56). Why the 5′ untranslated region of the HSP 60 pseudogenes, which is

presumably not subjected to selection pressure, is conserved in sequence and has remained identical to that of the expressed gene is not clear and represents an unsuspected observation.

Although there is no obvious precedent described, we would like to suggest the possibility that rare gene rearrangements or chromosomal translocations could involve one of the multiple HSP 60 pseudogenes and lead to transcriptional activation of all or part of its sequence. This could result in the synthesis of a fusion or truncated protein, as documented in the case of other gene rearrangements (56). Since hsp 60 is a strongly immunogenic protein, it is interesting to speculate that such an aberrant expression of a modified and truncated form of hsp 60 peptides, encoded by a pseudogene, could lead to aberrant immunoreactivity with pathological consequences.

In addition to the lack of introns and to their sequence diversity, the multiple HSP 60 pseudogenes are characterized by a G at position 760, whereas the HSP 60 cDNA has an A at that same position. We used this property of the expressed HSP 60 sequence as a tool for the identification of the genomic clones containing the expressed HSP 60 gene, in the context of a large excess of clones corresponding to pseudogenes Interestingly, other mammalian HSP 60 cDNAs have a G at position 760 (48–50,57), unlike the human cDNA, but like the human pseudogenes. We therefore suggest that the generation of all the pseudogenes came first and that subsequently, in man, the G to A substitution took place at the expressed HSP 60 locus In that prospect, it will be interesting to analyze if the HSP 60 genes of other primates have an A or a G at position 760 in the expressed HSP 60 gene.

Because of the obvious biological interest of the regulation of expression of the HSP 60 gene, we have analyzed the structure of its promoter and identified the transcription initiation sites. Like in the promoter of most housekeeping genes, there is no obvious TATAA box in the HSP 60 promoter, nor a CAT box. Other HSP genes do have a TATAA and a CAT box (58) The presence of two Sp1 sites in the HSP 60 promoter might play a role in recruiting other required transcription factors, as suggested in other cases (59). We have also detected typical HSE motives, palindromes involving the consensus sequence nGAAn and known to be the targets for the binding of factor HSF (60). In the HSP 60 promoter, there are four such sequences in tandem around position –219 to –233 (Fig. 7), forming two adjacent palindromes. The position of the HSE motives is more distant from the initiation site than in the HSP 70 promoter and in the latter, two HSE palindromes are separated by 20 bp (41) The human HSP 60 gene is indeed induced by heat shock. The availability of the DNA sequence of the HSP 60 promoter will now make it possible to study the molecular mechanism controlling this induction

The elucidation of the genetic complexity of the human HSP 60 gene opens the way to a number of different studies. Genetic linkage analysis can now be considered in order to correlate HSP 60 gene expression with disease susceptibility, in particular in the case of human and mouse insulin-dependent diabetes. The possibility of a disregulation of expression of the HSP 60 gene in particular tissues in various disease situations, as well as the search for functionally relevant polymorphic changes within the promoter can be considered.

Indeed, fluctuations in the magnitude of the hsp 60 response could result in aberrant immune responses and immunopathology. By homologous recombination and transgenic experiments, it will now also be possible to modify the normal HSP 60 sequence, or its pattern of expression *in vivo*, and study the pathological consequences.

Acknowledgements

Abbreviations

HSE   heat shock response element

References

1 Hemmingsen, S. M., Woolford, C., van der Vies, S M , Tilly, K , Dennis, D. T , Georgopoulos, C. P, Hendrix, R. W. and Ellis, R. J 1988 Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* 333 330.
2 Ellis, R. J. 1990 The molecular chaperone concept [Review]. *Semin Cell Biol* 1 1.
3 Manning Krieg, U. C , Scherer, P. E and Schatz, G. 1991 Sequential action of mitochondrial chaperones in protein import into the matrix *EMBO J.* 10 3273.
4 Gething, M. and Sambrook, J 1992. Protein folding in the cell *Nature* 355:33.
5 Hartl, F , Hlodan, R. and Langer, T. 1994. Molecular chaperones in protein folding the art of avoiding sticky situations. *Trends Biol. Sci.* 19·20.
6 Horwich, A L and Willison, K. R 1993 Protein folding in the cell. function of two families of molecular chaperone, hsp60 and TF55-TCP1. *Phil. Trans. Roy. Soc (Lond.) B* 339.313.
7 Hendrick, J. P. and Hartl, F -U. 1993 Molecular chaperone functions of heat-shock proteins. *Annu Rev. Biochem* 62·349.
8 Ellis, R. J. and van der Vies, S. M. 1991 Molecular chaperones [Review]. *Annu Rev. Biochem* 60·321
9 Cheng, M. Y., Hartl, F U and Horwich, A L 1990. The mitochondrial chaperonin hsp60 is required for its own assembly. *Nature* 348.455.
10 Martin, J., Langer, T., Boteva, R., Schramel, A , Horwich, A L and Hartl, F. U 1991. Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate [see comments]. *Nature* 352.36.
11 Creighton, T E. 1991. Unfolding protein folding. *Nature* 352:17.
12 Goloubinoff, P, Gatenby, A. A. and Lorimer, G. H. 1989 GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli. Nature* 337:44.
13 Azem, A., Kessel, M and Goloubinoff, P. 1994. Characterization of a functional GroEL14(GroES7)2 chaperonin hetero-oligomer. *Science* 265:653.
14 Trent, J. D., Nimmesgern, E., Wall, J S , Hartl, F and Horwich, A. L. 1991. A molecular chaperone from a thermophilic archaebacterium is related to the eukaryote protein t-complex polypeptide-1. *Nature* 354.490.
15 Phipps, B. M., Hoffmann, A., Stetter, K. O. and Baumeister, W. 1991. A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria. *EMBO J.* 10:1711
16 Young, D. B. 1990. Chaperonins and the immune response. *Semin Cell. Biol.* 1.27.
17 Kaufmann, S. H. E. 1992 The cellular immune response to heat shock proteins. *Experientia* 48:640.
18 Ottenhoff, T. H., Ab, B. K., Van Embden, J D., Thole, J. E. and Kiessling, R. 1988. The recombinant 65-kD heat shock protein of *Mycobacterium bovis* Bacillus Calmette-Guerin/*M. tuberculosis* is a target molecule for CD4$^+$ cytotoxic T lymphocytes that lyse

human monocytes. *J Exp. Med* 168·1947

19 Young, D., Lathigra, R., Hendrix, R., Sweetser, D and Young, R. A. 1988 Stress proteins are immune targets in leprosy and tuberculosis *Proc. Natl Acad Sci USA* 85·4267

20 Shinnick, T. M., Vodkin, M. H and Williams, J C. 1988. The *Mycobacterium tuberculosis* 65-kilodalton antigen is a heat shock protein which corresponds to common antigen and to the Escherichia coli GroEL protein. *Infect. Immun.* 56.446.

21 Kaufmann, S H., Vath, U, Thole, J E., Van Embden, J D and Emmrich, F. 1987. Enumeration of T cells reactive with *Mycobacterium tuberculosis* organisms and specific for the recombinant mycobacterial 64-kDa protein. *Eur. J. Immunol.* 17 351.

22 Holoshitz, J, Naparstek, Y., Ben Nun, A. and Cohen, I R 1983 Lines of T lymphocytes induce or vaccinate against autoimmune arthritis. *Science* 219:56

23 Holoshitz, J., Matitiau, A. and Cohen, I. R. 1984. Arthritis induced in rats by cloned T lymphocytes responsive to mycobacteria but not to collagen type II. *J. Clin. Invest.* 73.211

24 van Eden, W, Thole, J E, van der Zee, R., Noordzij, A., Van Embden, J. D., Hensen, E. J. and Cohen, I. R 1988 Cloning of the mycobacterial epitope recognized by T lymphocytes in adjuvant arthritis *Nature* 331:171

25 Elias, D, Markovits, D., Reshef, T., van der Zee, R and Cohen, I. R. 1990 Induction and therapy of autoimmune diabetes in the non-obese diabetic (NOD/Lt) mouse by a 65-kDa heat shock protein *Proc. Natl Acad. Sci USA* 87·1576

26 Elias, D., Reshef, T., Birk, O. S., van der Zee, R, Walker, M. D. and Cohen, I. R. 1991. Vaccination against autoimmune mouse diabetes with T-cell epitope of the human 65 heat shock protein *Proc. Natl Acad Sci USA* 88.3088.

27 Elias, D. and Cohen, I. R. 1994. Peptide therapy for diabetes in NOD mice *Lancet* 343·704

28 De Graeff Meeder, E. R., Rijkers, G. T., Voorhorst-Ogink, M M, Kuis, W., Van der Zee, R and Zegers, J M 1993 Antibodies to human HSP60 in patient with juvenile chronic arthritis, diabetes mellitus, and cystic fibrosis *Pediatr Res* 34·424

29 Anderton, S M, van der Zee, R, Prakken, B., Noordzij, A. and van Eden, W. 1995 Activation of T cells recognizing self 60-kD heat shock protein can protect against experimental arthritis. *J. Exp Med.* 181.943.

30 Yost, H. J, Petersen, R B. and S Lindquist. 1990 Posttranscriptional regulation of heat shock protein synthesis in *Drosophila.* In Morimoto, R I, Tissières, A and Georgopoulos, C., eds, *Stress Proteins in Biology and Medicine*, p. 379. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

31 Holmgren, R, Livak, K, Morimoto, R. I., Freund, R. and Meselson, M 1979. Studies of the cloned sequences from four *Drosophila* heat shock loci. *Cell* 18·1359.

32 Gros-Bellard, M., Oudet, P and Chambon, P. 1973. Isolation of a high molecular-weight DNA from mammalian cells *Eur J Biochem.* 36:32

33 Tiercy, J. M., Gorski, J, Bétuel, H, Freidel, A C, Gebuhrer, L, Jeannet, M. and Mach, B 1989. DNA typing of DRw6 subtypes· correlation with DRB1 and DRB3 allelic sequences by hybridization with oligonucleotide probes. *Hum. Immunol* 24·1

34 Sanger, F., Nicklen, S. and Coulson, A. R 1977 DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad. Sci. USA* 74·5463.

35 Wilkinson, M 1988. A rapid and convenient method for isolation of nuclear, cytoplasmic and total cellular RNA. *Nucleic Acids Res.* 16:10934.

36 Ugozzoli, L, Baldi, L, Delfini, C., Lucarelli, G, Wallace, R. B and Ferrara, G. B. 1989. Genotypic analysis of engraftment in thalassemia following bone marrow transplantation using synthetic oligonucleotides *Bone Marrow Transplant* 4:173

37 Bellocq, C. and Kolakofsky, D. 1987. Translational requirement for La Crosse virus S-mRNA synthesis a possible mechanism. *J. Virol.* 61:3960

38 Jindal, S, Dudani, A K., Singh, B., Harley, C. B. and Gupta, R S. 1989 Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and to the 65-kilodalton mycobacterial antigen. *Mol. Cell Biol.* 9·2279.

39 Lamb, J. R., Bal, V., Mendez-Samperio, P., Mehlert, A., So, A,

Rothbard, J, Jindal, S, Young, R A. and Young, D. B 1992. Stress proteins may provide a link between the immune response to infection and autoimmunity *Int. Immunol.* 1:191

40 Xiao, H and Lis, J T 1988. Germline transformation used to define key features of heat-shock response elements. *Science* 238 1247.

41 Amin, J., Ananthan, J and Voellmy, R. 1988 Key features of heat shock regulatory elements. *Mol. Cell. Biol.* 8:3761

42 Xiao, H, Perisic, O. and Lis, J T 1991. Cooperative binding of *Drosophila* heat shock factor to arrays of a conserved 5 bp unit *Cell* 64:585

43 Kimura, E, Enns, R E., Thiebaut, F and Howell, S. 1993 Regulation of HSP60 expression in a human ovarian carcinoma cell line. *Cancer Chemother Pharmacol* 32 279

44 Cohen, I. R 1991. Autoimmunity to chaperonins in the pathogenesis of arthritis and diabetes *Annu Rev. Immunol.* 9:567.

45 Holoshitz, J, Klajman, A, Drucker, I, Lapidot, Z., Yaretzky, A, Frenkel, A., van Eden, W and Cohen, I R 1986 T lymphocytes of rheumatoid arthritis patients show augmented reactivity to a fraction of mycobacteria cross-reactive with cartilage. *Lancet* 2 305

46 Karlsson Parra, A., Soderstrom, K., Ferm, M., Ivanyi, J, Kiessling, R and Klareskog, L 1990. Presence of human 65 kD heat shock protein (hsp) in inflamed joints and subcutaneous nodules of RA patients [corrected and republished with original paging, article originally printed in *Scand. J. Immunol* 1990 31 283]. *Scand. J. Immunol* 31·283

47 De Graeff Meeder, E. R, Voorhorst, M, van Eden, W, Schuurman, H J, Huber, J, Barkley, D, Maini, R N, Kuis, W, Rijkers, G T and Zegers, B J 1990 Antibodies to the mycobacterial 65-kd heat-shock protein are reactive with synovial tissue of adjuvant arthritic rats and patients with rheumatoid arthritis and osteoarthritis. *Am. J. Pathol.* 137.1013.

48 Venner, T J and Gupta, R S 1990 Nucleotide sequence of rat hsp60 (chaperonin, GroEL homolog) cDNA. *Nucleic Acids Res.* 18 5309.

49 Venner, T J., Singh, B and Gupta, R S 1990 Nucleotide sequences and novel structural features of human and Chinese hamster hsp60 (chaperonin) gene families. *DNA Cell Biol.* 9·545

50 Venner, T. J. and Gupta, R. S. 1990. Nucleotide sequence of mouse HSP60 (chaperonin, GroEL homolog) cDNA. *Biochim. Biophys. Acta* 1087.336

51 Hickey, E, Brandon, S E, Sadis, S, Smale, G. and Weber, L. A. 1986. Molecular cloning of sequences encoding the human heat-shock proteins and their expression during hyperthermia *Gene* 43.147.

52 Hickey, E, Brandon, S E, Potter, R., Stein, G., Stein, J and Weber, L. A. 1986. Sequence and organization of genes encoding the human 27 kDa heat shock protein [published erratum appears in *Nucleic Acids Res* 1986 14·8230] *Nucleic Acids Res.* 14·4127

53 Walter, T, Drabent, B., Krebs, H., Tomalak, M, Heiss, S and Benecke, B. J. 1989. Cloning and analysis of a human 86-kDa heat-shock-protein-encoding gene *Gene* 83·105

54 Sorger, P. K and Pelham, H. R 1987 Cloning and expression of a gene encoding hsc73, the major hsp70-like protein in unstressed rat cells *EMBO J* 6.993.

55 Dworniczk, B. and Mirault, M. E. 1987 Structure and expression of a human gene coding for a 70 kd heat shock cognate protein *Nucleic Acids Res.* 15.5181.

56 Weiner, A M, Deininger, P. L. and Efstratiadis, A 1986 Nonviral retroposons· genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev Biochem* 55:631.

57 Lotscher, E. and Allison, J. P. 1990. Nucleotide and deduced amino acid sequence of a murine cDNA clone encoding one member of the hsp65 multigene family. *Nucleic Acids Res* 18:7153.

58 Bienz, M. and Pelham, H. R. 1986. Heat shock regulatory elements function as an inducible enhancer in the *Xenopus* hsp70 gene and when linked to a heterologous promoter. *Cell* 45.753

59 Pugh, B. F and Tjian, R 1990. Mechanism of transcriptional activation by Sp1. evidence for coactivators. *Cell* 61:1187.

60 Perisic, O, Xiao, H and Lis, J. T. 1989 Stable binding of *Drosophila* heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. *Cell* 59.797