# Amino Acid Sequence of a Chicken Heat Shock Protein Derived from the Complementary DNA Nucleotide Sequence†

Markku S. Kulomaa,‡ Nancy L. Weigel, Don A. Kleinsek, Wanda G. Beattie, Orla M. Conneely, Carl March,§ Tanya Zarucki-Schulz, William T. Schrader, and Bert W. O'Malley*

*Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030*

*Received February 20, 1986; Revised Manuscript Received May 13, 1986*

ABSTRACT: The complete nucleotide sequence for a chicken heat shock protein (hsp108) was determined from cDNA clones isolated from hen oviduct and bursal lymphoma recombinant DNA libraries. This protein has certain biochemical similarities to the progesterone receptor, but it is clearly distinct from it. The initial cDNA clone, isolated from a chicken oviduct cDNA library, was detected by antibody screening and hybrid-selected translation [Zarucki-Schulz, T., Kulomaa, M. S., Headon, D. R., Weigel, N. L., Baez, M., Edwards, D. P., McGuire, W. L., Schrader, W. T., & O'Malley, B. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6358-6362]. The earlier clones were used to screen for additional cDNAs, and cDNAs that define the entire mRNA sequence of hsp108 have been obtained. The nucleotide sequence codes for peptides present in hsp108 as determined by protein microsequencing. The 5′ end of the mRNA was determined by primer extension studies. The mRNA contains a noncoding region of 101 nucleotides upstream from the predicted initiation codon. The 3′ untranslated region contains 244 nucleotides beyond the termination codon, and it contains a predicted polyadenylation signal 26 nucleotides from the end of the complete cDNA. The coding region of 2385 nucleotides corresponds to a polypeptide chain of 795 amino acids, giving a molecular weight of 91 555 for the hsp108 protein. In another paper, evidence is presented that hsp108 shows a high degree of amino acid sequence homology with two heat shock proteins, hsp90 (yeast) and hsp83 (*Drosophila*), and is indeed inducible by heat shock [Sargan, D. R., Tsai, M.-J., & O'Malley, B. W. (1986) *Biochemistry* (following paper in this issue)].

Recently we have prepared a monoclonal antibody that recognizes a protein with the same size and charge as the chicken progesterone receptor B protein (Edwards et al., 1984; Peleg et al., 1985) but which does not recognize the authentic hormone binder even in a denatured form. This protein, which we have now designated as hsp108[1] (heat shock protein, $M_r$ 108 000),[2] copurifies with the receptor B subunit when isolated from mature hens (Weigel et al., 1981) but can be chromatographically separated when isolated from diethylstilbestrol-treated chick oviducts (Peleg et al., 1985). Despite the apparent similarity in chromatographic properties, there is no evidence that hsp108 is physically associated with the progesterone receptor in vivo or during purification.

Due to certain chemical similarities between hsp108 and the hormone-binding progesterone receptor B subunit, we were misled initially that the two proteins could be similar in peptide structure. Instead, recent immunologic studies showed that they are not closely related, and peptide mapping studies also reveal no significant homologies. In order to study the regulation of hsp108, we have attempted to isolate the gene for this protein. Recently, we isolated a partial cDNA clone coding for hsp108 (Zarucki-Schulz et al., 1984) and now report the complete amino acid sequence of the protein deduced from a series of cDNAs and overlapping genomic clones encompassing the entire mature mRNA. The sequence has been confirmed for certain peptides of hsp108, which were isolated and subjected to protein microsequencing. We have also compared the sequence of this protein with the known sequences of other proteins and find that it has a high degree of homology with the sequences of two heat shock proteins, hsp90 from yeast (Farrelly & Finkelstein, 1984) and hsp83 from *Drosophila melanogaster* (Holmgren et al., 1979).

## Experimental Procedures

*Chemicals and Reagents.* Sequencing chemicals were purchased from Applied Biosystems, Inc. Tetrahydrofuran (UV cutoff 210 without preservative), 2-propanol, and acetonitrile were purchased from Burdick and Jackson. HPLC-grade acetone and general chemicals were purchased from Fisher. Pellicular CN packing was purchased from Applied Bioscience Division (State College, PA). Trifluoroacetic acid was redistilled before use. TPCK-treated trypsin was from Worthington. Deoxynucleoside triphosphates and dideoxynucleoside triphosphates for DNA sequencing were purchased from P-L Biochemicals. [$\alpha$-$^{35}$S]dATP was from Amersham Corporation, and [$\alpha$-$^{32}$P]dCTP and [$\gamma$-$^{32}$P]dATP were from ICN Pharmaceuticals. AMV reverse transcriptase was ob-

† This research was supported in part by National Institutes of Health Grants HD07857 and HD08188.

‡ Present address: University of Tampere, Department of Biomedical Science, SF-33101 Tampere, Finland.

§ Present address: Immunex Corp., Seattle, WA 98101.

[1] Abbreviations: hsp108, chicken 108-kDa heat shock protein; HPLC, high-performance liquid chromatography; PTH, phenylthiohydantoin; TFA, trifluoroacetic acid; SDS, sodium dodecyl sulfate; kDa, kilodalton; kb, kilobase; bp, base pair; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; PIPES, piperazine-*N*,*N*′-bis(2-ethanesulfonic acid).

[2] In the following paper, we show that this protein's abundance in cells is increased by heat shock (Sargan et al., 1986). Thus the nomenclature hsp108 is appropriate to indicate the protein's mass and sensitivity to heat shock. This same protein was defined by us earlier as PRB (Zarucki-Schulz et al., 1984; Edwards et al., 1984). Since this protein is immunologically distinct from progesterone receptor (Peleg et al., 1985) and bears little structural relationship to that protein, the present nomenclature is preferable.

0006-2960/86/0425-6244$01.50/0 

tained from Life Sciences, Inc. Cloning and restriction enzymes were obtained from Bethesda Research Laboratories and New England Biolabs. Polynucleotide kinase and T4 DNA ligase were from Boehringer Mannheim. All other chemicals were reagent grade.

*Purification of hsp108.* The protein was purified as previously described (Weigel et al., 1981) with gel filtration on a TSK 4000 column as the final step (Edwards et al., 1984). The protein was greater than 95% pure as judged by denaturing gel electrophoresis in 0.1% SDS.

*Trypsin Digestion.* The salt-free protein (~6 nmol) was dissolved in 1 mL of 0.08 M Tris-HCl, pH 8.8, containing 10 mM EDTA and 6 M guanidine hydrochloride. Dithiothreitol [8 μL of a 10% (w/v) solution] was added, and the sample was incubated for 50 min at 50 °C. Iodoacetic acid (50 μL of a 0.25 M solution) was added, and the sample was incubated for 15 min at 50 °C. The sample was neutralized and desalted on a TSK 4000 column equilibrated in 0.1 M $Na_xPO_4$, pH 7.0. The carboxymethylated hsp108 peak was pooled and incubated with 3 μg of trypsin in 6 mL of 0.1 M $Na_xPO_4$, pH 7.0, for 16 h at 37 °C.

*Separation of Tryptic Peptides.* A Beckman Model 334 HPLC[1] equipped with two Model 110 pumps and a Model 165 variable-wavelength detector was used for analysis of the peptides. The digest was applied to a 4.6 × 250 mm Vydac $C_{18}$ column equilibrated with 0.1% trifluoroacetic acid (TFA).[1] The peptides were eluted with a gradient of 0–50% 2-propanol in 0.1% TFA at $^1/_2$%/min and a flow rate of 1 mL/min. The peptides were detected at 210 nm.

*Sequencing of Peptides.* A Model 470A Applied Biosystems gas-phase sequencer was used for all analyses. PTH[1]-amino acids were identified with a Beckman Model 344 HPLC equipped with two Model 112 pumps and two Model 160 detectors. Samples were injected automatically with a Model M710B Waters Intelligent Sample Processor (WISP). PTH-amino acids were analyzed on a 5-μm pore size IBM cyano column (4.5 mm × 25 cm) with either a Du Pont ETH guard column or a pellicular CN guard column. The samples were analyzed at 35 °C with a gradient system consisting of buffer A (0.02 M sodium acetate, 5% tetrahydrofuran, pH 5.95, and 17 μL of HPLC-grade acetone/L) and buffer B (acetonitrile). The PTH-amino acids are eluted by increasing buffer B from 11% to 45% at 2%/min at a flow rate of 1 mL/min.

*Preparation of a 31-kDa Peptide from hsp108.* We have previously shown that partial trypsinization of the native hsp108 protein results in the production of a 31-kDa peptide as well as many peptides too small to be resolved by denaturing gel electrophoresis (Edwards et al., 1984). Purified hsp108 (100 μg/mL) was incubated with 2% (w/w) trypsin in 0.3 M NaCl, 10 mM Tris, pH 7.5, 1 mM EDTA, and 12 mM thioglycerol overnight at 37 °C. The 31-kDa fragment was separated from other peptides by gel filtration on a 60-cm TSK 2000 column equilibrated in 0.1 M $Na_xPO_4$, pH 7.0. About 7 nmol of 31-kDa peptide was dialyzed against 0.05M $NH_4HCO_3$, heated 5 min at 90 °C to denature the protein, and redigested with 2.5% trypsin overnight at 37 °C. The digest was lyophilized, redissolved in 0.1% TFA, and applied to the Vydac $C_{18}$ column equilibrated in 0.1% TFA (buffer A). The peptides were eluted with a gradient of buffer A and buffer B (2-propanol in 0.1% TFA) at 1 mL/min. Buffer B was brought to 5% in 2.5 min. At 10 min the percent buffer B was increased at $^1/_4$%/min until 40% buffer B was reached. Buffer B was then increased at $^1/_2$%/min until the final concentration was 60%. Peptide elutions were monitored at 210 nm.

*Construction of the Recombinant cDNA Libraries.* Poly(A)+ RNA was prepared from oviducts of 5-week-old white Leghorn chickens treated with diethylstilbestrol (DES) for 3 weeks or from white Leghorn laying hens. Either phenol/SDS extraction (Rosen et al., 1975) or guanidine thiocyanate (Ullrich et al., 1977) was used for RNA extraction, followed by two cycles of oligo(dT)–cellulose chromatography (Aviv & Leder 1972).

Double-stranded cDNA was synthesized as described (Gubler & Hoffman, 1983). Natural *Eco*RI sites in the double-stranded cDNA were protected by methylation with *Eco*RI methylase (Maniatis et al., 1978). The cDNA was repaired by the Klenow fragment of DNA polymerase I (Jacobsen et al., 1974) and synthetic *Eco*RI linkers (BRL) were ligated (Sugino et al., 1977) to the blunt ends of the cDNA. Following digestion with *Eco*RI, the cDNA was inserted into the single *Eco*RI site of bacteriophage λgt11, as described by Young and Davis (1983a,b), packaged in vitro into phage particles, and amplified by infection of the bacterial host *Escherichia coli* Y1088 (Maniatis et al., 1982). The chicken bursal lymphoma library was a gift from Steven McKnight (Carnegie Institute of Washington, Baltimore, MD). The natural *Eco*RI sites in the latter library were not protected by in vitro methylation.

*Screening of the Libraries.* The λgt11 bacteriophage recombinant libraries were screened at a density of 5000–10000 plaques/150-mm Petri dish with a nick-translated cDNA probe (Maniatis et al., 1982) prepared from recombinant clone pPRB-2, previously isolated in our laboratory (Zarucki-Schulz et al., 1984). The filters were hybridized (Maniatis et al., 1982) in a sealed bag overnight at 68 °C in the presence of (1–2) × $10^6$ cpm of $^{32}$P-labeled probe/filter [sp act. (1–5) × $10^8$ cpm/μg], washed with 2 × SSC/0.1% SDS (1 × SSC: 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), dried, and exposed to X-ray film (Kodak XR-Omat) with an intensifying screen at −70 °C. Phage DNA was isolated from single plaque recombinant clones by a rapid small-scale isolation procedure (Benson & Taylor, 1984), digested with *Eco*RI, and analyzed for insert size on a 1% agarose gel. Selected recombinant DNA molecules recovered from these plaques were then subcloned into the *Eco*RI site of vector pBR322. Plasmid DNAs containing the desired insert were isolated by a cleared lysate method (Birnboim & Doly, 1979; Katz et al., 1973).

*DNA and RNA Blot Hybridization Analysis.* DNA fragments were separated in an agarose gel, transferred to a nitrocellulose filter, and analyzed by the DNA blot hybridization method of Southern as previously described (Southern, 1975). The RNA blot hybridization (Northern) analysis was performed after separation of RNA molecules on a denaturing formaldehyde gel as described in detail elsewhere (Lehrach et al., 1977; Goldberg, 1980).

*5′ Primer Extension Study.* To map the 5′ end of the hsp108 mRNA, a single-stranded primer was prepared from the cDNA coding strand of the most 5′ recombinant clone pPRB-13, isolated in this study (see Results). An *Eco*RI–*Sph*I DNA fragment from pPRB-13 was subcloned into the m13mp18 cloning vector (Sanger et al., 1977) to generate an M13 recombinant phage containing the sense strand of the cDNA fragment. Single-stranded M13 recombinant DNA was isolated (Sanger et al., 1977; Messing, 1983) and hybridized to a 17-bp M13 sequencing primer (Maniatis et al., 1982) labeled at the 5′ end with [γ-$^{32}$P]ATP by using T4 DNA polynucleotide kinase (Maniatis et al., 1982). The DNA was extended in a reaction containing 50 μM dGTP, dATP, dCTP, and dTTP and the Klenow fragment of DNA polymerase I

by incubation at 30 °C for 30 min (Maniatis et al., 1982). The DNA was precipitated with ethanol, resuspended in an appropriate restriction endonuclease buffer, and digested with *Ava*I. After ethanol precipitation, the DNA was dissolved in formamide–dye (0.03% xylene cyanol FF/0.03% bromphenol blue/10 mM EDTA/99% deionized formamide), heat denatured at 90 °C for 3 min, and fractionated by electrophoresis through a 5% polyacrylamide gel. A 108-bp labeled single-stranded fragment was purified from the gel (Maniatis et al., 1982). The fragment contained 26 bases of M13 DNA at its 5′ end and 82 bases of hsp108 anticoding sequence at the 3′ end, complementary to a region of the mRNA spanning from +52 to +134 with respect to the cap site of the mRNA.

The 108-bp end-labeled [$^{32}$P]DNA primer was coprecipitated with 400 μg of poly(A)+ RNA obtained from estrogen-stimulated chick oviducts and dissolved in four aliquots (100 μL each) of 400 mM NaCl, 1 mM EDTA, and 10 mM Tris-HCl, pH 7.5. The sample was heated in a boiling water bath for 5 min and immediately transferred to a 68 °C water bath for 18 h. After this hybridization reaction, the sample was precipitated with ethanol, pelleted, washed in 70% ethanol, and dried under vacuum. The precipitate was then resuspended in 200 μL of extension buffer, and the primer DNA was extended by using AMV reverse transcriptase as previously described (Bina-Stein et al., 1979; Ghosh et al., 1980; McKnight et al., 1981). The extended DNA product was purified from a 10% nondenaturing polyacrylamide gel and sequenced by the method of Maxam and Gilbert (1977) (Sanger & Coulson, 1978). The size of the extended product was estimated by electrophoresis carried out in a 10% polyacrylamide gel containing 7 M urea, followed by autoradiography.

*S1 Nuclease Mapping.* To map the 3′ end of the hsp108 gene, a $^{32}$P end-labeled probe was prepared from genomic clone pHSP2.8, an *Eco*RI fragment isolated in our laboratory and inserted in pBR322 (D. Kleinsek, unpublished results). This 2.8-kb *Eco*RI fragment containing the 3′ portion of hsp108 DNA was isolated by digestion of pHSP2.8 with *Eco*RI, followed by slab gel electrophoresis. The staggered ends were filled in with the Klenow fragment of *E. coli* DNA polymerase I in the presence of 25 μCi of [α-$^{32}$P]dATP (3000 Ci/mmol) (Maniatis et al., 1982). The DNA was subsequently digested with HindIII, and the resulting 1.5-kb end-labeled probe was purified from a 6% polyacrylamide gel (Maniatis et al., 1982). This fragment contained a 3′ end label on the anti-sense strand. The S1 nuclease protection experiment was performed according to previously published procedures (Berk & Sharp, 1977). The 1.5-kb probe ($20 \times 10^3$ cpm) was coprecipitated with 25 μg of poly(A)+ RNA and resuspended in 20 μL of hybridization buffer (20 mM PIPES, pH 6.4, 1 mM EDTA, 400 mM NaCl, and 70% formamide). The hybridization reaction was incubated at 53 °C for 12 h, followed by incubation at 42 °C for 9 h. The S1 nuclease digestion of the RNA–DNA hybrids was performed as reported previously (Berk & Sharp, 1977). The size of the final RNA-protected DNA fragment was determined by electrophoresis on a 10% polyacrylamide gel containing 7 M urea.

## Results

*Isolation and Amino Acid Sequencing of Tryptic Peptides.* Intact hsp108 (250 pmol) was subjected to amino acid sequence analysis as described under Experimental Procedures. No sequence was detected, although even a 10% yield would have been readily detected. Previous sequencing trials using hsp108 electroeluted from denaturing gel electrophoresis also failed to yield any sequence. We concluded that the amino terminus of this protein is modified. Therefore, several tryptic peptides derived from the hsp108 protein were purified for amino acid sequencing.

Table I: Amino Acid Sequence of hsp108 Peptides

| | peptide no.[a] | | | | |
|---|---|---|---|---|---|
| cycle[b] | Ia (pmol) | Ib (pmol) | II (pmol) | III (pmol) | IV (pmol) |
| 1 | Val (504)[c] | Val (1370) | Gly (629) | Thr (+)[d] | Thr (+)[d] |
| 2 | Phe (381) | Phe (1783) | Leu (1120) | Val (116) | Val (442) |
| 3 | Ile (690) | Ile (1722) | Phe (876) | Trp (29) | Trp (146) |
| 4 | Thr (+)[d] | Thr (+)[d] | Asp (507) | Asp (39) | Asp (300) |
| 5 | Asp (263) | Asp (317) | Glu (430) | Trp (19) | Trp (85) |
| 6 | Asp (318) | Asp (924) | Tyr (740) | Glu (19) | Glu (97) |
| 7 | Phe (276) | Phe (527) | Gly (482) | Leu (29) | Leu (125) |
| 8 | His (62) | His (ND)[e] | ND[e] | Met (18) | Met (55) |
| 9 | Asp (91) | Asp (317) | Lys (218) | ND[e] | Asn (88) |

[a] Peptides Ia and Ib from hsp108 were analyses of two aliquots of the same HPLC peak I shown in Figure 1A. Peptide II was obtained from peak II of Figure 1B. Peptides III and IV were obtained for peaks III and IV of Figure 1C, respectively. [b] Amino acid sequencing by gas-phase automated sequencing as described under Experimental Procedures. [c] Numbers in parentheses are picomoles of amino acid detected. [d] Threonine detected by absorbance of the PTH derivative at 313 nm. [e] ND, no residue detected at this cycle.

Table II: Amino Acid Composition of Peptide II[a]

| amino acid | pmol of amino acid/pmol of lysine |
|---|---|
| Asp | 1.0 |
| Ser | 0.9 |
| Glu | 1.1 |
| Gly | 2.0 |
| Leu | 0.9 |
| Tyr | 0.9 |
| Phe | 0.9 |
| Lys | 1.0 |

[a] Determined by amino acid analysis after hydrolysis at 110 °C for 24 h in 5.7 N HCl with an LKB amino acid analyzer.

Figure 1A shows the separation of the tryptic peptides derived from hsp108 as described under Experimental Procedures. Peak I, indicated by the bold arrow, appeared to be well resolved from other peaks. One-third of the sample was subjected to automated sequencing as described under Experimental Procedures. The remaining two-thirds of the sample was analyzed separately in a subsequent experiment. The results of these two analyses are shown in Table I. The combined runs suggested that this peak contained at least 2.5 nmol of sequenced peptide derived from a digest of 6 nmol. A second UV-absorbing peak marked with a thin arrow as peptide II in Figure 1A was purified further by rerunning on a Vydac $C_{18}$ column with a different buffer as shown in Figure 1B.

This peptide (2.4 nmol) was also sequenced as shown in Table I. All positions except for position 8 were identified. Amino acid composition analysis of the peptide (Table II) indicates the presence of serine; thus we assigned serine tentatively to position 8. This identification was confirmed when the nucleotide sequence was obtained (see Figure 5).

The peptides described above derived from hsp108 were obtained from total trypsin digestion of the protein. In order to obtain sequence data for a region of the protein identifiable in the native structure, we prepared a 31-kDa fragment of hsp108, obtained by mild trypsin digestion under nondenaturing conditions. Figure 1C shows the separation of the peptides derived from the 31-kDa fragment. Two of these peptides were analyzed by automated sequencing. Eighty percent of the peptide fraction III in Figure 1C was analyzed, and the amino acid sequence is shown in Table I. Peptide IV was also analyzed. The sequence of this peptide is also shown

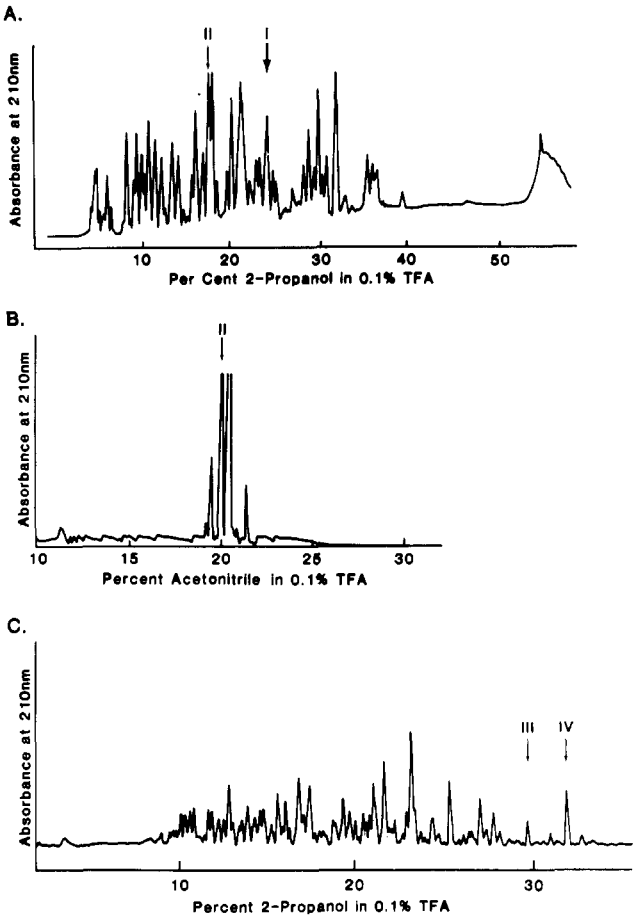


FIGURE 1: Reverse-phase HPLC chromatograms of peptides derived from exhaustive digestion of hsp108 with trypsin. Digests were prepared from purified hsp108 (A, B) or from a 31-kDa fragment (C) prepared as described under Experimental Procedures. Peptides were applied to a Vydac $C_{18}$ column and eluted with the indicated buffers shown on the abscissa. (A) Elution profile of total tryptic digest of hsp108. Peaks I (broad arrow) and II (thin arrow) were collected manually in siliconized plastic tubes and neutralized with 1.0 M $NH_4OH$. (B) Rechromatography of peak II from (A) on Vydac $C_{18}$ with a different solvent system. Peak II was evaporated to dryness, redissolved in 0.1% trifluoroacetic acid (TFA), and applied to the column equilibrated in 10% acetonitrile and 0.1% TFA. Peak II was pooled from this elution, neutralized, and used for subsequent amino acid sequencing. (C) Elution profile of the tryptic peptides derived from the 31-kDa fragment of hsp108 prepared as described under Experimental Procedures. Two peaks, III and IV, shown by the arrows were collected manually and used for amino acid sequencing.

in Table I. The N-terminal sequences of the two peptides were identical. Amino acid analysis of peptide III showed that it contained 2 mol of Lys/1 mol of Arg, indicating that it was an incompletely digested fragment. Inspection of the sequence derived from the cDNA (Figure 5) reveals both an Arg–Pro and a Lys–Pro bond, both of which may digest slowly with trypsin (Allen, 1981). Furthermore, the cDNA data predict only one copy of the first nine amino acids of peptide III, thereby eliminating the possibility that this sequence is repeated in two different regions of hsp108.

*Isolation of the cDNA Clones.* Three overlapping hsp108 cDNA clones (pPRB-1, pPRB-2, and pPRB-3) have been isolated previously from a recombinant DNA library constructed with size-selected mRNA (Zarucki-Schulz et al., 1984). Their sizes and relative positions are shown at the top of Figure 2. Clone PRB-2 was used for isolation of longer cDNAs in the present study.

The bursal lymphoma cDNA library contained cDNA inserts with an average size of about 1.5–2 kb. When 100 000 clones from the library were screened with a nick-translated *Hin*dIII/*Eco*RI fragment of pPRB-2, 150–200 clones, i.e., 0.15–0.2%, were recognized by the probe. Phage DNA from 10 randomly selected positive clones was isolated and digested with *Eco*RI, and DNA fragments were analyzed on a 1% agarose gel. All ten λ recombinant DNAs contained the same size inserts of about 1.9 kb. They all hybridized to $^{32}P$-labeled pPRB-2 when analyzed by Southern blot analysis (data not shown). The results suggested the presence of one or two natural *Eco*RI sites in the hsp108 cDNA that had not been protected by methylation (see Experimental Procedures). The 1.9-kb insert from one λ clone was subcloned into plasmid pBR322 and termed pPRB-6. This insert was isolated and analyzed by restriction enzyme mapping as shown in Figure 2, which shows its overlap with PRB-2. DNA sequencing data of PRB-6 are included in Figure 5. The fragment is 1928 nucleotides in length and has an open reading frame from one end to a stop codon, TAA, 56 nucleotides from the other end. This sequence also coded for hsp108 peptides I, II, and III of Table I. The sequenced restriction fragments are shown below the cDNA map in Figure 2.

Clone PRB-6 is bounded by two natural *Eco*RI sites. To obtain additional clones extending closer to the 5′ end of the coding sequence, we used as a probe a genomic DNA clone obtained by us in experiments to be reported elsewhere (D. Kleinsek, unpublished experiments). This genomic clone, pHSP3.3, contained the *Eco*RI site at the 5′ end of PRB-6. We prepared an *Eco*RI fragment of pHSP3.3 and used it as a probe for additional cDNAs. This genomic *Eco*RI fragment did not hybridize to the pPRB-6 insert when analyzed by DNA blot hybridization analysis, whereas it did hybridize to the same size RNA in RNA blot hybridization analysis of total poly(A)+ RNA isolated from the hen oviduct (not shown). These data indicated that the genomic DNA fragment contained additional nucleotide sequence toward the 5′ end of the hsp108 cDNA. The genomic probe identified an additional short cDNA clone when the bursal lymphoma library was rescreened by using this fragment. This new clone was then used to screen the λgt11 library. A new cDNA clone, designated PRB-13, was detected. The PRB-13 insert was subcloned into pBR322 vector, isolated, and analyzed by restriction mapping and sequencing. Its location relative to the other cDNA clones is shown in Figure 2. It was 609 nucleotides long and had an open reading frame throughout the entire nucleotide sequence (Figure 5).

*Definition of the 5′ and 3′ Ends of the hsp108 mRNA.* In order to define the 5′ terminus of the hsp108 mRNA, primer-directed cDNA synthesis was performed as described under Experimental Procedures. A labeled primer containing the *Eco*RI–*Sph*I 5′ fragment from PRB-13 was used for extension as shown in Figure 3. The size estimation of the extended product was about 160 bp. This extended product was sequenced, and its length was 160 nucleotides. As shown in Figure 5 the sequence begins with an adenine residue located 11 nucleotides upstream from the end of PRB-13. We concluded that this adenine is the first nucleotide of the hsp108 mRNA sequence and is thus the cap site.

The first AUG codon occurs 102 residues downstream from the cap site, thus defining the translation start site. This site is encoded 91 nucleotides in from the end of PRB-13 as shown in Figure 5.

Both PRB-13 and PRB-6 are bounded by an *Eco*RI site at their junction in the proposed map of Figure 2. None of the sequenced fragments crossed this junction. It was possible that an additional *Eco*RI fragment existed between these two

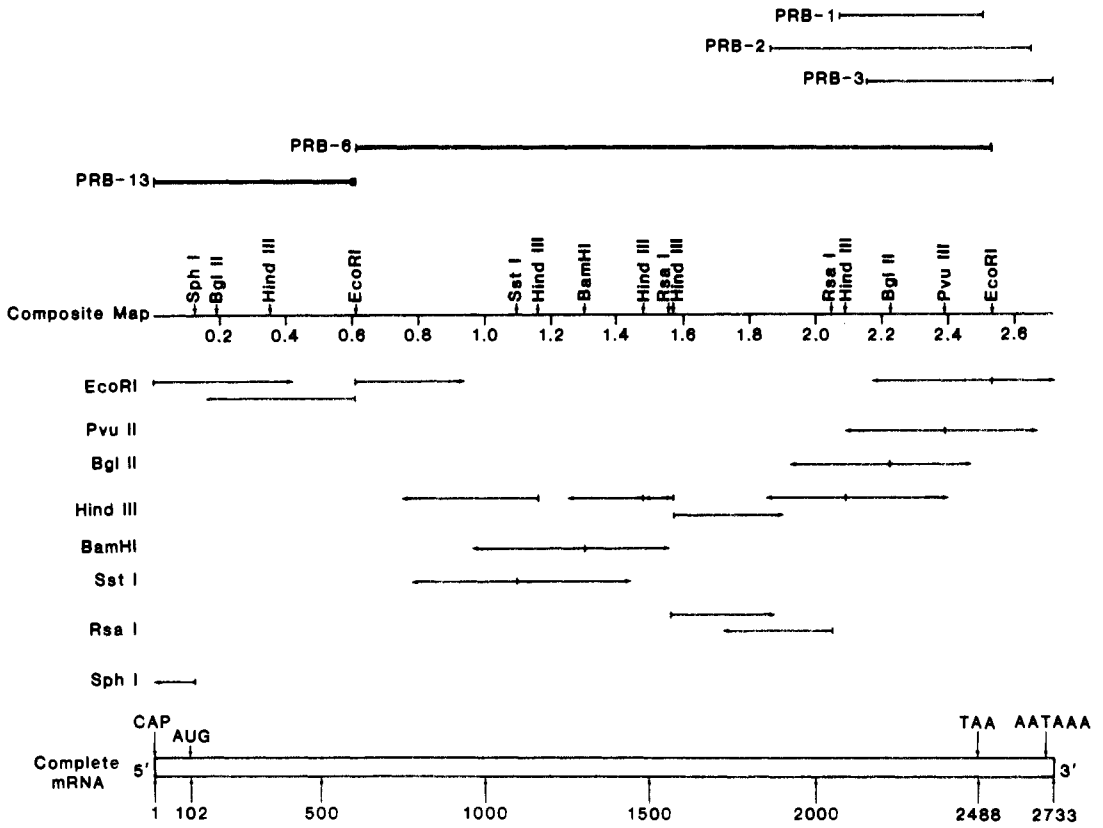


FIGURE 2: Recombinant cDNA clones used to define the hsp108 transcript. As described in the text and under Experimental Procedures, six cDNA clones were isolated and subcloned into pBR322 and designated as pPRB. Sizes of the cDNA are given in parentheses. The composite restriction map line shows placement of the various cDNAs and relevant restriction sites. Numbers below this line refer to kilobase pairs of fragment length, beginning at the cap site of the complete mRNA sequence. PRB-1, -2, -3, -6, and -13 are the separate clones that were sequenced to produce most of the sequence of the complete cDNA, which is represented by the composite map. The sequence of the cDNAs was determined by the dideoxy chain termination method (Sanger et al., 1977) on templates derived from M13 phages. The vertical lines for each enzyme at the bottom represent the position of restriction sites where fragments were generated for ligation into the multiple-cloning region of an M13 vector (Messing, 1983). The arrows indicate the direction and extent of sequence from these restriction sites. Sixty-two percent of the final sequence was determined from both strands. Twenty-five percent of the final sequence was determined at least twice on one strand from overlapping restriction sites, and the remaining 13 percent was determined at least twice on one strand from the same restriction site. Below the restriction enzyme sequencing strategy maps is a block diagram of the complete mRNA deduced from these clones together with the 5′ and 3′ end data of Figures 3 and 4. The only restriction site that has not been sequenced across is the *Eco*RI site at position 0.61 kb from the cap site. However, the cDNA sequence obtained is colinear with the genomic sequence (Kleinsek, unpublished results) obtained from a single clone of λ phage. In addition a genomic clone containing this site was digested with *Eco*RI and the DNA fragments were end-labeled with [$^{32}$P]dATP by using the Klenow fragment of *E. coli* DNA polymerase I. The fragments were separated on an 8% acrylamide gel, which would be able to detect lengths down to 15 bp. Only the expected high molecular weight *Eco*RI bands were obtained, which indicated that no further *Eco*RI fragments were present in this region of the cDNA. The locations of the initiation and stop codons and the polyadenylation signal are shown. The complete sequence is presented in Figure 5.

pieces. We tested this possibility using genomic clone λ147 (D. Kleinsek, unpublished experiments), which hybridizes to both PRB-6 and PRB-13 and should contain any missing *Eco*RI fragments. The genomic clone was digested with *Eco*RI, and the DNA fragments were end-labeled with [$^{32}$P]dATP, using the Klenow fragment of *E. coli* DNA polymerase I. The fragments were separated on an 8% acrylamide gel that would be able to detect lengths down to 15 bp. Only the expected high molecular weight *Eco*RI genomic fragments were obtained, which suggested that no further *Eco*RI fragments were present in this region of the cDNA (data not shown). Such a test does not exclude the presence of a short cDNA sequence of between 8 and 15 nucleotides in length.

The size of the 3′ untranslated region was determined by S1 nuclease mapping analysis. Chicken oviduct poly(A)+ RNA was hybridized with an *Eco*RI fragment isolated from genomic clone pHSP2.8 as described under Experimental Procedures. This fragment contains the 3′ end of the mRNA sequence and extends about 1.5 kb downstream of the translation termination codon. The DNA probe was labeled at the 3′ end of the anti-sense strand as described under Experimental Procedures. The DNA–RNA hybrids were digested with S1 nuclease and fractionated on a denaturing polyacrylamide gel. The results of this experiment are shown in Figure 4. About 194 nucleotides of the DNA probe were protected by RNA from S1 nuclease digestion. This is 7 residues longer than the cDNA clone pPRB-3 and is 26 residues downstream of the polyadenylation signal (AATAAA) shown in Figure 5.

The combined data from 5′ primer extension, 3′ S1 nuclease mapping, and DNA sequence analysis of the cDNA clones specify an mRNA of 2733 bases in length, excluding the poly(A)+ tail. The complete sequence is shown in Figure 5. Figure 5 shows the location of each cDNA clone and the derived amino acid sequence. From that sequence, the peptides I, II, and III/IV are located as shown. The presence of these three peptide sequences within that predicted from nucleotide data thus confirms the identity of the clones as being cDNA's coding for hsp108.

The size of the mRNA also has been estimated by Northern RNA blotting (Lehrach et al., 1977; Goldberg, 1980). Oviduct poly(A)+ RNA was analyzed by hybridization with PRB-2 cDNA. The size calculated by this method was 2750 bp

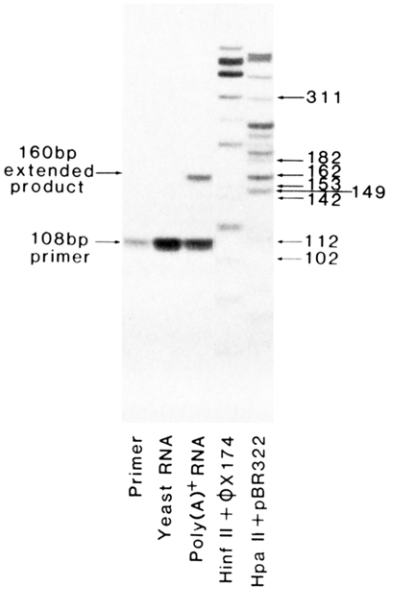


FIGURE 3: Primer extension study to determine the transcription initiation site. A 108-bp DNA primer was prepared from cDNA clone pPRB-13 by subcloning an *Ava*I–*Sph*I fragment into vector M13 as described under Experimental Procedures. This [$^{32}$P]DNA single-stranded probe was hybridized to oviduct poly(A)+ RNA. Hybrids were isolated, and the [$^{32}$P]DNA strand was extended by using AMV reverse transcriptase. The resulting DNA was analyzed for its size by gel electrophoresis as described under Experimental Procedures. Arrows on right of lanes denote DNA fragment lengths of DNA molecular weight standards run in the two right-hand lanes. Primer alone (108 bp; left lane) or primer after incubation with yeast RNA control (second lane from left) showed only a 108-bp DNA fragment. Middle lane is DNA size after hybridization and primer extension by using oviduct RNA. Upper band is identified on left as a 160-bp extended product. Lanes shown are an autoradiograph of a gel after electrophoresis. Two molecular weight standard lanes are shown; these are a *Hpa*II digest of pBR322 and a *Hin*fI digest of $\phi$X174 DNA.

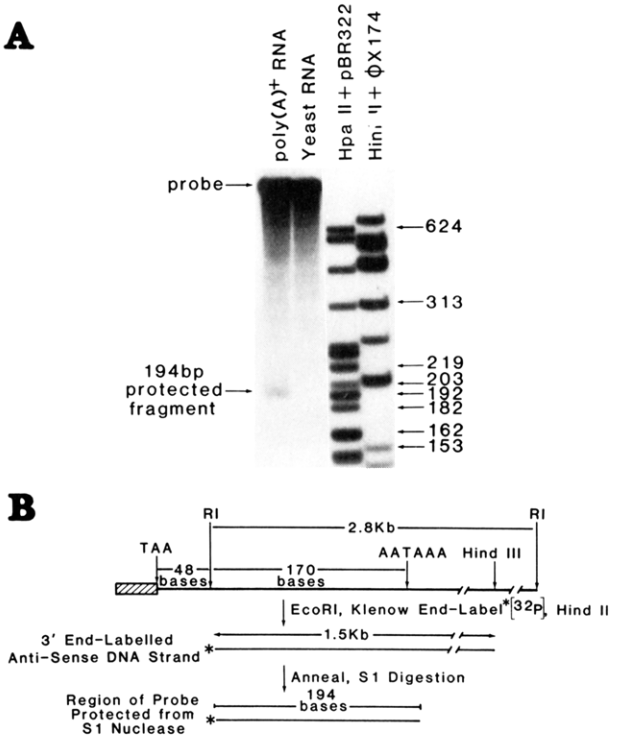


FIGURE 4: S1 nuclease mapping of the 3′ untranslated region of hsp108 mRNA. (A) Autoradiogram of gel electrophoresis experiment to determine the length of [$^{32}$P]DNA probe protected by mRNA. (B) Schematic map of the 3′ end of the gene showing strategy for preparation of hybridization probe. A 1.5 kb end-labeled DNA was prepared from genomic clone pHSP 2.8 by digestion with *Eco*RI and *Hin*dIII as described under Experimental Procedures. After hybridization to either 25 $\mu$g of poly(A)+ RNA from estrogen-treated chick oviducts [left lane, (A)] or 25 $\mu$g of yeast RNA (second lane from left), S1 nuclease digestion yielded undigested [$^{32}$P]DNA/mRNA hybrids. Size of the [$^{32}$P]DNA was determined in comparison to molecular weight standards shown by the arrows on the right. Numbers show DNA fragment length in base pairs with the same standards as used in Figure 3. In (B) the hatched block shows the extent of the mRNA translated region and the TAA stop codon.

(Zarucki-Schulz et al., 1984). These two determinations yielded comparable mRNA size estimates for PRB mRNA.

*Primary Structure of the hsp108.* The open reading frame defined by clones PRB-6 and PRB-13 extends for a length of 2385 nucleotides from the first AUG to the TAA stop codon. This sequence codes for a polypeptide chain of 795 amino acids with a molecular weight of 91 555. The predicted amino acid composition shows a relatively high amount of charged amino acid residues, corresponding to more than a third of the protein. Particularly, the predicted number of glutamic acid and lysine residues, 105 and 81, respectively, is high.

## Discussion

The complete sequence of the hsp108 protein has been determined by using a combination of protein sequencing and nucleotide sequencing. The amino acid sequences obtained agreed with those found by sequencing the cDNA. Evidence that this protein can be regulated by heat shock is presented in the following paper (Sargan et al., 1986).

The total length of the hsp108 protein cDNA was 2733 nucleotides, which was in good agreement with the size for the hsp108 mRNA from RNA blot hybridization analysis (Zarucki-Schulz et al., 1984; Kleinsek, unpublished results). The noncoding region at the 3′ end was 244 nucleotides long and contained the polyadenylation signal AATAAA (Benoist et al., 1980) 26 nucleotides from the end of the clone.

Since the clone containing the AATAAA sequence did not extend into the poly(A) tail, we cannot exclude the possibility that a second polyadenylation signal lies further downstream. However, from the good agreement in mRNA lengths obtained by sequencing and Northern blots, such a site must lie close to the one reported here. The S1 nuclease mapping data (Figure 5) are also consistent with use of the AATAAA sequence shown. The first ATG codon (nucleotides 102–104), corresponding to methionine, left an open reading frame of 101 nucleotides at the 5′ end of the cDNA. Thus, the hsp108 has an mRNA coding region of 2385 nucleotides able to encode a polypeptide chain of 795 amino acids with a molecular weight of 91 555. It was slightly smaller than expected from polyacrylamide gel electrophoresis in the presence of SDS (Porzio & Pearson, 1974), where the hsp108 protein migrated above a 97 400 molecular weight marker at 108 kDa (Zarucki-Schulz et al., 1984).

By computer research of a data base library (Browner & Lawrence, 1986), 48% homology was found between the amino acid sequence of the hsp108 protein and a yeast heat shock protein, hsp90 (Farrelly & Finkelstein, 1984). Although heat shock proteins are a group of proteins identified by their induction as a result of heat shock, the function or functions of these proteins are as yet unknown. They differ greatly in size, ranging from about 20 000 to 110 000 daltons, in subcellular localization, and in DNA binding properties (Nover, 1984; Schlesinger et al., 1982; Neidhardt et al., 1984). The functions of the hsp90 and hsp83 proteins are not known. No significant homology with medium-sized (68K–70K) heat shock proteins, with small (22K–27K) heat shock proteins, or with other proteins (Southgate et al., 1983) inserted into the base library

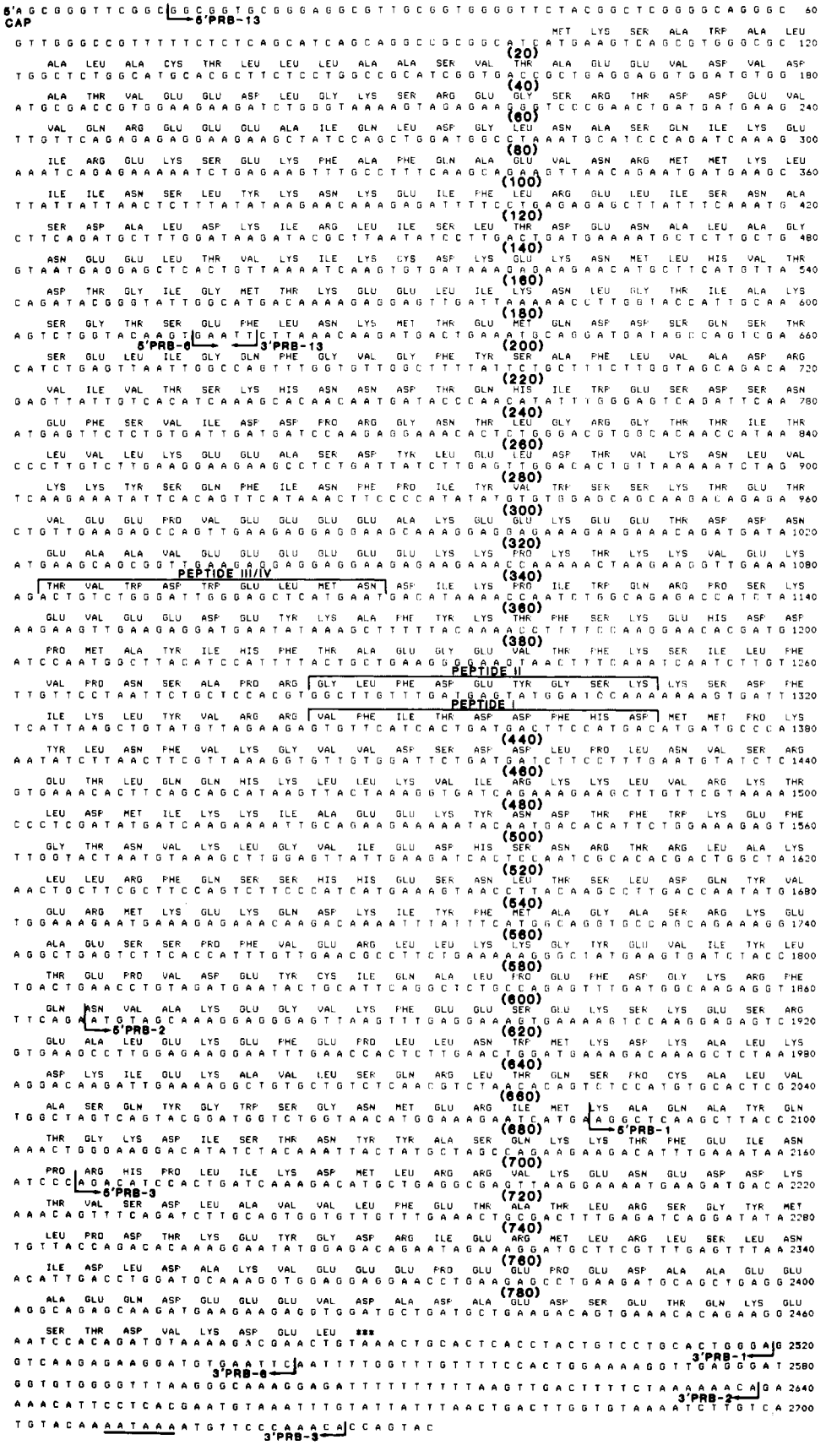


FIGURE 5: Nucleotide sequence and predicted amino acid sequence of hsp108. Complete nucleotide sequence of hsp108 mRNA and its predicted amino acid sequence. Nucleotides are numbered at the right of each line beginning with the 5′ end determined in Figure 3 and by sequencing of the primer-extended product. Arrows below the sequence show the location of the 5′ and 3′ end of each cDNA clone. Predicted amino acids are shown above the nucleotide sequence and are numbered in parentheses, beginning with the initiator methionine. Location of peptides I, II, and III/IV determined by microsequencing of tryptic fragments are shown by brackets above the amino acid sequence. The 3′ untranslated mRNA region determined in Figure 4 extends 7 nucleotides beyond the 3′ end of clone PRB-3. The putative polyadenylation signal is underlined. The nucleotide sequence shown is the sense strand of the cDNA.

was found by computer search. The yeast hsp90 protein appears to have a molecular weight of 90 000 by SDS gel electrophoresis but is only 81 000 when sequenced. Both proteins contain very acidic regions with long stretches of glutamic and aspartic acids. It is possible that these unusual sequences result in anomalous migration on SDS gels. An alternative possibility would be the presence of posttranslational modifications. However, the in vitro translation product of hsp108 from a rabbit reticulocyte lysate has a molecular weight of 108 000 (Zarucki-Schulz et al., 1984). Thus, it appears more likely that the anomalous migration is a function primarily of the amino acid sequence, not of a subsequent modification.

An interesting property of the hsp108 protein is the regulation of its transcription by estrogen (Zarucki-Schulz et al., 1984) and also by progesterone (M. Baez, unpublished results). One of the small heat shock protein genes in *D. melanogaster*, hsp22, was reported to be induced by physiological doses of molting hormone, ecdysone (Viteck & Berger, 1984).

Recently, a partial sequence of chicken hsp90 (Catelli et al., 1985), a protein that associates with the progesterone receptor in chicken cytosol, has been reported. This partial sequence of the protein also shows substantial homology with yeast hsp90, hsp83, and chicken hsp108 but is clearly distinguishable from all three proteins. Peptide mapping of chicken hsp108 and hsp90 (Birnbaumer et al., 1985) demonstrates that the two proteins are quite different. Moreover, monoclonal antibody 9G10 and a polyclonal antibody against hsp108 both fail to detect hsp90 (Weigel et al., submitted for publication). The induction of hsp108 and of other chicken heat shock proteins is examined in the following paper.

**Registry No.** DNA (chicken heat shock protein hsp108 messenger RNA complementary), 103958-78-7; protein hsp108 (chicken reduced), 103958-81-2; RNA (chicken protein hsp108-specifying messenger), 103958-82-3.

## References


Allen, G. (1981) in *Sequencing of Proteins and Peptides* (Work, T. S., & Burdon, R. H., Eds.) Elsevier, Amsterdam.

Aviv, M., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Benoist, C., O'Hare, K., Breathnach, R., & Chambon, P. (1980) *Nucleic Acids Res. 8*, 127–142.

Benson, S. A., & Taylor, R. K. (1984) *BioTechniques 2* (*May–June*), 126–127.

Berk, A. J., & Sharp, P. A. (1977) *Cell* (*Cambridge, Mass.*) *12*, 721–732.

Bina-Stein, M., Thoren, M., Salzman, N., & Thompson, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 731–735.

Birnbaumer, M., Bell, R. C., Schrader, W. T., & O'Malley, B. W. (1984) *J. Biol. Chem. 259*, 1091–1098.

Birnboim, H. D., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Browner, M. F., & Lawrence, C. B. (1986) *Nucleic Acids Res.* (in press).

Catelli, M. G., Binart, N., Feramisco, J. R., & Helfman, D. M. (1985) *Nucleic Acids Res. 13*, 6035–6047.

Edwards, D. P., Weigel, N. L., Schrader, W. T., O'Malley, B. W., & McGuire, W. L. (1984) *Biochemistry 23*, 4427–4435.

Farrelly, F. W., & Finkelstein, D. B. (1984) *J. Biol. Chem. 259*, 5745–5751.

Ghosh, P. K., Reddy, V. B., Piatak, M., Lebowitz, P., & Weissman, S. M. (1980) *Methods Enzymol. 65*, 580–595.

Goldberg, D. A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5794–5798.

Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 263–269.

Holmgren, R., Livak, K., Morimoto, R., Freund, R., & Meselson, M. (1979) *Cell* (*Cambridge, Mass.*) *18*, 1359–1370.

Jacobsen, H., Klenow, H., & Ovargaard-Hansen, K. (1974) *Eur. J. Biochem. 45*, 623–627.

Katz, L., Kingsbury, D. J., & Helinski, D. R. (1973) *J. Bacteriol. 114*, 577–591.

Lehrach, H. D., Diamond, J. M., Wozney, J. M., & Boedtker, H. (1977) *Biochemistry 16*, 4743–4751.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D. K., & Efstratiadis, A. (1978) *Cell* (*Cambridge, Mass.*) *15*, 687–701.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Maniatis, T., Fritsch, E. F., & Sambrook, J., Eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maxam, A. M., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560–564.

McKnight, S. L., Gavis, E. R., Kingsbury, R., & Axel, R. (1981) *Cell* (*Cambridge, Mass.*) *25*, 385–398.

Messing, J. (1983) *Methods Enzymol. 101*, 20–78.

Neidhardt, F. C., VanBogelen, R. A., & Vaugn, V. (1984) *Annu. Rev. Genet. 18*, 295–329.

Nover, L., Ed. (1984) *Heat Shock Response of Eukaryotic Cells*, Springer-Verlag, Berlin.

Peleg, S., Schrader, W. T., Edwards, D. P., McGuire, W. L., & O'Malley, B. W. (1985) *J. Biol. Chem. 260*, 8492–8501.

Porzio, M. A., & Pearson, A. M. (1974) *Biochim. Biophys. Acta 490*, 2734.

Rosen, J. H., Woo, S. L. C., Holder, J. W., Means, A. R., & O'Malley, B. W. (1975) *Biochemistry 14*, 69–78.

Sanger, F., & Coulson, A. R. (1978) *FEBS Lett. 87*, 107–110.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Sargan, D. R., Tsai, M.-J., & O'Malley, B. W. (1986) *Biochemistry* (following paper in this issue).

Schlesinger, M. J., Ashburner, M., & Tissieres, A., Eds. (1982) *Heat Shock: From Bacteria to Man*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Southgate, R., Ayme, A., & Voellmy, R. (1983) *J. Mol. Biol. 165*, 35–57.

Sugino, A., Goodman, A. H., Heyneker, H. L., Shine, J., Boyer, H. W., & Cozzarelli, N. R. (1977) *J. Biol. Chem. 252*, 3987–3994.

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R. E., Tischer, E., Butler, W. J., & Goodman, H. M. (1977) *Science* (*Washington, D.C.*) *196*, 1313–1316.

Vitec, M. P., & Berger, E. M. (1984) *J. Mol. Biol. 178*, 173–189.

Weigel, N. L., Tash, J. S., Means, A. R., Schrader, W. T., & O'Malley, B. W. (1981) *Biochem. Biophys. Res. Commun. 102*, 513–519.

Weigel et al. (1986) *J. Biol. Chem.* (submitted for publication).

Young, R. A., & Davis, R. W. (1983a) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1194–1198.

Young, R. A., & Davis, R. W. (1983b) *Science* (*Washington, D.C.*) *222*, 778–782.

Zarucki-Schulz, T., Kulomaa, M. S., Headon, D. R., Weigel, N. L., Baez, M., Edwards, D. P., McGuire, W. L., Schrader, W. T., & O'Malley, B. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6358–6362.