MOLECULAR AND CELLULAR BIOLOGY, Aug. 1991, p. 4036–4044
0270-7306/91/084036-09$02.00/0

Vol. 11, No. 8

HSP47: a Tissue-Specific, Transformation-Sensitive, Collagen-Binding Heat Shock Protein of Chicken Embryo Fibroblasts

KAZUNORI HIRAYOSHI,[1] HITOSHI KUDO,[1] HAJIME TAKECHI,[1] AKIRA NAKAI,[1] AKIHIRO IWAMATSU,[2] KENNETH M. YAMADA,[3] AND KAZUHIRO NAGATA[1]*

Department of Cell Biology, Chest Disease Research Institute, Kyoto University, Kyoto 606-01,[1] and Central Laboratories for Key Technology, Kirin Brewery Co., Ltd., Maebashi 371,[2] Japan, and Laboratory of Developmental Biology, National Institute of Dental Research, Bethesda, Maryland 20892[3]

Received 11 February 1991/Accepted 14 May 1991

We report the isolation and characterization of a cDNA clone encoding HSP47, a transformation-sensitive heat shock protein that binds to collagen. A cDNA library was prepared from total RNA isolated from heat-shocked chicken embryo fibroblasts and screened by using oligonucleotide mixtures prepared on the basis of the N-terminal amino acid sequence of biochemically purified HSP47. The cDNA insert contained 3,278 bp, which encoded a 15-amino-acid signal peptide and a mature protein coding region consisting of 390 amino acid residues; it also included part of the 5′ noncoding region and a long 3′ noncoding region. The deduced amino acid sequence revealed an RDEL sequence at the C terminus, which is a variant of the KDEL retention signal for retention of proteins in the endoplasmic reticulum. Northern (RNA) blot analyses and nuclear run-on assays established that the induction of HSP47 by heat shock and its suppression after transformation of chicken embryo fibroblasts by Rous sarcoma virus are regulated at the transcriptional level. A homology search revealed that this protein belongs to the serpin family, the superfamily of plasma serine protease inhibitors. Although structurally homologous to the serpins, HSP47 lacks the active site thought to be essential for the inhibition of proteases and does not appear to bind to intracellular proteases. HSP47 is the first heat shock protein found to be a member of the serpin superfamily. Conversely, it is the first serpin family member that is not secreted from cells, which could be explained by acquisition of the RDEL retention signal during evolution.

When living cells of organisms ranging from bacteria to humans are exposed to temperatures 5 to 10°C above the optimum for growth, they respond by synthesizing a group of proteins called heat shock proteins (HSPs). Because these proteins are induced in response to other stresses, such as heavy metals, amino acid analogs, glucose starvation, ethanol, and hypoxia, they are more generally termed stress proteins. In mammalian cells, three major groups of HSPs have been well characterized, the HSP90 and HSP70 families, and a small HSP, HSP28 (22, 43). In addition to responding to stresses, some of these proteins are synthesized constitutively, and their roles under normal conditions have been shown to be essential for physiological functions, including membrane transport of proteins, as well as folding and unfolding of proteins (9). These functions of stress proteins have led to their designation as molecular chaperones.

A novel heat shock glycoprotein of M_r = 47,000 has recently been identified in chicken embryo fibroblasts (CEF), in addition to the well-known HSPs described above (22, 43). This heat shock protein (HSP47) is characteristic in its capacity to bind to collagen (28) and its sensitivity to malignant transformation (30, 31). The synthesis of HSP47 is decreased in CEF transformed by Rous sarcoma virus (RSV) and in BALB/c 3T3 cells transformed by simian virus 40 (30, 31). The degree of phosphorylation of HSP47 increases by a factor of 7 after transformation of CEF by RSV, although the HSP47 of BALB/c 3T3 cells is not phosphorylated (30, 31). Quantities of HSP47 increase markedly during differentiation of F9 mouse teratocarcinoma cells induced by retinoic acid or retinoic acid plus dibutyryl cyclic AMP (cAMP) (38a). Interestingly, the expression of HSP47 is always closely related to that of collagen in various cell lines (35a), and tissue specificities for expression were reported previously (33). Thus, HSP47 has tissue specificity in expression and substrate specificity in binding. In vitro translation experiments using mRNAs purified from CEF reveal that the induction of HSP47 by heat shock and the decrease of HSP47 synthesis in RSV-transformed CEF can be accounted for by the level of total HSP47 mRNA (27).

The cellular localization of HSP47 is restricted to the endoplasmic reticulum (ER), as determined from immunoelectron microscopy and immunofluorescence studies (35). The intracellular localization of HSP47 coincides well with that of procollagen in fibroblastic cells. Immunoprecipitation and immunofluorescence studies using BALB/c 3T3 cells demonstrate the direct intracellular association of HSP47 with procollagen (31, 32a). Binding of HSP47 to collagen at neutral pH is not disrupted by even 2 M NaCl. However, HSP47 is readily eluted from gelatin- or native collagen-conjugated affinity columns when the pH is decreased to pH 6.3 (35), suggesting that the binding between these two proteins is regulated by physiological changes in pH of intracellular compartments.

HSP47 may function as a molecular chaperone for collagen, but the mechanism of its interaction with collagen is not yet clear. To clarify the molecular structure and regulation of HSP47, we attempted to clone HSP47 cDNA. We report the cloning and sequence of cDNA encoding the entire HSP47 molecule. The induction of HSP47 after heat shock and the repression after transformation were examined by Northern (RNA) blot and nuclear run-on analyses. Unexpectedly,

* Corresponding author.

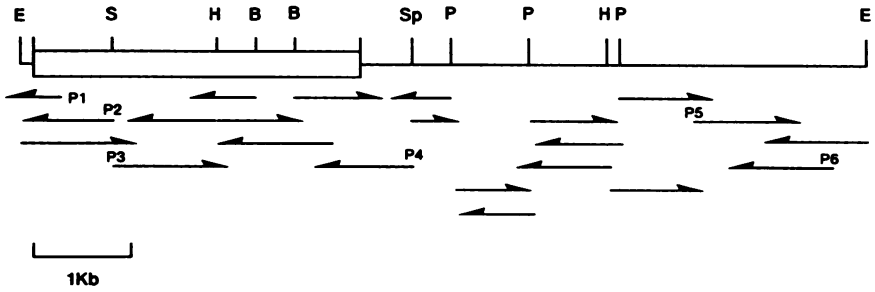

FIG. 1. Physical map of pCH471 and nucleotide sequencing strategy. The diagram indicates the physical map of the pCH471 cDNA clone for chicken HSP47. The open box shows the open reading frame of pCH471. Capital letters show cutting sites of restriction endonucleases (B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sac*I; Sp, *Sph*I). The direction and length of sequence determinations are indicated by horizontal arrows under the physical map. P1 to P6 show the positions of oligonucleotides used as sequence primers.

HSP47 was found to belong to the superfamily of plasma serine protease inhibitors.

MATERIALS AND METHODS

Cell culture. CEF were maintained in Vogt's GM medium (40) and passaged by using 0.25% trypsin and 1 mM EDTA. Transformed cells were established by infecting primary cultures of CEF with the Schmidt-Ruppin strain of RSV.

Oligonucleotide synthesis. Oligonucleotide probes for HSP47 were synthesized by a DNA synthesizer (model 370A; Applied Biosystems, Foster City, Calif.) on the basis of amino acid sequences of the amino terminus and of lysyl endopeptidase-generated fragments determined by a model 470A gas-phase protein sequencer (Applied Biosystems). The amino acid sequence used for the initial probe was Met-Ala-Lys-Asp-Lys-Asn-Met (KAZ0; 21-mer 64-oligonucleotide mixture), and the probes for further screening were Asn-Glu-Trp-Ala-Ala-Glu-Thr (KAZ1; 20-mer 8-oligonucleotide mixture) and Asn-Phe-Ala-Asp-Asp-Phe-Val (KAZ2; 20-mer 32-oligonucleotide mixture). In KAZ1 and KAZ2, inosine was used as the third base in the codon for Ala to decrease the redundancy of the mixture.

Preparation of mRNA. Total RNA was isolated from CEF, heat-shocked CEF, and CEF transformed with RSV by the protocol of Adams et al. (1). In brief, after being washed once with cold phosphate-buffered saline (PBS), cells were harvested by trypsinization, washed twice with cold PBS, and homogenized with a Polytron in a solution of 8 M guanidine thiocyanate according to the method of Chirgwin et al. (6). After ultracentrifugation at 36,000 rpm (100,000 × *g*) for 12 h in a RPS55T-2 rotor (Hitachi, Tokyo, Japan), total RNA was recovered, and poly(A)$^+$ RNA was isolated by using oligotex dT30 and the protocol of Daiichi Pure Chemicals Co. Ltd. (Tokyo, Japan).

Preparation of the cDNA library. The first strand of cDNA was synthesized from poly(A)$^+$ RNA prepared from CEF after heat shock treatment at 45°C for 4 h, using the cDNA cloning system plus from Amersham International, Inc. (Buckinghamshire, England), which is based on the method of Gubler and Hoffman (14). Double-stranded DNA was size separated by Sepharose 4B (Pharmacia, Uppsala, Sweden) column chromatography. Fragments estimated to be ≥2 kb in length were pooled and ligated to phage λgt11 (45) arms (Amersham). The efficiency of this library was 10^7 PFU/μg of cDNA, and the average size of the cDNA in this library was approximately 2.5 kb.

Screening of the cDNA library. A total of 5×10^4 independent plaques were screened by using the KAZ0 oligonucleotide probe derived from the N-terminal amino acid sequence. This probe was labeled with T4 polynucleotide kinase (New England BioLabs, Beverly, Mass.) and [γ-^{32}P]ATP (specific activity, 3,000 Ci/mol), resulting in a labeled probe with a specific activity of 8×10^6 cpm/pmol, and it was used to screen the cDNA library prepared from heat-shocked CEF. Then 5×10^4 recombinants were screened at 5×10^3 plaques per 150-mm plate by hybridizing duplicate nitrocellulose filters with the KAZ0 probe (2×10^6 cpm/ml) at 37°C for 16 h in 6× SSC (1× SSC is 150 mM sodium chloride plus 15 mM sodium citrate, pH 7.0)–1× Denhardt's solution (0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin, 0.02% Ficoll)–100 μg of sonicated salmon sperm DNA per ml. Filters were washed with 6× SSC–0.1% sodium dodecyl sulfate (SDS) at room temperature and then at 37°C. Plaques positive on both filters were isolated for secondary screening following amplification. Further screening was performed with the KAZ1 and KAZ2 probes under the same conditions as the first screening.

DNA sequencing. A 3.3-kb cDNA clone (pCH471) that hybridized with all three probes was digested with appropriate restriction enzymes. The digested fragments were subcloned into pUC19 for DNA sequencing. Double-stranded plasmid DNA was prepared according to Maniatis et al. (24), and plasmid DNA was further purified by polyethylene glycol precipitation or $CsCl_2$ gradient centrifugation. Double-stranded plasmid DNA was denatured and used for sequencing (16). Dideoxynucleotide sequencing was performed by using the Sequenase kit version 2.0 (U.S. Biochemical Corp., Cleveland, Ohio). To sequence each cDNA subclone, an M13 universal primer and a reverse sequencing primer were used. Oligomers KAZ0 (P1) and KAZ2 (P2), oligomers complementary to positions 1202 to 1218 (P4) and 3016 to 3035 (P6), and oligomers corresponding to positions 384 to 401 (P3) and 2584 to 2603 (P5) were used as sequencing primers (Fig. 1). Sequences were determined by complete sequencing of both cDNA strands.

Sequence analysis. The deduced amino acid sequence was analyzed by using the PC/GENE sequence analysis system (Amos Bairoch, University of Geneva, Geneva, Switzerland/IntelliGenetics, Inc., Mountain View, Calif.). Amino acid homology searches were performed with the Kanehisa algorithm (12) as well as with the FASTP algorithm of Lipman and Pearson (23), using EMBL/GenBank protein data bank. Searches for potential phosphorylation sites and other sites of potential posttranslational modification were conducted by using the PROSITE program of PC/GENE (3). Transmembrane prediction programs included those of

Klein et al. (18), Rao and Argos (34), and Eisenberg et al. (8). Hydropathy (hydrophobicity) profiles were determined by the method of Kyte and Doolittle (19).

Northern blot analysis. Total RNAs (10 μg) were separated on 1% agarose gels containing formaldehyde (20) and transferred to nylon membrane filters (GeneScreen Plus; Du Pont-New England Nuclear, Boston, Mass.) as recommended by the manufacturer. Blotted filters were prehybridized overnight at 42°C in 5× SSC containing 50% formamide, 1× Denhardt's solution, 1% SDS, and 100 μg of denatured salmon sperm DNA per ml. A cDNA probe for HSP47, the *Eco*RI-*Eco*RI fragment of pCH471, and a cDNA probe for β-actin, the *Bam*HI-*Bam*HI fragment of pHfβA-1 (kindly provided by R. Morimoto, Northwestern University), were labeled with [α-^{32}P]dCTP by the multiprime labeling method of Feinberg and Vogelstein (10). The oligonucleotide probe for the N-terminal amino acid sequence (KAZ0) was labeled by the method of Wallace et al. (42). Labeled oligonucleotides were separated from unreacted [γ-^{32}P]ATP by chromatography on a Whatman DE52 cellulose column with an elution buffer containing 2 M NaCl and were used directly for hybridization in a solution of 3× NET (1× NET is 150 mM NaCl, 1 mM EDTA, and 0.1 M Tris-HCl, pH 8.0) and 1× Denhardt's solution. Hybridized membranes were washed with 6× SSC–0.1% SDS at room temperature and then at 65°C with the same solution.

Nuclear run-on assay. CEF and RSV-transformed CEF with or without prior heat shock treatment were washed three times with ice-cold 1× SSC and incubated in Nonidet P-40 lysis buffer (10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM $MgCl_2$, 0.5% [vol/vol] Nonidet P-40) at 4°C for 10 min. After centrifugation at 500 × *g* for 5 min, the nuclear pellets were washed twice with the same lysis buffer and then frozen in 50 mM Tris-HCl (pH 8.3)–40% (vol/vol) glycerol–5 mM $MgCl_2$–0.1 mM EDTA. After thawing, nuclei were incubated at 30°C for 20 min in a reaction mixture containing 0.1 mCi of [α-^{32}P]UTP (specific activity, 3,000 Ci/mol) as described by Greenberg and Ziff (13). Equal amounts of radioactivity from each sample were hybridized with cDNAs or mock plasmid DNA immobilized on nitrocellulose membranes. For probe DNAs, plasmids linearized by restriction enzyme digestion were treated with NaOH and spotted onto nitrocellulose filters, using a vacuum blot apparatus; 40 μg of DNA was blotted per slot. cDNAs used as probes were whole pCH471 and human β-actin.

Nucleotide sequence accession number. The sequence reported here has been deposited in the EMBL data bank under accession number X57157.

RESULTS

Cloning and analysis of cDNA encoding HSP47. A cDNA encoding HSP47 was isolated by screening a cDNA library prepared from CEF treated at 42°C for 4 h, using an oligonucleotide probe (KAZ0; 21-mer 64-oligonucleotide mixture) that was synthesized on the basis of N-terminal amino acid sequence of purified HSP47. When KAZ0 was used as a probe for Northern blot analysis of CEF, a single 4.5-kb band was detected under stringent conditions using 6× SSC–0.1% SDS at 65°C (see Fig. 4). Since other oligonucleotides such as KAZ1 and KAZ2 did not detect a single band by Northern blot analysis under these high-stringency conditions, we used KAZ0 as a probe for the first screening of the cDNA library. A total of 5 × 10^4 independent plaques were screened as described in Materials and Methods, resulting in the isolation of three positive clones. For the second screening, KAZ1 and KAZ2 were used as probes. After three cycles of screening, one positive clone was isolated.

This clone (pCH471) was sequenced by the strategy shown in Fig. 1. All fragments were sequenced twice in both directions. This clone contained the correct coding sequence, since an N-terminal amino acid sequence deduced from the base sequence coincided exactly at each of 36 residues with the results of direct amino acid sequencing (Val-Pro-Ser-Glu-Asp-Arg-Lys-Leu-Ser-Asp-Lys-Ala-Thr-Thr-Leu-Ala-Asp-Arg-Ser-Thr-Thr-Leu-Ala-Phe-Asn-Leu-Tyr-His-Ala-Met-Ala-Lys-Asp-Lys-Asn-Met; underlined in Fig. 2). Exact matches with amino acid sequences determined from fragments generated with lysyl endopeptidase were also detected in the sequence (Ser-Ile-Asn-Glu-Trp-Ala-Ala-Gly-Thr-Thr-Asp-Gly-Lys-Leu-Pro-Glu-Val-Thr-Lys and Ile-Gly-Asn-Arg-Leu-Tyr-Gly-Pro-Ala-Ser-Ile-Asn-Phe-Ala-Asp-Asp-Phe-Val; underlined and marked by KAZ1 and KAZ2 in Fig. 2). The molecular weight of this protein (45,681) predicted from the deduced 405 amino acids containing a signal sequence (see below) was very close to results obtained by SDS-polyacrylamide gel electrophoresis after gelatin purification of the in vitro-translated product of mRNA from CEF (27).

The 3.3-kb sequence contained a single open reading frame which extended from the 5′ terminus at position 60 to nucleotide 1275 (Fig. 2). The results of primer extension using KAZ0 as the primer revealed that the 5′ noncoding region was about 800 bp in length (data not shown). These results indicated that this clone lacked about 700 bp of 5′ noncoding region. At the 5′ end of the coding region, this clone displayed 15 amino acids comprising a typical hydrophobic signal sequence (Fig. 2 and 3). According to the eukaryotic signal sequence criteria of von Heijne (41), this sole candidate signal sequence had a favored cleavage site located between the Ala at residue 15 and the following Val. According to amino acid sequencing, this Val residue was in fact the first amino acid of mature HSP47. There was no unequivocal evidence for a transmembrane domain. Hydrophobicity (hydropathy) analyses gave equivocal results with a possible region of increased hydropathy (Fig. 3) according to the method of Kyte and Doolittle (19). Programs for predicting transmembrane domains were also not in entire agreement: the method of Rao and Argos (34) predicted that the sequence Ile-Leu-Leu-Ser-Pro-Val-Val-Val-Ala-Ser-Ser-Leu-Gly-Leu-Val-Ser-Leu-Gly-Gly at positions 54 to 72 would form a transmembrane helix, and the method of Klein et al. (18) similarly identified this region as a predicted transmembrane segment; in contrast, the method of Eisenberg et al. (8) predicted no membrane-associated helices in HSP47. At the C terminus of the polypeptide, an RDEL (Arg-Asp-Glu-Leu) sequence was detected. This sequence can theoretically function as a retention signal for retaining proteins in the lumen of the ER (2).

Two carbohydrate chains were previously reported to be present in mature HSP47 (17). In our deduced sequence, two Asn-X-Thr sequences were found at residues 107 and 112. Because HSP47 phosphorylation is regulated by malignant transformation, potential phosphorylation sites were mapped. Possible cAMP/cGMP-dependent protein kinase phosphorylation sites were found at serine residues 24 and 288. Possible protein kinase C phosphorylation sites were found at serines 24, 137, and 328 and at threonines 109 and 114. Possible casein kinase II phosphorylation sites were found at threonines 29 and 320 and at serine 159. Finally, possible tyrosine sulfation sites were found at residues 232, 233, 353,

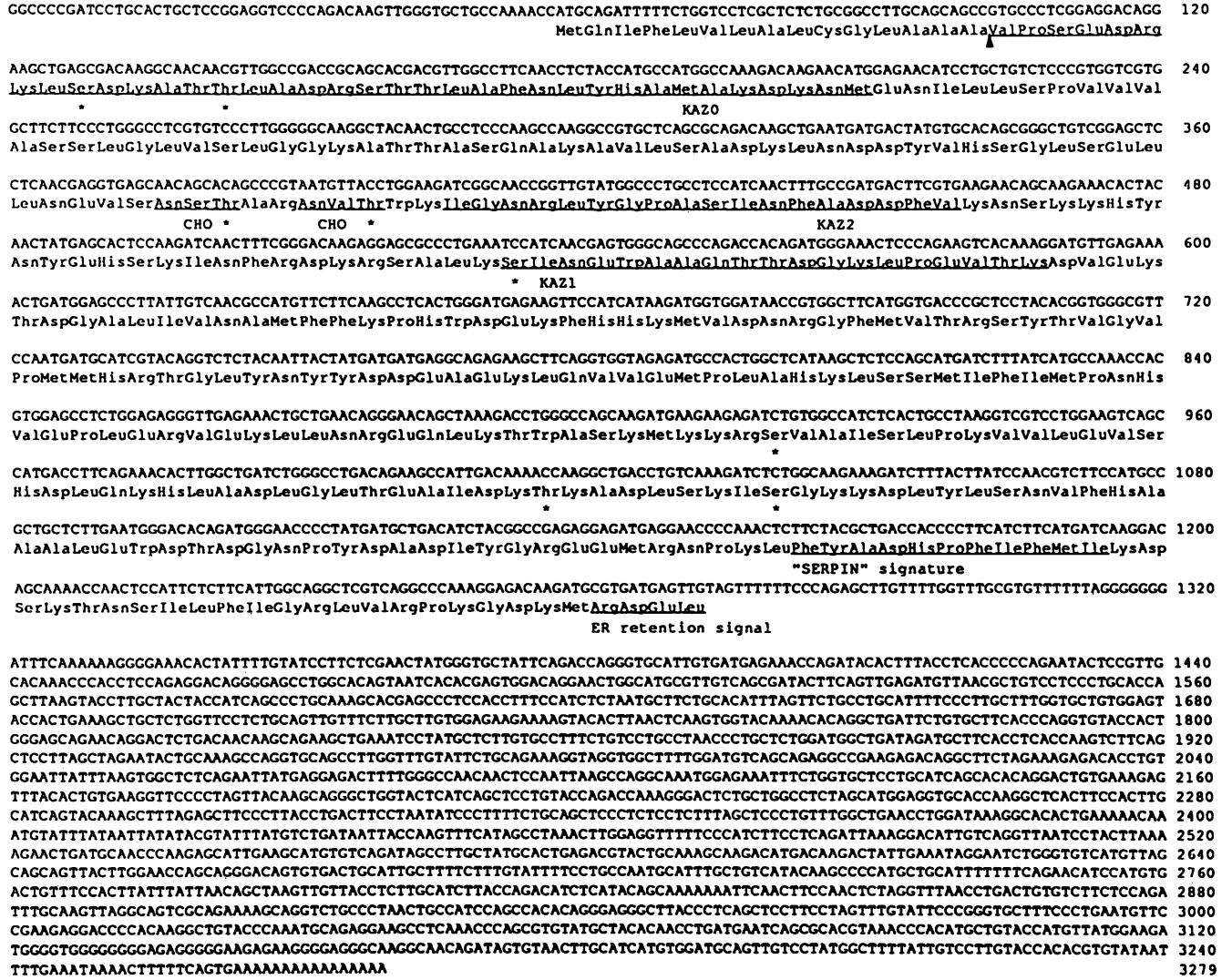

FIG. 2. Nucleotide sequence and encoded amino acid sequence of chicken HSP47. First line, nucleotide sequence of cDNA for chicken HSP47; second line, amino acid sequence encoded by this cDNA. The underlined sequence labeled KAZ0 indicates the deduced amino acid sequence coincident with the chemically determined amino acid sequence (corresponding to the oligonucleotide KAZ0). The underlined sequences KAZ1 and KAZ2 indicate amino acid sequences coinciding with amino acid sequences from lysyl endopeptidase-digested fragments (corresponding to oligonucleotides KAZ1 and KAZ2). Predicted N-linked oligosaccharide chain binding sites are indicated by CHO-labeled underlines. The arrow indicates the predicted cleavage site of signal endopeptidase. Asterisks indicate the putative phosphorylation sites. The underlined sequence labeled "SERPIN" signature near the C terminus of the amino acid sequence indicates residues that match the consensus sequence for the serpin signature of Bairoch (3). The C-terminal RDEL sequence, a theoretical retention signal for retaining proteins in the lumen of the ER, is underlined and labeled ER retention signal.

and 358, and a possible amidation site was found at serine 328.

The 3′ noncoding region was unusually long, and one typical poly(A) addition signal, AATAAA, was located at position 3257, which was 15 bases upstream from the poly(A) tail.

Although HSP47 binds to collagen (28), no homology was observed with well-known collagen-binding proteins such as fibronectin, laminin, vitronectin, and von Willebrand factor.

Regulation of the expression of HSP47. The induction of HSP47 after heat shock and the decrease in HSP47 synthesis after transformation with RSV, which we previously established at the protein level (30), were confirmed at the mRNA level by performing Northern blot analysis of RNAs prepared from CEF (Fig. 4). Both the oligonucleotide mixture probe (KAZ0) and clone pCH471 recognized the same-sized band, and the extent of induction of HSP47 mRNA after heat shock was similar, regardless of whether KAZ0 or pCH471 was used as the probe.

Nuclear run-on assays (Fig. 5) demonstrated that the expression of HSP47 was regulated at the transcriptional level after heat shock as well as after transformation. Both the extent of induction after heat shock and the amount of repression after transformation observed in HSP47 transcripts in the nuclear run-on assay were similar to those observed in Northern assays or in HSP47 protein synthesis. To examine the kinetics of expression of HSP47, we performed nuclear run-on assays and Northern blot analyses in

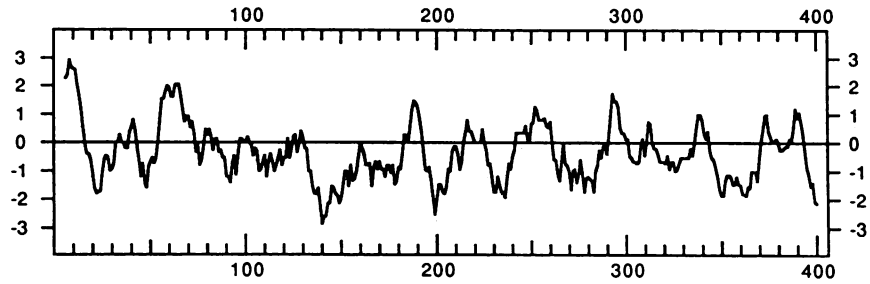

FIG. 3. Hydrophobicity pattern of chicken HSP47. Averages of the hydrophobicity of 10 amino acids are calculated by the method of Kyte and Doolittle (19). Hydrophobic amino acids are plotted as positive; hydrophilic amino acids are plotted as negative.

parallel during heat treatment (45°C) and a recovery period at 37°C after the heat treatment. Transcripts of the actin gene decreased gradually after the heat treatment and then returned to normal levels during the recovery period (Fig. 6A), as is often observed under severe heat shock conditions. The transcripts of HSP47 increased rapidly after the initiation of heat treatment at 45°C and rapidly returned to basal levels during the recovery period (Fig. 6A). The relative increase of the ratios of transcripts of HSP47 to those of actin was over 10-fold at 2 h after the initiation of heat treatment. Quantities of total HSP47 mRNA gradually increased after the temperature shift to 45°C and reached a plateau at 6 to 8 h (Fig. 6B). The results shown in Fig. 6 suggest that the half-life of HSP47 mRNA is long.

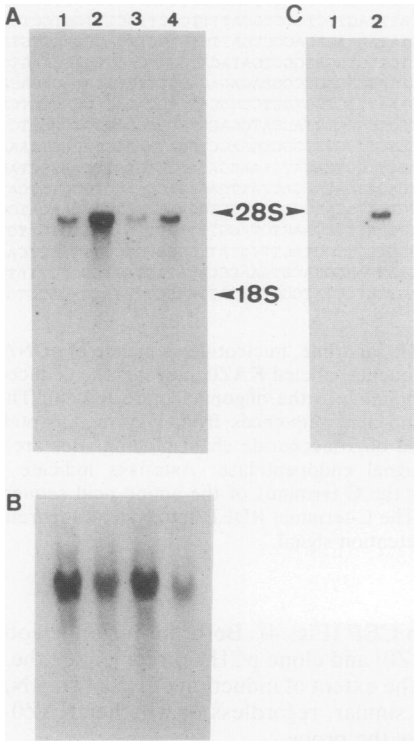

FIG. 4. Northern blot hybridization analysis of HSP47 mRNA. Northern blot analysis of RNA was performed with an HSP47 cDNA probe compared with an oligonucleotide probe (KAZ0) as described in Materials and Methods. Arrows indicate the migration positions of 18S and 28S rRNAs. (A) Hybridization of pCH471 with total RNAs from CEF (lane 1), heat-treated CEF (lane 2), RSV-transformed CEF (lane 3), and heat-treated, RSV-transformed CEF (lane 4). (B) Hybridization of β-actin with the same RNAs used for panel A. (C) Hybridization of oligonucleotide KAZ0 with poly(A)$^+$ RNAs from CEF (lane 1) and heat-treated CEF (lane 2).

HSP47 is homologous to the serpin family. A comparison of the deduced HSP47 amino acid sequence with those in the EMBL/GenBank protein data bank revealed a homology with members of the serpin protein family. The serpin family includes a group of serine protease inhibitors with a common tertiary structure and functional domains often containing a reactive center (5). A peptide sequence that corresponded to a characteristic, conserved site comprising part of the reactive center of the serpin family was detected at amino acid residues 369 to 379 in HSP47 (Phe-Tyr-Ala-Asp-His-Pro-Phe-Ile-Phe-Met-Ile). Moreover, conserved secondary structures in the serpin family were also observed in HSP47. Figure 7 shows the alignments of the sequences of HSP47 and predicted representative members of the serpin family. Squares show secondary structure (predicted α helix and β sheets) that were conserved in the serpin family and in HSP47. The results shown in Fig. 6 also suggest that the tertiary structure of HSP47 may be similar to that of the serpin family. HSP47 had the highest homology to a human plasma serine protease (protein C) inhibitor (38). The amino acid sequence identity between HSP47 and protein C inhibitor was 31%, and that between HSP47 and other members in serpin family ranged from 10 to 30%.

DISCUSSION

A major step toward an understanding of the function of the HSP47 collagen-binding HSP is to identify the cDNA encoding HSP47 and to characterize the primary structure of

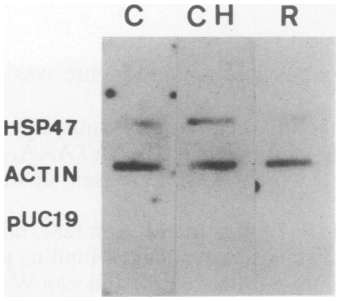

FIG. 5. Nuclear run-on assay for HSP47 gene transcription. HSP47 cDNA, actin cDNA, and vector (pUC19) bound to nitrocellulose membrane filters were hybridized with ^{32}P-labeled run-on transcripts in nuclei isolated from control CEF (C), heat-treated CEF (CH), and RSV-transformed CEF (R).

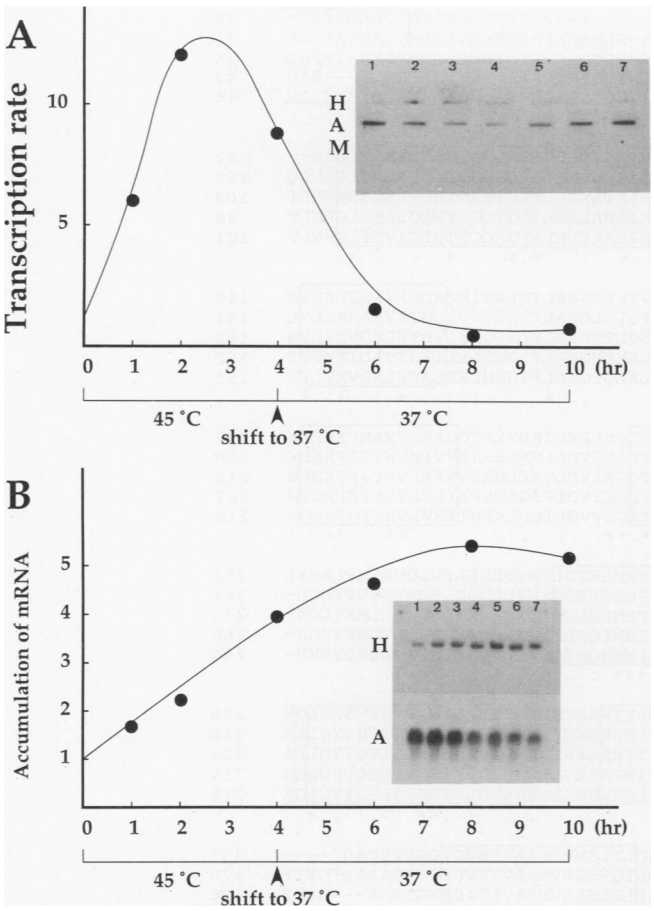

FIG. 6. Time course of the expression of HSP47 during heat treatment and recovery periods at 37°C. (A) Nuclear run-on assays for HSP47 transcripts. Nuclei were prepared from CEF that were heat treated at 45°C for 1 (lane 2), 2 (lane 3), and 4 (lane 4) h. Nuclei were also prepared from CEF that were treated at 45°C for 4 h and then incubated at 37°C for 2 (lane 5), 4 (lane 6), and 6 (lane 7) h. Nuclei from CEF without heat treatment were prepared to serve as the control (lane 1). Run-on transcripts were hybridized with HSP47 cDNA (H), actin cDNA (A), and the pUC19 vector (M) bound to nitrocellulose membrane filters. The relative density of each band was determined by using a densitometer (CS9000; Shimadzu, Kyoto, Japan), and the HSP47/β-actin ratio at each time point was plotted by normalizing the value of lane 1 to 1 U. (B) Northern blot hybridization analysis of HSP47 mRNA. RNAs were extracted from CEF treated as for panel A. Hybridizations were performed with HSP47 cDNA (H) and β-actin cDNA (A) probes. Lanes are the same as in panel A. The HSP47/β-actin ratios were determined and plotted as described for panel A.

this protein. In this study, we have cloned and analyzed the cDNA encoding chicken HSP47. Several results, including complete matches with three sets of amino acid sequences, indicate that the entire coding region of HSP47 mRNA is included in the pCH471 open reading frame. Northern blot analysis and nuclear run-on assays demonstrated that HSP47 mRNA is induced by heat shock and repressed after transformation of CEF. Enhancement of HSP47 protein synthesis by heat shock can therefore be ascribed to transcriptional control.

Sequencing data revealed that HSP47 has a signal peptide at its N terminus. It also has a potential transmembrane domain. HSP47 was previously localized to the ER by immunofluorescence and immunoelectron microscopic studies. It is not clear, however, whether HSP47 is a luminal or transmembrane protein of the ER. This report demonstrates the presence of an RDEL sequence at the C terminus of HSP47. All of the ER luminal proteins described to date, such as GRP78, GRP94, and protein disulfide isomerase, have a KDEL sequence at their C terminus (26, 33). This sequence reportedly functions as a retention signal determining that a protein is to reside in the ER. Recently, Andres et al. showed that for neuropeptide Y, the sequence RDEL can be substituted experimentally for KDEL as its ER retention signal (2). Our report provides the first example of use of the RDEL sequence in a naturally occurring protein. The existence of a potential transmembrane sequence in HSP47 will require further evaluation, since even though two prediction programs suggested such a region, a third did not confirm the existence of a predicted intramembrane α helix. Such a highly hydrophobic region could conceivably also be buried in the center of a globular nonmembrane protein. It is not yet clear whether the RDEL sequence could also function to retain integral membrane proteins in the ER, since C-terminal KDEL sequences have been characterized to date only on intraluminal ER proteins. Recently, receptors for KDEL and HDEL were identified in mammalian cells and yeast cells, respectively (21, 25, 39). It will thus be interesting to determine whether the RDEL sequence is recognized by a KDEL receptor or by an independent receptor specific for the RDEL sequence.

The biological significance of HSP47 membership in the serpin superfamily is not yet clear. The evolutionary relationship of members of the serpin superfamily of plasma serine protease inhibitors was first noted when the structure of antithrombin was compared with that of α1-antitrypsin (4). Subsequently, diverse groups of homologous proteins have been reported to belong to this superfamily. Serpins also include homologous proteins without inhibitory activity for serine proteases such as ovalbumin and angiotensinogen. General features of this superfamily are (i) small glycoproteins (M_r 40,000 to 60,000) consisting of a single polypeptide with a variable number of oligosaccharide side chains, (ii) homologous amino acid sequences, (iii) common secondary and probable common tertiary structures, (iv) a conserved reactive center, (v) conserved flanking sequences around the reactive center, and (vi) often (but not always) the ability to form a tight complex with a protease involving proteolytic cleavage of the serpin P1-P1′ peptide bond in the reaction center. This P1 site acts as a pseudosubstrate for the target protease. HSP47 shares with the serpins at least features i, ii, and, to some extent v. Besides an overall amino acid sequence homology with serpins, HSP47 displays the distinctive serpin signature described by Bairoch (3). This diagnostic pattern consists of a characteristic sequence centered about a well-conserved Pro-Phe sequence located approximately 10 to 15 residues toward the C terminus from the reactive bond of the serpins. Moreover, a 12-amino-acid stretch in this region shows 75% identity with the homologous site in ovalbumin. Interestingly, however, HSP47 appears to lack the reactive P1 bond itself: although an Arg residue that is frequently found in serpins is located at the putative P1 site at residue 364, the Ser residue found in almost all serpins is absent from the P1′ site and is replaced by an Asn. Moreover, the flanking sequence toward the N terminus shows no homology with this region of ovalbumin or other serpins; this region may serve an important function as an α-helical stressed loop in the classical serpins (5, 37).

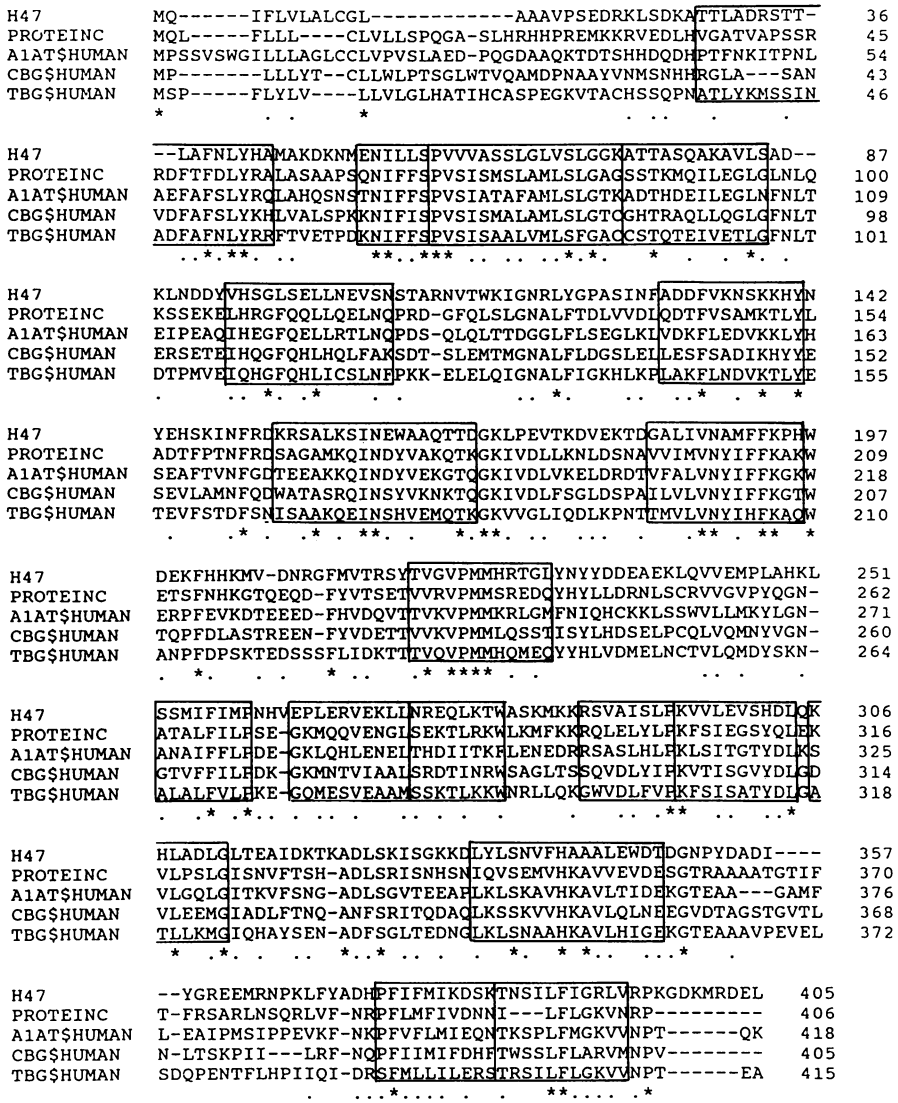

FIG. 7. Alignment of shared amino acid sequences between chicken HSP47 and members of the serpin family. The standard one-letter amino acid code is used. Symbols indicate that a position in the alignment is perfectly conserved (*) or well conserved (.). Boxes show conserved regions of predicted secondary structure. Human plasma serine protease (protein c) inhibitor (PROTEINC) (38), α1-antitrypsin precursor (α1 protease inhibitor) (A1AT$HUMAN) (7), corticosteroid-binding globulin (CBC$HUMAN) (15), and thyroxine-binding globulin (TBG$HUMAN) (11) were aligned with HSP47 (H47).

A recent in vivo cross-linking study using a membrane-permeable and thiol-cleavable cross-linker and immunoprecipitation analysis demonstrated that HSP47 binds to procollagen within cells (32a). In these experiments, anti-HSP47 antibody coprecipitated three major bands corresponding to the α_1(I) band of type I procollagen, GRP78, and protein disulfide isomerase, plus several faint bands. The mechanism of HSP47 binding to procollagen is different, however, from that of the serpins to serine proteases: serpins are cleaved at the reactive center by the target protease, and then they bind to the protease through a covalent bond; in contrast, HSP47 can be dissociated readily from procollagen by changing the pH to 6.3 (35).

Although HSP47 therefore lacks such a serpin reactive center, its retention of the serpin signature and its overall homology to serpins are intriguing. It remains possible that the binding function of serpins to serine proteases has been transformed in some fashion into the capacity of HSP47 to bind to collagen and fetuin (32). This possibility could be tested by site-directed mutagenesis of the serpinlike sequences in HSP47.

During the preparation of this report, Wang and Gudas (44) published a report describing a protein belonging to the serpin family that was induced after mouse F9 teratocarcinoma cells were treated with retinoic acid. By comparing amino acid sequences, we found a strong homology (80% sequence identity) between chicken HSP47 and this F9 cell protease inhibitor homolog. Moreover, we have observed that HSP47 is also markedly induced by treating F9 cells with retinoic acid alone or in combination with dibutyryl cAMP (38a). Although their protease inhibitor from F9 cells is similar to our HSP47, the N-terminal region in their protein was completely different from that of HSP47 over a span of 41 amino acids, and HSP47 lacks its serpin reactive

center. Cloning of the mouse homologs of HSP47 should elucidate these differences.

ACKNOWLEDGMENTS

We are grateful to H. Mori (Institute for Virus Research, Kyoto University) for technical advice about computer sequence analyses and to R. Morimoto (Northwestern University) for providing actin cDNA.

This work was supported in a part by a Grant-in-Aid for Cancer Research and for Scientific Research from the Ministry of Education, Science, and Culture of Japan and by grants from the Naito Foundation and from the Toray Science Foundation.

REFERENCES

1. **Adams, S. L., M. E. Sobel, B. H. Howard, K. Olden, K. M. Yamada, B. de Crombrugghe, and I. Pastan.** 1977. Levels of translatable mRNAs for cell surface protein, collagen precursors, and two membrane proteins are altered in Rous sarcoma virus-transformed chick embryo fibroblasts. Proc. Natl. Acad. Sci. USA **74**:3399–3403.
2. **Andres, D. A., I. M. Dickerson, and J. E. Dixon.** 1990. Variants of the carboxyl-terminal KDEL sequence direct intracellular retention. J. Biol. Chem. **265**:5952–5955.
3. **Bairoch, A.** 1990. PROSITE: a dictionary of protein sites and patterns, fifth release. University of Geneva, Geneva, Switzerland.
4. **Carrell, R., M. Owen, S. Brennan, and L. Vaugham.** 1979. Carboxy terminal fragment of human α1-antitrypsin from hydroxylamine cleavage; homology with antithrombin-III. Biochem. Biophys. Res. Commun. **91**:1032–1037.
5. **Carrell, R. W., and D. R. Boswell.** 1986. Serpins: the superfamily of plasma serine proteinase inhibitors, p. 403–420. *In* A. Barrett and G. Salvesen (ed.), Proteinase inhibitors. Elsevier/North-Holland Biomedical, Amsterdam.
6. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
7. **Ciliberto, G., L. Denta, and R. Cortese.** 1985. Cell-specific expression of a transfected human alpha-1-antitrypsin gene. Cell **41**:531–540.
8. **Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall.** 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179**:125–142.
9. **Ellis, J.** 1987. Proteins as molecular chaperones. Nature (London) **328**:378–379.
10. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
11. **Flink, I. L., T. J. Bayley, T. A. Gustafson, B. E. Markham, and E. Morkin.** 1986. Complete amino acid sequence of human thyroxine-binding globulin deduced from cloned DNA: close homology to the serine antiproteases. Proc. Natl. Acad. Sci. USA **83**:7708–7712.
12. **Goad, B. W., and M. I. Kanehisa.** 1982. Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries. Nucleic Acids Res. **10**:247–263.
13. **Greenberg, M. E., and E. B. Ziff.** 1984. Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. Nature (London) **311**:433–438.
14. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
15. **Hammond, G. L., C. L. Smith, I. S. Goping, D. A. Underhill, M. J. Harley, J. Reventos, N. A. Musto, G. L. Gunsalus, and C. W. Bardin.** 1987. Primary structure of human corticosteroid binding globulin deduced from hepatic and pulmonary cDNAs exhibits homology with serine protease inhibitors. Proc. Natl. Acad. Sci. USA **84**:5153–5157.
16. **Hattori, M., and Y. Sakaki.** 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152**:232–238.
17. **Hughes, R. C., S. Tailor, H. Sage, and B. L. M. Hogan.** 1987. Distinct pattern of glycosylation of colligan, a collagen-binding glycoprotein, and SPARC (osteonectin), a secreted Ca^{2+}-binding glycoprotein. Eur. J. Biochem. **163**:57–65.
18. **Klein, P., M. Kanehisa, and C. Delisi.** 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta **815**:468–476.
19. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
20. **Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16**:4743–4751.
21. **Lewis, M. J., and H. R. B. Pelham.** 1990. A human homologue of the yeast HDEL receptor. Nature (London) **348**:162–163.
22. **Lindquist, S., and E. A. Craig.** 1988. The heat shock proteins. Annu. Rev. Genet. **22**:631–677.
23. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
24. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 368–369 Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Michael, J. A., D. J. Sweet, and H. R. B. Pelham.** 1990. The ERD2 gene determines the specificity of the luminal ER protein retention system. Cell **61**:1359–1363.
26. **Munro, S., and H. R. B. Pelham.** 1987. A C-terminal signal prevents secretion of luminal ER proteins. Cell **48**:899–907.
27. **Nagata, K., K. Hirayoshi, M. Obara, S. Saga, and K. M. Yamada.** 1988. Biosynthesis of a novel transformation-sensitive heat-shock protein that binds to collagen. J. Biol. Chem. **263**:8344–8349.
28. **Nagata, K., S. Saga, and K. M. Yamada.** 1986. A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein. J. Cell Biol. **103**:223–229.
29. **Nagata, K., S. Saga, and K. M. Yamada.** 1988. Characterization of a novel transformation-sensitive heat-shock protein (HSP47) that binds to collagen. Biochem. Biophys. Res. Commun. **153**:428–434.
30. **Nagata, K., and K. M. Yamada.** 1986. Phosphorylation and transformation sensitivity of a major collagen-binding protein of fibroblasts. J. Biol. Chem. **261**:7531–7536.
31. **Nakai, A., K. Hirayoshi, and K. Nagata.** 1990. Transformation of BALB/3T3 cells by simian virus 40 causes a decreased synthesis of a collagen-binding heat-shock protein (HSP47). J. Biol. Chem. **265**:992–999.
32. **Nakai, A., K. Hirayoshi, S. Saga, K. M. Yamada, and K. Nagata.** 1989. The transformation-sensitive heat shock protein (HSP47) binds specifically to fetuin. Biochem. Biophys. Res. Commun. **164**:259–264.

32a. **Nakai, A., et al.** Unpublished data.

33. **Pelham, H. R. B.** 1989. Control of protein exit from the endoplasmic reticulum. Annu. Rev. Cell Biol. **5**:1–23.
34. **Rao, M. J. K., and P. Argos.** 1986. A conformational preference parameter to predict helices in integral membrane proteins. Biochim. Biophys. Acta **869**:197–214.
35. **Saga, S., K. Nagata, W.-T. Chen, and K. M. Yamada.** 1987. pH-dependent function, purification, and intracellular location of a major collagen-binding glycoprotein. J. Cell Biol. **105**:517–527.

35a. **Satoh, M., et al.** Unpublished data.

36. **Semenza, J. C., K. G. Hardwick, D. Neta, and H. R. B. Pelham.** 1990. ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. Cell **61**:1349–1357.
37. **Stein, P. E., A. G. W. Leslie, J. T. Finch, W. G. Turnell, P. J. McLaughlin, and R. W. Carrell.** 1990. Crystal structure of ovalbumin as a model for the reactive center of serpins. Nature (London) **347**:99–102.
38. **Suzuki, K., Y. Deyashiki, J. Nishioka, K. Kurachi, M. Akira, S. Yamamoto, and S. Hashimoto.** 1987. Characterization of a cDNA for human protein C inhibitor. J. Biol. Chem. **262**:611–616.

38a. **Takechi, H., et al.** Unpublished data.

39. **Vaux, D., J. Tooze, and S. Fuller.** 1990. Identification by anti-idiotype antibodies of an intracellular membrane protein that recognize a mammalian endoplasmic reticulum retention signal. Nature (London) **345**:495–502.
40. **Vogt, P. K.** 1969. Focus assay of Rous sarcoma virus, p. 66–71. *In* K. Habel and N. P. Salzman (ed.), Fundamental techniques in virology. Academic Press, New York.
41. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14**:4683–4690.
42. **Wallace, R. B., J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura.** 1979. Hybridization of synthetic oligodeoxyribonucleotides to $\psi\chi$174 DNA: the effect of single base pair mismatch. Nucleic Acids Res. **6**:3543–3557.
43. **Welch, W. J.** 1990. The mammalian stress response: cell physiology and biochemistry of stress proteins, p. 223–278. *In* R. I. Morimoto, A. Tissières, and C. Georgopoulous (ed.), Stress proteins in biology and medicine. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. **Wang, S., and L. J. Gudas.** 1990. A retinoic acid-inducible mRNA from F9 teratocarcinoma cells encodes a novel protease inhibitor homologue. J. Biol. Chem. **265**:15818–15822.
45. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80**:1194–1198.