Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2360–2364, April 1982
Genetics

Four small *Drosophila* heat shock proteins are related to each other and to mammalian α-crystallin

(DNA sequence/sequence comparisons/gene family)

THOMAS D. INGOLIA AND ELIZABETH A. CRAIG

Department of Physiological Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706

Communicated by Daniel E. Koshland, Jr., January 18, 1982

ABSTRACT **The primary base sequence of the protein coding regions of the four small heat shock genes of *Drosophila melanogaster* present at cytological locus 67B has been determined. A single open reading frame large enough to encode a small heat shock protein is found for each gene. The molecular weights of the predicted proteins are in good agreement with experimentally determined values obtained from gel electrophoresis. The predicted amino acid sequences of the four small heat shock genes show striking homologies over ≈50% of their lengths. This region of extensive homology extends from about amino acid 85 to amino acid 195 out of a total of ≈200 amino acids. Comparison of the predicted sequence with the known sequences of other proteins revealed a remarkable similarity between this region of homology and the corresponding region of mammalian α-crystallin. The possible functional significance of this structural similarity is discussed.**

Heat shock and a variety of other stimuli result in a dramatic change in the pattern of gene expression in *Drosophila melanogaster* (for review, see ref. 1). Transcription of most genes is suppressed and expression of a small set of previously inactive genes is enhanced. A similar response is observed in cells of a wide variety of species including yeast (2, 3), *Dictyostelium* (4), soy bean (5), and chicken and man (6). Little is known about the physiological role of the heat shock response or the induced proteins. However, the heat shock response has been shown to provide protection from thermal killing in *Drosophila* (7), yeast (8), and *Dictyostelium* (4). Cytological and cell fractionation studies indicate that many of the *Drosophila* heat shock polypeptides reside in the nucleus (9–11) associated with either the chromosomes (9) or the nucleoskeleton (11). One of the heat shock polypeptides of chicken will associate with the *src* gene product of Rous sarcoma virus (12) and another is found associated with the cytoskeleton (13).

Four of the *Drosophila* heat shock-activated genes, encoding proteins of M_r 27,000 (hsp 27), 26,000 (hsp 26), 23,000 (hsp 23), and 22,000 (hsp 22), are located at chromosomal subdivision 67B. The four genes contained within an 11-kilobase region are not transcribed in the same direction and appear to contain no intervening sequences (14–17). Little information concerning the structure of these four heat shock proteins is available. The size of the proteins has been estimated from $NaDodSO_4$ gel electrophoresis data, and isoelectric focusing data indicate that hsp 27 and hsp 26 are relatively basic with isoelectric points of ≈7.5 and hsp 23 and hsp 22 are more acidic with isoelectric points of ≈5.9 and 5.8, respectively (18, 19). Here we report the primary sequences of the protein coding regions of these four small heat shock genes and the deduced amino acid sequences of the four proteins.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Restriction enzyme digestions, agarose and acrylamide gel electrophoresis, plasmid DNA isolation, end labeling by T4 polynucleotide kinase, and Maxam and Gilbert DNA sequence analyses were carried out as described (20, 21). Sequence analysis of the *hsp 27*, *hsp 26*, and *hsp 23* genes was carried out by using DNA from clone J1; analysis of the *hsp 22* gene was carried out by using DNA from clone T6. J1 and T6 have been described (15).

RESULTS

Primary DNA Sequence of Protein Coding Regions. The primary sequence of the protein coding regions of the *hsp 27*, *hsp 26*, *hsp 23*, and *hsp 22* genes has been determined. An overview of the organization of the four heat shock genes at locus 67B is presented in Fig. 1 and the strategy followed in the sequence determination is summarized in Fig. 2. The DNA sequences obtained are shown in Fig. 3. One open reading frame large enough to encode the appropriate heat shock protein was found in each segment of DNA that hybridized to mRNA from heat-treated cells. Previously, we reported the primary sequence of the regions surrounding the 5′ end of the mRNAs (20). The ATG that begins each sequence is the first ATG after the proposed site of initiation of transcription identified previously. The open reading frame following the initiation triplet is translated in Fig. 2. All alternative reading frames contain multiple stop codons and could not encode a protein as large as the small heat shock proteins.

Analysis of Predicted Amino Acid Sequences. The predicted amino acid compositions of the four small heat shock proteins are consistent with the limited experimental data available on the proteins. The predicted molecular weights of the proteins based on the amino acid composition (Table 1) match well with the size estimates from $NaDodSO_4$ gels (17, 18); hsp 27 is 27,091 daltons, hsp 26 is 26,618 daltons, hsp 23 is 23,544 daltons, and hsp 22 is 22,666 daltons.

Radioactive labeling experiments by Arrigo *et al.* (10) demonstrated that hsp 23 was not labeled with [^{14}C]tryptophan whereas the other three small heat shock genes could be labeled with this amino acid. This result indicates that hsp 23 contains no tryptophan and hsp 27, hsp 26, and hsp 22 contain tryptophan. The amino acid composition data in Table 1 are consistent with this finding since only hsp 23 lacks tryptophan.

Comparison of Predicted Amino Acid Sequences. A comparison of the four amino acid sequences revealed that the four are closely related over a large part of their sequences. The four sequences are aligned in Fig. 3; the four amino acid sequences are most homologous from position 85 to position 195. In this region, the same amino acid is used for at least three of the proteins at 71% of the positions, and the same amino acid is used

Abbreviation: hsp, heat shock protein.

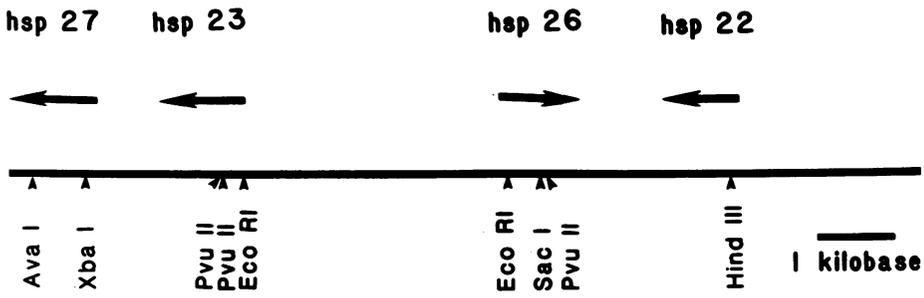

FIG. 1. General structure of the portion of the 67B cytological locus encoding the four small heat shock proteins. The direction of the arrows indicates the direction of transcription of each mRNA. The length of the arrows corresponds to the size of the transcribed regions. Data are from ref. 15.

for at least two of the proteins at 92% of the positions. A secondary region of homology exists from amino acid positions 1–14, where one amino acid is used by at least three of the proteins at 50% of the positions, and one amino acid is used by at least two of the proteins at 93% of the positions.

Hsp 22 and hsp 23 have deletions of 27 and 20 amino acids relative to hsp 26 and hsp 27; these deletions which occur toward the amino terminus relative to major regions of homologies account for most of the molecular weight differences among the four small heat shock proteins. When aligned as shown in Fig. 2, hsp 22 is missing amino acids 54 through 80, and hsp 23 is missing amino acids 66 through 86. The remainder of the differences in molecular weight are due to truncated carboxy termini. Hsp 22 has 13 fewer amino acids at the carboxy terminus than hsp 27; 9 fewer than hsp 26 and 5 fewer than hsp 23.

The Heat Shock Proteins Are Similar to α-Crystallin. The sequences of the heat shock proteins as presented in Fig. 3 were compared with known amino acid sequences (24). The amino acid sequence of bovine α-crystallin was found to be similar to that of the heat shock proteins over ≈40% of their lengths. The region of greatest homology is shown in Fig. 4 and extends from amino acid 89 to amino acid 164 (Fig. 3) and from amino acid 72 to amino acid 145 (out of 173) in the α-crystallin chain (25, 26). Over this region of 76 amino acids, the same amino acid is used by α-crystallin and at least one of the heat shock genes at 53 (68%) of the positions. Over this same region, the same amino acid is used by α-crystallin and at least three of the heat shock genes at 39 (51%) of the positions. However, in this stretch of 76 amino acids, there are only 59 positions at which the same amino acid is used by at least three of the heat shock genes; α-crystallin matches at least three of the heat shock genes at 39 out of 59 (66%) of these positions. In addition, similar but nonidentical amino acids are found at nine additional positions in this region (111, 116, 128, 140, 143, 148, 152, and 164 in Fig. 4).

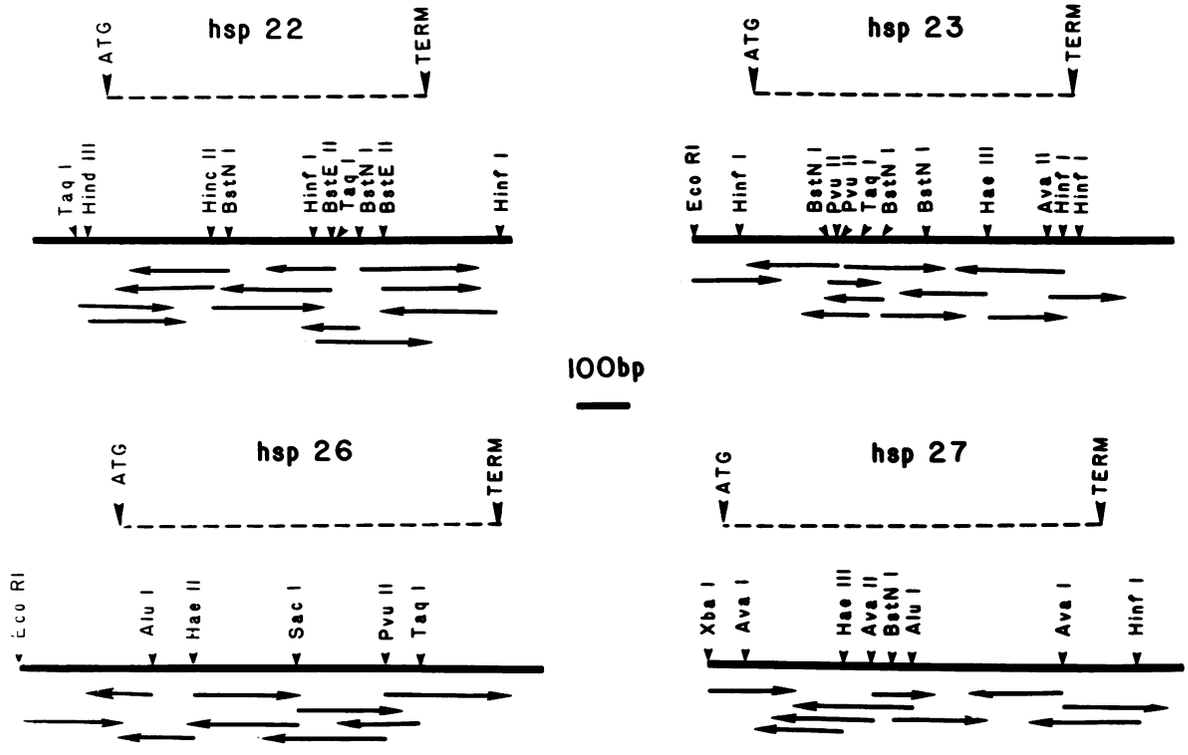

FIG. 2. Strategy used to determine the nucleotide sequences of the protein coding regions of the small heat shock genes. Only relevant restriction sites are indicated (for a complete restriction map of the clones, see refs. 22, 20). —, translated region; TERM, termination signal. DNA sequence analysis using the chemical cleavage method of Maxam and Gilbert (23) was carried out from each of the sites indicated (◄) at the bottom of the figure. The length of the arrows corresponds to the number of nucleotides actually determined from each start. bp, Base pair(s).

DISCUSSION

The coding regions of the four small heat shock proteins have been identified by primary sequence analysis. Translation from the first ATG of the open reading frame to the first stop codon would produce proteins of M_r 27,100, 26,600, 23,500, and 22,700. These values are in good agreement with the experimentally determined values of 27,000, 26,000, 23,000, and 22,000. Our data predict that only hsp 23 lacks tryptophan, and Arrigo *et al.* (10) found that only hsp 23 could not be labeled with radioactive tryptophan.

The Four Small Heat Shock Proteins Have Similar Sequences. The amino acid sequences of the small heat shock proteins are similar for more than half of their length. The most extensive homology extends from position 85 to position 195

```
                                                                                                          aa25
hsp27         met ser ile ile pro leu leu his leu ala arg glu leu asp his asp tyr arg thr asp trp gly his leu
              ATG TCA ATT ATA CCA CTG CTG CAC TTG GCC CGG GAG TTG GAT CAT GAC TAC CGC ACC GAC TGG GGG CAT TTG

hsp26         met ser leu ser thr ser leu ser leu val asp glu leu     gln glu pro arg ser pro ile tyr glu leu
              ATG TCG CTA TCT ACT TCG CTT TCG CTT GTG GAT GAA CTC     CAG GAC CCC CGC AGC CCC ATC TAC CAG CTT

hsp23     met ala asn ile pro leu leu leu ser leu ala asp asp leu gly arg met ser met val pro phe tyr glu leu
          ATG GCA AAT ATT CCG TTG TTG TTG AGC CTT GCC GAC GAT TTG GGC CGA ATG TCC ATG GTG CCC TTC TAT GAG CCC

hsp22     met arg ser leu pro met phe trp arg met ala glu glu met ala arg met pro arg leu ser ser pro phe his
          ATG CGT TCC TTA CCG ATG TTT TGG CGC ATG GCC GAG GAG ATG GCA CGG ATG CCA CGC CTC TCC TCG CCC TTT CAC


                                                                                                          aa50
hsp27     leu glu asp asp phe gly phe gly val his ala his asp leu phe asn pro arg arg leu leu leu pro asn thr
          CTG GAG GAT GAC TTC GGT TTT GGC GTC CAT GCC CAC GAT CTG TTC AAT CCG CGT CGC CTG CTA CTG CCC AAC ACC

hsp26     gly leu gly leu his pro his ser arg tyr val leu leu pro leu gly thr gln gln arg arg ser ile asn gly
          GGA CTG GGA TTG CAT CCG CAT TCC CGC TAC GTG CTG CTG CCC CTT GGC ACT CAG CAG CGC CGT TCC ATC AAC GGA

hsp23     tyr tyr cys gln arg arg gly ile pro tyr leu ala leu val gly pro met glu gln gln leu arg gln leu glu
          TAC TAC TGC CAG CGC CGA CGA ATC CCC TAC TTG GCC CTG GTT GGA CCG ATG GAG CAG CAG CTG CGC CAG CTG GAG

hsp22     ala phe phe his glu pro pro val trp ser val ala leu pro arg asn     trp gln gln ile ala arg trp gln
          GCC TTC TTC CAC GAG CCG CCC GTT TGG AGT GTG GCG CTA CCG AGG AAC     TGG CAG CAT ATT GCC CGC TGG CAG


                                                                                                          aa75
hsp27     leu gly leu gly arg arg arg tyr ser pro tyr glu arg ser his gly pro his asn gln met ser arg arg ala
          CTG GGA CTG GGT CGT CGT CGC TAT TCG CCG TAC GAG AGG AGC CAT GGC CCC CAC AAT CAA ATG TCA CGT CGC GCG

hsp26     cys pro cys ala ser pro ile cys pro ser ser pro ala gly gln val leu ala leu arg arg glu met ala asn
          TGT CCT TGC GCA TCG CCG ATA TGC CCA TCG TCG CCC GCT GGC CAG GTT TTG GCT TTA CGG CGC GAG ATG GCC AAC

hsp23     lys gln val gly ala ser ser gly ser ser gly ala val ser lys
          AAA CAG GTG GGC GCC TCG TCG GGA TCG TCG GGA GCC GTC TCG AAA

hsp22     glu gln glu
          GAG CAG GAG


                                                                                                         aa100
hsp27     ser gly gly pro asn ala leu leu pro ala     val gly lys asp gly phe gln val cys ile asp val ser gln
          TCG GGA GGT CCA AAC GCT CTG CTG CCC GCC     GTG GGC AAA GAT GGC TTC CAG GTG TGC ATC GAT GTG TCG CAG

hsp26     arg asn asp ile his trp pro ala thr ala his val gly lys asp gly phe gln val cys met asp val ala gln
          CGC AAC GAC ATC CAC TGG CCG GCA ACC GCC CAT GTG GGC AAG GAT GGA TTC CAG GTG TGC ATG GAC GTC GCC CAG

hsp23                                                 ile gly lys asp gly phe gln val cys met asp val ser his
                                                      ATC GGA AAG GAT GGC TTC CAG GTC TGC ATG GAT GTG TCG CAC

hsp22                         leu pro pro pro ala thr val asn lys asp gly tyr lys leu thr leu asp val lys asp
                              TTG CCT CCG CCG GCC ACC GTG AAC AAG GAT GGC TAC AAA CTC ACC CTG GAC GTC AAG GAC


                                                                                                         aa125
hsp27     phe lys pro asn glu leu thr val lys val val asp asn thr val val     val glu gly lys his glu glu arg
          TTC AAG CCC AAC GAG CTG ACC GTC AAG GTG GTG GAC AAC ACC GTG GTG     GTA GAG GGG AAG CAC GAG GAG CGC

hsp26     phe lys pro ser glu leu asn val lys val val asp ala ser ile leu     val glu gly lys his glu glu arg
          TTC AAG CCC AGT GAG CTC AAC GTG AAG GTG GTG GAC GCC TCC ATT TTG     GTC GAG GGC AAG CAT GAG GAA CGC

hsp23     phe glu pro ser glu leu val val gly val gln asp asn ser val leu     val glu gly asn his glu glu arg
          TTC GAG CCC AGC GAA CTG GTG GTC GGA GTG CAG GAC AAC TCC GTG GTG     GTG GAG GGC AAC CAT GAG GAG CGC

hsp22     tyr         ser glu leu lys val lys val leu asp gly ser val val leu val gly gly lys ser glu gln gln
          TAC         AGC GAG CTG AAG GTC AAG GTG CTG GAC GAG AGC GTT GTC CTG GTG GAG GCA AAA TCG GAG CAG CAG


                                                                                                         aa150
hsp27     gly asp gly his gly met ile     gln arg his phe val arg lys tyr thr leu pro lys gly leu thr pro thr
          GAG GAC GGC CAT GGA ATG ATC     CAG CGT CAC TTT GTG CGC AAG TAT ACC CTG CCC AAG GGC TTG ACC CCA ACG

hsp26     gln asp asp his gly his ile     met arg his phe val arg arg tyr lys val pro asp gly tyr lys ala glu
          CAG GAC GAC CAT GGT CAC ATC     ATG GGC CAC TTT GTG CGC CGC TAC AAG GTC CCC GAT GGC TAC AAG GCG GAG

hsp23     glu asp asp his gly phe ile     thr arg his phe val arg arg tyr ala leu pro pro gly tyr glu ala asp
          GAG GAT GAC CAT GGC TTC ATC     ACT CGT CAC TTT GTC CGC CGC TAT GCT CTG CCA CCC GGT TAT GAG GCT GAT

hsp22     glu ala glu gln gly gly tyr ser ser arg his phe leu arg arg phe val leu pro glu gly tyr glu ala asp
          GAG GCC GAA CAA GGT GGC TAT AGC TCC AGG CAC TTC CTC CGC CGA TTC GTT CTG CCG GAA GGA TAC GAG GCG GAC
```

FIG. 3. Nucleotide sequences of and amino acid sequences encoded by the four small heat shock genes. The ATG at the beginning of each sequence is the first ATG after the proposed site of initiation of transcription identified previously (20). There are multiple stop codons in the other reading frames of each gene. The methionine encoded by *hsp 23* and *hsp 22* was arbitrarily defined as amino acid 1 for the four proteins.

(Fig. 3). The same amino acid is used for at least three of the four proteins at 71% of the positions. A secondary region of homology exists in the first 14 amino acids of the four proteins. The four proteins are relatively dissimilar from amino acid 15 to amino acid 84 and at the carboxyl termini. The former region also includes deletions in hsp 22 and hsp 23 relative to hsp 26 and hsp 27. These deletions account for most of the difference in molecular weight between the two groups of proteins.

Table 1. Proposed amino acid compositions of the four small heat shock proteins

Amino acid	Residues per molecule			
	hsp 27	hsp 26	hsp 23	hsp 22
Ala	12	12	11	11
Arg	19	12	9	10
Asn	7	8	6	3
Asp	12	13	11	7
Cys	1	4	2	0
Gln	6	11	10	10
Glu	13	15	16	18
Gly	22	15	18	10
His	11	9	5	3
Ile	5	10	8	3
Leu	20	18	16	14
Lys	11	14	10	12
Met	5	4	5	5
Phe	6	3	5	6
Pro	17	15	13	15
Ser	13	17	14	15
Thr	11	4	3	9
Trp	1	1	0	4
Tyr	4	4	6	4
Val	17	20	18	15

As expected, the four heat shock genes also have homologous DNA sequences. In the alignment of Fig. 3, from position 85 to position 195, the same nucleotide is found in at least three of the four genes at 77% of the positions and the same nucleotide is found in all four genes at 37% of the positions. The partial homology of these four genes, which are located within an 11-kilobase region, suggests that these genes arose from duplication of a single gene.

The similarity among the four small heat shock genes allows the heat shock proteins of *Drosophila* to be grouped into three categories. One group comprises the small heat shock gene family. A second group includes hsp 68 and hsp 70, which have been shown to be partially homologous (27). DNA sequence analysis suggests that hsp 68 and hsp 70 are also similar at the amino acid level (unpublished data). The sole member of the third group is hsp 83. Thus, it is possible that the variety of functions carried out by the products of the heat shock genes is more limited than initially envisioned when it appeared that the synthesis of seven unrelated proteins was induced by heat shock.

α-Crystallin Is Remarkably Similar to the Small Heat Shock Proteins. A comparison was made between the heat shock proteins and known protein sequences, in hopes that a similarity in sequence might be due to a similarity in function. If a protein of known function were found that was structurally similar to the heat shock proteins, a clue to the function of the heat shock proteins might be provided.

The amino acid sequence of mammalian α-crystallin is surprisingly similar to those of the four small heat shock proteins. For a region containing 76 amino acids, the same amino acid is used by α-crystallin and at least three of the heat shock proteins at 39 positions and at least one of the heat shock proteins

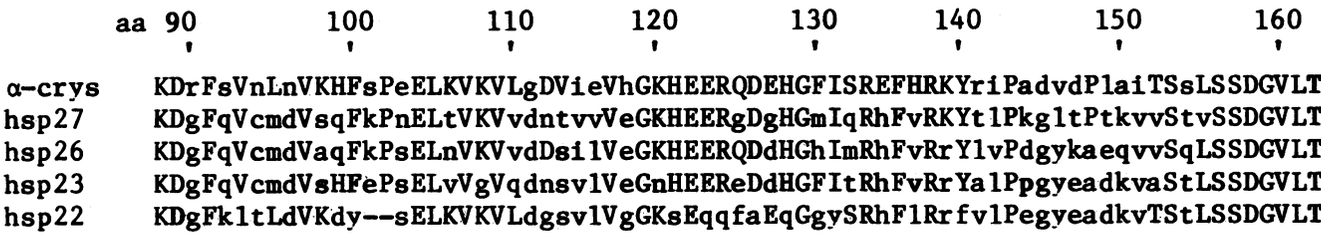

FIG. 4. Comparison of deduced amino acid sequences of the heat shock proteins and α-crystallin. The sequence of bovine α-crystallin (α-crys) is that of the B_2 chain (25). The numbering is as in Fig. 2 for the heat shock genes. The α-crystallin sequence comprises amino acids 72–145 (out of 173). Matches between α-crystallin and the heat shock proteins are indicated by capital letters. A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

at 53 positions. The homologies are contained in the same 40% of the amino acid sequences, near but not extending to the carboxyl termini. This region of the heat shock proteins that is homologous to α-crystallin is the same region at which the heat shock proteins are most similar to one another. Thus, it seems likely that this region represents a domain that is important for some function of these proteins.

Other intriguing similarities exist between mammalian α-crystallins and the four small heat shock genes of *Drosophila* besides the homologies described above. There are four closely related types of α-crystallin ≈ 20,000–30,000 daltons in size. There are four small heat shock polypeptides, ranging in size from 22,000 to 27,000 daltons. Two of the α-crystallins have isoelectric points of 7.4 and 7.1 and have been designated in the literature as B types (for basic). The other two have isoelectric points of 5.92 and 5.6 and have been designated as A types (for acidic) (28). Two of the heat shock proteins, hsp 27 and hsp 26, are relatively basic and have isoelectric points of ≈7.5. Hsp 23 and hsp 22 are more acidic and have isoelectric points of ≈5.9 and 5.8 (18, 19).

α-Crystallins are a major component of the vertebrate eye lens, comprising 35% of the protein of the lens (for review, see ref. 29). About 30% of the protein being synthesized in the cells of the lens is α-crystallin. α-Crystallins are highly specific polymeric proteins. The four closely related types form aggregates with an average molecular weight of 800,000 (30, 31). These aggregates perform a major structural role in determining the unique properties of the eye lens. A possible reason for the similar amino acid sequence in α-crystallin and the heat shock proteins is that this domain serves to facilitate aggregation. One testable hypothesis is that the heat shock proteins aggregate into complexes similar to the M_r 800,000 α-crystallin complexes. One could imagine that these complexes could serve a structural role in the nucleus, perhaps stabilizing and protecting the DNA against the traumas that are known to induce synthesis of the heat shock proteins.

We thank S. Vernick for expert technical assistance and R. Metzenberg for thoughtful reading of the manuscript. This research was supported by a grant from the U.S. Public Health Service to E.A.C. and by a postdoctoral fellowship from the Leukemia Society of America (T.D.I.). E.A.C. was a recipient of U.S. Public Health Service Research Career Development Award.

1. Ashburner, M. & Bonner, J. J. (1979) *Cell* **17**, 244–254.
2. Muller, M. J., Xuong, N. & Geiduschek, E. P. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5222–5225.
3. McAlister, L., Strausberg, S., Kulaga, A. & Finkelstein, D. B. (1979) *Curr. Genet.* **1**, 63–74.
4. Loomis, W. T. & Wheeler, S. (1980) *Dev. Biol.* **79**, 399–408.
5. Barnett, T., Altschuler, M., McDaniel, C. & Mascarenhas, J. (1980) *Dev. Genet.* **1**, 331–340.
6. Kelley, P. & Schlessinger, M. (1978) *Cell* **15**, 1277–1286.
7. Mitchell, H. K., Moller, G., Peterson, N. S. & Lipps–Sarimento, L. (1979) *Dev. Genet.* **1**, 181–192.
8. McAlister, L. & Finkelstein, D. B. (1980) *J. Bacteriol.* **143**, 603–612.
9. Velazquez, J. M., DiDomenico, B. J. & Lindquist, S. (1980) *Cell* **20**, 679–689.
10. Arrigo, A. P., Fakan, S. & Tissieres, A. (1980) *Dev. Biol.* **78**, 86–103.
11. Sinibaldi, R. M. & Morris, P. W. (1981) *J. Biol. Chem.* **21**, 10735–10738.
12. Oppermann, H., Levinson, W. & Bishop, J. M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1067–1071.
13. Wang, C., Gomer, R. H. & Lazarides (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3531–3535.
14. Wadsworth, S., Craig, E. A. & McCarthy, B. J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2134–2137.
15. Craig, E. A. & McCarthy, B. J. (1979) *Nucleic Acids Res.* **8**, 4441–4457.
16. Corces, V., Holmgren, R., Freund, R., Morimoto, R. & Meselson, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5390–5393.
17. Voellmy, R., Goldschmidt–Clermont, M., Southgate, R., Tissieres, A., Levis, R. & Gehring, W. (1981) *Cell* 261–271.
18. Storti, R. V., Scott, M. P., Rich, A. & Pardue, M. L. (1980) *Cell* **22**, 825–835.
19. Mirault, M.-E., Goldschmidt–Clermont, M., Moran, L., Arrigo, P. A. & Tissieres, A. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 819–827.
20. Ingolia, T. D. & Craig, E. A. (1981) *Nucleic Acids Res.* **9**, 1627–1642.
21. Craig, E. A., McCarthy, B. J. & Wadsworth, S. C. (1979) *Cell* **16**, 575–585.
22. Ingolia, T. D., Craig, E. A. & McCarthy, B. J. (1980) *Cell* **21**, 669–679.
23. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **62**, 560–564.
24. Dayhoff, M. O., Barler, W. C., Hont, L. T. & Schwartz, R. M. (1978) *Protein Segment Dictionary 78* (National Biomedical Research Foundation, Washington, D.C.).
25. Vander Ouderaa, F. J., de Jong, W. W. & Bloemendal, H. (1973) *Eur. J. Biochem.* **39**, 207–211.
26. Vander Ouderaa, F. J., de Jong, W. W. & Bloemendal, H. (1974) *Eur. J. Biochem.* **49**, 157–161.
27. Holmgren, R., Livak, K., Morimoto, R., Freund, R. & Meselson, M. (1979) *Cell* **18**, 1359–1370.
28. Schoenmakers, J. G. G. & Bloemendal, H. (1968) *Nature (London)* **220**, 790–791.
29. Bloemendal, H. (1977) *Science* **197**, 127–138.
30. Bloemendal, H., Berns, T., Zweers, A., Hoenders, H. & Bendetti, E. L. (1972) *Eur. J. Biochem.* **24**, 401–409.
31. Li, L. K. (1974) *Exp. Eye Res.* **18**, 381–389.