Biochemistry

 Volume 29, Number 36 September 11, 1990

*Accelerated Publications*

# Secondary Structure of the Mammalian 70-Kilodalton Heat Shock Cognate Protein Analyzed by Circular Dichroism Spectroscopy and Secondary Structure Prediction†

Seth Sadis, Krishnamurthy Raghavendra,* and Lawrence E. Hightower*
*Department of Molecular and Cell Biology, The University of Connecticut, Storrs, Connecticut 06269-3044*
*Received May 31, 1990; Revised Manuscript Received July 2, 1990*

ABSTRACT: Heat shock proteins are rapidly synthesized when cells are exposed to stressful agents that cause protein damage. The 70-kDa heat shock induced proteins and their closely related constitutively expressed cognate proteins bind to unfolded and aberrant polypeptides and to hydrophilic peptides. The structural features of the 70-kDa heat shock proteins that confer the ability to associate with diverse polypeptides are unknown. In this study, we have used circular dichroism (CD) spectroscopy and secondary structure prediction to analyze the secondary structure of the mammalian 70-kDa heat shock cognate protein (hsc 70). The far-ultraviolet CD spectrum of hsc 70 indicates a large fraction of $\alpha$-helix in the protein and resembles the spectra one obtains from proteins of the $\alpha/\beta$ structural class. Analysis of the CD spectra with deconvolution methods yielded estimates of secondary structure content. The results indicate about 40% $\alpha$-helix and 20% aperiodic structure within hsc 70 and between 16–41% $\beta$-sheet and 21–0% $\beta$-turn. The Garnier–Osguthorpe–Robson method of secondary structure prediction was applied to the rat hsc 70 amino acid sequence. The predicted estimates of $\alpha$-helix and aperiodic structure closely matched the values derived from the CD analysis, whereas the predicted estimates of $\beta$-sheet and $\beta$-turn were midway between the CD-derived values. Present evidence suggests that the polypeptide ligand binding domain of the 70-kDa heat shock protein resides within the C-terminal 160 amino acids [Milarski, K. L., & Morimoto, R. I. (1989) *J. Cell Biol. 109*, 1947–1962]. Much of this region in hsc 70 is unusually rich in charged and polar amino acids and is predicted to have a high $\alpha$-helical content. A model for polypeptide binding based on these structural features is proposed.

A wide variety of environmental perturbations induce cells to rapidly synthesize a group of polypeptides known as the heat shock (stress) proteins [reviewed by Lindquist and Craig (1988)]. The diverse treatments that activate stress protein synthesis have in common the potential to cause the intracellular accumulation of denatured or aberrant protein (Hightower, 1980; Finley et al., 1984; Ananthan et al., 1986; Edington et al., 1989). Eukaryotic cells contain a multigene family that encodes several closely related and evolutionarily conserved 70-kDa[1] stress proteins (the hsp 70 family) that differ in their intracellular location and regulation (Pelham, 1986; Lindquist & Craig, 1988). The constitutively expressed hsc 70 protein and stress-inducible hsp 70 are cytoplasmic/nuclear proteins, whereas grp 78/BiP and grp 75 reside in the lumen of the endoplasmic reticulum and in the matrices of mitochondria, respectively (Lindquist & Craig, 1988; Mizzen et al., 1989). Like hsc 70, grp 78/BiP and grp 75 are also

† This work was supported by grants from the National Institutes of Health (GM35334) and the National Science Foundation (DCB-8916418).
* Corresponding authors.

[1] Abbreviations: ATP, adenosine 5′-triphosphate; CD, circular dichroism; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; FPLC, fast protein liquid chromatography; Gdn-HCl, guanidine hydrochloride; grp 75, 75-kDa glucose-regulated protein; grp 78/BiP, 78-kDa glucose-regulated protein/immunoglobulin heavy-chain binding protein; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; hsc 70, 70-kDa heat shock cognate protein; hsp 70, 70-kDa heat shock protein; kDa, kilodalton; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; UV, ultraviolet.

0006-2960/90/0429-8199$02.50/0 

constitutively expressed and are abundant in the absence of stress, but can be induced further by agents capable of perturbing protein maturation in their respective compartments.

Insights into the function of the 70-kDa stress proteins have arisen recently from combined biochemical and genetic analyses. Both hsp 70 and hsc 70 accumulate in the granular region of the nucleolus during heat treatments that cause visible alterations in nucleolar morphology (Pelham, 1984; Welch & Feramisco, 1984; Welch & Mizzen, 1988). Cells that synthesize excess hsp 70 from transfected hsp 70 genes display accelerated recovery from heat-induced nucleolar damage (Pelham, 1984; Milarski & Morimoto, 1989). In unstressed cells, hsc 70 has been shown to play an important role in the translocation of certain proteins across organellar membranes (Deshaies et al., 1988a; Chirico et al., 1988). Hsc 70 may also catalyze removal of clathrin from coated vesicles (Ungewickell, 1985; Rothman & Schmid, 1986; Chappell et al., 1986). In addition, hsc 70 binds to calmodulin (Stevenson & Calderwood, 1990) and to apocytochrome *c* (Sadis et al., 1990), and both hsc 70 and grp 78/BiP bind to several hydrophilic peptides (Flynn et al., 1989; Vanbuskirk et al., 1989). A portion of hsc 70 is also found associated with microtubules (Green & Liem, 1989) and with the avian progesterone receptor (Kost et al., 1989). Immunoprecipitation analyses using antisera specific for viral or cellular oncogene products indicate that hsc 70 and hsp 70 associate with mutated forms of the cellular p53 protein (Pinhasi-Kimhi et al., 1986; Finlay et al., 1988), the polyomavirus middle T-antigen mutant NG59 (Walter et al., 1987), as well as the wild-type large T-antigen of simian virus 40 (Sawai & Butel, 1989), and the E1A product of adenovirus (White et al., 1988). The use of antisera specific for the 70-kDa stress proteins coprecipitates a variety of as yet unidentified cellular proteins, some of which associate in a cell-cycle-specific manner (Milarski et al., 1989). The results of a detailed immunoprecipitation analysis by Welch and co-workers suggest that hsc 70 and hsp 70 interact transiently with nascent polypeptides in cells under normal growth conditions (Beckmann et al., 1990). By a similar approach, grp 78/BiP has been shown to bind to immunoglobulin heavy chain and to a variety of misfolded proteins in the lumen of the endoplasmic reticulum (Bole et al., 1986; Dorner et al., 1987; Kassenbrock et al., 1988; Kozutsumi et al., 1988; Hurtley et al., 1989). The 70-kDa stress proteins also bind to and hydrolyze ATP, and in many cases, the addition of ATP to complexes formed between heat shock proteins and other polypeptides causes disruption of the complex (Munro & Pelham, 1986; Clarke et al., 1989; Flynn et al., 1989; Hurtley et al., 1989; Kost et al., 1989; Vanbuskirk et al., 1989; Beckmann et al., 1990). Taken together, these observations suggest that members of the hsp 70 family associate with misfolded proteins that accumulate during stress and may further assist certain protein folding and trafficking pathways for nascent polypeptides in nonstressed cells (Pelham, 1986; Deshaies et al., 1988b; Ellis & Hemmingsen, 1989; Rothman, 1989; Fischer & Schmid, 1990; Beckmann et al., 1990).

A major unanswered question is how a single heat shock protein can associate with many other polypeptides that lack extensive sequence similarity but may share similar conformational or structural features. A better understanding of the heat shock protein–polypeptide ligand interaction is currently prevented by our fragmentary knowledge of heat shock protein structure. Although the relative distribution of binding domains for ATP, calmodulin, and misfolded polypeptide ligands has been investigated, there have been few biophysical studies of heat shock protein structure. Currently, only the N-terminal 44-kDa fragment of hsc 70 has been crystallized in preparation for X-ray diffraction studies (DeLuca-Flaherty et al., 1988).

We have used circular dichroism (CD) spectroscopy to assess the types and amounts of secondary structural elements within the 70-kDa stress proteins. In a preliminary study, we compared the far-UV CD spectrum obtained for mammalian hsc 70 to that obtained for the bacterial DnaK protein, the prokaryotic hsc 70 homologue. Qualitatively, the spectra were quite similar, suggesting that the overall folding pattern of the 70-kDa stress proteins has been conserved throughout evolution (Sadis et al., 1990). In this report we analyze the CD spectra obtained for mammalian hsc 70 in greater detail to determine secondary structure content. In conjunction, we also estimate secondary structure with prediction algorithms applied to the hsc 70 amino acid sequence.

## Materials and Methods

*Purification of hsc 70 from Bovine Brain.* The following protocol was adapted from the methods described by Schlossman et al. (1984) and Ungewickell (1985). Dialyzed bovine brain cytosol was prepared as described (Schmid & Rothman, 1985) and fractionated by anion-exchange chromatography on a 225-mL column of Q-Sepharose fast flow (Pharmacia-LKB, Piscataway, NJ) at a flow rate of 1 mL/min. All chromatograpic steps were carried out at 4 °C. Protein was eluted with a 600-mL linear gradient of 0–0.6 M KCl in buffer A (25 mM Tris-HCl, 0.1 mM EDTA, pH 7.0). Hsc 70 enriched fractions, identified after each chromatographic step by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and silver staining, were pooled and dialyzed overnight against buffer C [20 mM Hepes–KOH, 25 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1 mM EDTA, 1 mM DTT, pH 7.0]. The sample was clarified by ultracentrifugation, adjusted to 2 mM magnesium acetate, and subjected to ATP–agarose chromatography as described (Schlossman et al., 1984). Hsc 70 enriched fractions were pooled and dialyzed overnight against buffer A plus 1 mM DTT and then applied to a 1-mL FPLC Mono Q (Pharmacia-LKB, Piscataway, NJ) anion-exchange column equilibrated in the same buffer without DTT. The sample was fractionated at a flow rate of 1 mL/min by the application of a 30-mL linear gradient of 0–0.6 M KCl in buffer A. Hsc 70 enriched fractions were dialyzed overnight against buffer C and then concentrated to 1–2 mL with a Centricon 30 ultrafiltration unit (Amicon, Danvers, MA). Approximately 10–15 mg of hsc 70 was obtained from 400 to 500 g of brain tissue, representing a 25–35% yield of hsc 70. The purified protein was stored at −70 °C prior to analysis.

*Gel Electrophoresis.* One-dimensional SDS–PAGE was performed as described by Laemmli (1970) using the gel sample buffer described by Hightower (1980). Proteins were visualized by silver staining according to a modification of the manufacturer's protocol (Sigma Chemical Co., St. Louis, MO), which gave more reliable staining of hsc 70. After gel fixation, the gels were incubated in the reducer solution first, followed by the silver equilibration solution and then the developer. Incubation times were not changed.

*Absorption Spectroscopy.* Absorption measurements of purified hsc 70 were obtained with an Aviv 14 DS spectrophotometer using a 1.00-cm path-length quartz cuvette. Extinction coefficients of hsc 70 were estimated by the methods of Scopes (1974) and Gill and von Hippel (1989). For the former method, the hsc 70 preparation was adjusted to 5 mM $Na_3PO_4$–50 mM $Na_2SO_4$, pH 7.0, by passage through an Econo-Pac 10DG desalting column (Bio-Rad Laboratories, Richmond, CA) equilibrated in the same buffer, followed by

concentration with a Centricon 30 ultrafiltration unit. To obtain denatured protein spectra, hsc 70 was adjusted to 4 M Gdn-HCl by the addition of an equal volume of 8 M Gdn-HCl, followed by gentle mixing on ice.

*Circular Dichroism Spectroscopy.* CD spectra were obtained on an Aviv 60 DS circular dichroism spectrometer calibrated by the procedure of Adler et al. (1973). Spectra of native hsc 70 were obtained in buffer C containing 2 mM magnesium acetate. Hsc 70 was combined with Gdn-HCl as described above to obtain denatured protein spectra. All spectra were recorded with thermostated quartz cylindrical cells (Hellma, Jamaica, NY) at 20 °C (±0.1 °C). Near-UV CD spectra were obtained with a 5.00-mm path-length cell in the wavelength region of 310–240 nm. Far-UV CD spectra were obtained with a 0.10-mm path-length cell in the wavelength region of 250–184 nm for native protein and in the region of 250–204 nm for denatured protein (as Gdn-HCl absorbs strongly below 204 nm). Data were recorded at 0.2- and 0.1-nm intervals for near- and far-UV measurements, respectively, with a time constant of 3.0 s and a 1.5-nm constant spectral bandwidth. Data were averaged over four repetitive scans for the near-UV measurements and over three repetitive scans for the far-UV measurements. Data were smoothed and normalized for protein concentration and optical path length. Protein concentration, determined from the UV absorption spectrum of the same sample used for CD measurements, varied between 0.7 and 1.0 mg/mL for different preparations. The mean residue ellipticity, $[\theta]$, was calculated by using a mean residue weight of 110.9 and expressed in deg·cm$^2$/dmol. The data obtained in this way for three different protein preparations were averaged.

*CD Data and Sequence Analysis.* Far-UV CD data of the native protein were analyzed by the spectral deconvolution methods of Manalavan and Johnson (1987; referred to as method 1) and Yang and co-workers (Yang et al., 1986; referred to as method 2). Computer programs were obtained from the authors and were used on an AT&T personal computer. $[\theta]$ values in the region of 260–184 nm at intervals of 0.5 nm were used for method 1 and $[\theta]$ values in the region of 250–190 nm at intervals of 1.0 nm were used for method 2. The results generated by method 1 were subject to the following constraints as suggested by the authors of this method: the sum of the fractional content of all secondary structural elements must lie between 0.99 and 1.01, and the root mean square value must be less than 0.22. Secondary structure was also estimated by the predictive methods of Garnier et al. (1978) and Chou and Fasman (1974) applied to the rat hsc 70 amino acid sequence (O'Malley et al., 1985). The Garnier–Osguthorpe–Robson method was obtained directly from the authors and was implemented on the University of Connecticut main-frame IBM 3090 system. The Chou–Fasman method was part of the University of Michigan software package (developed by Drs. S. D. Black and J. C. Glorioso) and was used on an AT&T personal computer.

## Results

Initially, hsc 70 was purified from bovine brain cytosol by a method utilizing affinity chromatography on ATP–agarose as the major purification step (Schlossman et al., 1984). In our hands, these preparations were typically 90–95% pure, as determined by laser densitometry of one-dimensional polyacrylamide gels stained with silver. To improve the purity of these preparations, we modified the original procedure by omitting the hydroxyapatite column step and including FPLC anion-exchange chromatography. A summary of the efficacy of each step in the purification is shown in Figure 1. Fractions

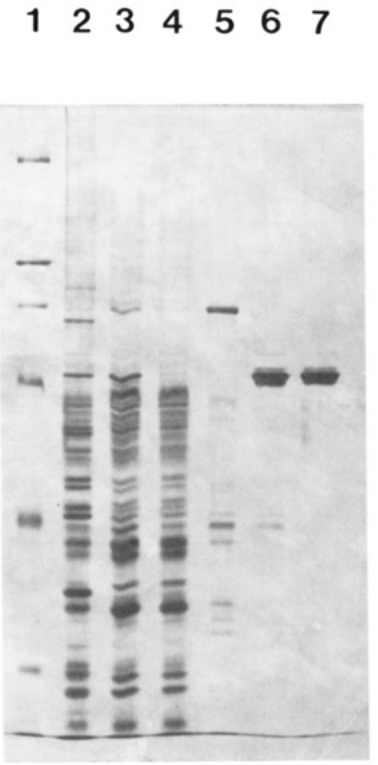


FIGURE 1: Analysis of bovine brain hsc 70 purification steps by SDS–PAGE. Aliquots from the major steps in the purification were fractionated on a 7.5% acrylamide gel and visualized by silver staining. Lane 1 contains a mixture of six proteins of known molecular mass: myosin, 205 kDa; $\beta$-galactosidase, 116 kDa; phosphorylase *b*, 97.4 kDa; bovine albumin, 66 kDa; ovalbumin, 45 kDa; carbonic anhydrase, 29 kDa. Lanes 2–7 contain portions of the pooled fractions of cytosol, Q-Sepharose fast flow, ATP–agarose flow-through, ATP–agarose salt wash, ATP elution, and Mono Q, respectively.

from each chromatographic run were analyzed by one-dimensional SDS–PAGE, and the resulting gels were visualized by silver staining. The final preparation was greater than 95% pure. In other work (data not shown), we characterized hsc 70 preparations further by measuring their capacity to catalyze ATP hydrolysis in the presence or absence of unfolded polypeptide ligands such as apocytochrome *c*. The specific activities of these preparations, approximately 3–5 $\mu$mol min$^{-1}$ mg$^{-1}$ in the presence of apocytochrome *c* and 0.5–1.0 $\mu$mol min$^{-1}$ mg$^{-1}$ in its absence, are similar to the values reported previously (Braell et al., 1984; Flynn et al., 1989).

Ultraviolet absorption spectra of native and denatured hsc 70 are presented in Figure 2. The number of aromatic amino acid residues within bovine hsc 70 that contribute to the spectra was estimated by amino acid analysis of the purified protein. Overall, the results (data not shown) were similar to the amino acid composition predicted for the human (Dworniczak & Mirault, 1987), rat (O'Malley et al., 1985), and mouse (Giebel et al., 1988) hsc 70 proteins by characterization of their DNA coding sequences. In particular, the analysis detected approximately 24 phenylalanine, 15 tyrosine, and 2 tryptophan residues in bovine hsc 70, the same as in human, rat, and mouse hsc 70. The conservation of aromatic amino acid content was expected, as the entire hsc 70 amino acid sequence has been highly conserved during mammalian evolution. For example, comparison of the mouse and human hsc 70 amino acid sequences shows only one amino acid substitution out of 646 amino acids (Giebel et al., 1988). The amino acid analysis suggests that the bovine hsc 70 sequence is also highly conserved. The maximum absorbance of hsc 70 typically occurred at 276 nm. A theoretical $A_{280}/A_{260}$ ratio of

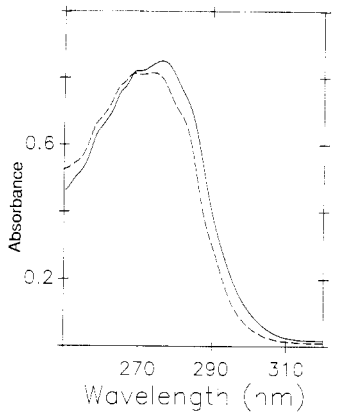


FIGURE 2: Absorption spectra of native (—) and denatured (---) hsc 70. Native hsc 70 was in 5 mM $Na_3PO_4$ and 50 mM $Na_2SO_4$, pH 7.0. Denatured hsc 70 was in the same buffer plus 4 M Gdn-HCl. The spectra were taken at 25 °C. The protein concentration was 1.6 mg/mL.

1.63 was calculated for pure protein on the basis of the extinction coefficients of Trp, Tyr, and Phe (Edelhoch, 1967; Teale & Weber, 1957) and their abundance in bovine hsc 70. The observed $A_{280}/A_{260}$ ratio of purified hsc 70 was 1.27 ± 0.02 ($n$ = 3), suggesting the presence of additional $A_{260}$ absorbing material (ATP, other nucleotides, or RNA) bound to hsc 70. Extensive dialysis against EDTA did not alter the $A_{280}/A_{260}$ ratio significantly (data not shown). We calculated that a molar ratio of approximately 0.5 ATP to 1.0 hsc 70 monomer could account for the discrepancy between the observed and the predicted $A_{280}/A_{260}$ (data not shown). Upon denaturation in 4 M Gdn-HCl, the absorbance spectrum of hsc 70 underwent a slight blue shift, a phenomenon commonly observed with many proteins (Edelhoch, 1967, and references cited therein).

Extinction coefficients for hsc 70 at 205 and 280 nm were estimated by the spectral method of Scopes (1974). Three independent measurements using different hsc 70 preparations yielded an $\epsilon_{205nm}^{1mg/mL}$ of 29.2 ± 0.09 and an $\epsilon_{280nm}^{1mg/mL}$ of 0.54 ± 0.02. These values fall within the range observed for other proteins (Scopes, 1974). For comparison, we also derived the extinction coefficient at 280 nm by the recently described theoretical method of Gill and von Hippel (1989). This method requires only knowledge of the amino acid composition and the relative absorbance of a protein in its native and denatured state. By this method, $\epsilon_{280nm}^{1mg/mL}$ was calculated to be 0.50. Both values are in general agreement with a recent report that estimated an $\epsilon_{280nm}^{1mg/mL}$ of 0.62 for bovine hsc 70 by measuring amino acid composition and absorbance of the same sample (Green & Eisenberg, 1990).

Circular dichroism spectra of native and denatured hsc 70, taken in both the far- and the near-UV wavelength regions, are shown in Figure 3. The far-UV CD spectrum of native hsc 70, in Figure 3A, was characterized by a strong maximum in the 190–195-nm region and minima at 208 and 222 nm. These features indicate the presence of a substantial fraction of α-helix in the native protein. The flattening of the spectrum between the nearly equivalent minima at 208 and 222 nm is a characteristic feature found in proteins of the α/β structural class that contain domains comprised of intermixed α-helical and β-sheet structures (Brahms & Brahms, 1980; Manavalan & Johnson, 1983). The nucleotide binding domains of several proteins contain such structures (Rao & Rossman, 1973; Levitt & Chothia, 1976). As expected, the far-UV CD spectrum of hsc 70 obtained in the presence of 4 M Gdn-HCl lacks minima at 208 or 222 nm, indicating that the structural elements

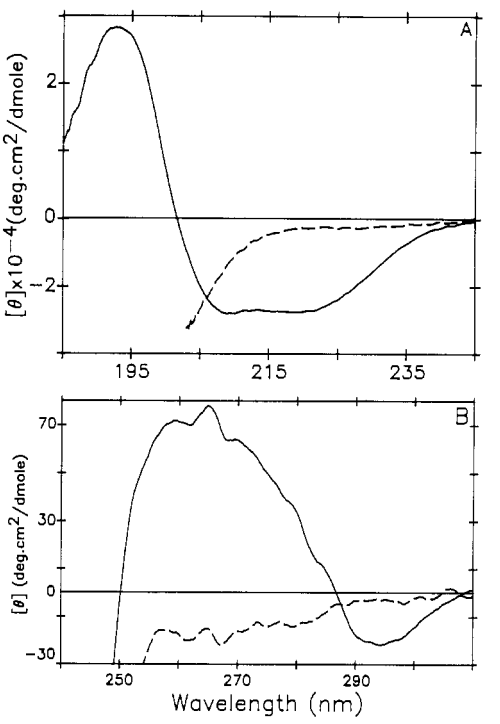


FIGURE 3: Circular dichroism spectra of native (—) and denatured (---) hsc 70. (Panel A) Far-UV CD spectra. (Panel B) Near-UV spectra. Native hsc 70 was in buffer C containing 2 mM magnesium acetate; denatured hsc 70 also contained 4 M Gdn-HCl. The spectra were taken at 20 °C and were base-line subtracted, smoothed, and normalized for cell path length and protein concentration. Experimental details are provided under Materials and Methods.

Table I: Abundance of Secondary Structure Elements in hsc 70

| | secondary structure (% of total) | | | |
|---|---|---|---|---|
| CD method | α-helix | β-sheet | β-turn | aperiodic |
| method 1[a] | 42 | 16 | 21 | 20 |
| method 2[b] | 40 | 41 | 0 | 19 |
| predictive method | α-helix | β-sheet | β-turn | aperiodic |
| GOR[c] | 40 | 24 | 13 | 23 |
| CF[d] | 28 | 23 | 38 | 11 |

[a] Far-UV CD spectrum was analyzed by the method of Manavalan and Johnson (1987). [b] Far-UV CD spectrum was analyzed by the method of Yang et al. (1986). [c] GOR refers to the Garnier–Osguthorpe–Robson method (1978); the decision constants were set at zero. [d] CF refers to the Chou–Fasman method (1974). The rat hsc 70 amino acid sequence was taken from O'Malley et al. (1985).

present in the native protein have adopted a random coil conformation in this denaturing solvent.

The near-UV CD spectrum of native hsc 70, shown in Figure 3B, exhibits fine structure due to aromatic Cotton effects arising from Phe chromophores in the region between 250 and 270 nm and Tyr and Trp chromophores between 270 and 290 nm. This fine structure suggests that many of the Phe, Tyr, and Trp side chains are ordered in the native protein. However, in 4 M Gdn-HCl, the near-UV CD spectrum of hsc 70 is featureless, indicating that the aromatic side chains are exposed to solvent and are disordered as expected when the protein unfolds.

The far-UV CD spectrum was analyzed by the methods of Manavalan and Johnson (1987) and Yang et al. (1986) to estimate the abundance of secondary structural elements within native hsc 70. The results are shown in Table I. Both methods estimated similar amounts of α-helix and aperiodic elements in hsc 70. Furthermore, the α-helix was the predominant secondary structural element within hsc 70, comprising about

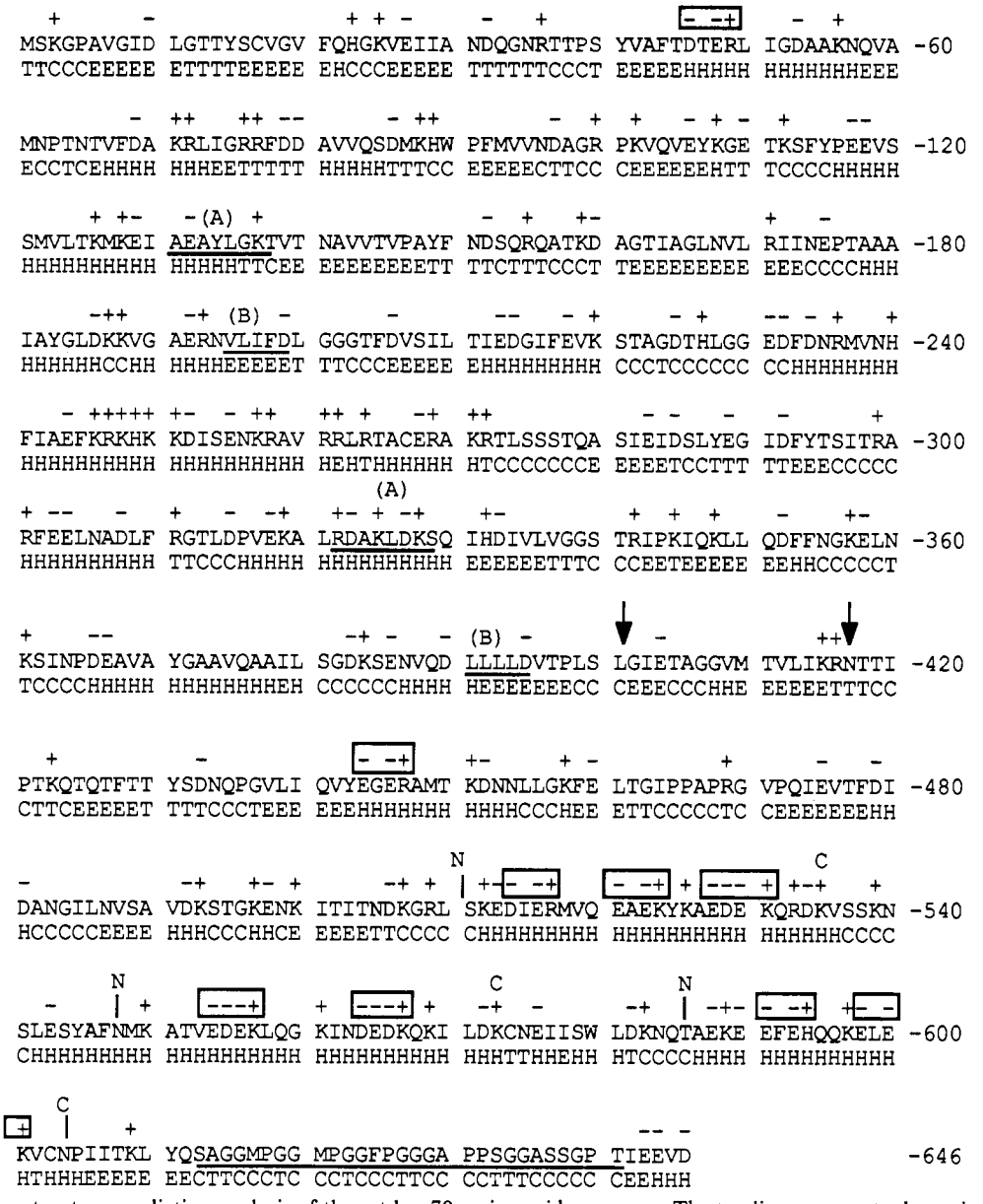


FIGURE 4: Secondary structure prediction analysis of the rat hsc 70 amino acid sequence. The top line represents the amino acid sequence, taken from O'Malley et al. (1985). Below the sequence lies the prediction; H, E, T, and C represent $\alpha$-helix, $\beta$-sheet, $\beta$-turn, and coil (aperiodic structure), respectively. The prediction was generated according to the methods of Garnier et al. (1978). See the text for further details.

40% of the total. Approximately 20% of hsc 70 was estimated to form aperiodic structures. The two methods differed in their estimation of $\beta$-sheet and $\beta$-turn content. The Manavalan and Johnson method estimated the presence of 16% $\beta$-sheet and 21% $\beta$-turn whereas the method of Yang et al. estimated the presence of 41% $\beta$-sheet and 0% $\beta$-turn. Typically, $\alpha$-helix content is estimated from CD data with the highest accuracy, followed, in order, by aperiodic structure, $\beta$-sheet, and $\beta$-turn (Manavalan & Johnson, 1987; Yang et al., 1986). For comparison, we averaged secondary structure content for 14 $\alpha/\beta$ proteins on the basis of the secondary structures determined by X-ray crystallography (Levitt & Greer, 1977). The "average" $\alpha/\beta$ protein is composed of 39% $\alpha$-helix (±11%), 29% $\beta$-sheet (±8%), 21% $\beta$-turn (±5%), and 11% aperiodic structure (±3%). Thus, the amount of secondary structure in hsc 70 estimated by CD analysis is in general agreement with the values derived for typical $\alpha/\beta$ proteins.

The abundance of secondary structural elements in hsc 70 was also estimated by secondary structure prediction using the methods of Garnier, Osguthorpe, and Robson (Garnier et al., 1978; referred to as the GOR method) and Chou and Fasman (1974) applied to the rat hsc 70 amino acid sequence (O'Malley et al., 1985). The results are shown in Table I. The content of secondary structural elements in hsc 70 predicted by the GOR method is in good overall agreement with the results obtained from the CD analysis. The predicted estimates of $\alpha$-helix and aperiodic structure are very close to those values estimated from the CD data, whereas the predicted content of $\beta$-sheet and $\beta$-turn is intermediate between the values derived from the two CD analysis methods. In contrast, the automated Chou–Fasman method differed significantly, underpredicting helix and aperiodic structure in favor of a very high content of $\beta$-turn. Therefore, the GOR method was chosen to examine the rat hsc 70 amino acid sequence.

The secondary structure prediction for the rat hsc 70 amino acid sequence (O'Malley et al., 1985) is presented in Figure 4. Immediately above the amino acid sequence are + and − symbols, which indicate the positions of positively and negatively charged amino acid residues (at physiological pH). Below the sequence lies the secondary structure prediction, using the symbols H for $\alpha$-helix, E for $\beta$-sheet, T for $\beta$-turn, and C for coil (aperiodic structure). The two arrows at amino

acids 401 and 417 mark the approximate positions of chymotrypsin and trypsin cleavage sites, respectively, that yield stable 44–45-kDa N-terminal fragments (Chappell et al., 1987; Milarski & Morimoto, 1989). This N-terminal region contains an ATP-binding domain that can fold and hydrolyze ATP independently of the C-terminal portion of the protein (Chappell et al., 1987; Milarski & Morimoto, 1989). The predicted secondary structure of this N-terminal region, for the most part, contains alternating stretches of $\alpha$-helices and $\beta$-sheets with shorter, interspersed areas of $\beta$-turn and aperiodic structure. The analysis predicts an extensive $\alpha$-helical region between amino acids 239 and 272. Portions of this sequence have the potential to form positively charged, amphiphilic $\alpha$-helices that have been shown to bind calmodulin in vitro (Stevenson & Calderwood, 1990).

The hsc 70 protein sequence C-terminal to the chymotrypsin and trypsin cleavage sites can be divided into three distinct regions on the basis of secondary structure predictions. The region between amino acids 420 and 510 is predicted strongly to contain $\beta$-sheet, $\beta$-turn, and aperiodic structures. In contrast, the predictions of $\alpha$-helices in this region are relatively weak and may be considered marginal. The next region between amino acids 511 and 604 is noteworthy because it is predicted to be extensively $\alpha$-helical. This region in hsc 70 has an unusually high content of charged and polar amino acids: 51% of the residues in this region are Glu, Lys, Asp, or Gln. The relative concentration of charged and polar residues would be expected to contribute, in part, to the strength of the $\alpha$-helical predictions (Chou & Fasman, 1974; Garnier et al., 1978). Indeed, the extent and magnitude of the $\alpha$-helix predictions are greater in this region of the hsc 70 amino acid sequence than in any other (data not shown).

The region in hsc 70 between amino acids 511 and 604 is predicted to contain three long contiguous $\alpha$-helices separated by short aperiodic structures. Possible amino and carboxy termini of the predicted $\alpha$-helices, designated in Figure 4 as N and C, respectively, were chosen for the sake of discussion according to the known preferences of certain amino acids to occupy these positions (Richardson & Richardson, 1988). During inspection of the sequence, we noted the presence of several copies of an element defined by the sequence E/D-x-E/D-K/R/H (acidic-x-acidic-basic). This "ionic" motif is identified in Figure 4 by using boxes to surround the charge symbols above the amino acid sequence. Nine copies of the ionic motif are present in the rat hsc 70 amino acid sequence. Seven are found clustered within the 90 amino acid segment beginning at amino acid 514. In the first and the third predicted $\alpha$-helices (numbered from the N-terminus of hsc 70), the clusters of ionic motifs repeat every seven residues. The spacing enables each amino acid side chain within the motif unit to extend from the same face of the predicted $\alpha$-helix. We also note that the spacings of the few hydrophobic residues present in the first and the third predicted $\alpha$-helices confine these residues to two or three adjacent faces. The region in hsc 70 containing these ionic motifs and $\alpha$-helices occupies a major portion of the polypeptide ligand binding domain as it is currently defined (Chappell et al., 1987; Milarski & Morimoto, 1989). Therefore, the ionic motifs, and the predicted $\alpha$-helices they reside in, may play a role in polypeptide binding.

Finally, the region in hsc 70 beginning at amino acid 605 and extending to the C-terminus is predicted to contain mostly alternating aperiodic structures and $\beta$-turns. Within this unusual glycine- and proline-rich segment, identified in Figure 4 as the underlined sequence at the extreme C-terminus, are four repeats of the sequence G-G-x-P, where x is Met, Phe, or Ala. Three of the tetrapeptide repeats are contiguous; the fourth is separated from the others by a Gly residue. The role of this sequence in hsc 70 function is unknown. In search of clues to possible structural and functional roles, we used the sequence GGMPGGMPGGFPGGGAP to search the National Biomedical Research Foundation protein sequence database. Although no identical matches were found, similar sequences containing contiguous tetrapeptide repeats were found in elastin. The isologous segment in elastin resides at the extreme amino terminus of the mature (processed) protein and has the sequence GGVPGAVPGGVPGGVFFP (Raju & Anwar, 1987). This segment is highly conserved in known mammalian and avian elastin sequences (Raju & Anwar, 1987); however, its secondary structure and/or function is not known.

## Discussion

The CD spectrum of the hsc 70 protein is similar to the spectra one obtains from other members of the $\alpha/\beta$ structural class. This evidence complements previous studies that characterized the ATP-binding properties of the 70-kDa stress proteins (Schlossman et al., 1984; Welch & Feramisco, 1985; Kassenbrock & Kelly, 1989). The nucleotide-binding domains of many proteins consist of $\beta\alpha\beta$ supersecondary structures (Rao & Rossmann, 1973; Levitt & Chothia, 1976) that generate characteristic CD spectra (Brahms & Brahms, 1980; Manavalan & Johnson, 1983). Therefore, the nucleotide-binding domain(s) of hsc 70 may be generally similar in structure to the nucleotide-binding domains within other proteins of the $\alpha/\beta$ structural class.

The localization of an ATP-binding site within the N-terminal fragment of the 70-kDa stress proteins (Chappell et al., 1987; Milarski & Morimoto, 1989) and the similarity of the hsc 70 CD spectrum to the spectra of other nucleotide binding proteins led us to inspect the amino acid sequence for patterns that matched the consensus elements for ATP binding as identified by Walker et al. (1982). The ATP-binding consensus sequence is divided into two short elements that are termed motifs A and B: motif A has the sequence A/G-$x_4$-G-K-T/S and motif B consists of an aspartate preceded by four hydrophobic residues. The sequences AEAYLGKT and VLIFD, which begin at amino acids 131 and 195, respectively, match exactly to the ATP-binding consensus. These sequences are identified in Figure 4 by underlined segments beneath the symbols (A or B) in the N-terminal portion of the hsc 70 amino acid sequence. This placement is consistent with the results of Milarski and Morimoto (1989), who find that deletion of amino acids 122–264 in the human hsp 70 gene abolishes the ability of the resulting gene product to bind ATP–agarose. Because biochemical evidence indicates the presence of two distinct ATP-binding sites (Rothman & Schmid, 1986), we sought additional matches to the ATP-binding consensus sequence. Although no other exact matches were identified, the sequences RDAKLDKS and LLLLD, which begin at amino acids 332 and 391, respectively, are perhaps the best candidates for a partial match to the consensus sequence.

The CD spectrum of bovine hsc 70 was analyzed by two different methods to estimate the content of secondary structural elements. Both methods generated similar estimates of $\alpha$-helical and aperiodic structures. Although the individual $\beta$-sheet and $\beta$-turn structure estimates differed, we note that the sum of the total $\beta$-structure is similar (37% for the Manavalan and Johnson method and 41% for the Yang et al. method). Whereas the Yang et al. method estimated a very

low value for β-turn (0%), the β-turn estimate of the Manavalan and Johnson method (21%) agrees closely with the average β-turn content in other α/β proteins (Levitt & Greer, 1977). The use of both methods was necessary in order to establish the range inherent in the individual secondary structure estimates obtained from the hsc 70 CD data. The GOR analysis of the hsc 70 amino acid sequence generated secondary structure estimates that were in good overall agreement with the results of the CD analysis. Although the Chou–Fasman method significantly underestimated the α-helical content of hsc 70, the predicted locations of the α-helices in the hsc 70 sequence were often coincident with those generated by the GOR method but the helices were simply not as long (data not shown). For example, the region between amino acids 511 and 604 was predicted as predominantly α-helical by both methods.

The 70-kDa stress proteins bind to aberrant or unfolded polypeptide chains, and it is this property that suggests their involvement in protein folding and assembly reactions (Pelham, 1986; Rothman, 1989; Fischer & Schmid, 1990; Beckmann et al., 1990). Previous studies suggest that the polypeptide ligand binding domain lies within the C-terminal 160 amino acids (Chappell et al., 1987; Milarski & Morimoto, 1989). This C-terminal region of hsc 70 has several distinctive properties including a high content of charged and polar amino acids. Earlier studies showing that high salt washes do not remove hsp 70 proteins bound to nuclei isolated from heat-shocked cells suggested a hydrophobic interaction between the heat shock protein and its substrate (Lewis & Pelham, 1985). However, this evidence is indirect, and more recent studies clearly demonstrate that hsc 70 and grp 78/BiP bind to a variety of hydrophilic peptides (Flynn et al., 1989; Vanbuskirk et al., 1989).

The results of the secondary structure prediction analysis of the rat hsc 70 amino acid sequence show a strong potential for α-helix formation between amino acids 511–604. Three α-helices were predicted for this region (Figure 4). Recently, analysis of other proteins that bind diverse polypeptide sequences has implicated α-helices as important structural components within the binding domain. In the first case, the structure of the human class I major histocompatability antigen determined by X-ray crystallography shows that its binding pocket for diverse peptide antigens consists of a groove formed between two parallel α-helices positioned above an antiparallel β-sheet platform (Bjorkman et al., 1987a,b). Structural analysis of calmodulin indicates a binding domain comprised of relatively short, clustered α-helices whose side chains form a hydrophobic patch flanked by negatively charged amino acids (O'Neil & DeGrado, 1990). These features assist calmodulin in binding to basic amphiphilic α-helices independent of their detailed sequences. In the case of the 54-kDa protein of the signal recognition particle, predicted α-helices unusually rich in Met residues are postulated to bind to the diverse but generally hydrophobic signal peptide sequences (Bernstein et al., 1989). Unlike the polypeptide binding domains of the signal recognition particle and calmodulin, the predicted α-helices in hsc 70's binding domain contain predominantly charged and polar amino acid residues. Therefore, hsc 70 might be expected to have greatest affinity for peptide ligands having hydrophilic side chains. Since hydrogen bonding to the main-chain amides and carbonyl oxygens of polypeptides with hydrophobic residues may also occur, hsc 70 may also bind these ligands but with lesser affinity. In the only study that has directly compared the interactions of different peptide sequences with the 70-kDa stress proteins, Flynn et al. (1989) find that different hydrophilic peptides do indeed stimulate the ATPase activity of hsc 70 and grp 78/BiP with varying efficiencies. However, the number of peptides tested in this study was too small to conclude whether overall hydrophilicity is the primary factor in determining binding affinities. Finally, hsc 70's binding domain appears to select for peptide sequences present in misfolded proteins. We suggest that a channel- or groove-like domain formed by α-helices may restrict binding to relatively extended or disordered polypeptides. Further studies are necessary to determine the three-dimensional structure of hsc 70's binding domain, which should aid our understanding of how hsc 70 associates with diverse peptide sequences.

ACKNOWLEDGMENTS

We thank Janet L. Frost for assistance with the secondary structure prediction algorithms, Myron Crawford at the Yale Nucleic Acid and Protein Chemistry Facility for performing the amino acid analysis, and Drs. Judith A. Kelly and Todd M. Schuster for many helpful discussions and for critically reading the manuscript. We acknowledge the Macromolecular Characterization Facility, Biotechnology Center, for use of the Aviv 14 DS.

**Registry No.** ATP, 56-65-5.

REFERENCES

Adler, A. J., Greenfield, N. J., & Fasman, G. D. (1973) *Methods Enzymol. 27*, 675–735.

Ananthan, J., Goldberg, A. L., & Voellmy, R. (1986) *Science 232*, 522–524.

Beckmann, R. P., Mizzen, L. A., & Welch. W. J. (1990) *Science 248*, 850–854.

Bernstein, H. D., Poritz, M. A., Strub, K., Hoben, P. J., Brenner, S., & Walter, P. (1989) *Nature (London) 340*, 482–486.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., & Wiley, D. C. (1987a) *Nature (London) 329*, 506–512.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., & Wiley, D. C. (1987b) *Nature (London) 329*, 512–519.

Bole, D. G., Hendershot, L. M., & Kearney, J. F. (1986) *J. Cell Biol. 102*, 1558–1566.

Braell, W. A., Schlossman, D. M., Schmid, S. L., & Rothman, J. E. (1984) *J. Cell Biol. 99*, 734–741.

Brahms, S., & Brahms, J. (1980) *J. Mol. Biol. 138*, 149–178.

Chappell, T. G., Welch, W. J., Schlossman, D. M., Palter, K. B., Schlesinger, M. J., & Rothman, J. E. (1986) *Cell 45*, 3–13.

Chappell, T. G., Konforti, B. B., Schmid, S. L., & Rothman, J. E. (1987) *J. Biol. Chem. 262*, 746–751.

Chirico, W. J., Waters, M. G., & Blobel, G. (1988) *Nature (London) 332*, 805–810.

Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 222–245.

Clarke, C. F., Cheng, K., Frey, A. B., Stein, R., Hinds, P. W., & Levine, A. J. (1989) *Mol. Cell. Biol. 8*, 1206–1215.

DeLuca-Flaherty, C., Flaherty, K. M., McIntosh, L. J., Bahrami, B., & McKay, D. B. (1988) *J. Mol. Biol. 200*, 749–750.

Deshaies, R., Koch, B., Werner-Washburne, M., Craig, E. A., & Scheckman, R. (1988a) *Nature (London) 332*, 800–805.

Deshaies, R. J., Koch, B. D., & Scheckman, R. (1988b) *Trends Biochem. Sci. 13*, 384–388.

Dorner, A. J., Bole, D. G., & Kaufman, R. J. (1987) *J. Cell Biol. 105*, 2665–2674.

Dworniczak, B., & Mirault, M.-E. (1987) *Nucleic Acids Res. 15*, 5181–5197.
Edelhoch, H. (1967) *Biochemistry 6*, 1948–1954.
Edington, B. V., Whelan, S. A., & Hightower, L. E. (1989) *J. Cell. Physiol. 139*, 219–228.
Ellis, R. J., & Hemmingsen, S. M. (1989) *Trends Biochem. Sci. 14*, 339–343.
Finlay, C. A., Hinds, P. W., Tan, T.-H., Eliyahu, D., Oren, M., & Levine, A. J. (1988) *Mol. Cell. Biol. 8*, 531–539.
Finley, D, Ciechanover, A., & Varshavsky, A. (1984) *Cell 37*, 43–55.
Fischer, G., & Schmid, F. X. (1990) *Biochemistry 29*, 2205–2212.
Flynn, G. C., Chappell, T. G., & Rothman, J. E. (1989) *Science 245*, 385–390.
Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.
Giebel, L. B., Dworniczak, B. P., & Bautz, E. K. F. (1988) *Dev. Biol. 125*, 200–207.
Gill, S. C., & von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319–326.
Green, L. A. D., & Liem, R. K. H. (1989) *J. Biol. Chem. 264*, 15210–15215.
Greene, L. E., & Eisenberg, E. (1990) *J. Biol. Chem. 265*, 6682–6687.
Hightower, L. E. (1980) *J. Cell. Physiol. 102*, 407–427.
Hurtley, S. M., Bole, D. G., Hoover-Litty, H., Helenius, A., & Copeland, C. S. (1989) *J. Cell Biol. 108*, 2117–2126.
Kassenbrock, C. K., & Kelly, R. B. (1989) *EMBO J. 8*, 1461–1467.
Kassenbrock, C. K., Garcia, P. D., Walter, P., & Kelly, R. B. (1988) *Nature (London) 333*, 90–93.
Kost, S. L., Smith, D. F., Sullivan, W. P., Welch, W. J., & Toft, D. O. (1989) *Mol. Cell. Biol. 9*, 3829–3838.
Kozutsumi, Y., Segal, M., Normington, K. Gething, M.-J., & Sambrook, J. (1988) *Nature (London) 332*, 462–464.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Levitt, M., & Chothia, C. (1976) *Nature (London) 261*, 552–557.
Levitt, M., & Greer, J. (1977) *J. Mol. Biol. 114*, 181–293.
Lewis, M. J., & Pelham, H. R. B. (1985) *EMBO J. 4*, 3137–3143.
Lindquist, S., & Craig, E. A. (1988) *Annu. Rev. Genet. 22*, 631–677.
Manavalan, P., & Johnson, W. C., Jr. (1983) *Nature (London) 305*, 831–832.
Manavalan, P., & Johnson, W. C., Jr. (1987) *Anal. Biochem. 167*, 76–85.
Milarski, K. L., & Morimoto, R. I. (1989) *J. Cell Biol. 109*, 1947–1962.
Milarski, K. L., Welch, W. J., & Morimoto, R. I. (1989) *J. Cell Biol. 108*, 413–423.
Mizzen, L. A., Chang, C., Garrels, J. I., & Welch, W. J. (1989) *J. Biol. Chem. 264*, 20664–20675.
Munro, S., & Pelham, H. (1986) *Cell 46*, 291–300.
O'Malley, K., Mauron, A., Barchas, J. D., & Kedes, L. (1985) *Mol. Cell. Biol. 5*, 3476–3483.
O'Neil, K. T., & DeGrado, W. F. (1990) *Trends Biochem. Sci. 15*, 59–64.
Pelham, H. R. B. (1984) *EMBO J. 3*, 3095–3100.
Pelham, H. R. B. (1986) *Cell 46*, 959–961.
Pinhasi-Kimhi, O., Michalovitz, D., Ben-Zeev, A., & Oren, M. (1986) *Nature (London) 320*, 182–185.
Raju, K., & Anwar, R. A. (1987) *J. Biol. Chem. 262*, 5755–5762.
Rao, S. T., & Rossmann, M. G. (1973) *J. Mol. Biol. 76*, 241–256.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648–1652.
Rothman, J. E. (1989) *Cell 59*, 591–601.
Rothman, J. E., & Schmid, S. L. (1986) *Cell 46*, 5–9.
Sadis, S., Raghavendra, K., Schuster, T. M., & Hightower, L. E. (1990) in *Current Research in Protein Chemistry* (Villafranca, J. J., Ed.) pp 339–347, Academic Press, New York.
Sawai, E. T., & Butel, J. S. (1989) *J. Virol. 63*, 3961–3973.
Schlossman, D. M., Schmid, S. L., Braell, W. A., & Rothman, J. E. (1984) *J. Cell Biol. 99*, 723–733.
Schmid, S. L., & Rothman, J. E. (1985) *J. Biol. Chem. 260*, 10050–10056.
Scopes, R. K. (1974) *Anal. Biochem. 59*, 277–282.
Stevenson, M. A., & Calderwood, S. K. (1990) *Mol. Cell. Biol. 10*, 1234–1238.
Teale, F. W. J., & Weber, G. (1957) *Biochem. J. 65*, 476–482.
Ungewickell, E. (1985) *EMBO J. 4*, 3385–3391.
Vanbuskirk, A., Crump, B. L., Margoliash, E., & Pierce, S. K. (1989) *J. Exp. Med. 170*, 1799–1809.
Walker, J. E., Saraste, M., J., & Gay, N. J. (1982) *EMBO J. 1*, 945–951.
Walter, G., Carbone, A., & Welch, W. J. (1987) *J. Virol. 61*, 405–410.
Welch, W. J., & Feramisco, J. R. (1984) *J. Biol. Chem. 259*, 4501–4513.
Welch, W. J., & Mizzen, L. A. (1988) *J. Cell Biol. 106*, 1117–1130.
White, E., Spector, D., & Welch, W. (1988) *J. Virol. 62*, 4153–4166.
Yang, J. T., Wu, C.-S. C., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208–269.