FEBS Letters 380 (1996) 68–72 FEBS 16625

The 170 kDa glucose regulated stress protein is a large HSP70-, HSP110-like protein of the endoplasmic reticulum

Xing Chen[a], Douglas Easton[b], Hyun-Ju Oh[a], Dong-Sin Lee-Yoon[a,**], Xiaoguang Liu[a], John Subjeck[a,*]

[a]*Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY 14263, USA*
[b]*Department of Biology, State University of New York College, Buffalo, NY 14222, USA*

Received 6 December 1995

Abstract The existence of a family of unusually large and highly diverged hsp70-like proteins (the hsp110/SSE family) has recently been described. The 170 kDa glucose regulated stress protein (grp170) is a retained endoplasmic reticulum glycoprotein that may be involved in immunoglobulin folding and/or assembly. We describe here the cloning of the cDNA for grp170 and show that it, like hsp110, is a large and highly diverged hsp70-like polypeptide which shares specific features with hsp70 (the dnaK family) and the hsp110/SSE family, while also differing from both. Grp170 contains an ATP binding domain and binds ATP, it possesses a carboxyl terminal NDEL sequence, and its mRNA is anoxia inducible.

Key words: Heat shock protein; Glucose regulated protein; Anoxia

1. Introduction

The stress response and the attending stress proteins have been implicated in numerous diverse areas of biomedical investigation, including cancer hyperthermia, tumor biology, resistance to chemotherapy, autoimmunity, ageing, the febrile response, stroke and cerebral ischemia, myocardial infarction, and several tissue disease states (cf. [1–3]). These proteins are also expressed in the absence of stress and have been shown to perform essential functions in the non-stressed cell. As a consequence of their fundamental involvement in cellular activities, the stress proteins are widely distributed phylogenetically and are evolutionarily conserved.

At the mammalian level, the principal stress proteins are separable into two interrelated groups; the heat shock proteins and the glucose regulated proteins. The glucose regulated proteins (grps) are localized in the endoplasmic reticulum/nuclear envelope and are co-induced by a unique set of stresses, including exposure to anoxia and reagents which perturb calcium homeostasis and/or glycosylation [4–8]. The principal grps of 78 kDa, 94 kDa, and 150 to 170 kDa size have been described (designated as grp78 or BiP, grp94 and grp170, respectively). Grp78 and grp94 are members of the 70 and 90 kDa families of heat shock proteins, bind immunoglobulin and have been implicated in the processing of proteins which are traversing the endoplasmic reticulum [4–8]. Grp170, like grp78 and grp94, has been long recognized. However, it has only recently been studied in detail. This protein has been found to be complexed with immunoglobulin in several human and mouse cell lines, regardless of the type of immunoglobulin chain(s) produced [9] and therefore may, like grp78 and 94, function as a molecular chaperone in the protein secretion pathway.

Recently, a family of large stress proteins has been described [10]. This family, referred to as the hsp110/SSE family, contains proteins whose sequences are the most distantly related known relatives of the well studied hsp70 family. Members of this group are significantly larger in size and contain sequence not present in members of the hsp70 family. Presently, all identified hsp70-like sequences can be categorized as being members of either one of these two groups; i.e. the dnaK (i.e. hsp70) family or the hsp110/SSE family [10]. The identification of hsp110 and its family demonstrates that the hsp70-related proteins have diverged more than had been previously believed. This suggests that the extent of this divergence is not yet fully evident and that additional divergent hsp70-like molecules may exist.

We describe here the cloning of a cDNA coding for grp170. We show that like members of the hsp110 family, grp170 is also a large and highly diverged hsp70-like protein; the first identification of such a protein resident in the endoplasmic reticulum. However, unlike all previously identified large hsp70-like sequences of this nature, grp170 is only distantly related to the hsp110/SSE family, although it is shown here to possess some structural features that are common to this family. Thus, grp170 and an additional sequence, an unidentified open reading frame from *Caenorhabditis elegans* cosmid T24H7, represent a third statistically distinct subfamily of unusual hsp70 related sequences.

2. Materials and methods

2.1. cDNA expression library screening

The λgt11 cDNA library was constructed using mRNA purified on an oligo-dT column from Chinese Hamster Ovary (CHO) cells and has been described previously [10]. Approximately 2×10^6 primary plaques were screened with the anti-CHO grp170 polyclonal antibody [9]. Positive clones were identified and inserts were isolated using appropriate restriction enzymes and subcloned into pBluescriptIIKS (Stratagene, La Jolla, CA).

2.2. Northern blot analysis

Exponentially growing CHO cells were used as controls or were exposed to an anoxic environment for 24 h as described previously [11], following which total RNA was extracted using lithium chloride (LiCl) precipitation method [12]. Up to 20 μg of RNA was electrophoresed using the glyoxal/DMSO method [13] and transferred to a nylon membrane (Zeta-probe, Bio-Rad, Richmond, CA). The membrane was hybridized with the positive clones which were labeled by random hexamer priming in hybridization buffer containing 0.5 M $NaHPO_4$, 1 mM ethylenediaminetetraacetic acid (EDTA), 1% BSA, and 7% SDS at pH

*Corresponding author. Fax: (1) (716) 845 8389.

***Present address:* Korea Inst. of Science and Technology, Biomedical Research Center, P.O. Box 131, Cheongryang, Seoul 130-165, Korea.

SSDI 0014-5793(96)00011-7

7. After 20 h of hybridization at 65°C, the membrane was washed twice in a solution containing 0.5% BSA, 1 mM EDTA, 5% SDS, 40 mM NaH_2PO_4 (pH 6.8) and a solution containing 1 mM EDTA, 1% SDS, 40 mM NaH_2PO_4 (pH 6.8), each for 30 min at 65°C. The membrane was exposed to X-ray film, Kodak XAR, for 24 h at −70°C.

2.3. DNA sequencing

Both strands of the clones were sequenced by the dideoxynucleotide chain termination method using Sequenase 2.0 (United States Biochemical) and by the PCR-based primer extension method [the latter using dsDNA Cycle Sequencing System (Gibco-BRL) or the *fmole* DNA sequencing system (Promega), according to manufacturer instructions]. Primers T3, T7, lambda gt11-forward, lambda gt11-reverse were used. Additionally, primers designed from determined sequence were used for further sequencing.

2.4. Rapid amplification of cDNA ends (RACE)

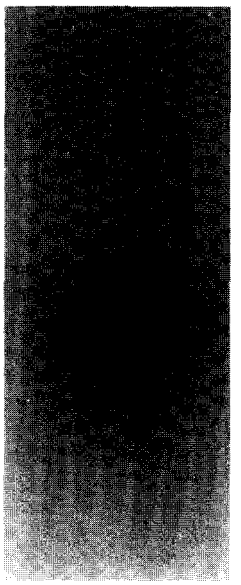
The procedure [14] follows the instructions of the 5′-AmpliFINDER RACE Kit which was purchased from Clontech laboratories, Inc. (Palo Alto, CA). The amplified RACE products were subcloned into plasmid pBluescript IIKS (Stratagene, La Jolla, CA) for sequencing.

2.5. Sequence analysis

Computer-based sequence analysis utilizes the Clustal method [15] as implimented by Lasergene (DNASTAR, Inc.; Madison, WI). Sequence comparisons using this method yield similarity values qualitatively but not quantitatively similar to those obtained from MacVector. The settings and parameters used are indicated in the figure legend. Sequence searches of Genbank were carried out using the BLAST program [16].

3. Results

An anti-GRP170 antibody, prepared and characterized in this laboratory [9], was used to screen a hamster λgt11 expression library. Four positive clones were identified and all of their cDNAs were found to hybridize, on a Northern blot of total RNA, to an anoxia-inducible message of approximately 4 kb. Fig. 1 shows this result for one of these cDNAs, the others being identical (lane 1, control; lane 2, 24 h of anoxia). These cDNAs were sequenced and found to significantly overlap, all encoding the same open reading frame which included a carboxyl terminal NDEL sequence. The latter is nearly identical to the C-terminal endoplasmic reticulum retention signal (KDEL) of GRP78 and GRP94 with N (asn) serving as a conservative substitution for K (lys). Since the largest of these cDNAs was 2 kb, the remainder of the open reading frame was obtained by the RACE method [14]. A total of 3.9 kb of sequence was obtained. An initiation codon defining the amino-terminus of the single open reading frame was found at position 55. It is flanked on its 5′ end by 4 of the 6 nucleotides required for optimal initiation of translation and has the predicted G residues at positions −3 and +4 [17]. The 3′ end of the cDNA contained a termination codon (TAA) at 3052 which was followed at position 3181 by a second termination codon (TGA). (This sequence is available from GenBank™/EMBL/DDBL under accession number U34206). A polyadenylation signal AATAAA [18] was identified at position 3860. The resulting cDNA contains a single open reading frame coding for a 999 residue polypeptide with a calculated molecular weight of 111,279 Da. The N-terminal 34 amino acids appears to comprise a typical leader sequence for targeting the protein to the endoplasmic reticulum. After removal of the putative leader, the protein has a (unglycosylated) predicted molecular weight of 107,464. The carboxyl terminal NDEL, the predicted leader sequence, the anoxia inducibility of the mRNA, and the sequence similarity to other stress proteins, in concert with the antibody screening, provide strong evidence that the clone is derived from the mRNA coding for the large (170 kDa) endoplasmic reticulum resident stress glycoprotein grp170.



1 2

Fig. 1. Northern blot analysis. Total RNA purified from control cells (lane 1) and from CHO cells exposed to 24 hrs of anoxia (lane 2) was electrophoresed, blotted to a nylon membrane and probed with the 2 kb cDNA obtained from antibody screening (see text). Similar analyses using cDNAs derived from three other positive clones resulted in indistinguishable results.

Table 1

Percent Similarity

	1	2	3	4	5	6	7	8	9	10	11	12		
1		42.3	22.9	56.8	24.9	74.1	24.1	23.8	25.2	24.3	20.9	22.9	1	BOV70C.PRO
2	51.2		21.0	43.2	21.8	45.1	20.5	21.8	22.4	17.8	19.4	20.8	2	DNAK.PRO
3	71.5	73.9		23.9	22.8	23.0	20.3	18.6	21.5	20.6	30.6	21.4	3	GRP170.PRO
4	39.0	51.3	71.3		24.6	55.8	21.6	23.6	25.8	22.4	20.0	23.4	4	GRP78.PRO
5	72.2	76.0	74.1	71.2		25.3	42.5	33.5	61.9	41.5	19.0	36.1	5	HSP110.PRO
6	23.3	48.6	73.1	41.0	69.8		25.0	25.8	26.6	24.4	22.5	25.2	6	SSA2.PRO
7	70.3	73.3	75.1	72.8	55.3	69.3		30.6	47.1	37.5	18.0	31.7	7	SUSS_T.PRO
8	72.4	74.7	77.7	75.2	64.4	72.1	65.6		34.2	30.8	19.2	75.5	8	SSE2.PRO
9	69.4	73.0	71.9	69.1	35.3	66.8	50.5	59.7		40.4	20.4	34.9	9	HSP70RY.PRO
10	70.8	74.5	75.4	73.0	55.6	70.1	59.7	66.0	53.4		17.3	33.3	10	C30.PRO
11	73.9	73.2	63.5	74.3	77.2	72.0	76.9	76.2	74.7	77.7		20.3	11	ELGNACII.PRO
12	72.9	74.8	75.0	74.3	62.0	72.7	65.2	24.2	59.6	64.0	75.0		12	SSE1.PRO
	1	2	3	4	5	6	7	8	9	10	11	12		

Percent Divergence

The predicted amino acid sequence was used to search the GenBank database with the Blast program [13]. Queries based on the whole grp170 peptide returned the following sequences: (1) an unidentified open reading frame from *Caenorhabditis elegans* cosmid T24H7; (2) members of the hsp110/SSE family intermixed with hsp70 sequences. Using multiple pair-wise alignments by the Clustal method, we compared the predicted amino acid sequence for grp170 with that of members of the hsp110/SSE group of proteins (hsp110 from hamster, sea urchin egg receptor for sperm, SSE1 and SSE2 from yeast, an open reading frame from *C. elegans* cosmid C30c11 and hsp70RY from man), with members of the dnaK (i.e. hsp70) family (Bovine hsc70, dnaK from *E. coli.*, grp78 (BiP) from

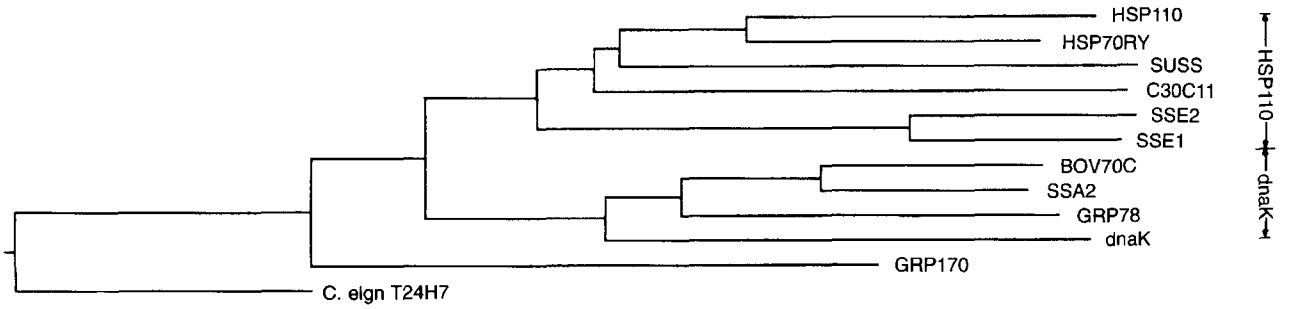


Fig. 2. Family tree of grp170 and related sequences. The unbalanced cladogram was determined by the Clustal method using the DNASTAR program. Branch distance corresponds to sequence divergence. Grp170 is seen to be most similar to the *C. elegans* open reading frame from cosmid T24H7. HSP110/SSE and dnaK subfamilies are indicated. Abbreviations and accession numbers are HSP110: 110 kDa heat shock protein from hamster, Z47807; HSP70RY: human B lymphocyte hsp70-like protein, L12723; SUSS: sea urchin egg receptor for sperm, L04969; C30C11: *C. elegans* hypothetical protein of 86.9 kDa from cosmid C30c11, Q05036; SSE1 and SSE2: *S. cerevisiae* hsp70-like proteins, D13908 and D13909; BOV70C: bovine hsc70, X53827; SSA2, *S. cerevisiae* hsp70 family member, P10592; GRP78: grp78 from yeast, P22010; dnaK, K01298; C.elgn T24H7, Z67884 (noted as ELgnacII in Table 1).

yeast, and SSA2 from yeast), and to the *Caenorhabditits elegans* cosmid T24H7 open reading frame. Table 1 shows the pair distances between these polypeptide sequences. Among this group of proteins, significantly closer relationship of grp170 to the *C. elegans* T24H7 open reading frame (ElgnacII in table 1) is clearly evident while grp170 appears to be equally distant from either hsp110 or dnaK family members. These data were then used to construct a phylogenetic relationship among these sequences (Fig. 2). In this family tree (unbalanced cladogram), branch distances correspond to sequence divergence. The protein sequences compared fall into three subfamilies; a dnaK subfamily (composed of hsp70 sequences) and the HSP110/SSE subfamily, as described previously [10], and a third previously unidentified subfamily, which contains grp170 and the *C. elegans* T24H7 sequence.

To further examine the relationship between grp170 and members of both the dnaK and hsp110 families, we performed a dot plot analysis. Fig. 3a shows a comparison of grp170 and hsp110 sequences while Fig. 3b shows a comparison of grp170 and bovine hsc70. While grp170 is only distantly related to the hsp110 family, as supported by the data in table 1 (analyzed in Fig. 2), the structure of grp170 in its carboxyl terminal one third of its sequence is clearly more like hsp110 than it is to hsp70. This region of the hsp110/SSE family contains a pattern of conserved sequences which defined this group [10]. Most of these defining segments of the hsp110/SSE family are present in grp170 and in correct register, but they are substantially diminished in similarity to corresponding segments in members of the hsp110 family. In addition, a large central non-homologous region (residues 500–700 in grp170), typical of the hsp110 family, is also present. However, in the case of grp170, this region is greatly expanded. The apparent 'expansion' of this region can account for much of the increase in size of grp170 relative to hsp110.

There are several stretches of sequence within the N-terminal 400 residues, 'ATPase domain', of grp170 which are similar to the corresponding N-terminal 400 residues of both hsp70 and hsp110 (indicated in figures 3a and b). While the carboxyl terminal one-third of grp170 is more similar to hsp110, analysis of the N-terminal 'ATPase domain' of grp170 shows that it is more generally similar to the Bovine hsc70 ATPase domain than it is to the corresponding region of hsp110. Comparison of residues 1–400 shows that grp170 exhibits a 30% similarity in amino acids with Bovine hsc70 and 25% with hsp110. More importantly, Bork et al. [19] have identified (by alignment of a large number of ATPase domains from sugar kinases, actin and HSP70s) a general pattern of residue types which define five elements of the ATP binding pocket in these proteins. We compared the percentage of residues in grp170, hsc70 and hsp110 which align with these five motifs for ATP binding. As indicated in Table 2, there is significant agreement between the specifications of Bork et al. and the sequences of grp170 and hsc70. A good agreement is also obtained with hsp110, although these motifs are less accurately represented in this molecule.

In Fig. 3c, we compared grp170 with the *C. elegans* open reading frame from cosmid T24H7. These two proteins are seen to exhibit similarity through their entire length, supporting the data presented in Fig. 2 and Table 1. Moreover, the *C. elegans* sequence also encodes a carboxyl terminal KTEL which, like the carboxyl terminal of grp170, strongly resembles an endoplasmic reticulum retention sequence.

Table 2
The compliance of 'ATPase' domains of grp170, hsp110 and hsc70 with ATP-binding motif specifications of Bork et al.

Motif[a] Protein	Phosphate 1 (19)	Phosphate 2 (15)	Connecting 1 (18)	Connecting 2 (15)	Adenosine (24)	Ave. %
grp170	(17) 89%	(13) 86%	(16) 88%	(11) 73%	(19) 79%	83%
hsp110	(19) 100%	(12) 80%	(13) 72%	(9) 60%	(13) 54%	73%
hsc70	(14) 74%	(14) 93%	(16) 88%	(10) 66%	(20) 83%	81%

[a]Motifs were located in 170 and 110 by alignment with the appropriate segments of bovine hsc70.

4. Discussion

The hsp70 family of proteins has long been considered to be a highly homologous family of proteins of approximately the same size and expressing the same general characteristics. However, the recent cloning of the cDNAs of a few proteins, such as the sea urchin egg receptor for sperm and mammalian hsp110, has demonstrated that much more greatly diverged and unusual relatives of the hsp70 family exist. This small group of proteins has been shown to comprise a statistically significant subfamily of hsp70 related sequences, i.e. the hsp110/SSE subfamily, which are the most distantly related known members of the hsp70 family [10].

A category of stress proteins in all eukaryotes are the glucose regulated proteins or grps. Two of the principal members of this family, grp78 and grp94 (of 78 and 94 kDa), are resident in the endoplasmic reticulum under normal conditions and have been strongly implicated in the processing of proteins traversing the secretory pathway [4–7]. In addition to these two grps, a third grp of approximately 150 to 170 kDa has been long observed, but has only recently been characterized [9]. Biochemical analysis of grp170 indicates that, like grp78 and grp94, it may also be involved in the processing of proteins in the secretory pathway. The discovery that grp170 can associate directly or indirectly with grp78, grp94 and immunoglobulin suggests that all three grps may function as a multimeric complex, either in unison or sequentially, in assembling immunoglobulin in vivo. However, grp170 has not been previously cloned.

We have described in this report the initial cloning of grp170 and demonstrate that, like members of the recently identified hsp110 family, it is also a large and highly diverged relative of the hsp70 family. Grp170 is the first example of such a protein resident in the endoplasmic reticulum/nuclear envelope and exhibiting the unique stress responsiveness characteristic of the grp class of stress proteins (e.g. induction during exposure to anoxia). Grp170 is the second example of an hsp70-like protein which is glycosylated (sea urchin egg receptor for sperm being the other). However, despite its resemblance to an endoplasmic reticulum member of the hsp110 family, it does not statistically align with either of the previously defined dnaK or hsp110 families. That this unique hsp70-, hsp110-like protein is phylogenetically conserved is suggested by the existence of the *C. elegans* open reading frame of cosmid T24H7 which is the most similar sequence to grp170 in the GenBank™ database. Interestingly, this *C. elegans* protein has a carboxyl terminal KTEL, which like the NDEL of grp170, strongly resembles an endoplasmic reticulum retention sequence. The existence of a similar protein of the endoplasmic reticulum which is conserved between *C. elegans* and mammals would suggest that it is an important element in the functioning of this organelle.

In addition to differing phylogenetically from either of the dnaK or hsp110 families, the evidence presented in this report also suggests that grp170 shares characteristics with each of these families. It has a carboxyl half which is closer to the hsp110 family, while it has an amino-terminal domain which exhibits greater similarity in sequence to members of the dnaK family. The carboxyl regions of the hsp70s are responsible for the peptide binding properties of these molecules. It has been proposed that a pattern of unique, conserved sequences in the carboxyl halves of hsp110 family members correspondingly determine unique peptide/protein binding characteristics which are anticipated to differ from those of the hsp 70's [10]. Moreover, grp170 also contains a nonhomologous expanded region (amino acid residues 500–700) which is similar to that observed in the hsp110 related sequences and which is also responsible for much of the increased size of these proteins. The amino-terminal halves of the hsp70's have been shown to strongly bind ATP, the binding and hydrolysis of which is believed to power

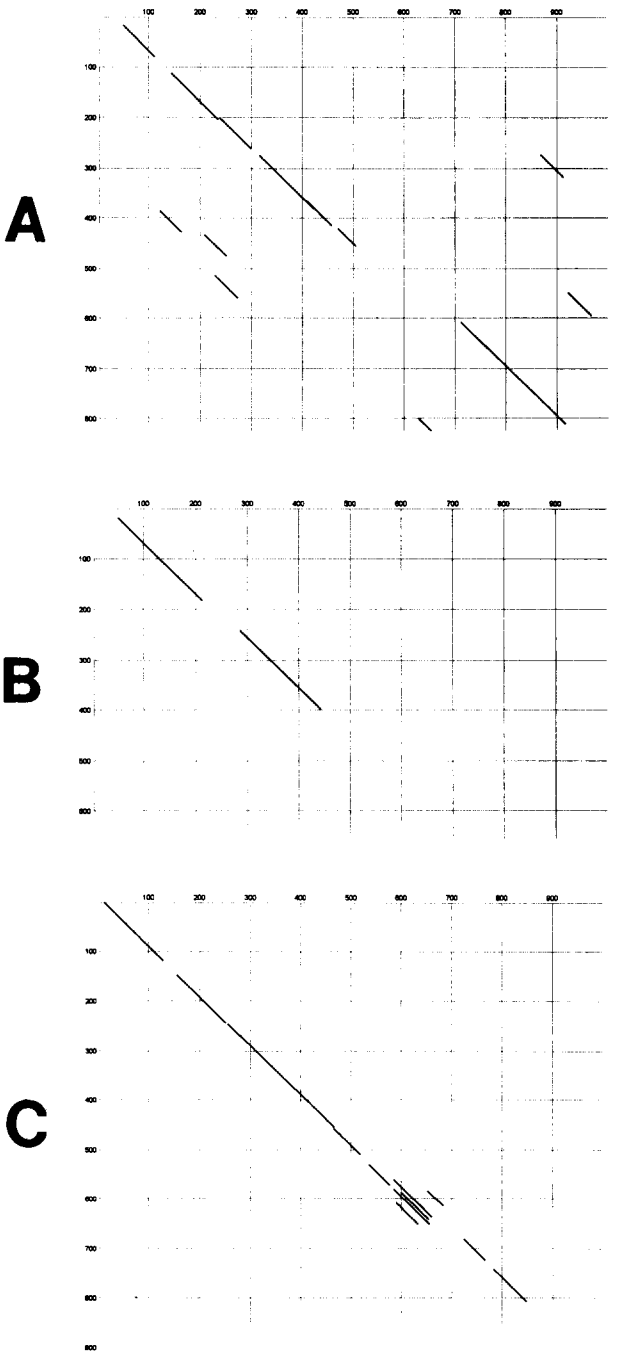


Fig. 3. Dot matrix plot comparisons of: (A) grp170 (X axis) vs. hsp110 (Y axis), (B) Grp170 vs. bovine hsc70, (C) grp170 vs. *C. elegans* hypothetical protein of cosmid T24H7. Grp170 is seen to be more similar to hsp110 in the carboxyl terminal region, but is most similar to the the *C. elegans* hypothetical protein of cosmid T24H7. In these plots a minimum similarity score of 25% and a window size of 30 was used. The 250 PAM matrix was used to score similarities. Accession numbers are presented in Fig. 2.

the peptide binding and release at the carboxyl end of the molecule. Sequence comparisons of the amino-terminal halves of these proteins indicates that grp170 more closely resembles the members of the dnaK family. Indeed, we have found that both hsp70 and grp170 avidly bind ATP agarose while hsp110 appears to exhibit little or no ATP binding capacity (data not shown). Moreover, it has been proposed that the ATP binding domain of one member of the hsp110 family (the sea urchin egg receptor) may be non-functional [20].

The hsp70 family has long been considered to be a highly conserved group of proteins, all exhibiting approximately the same molecular weight and sharing, at least, 50% identities in amino acid composition (cf. [10,21]). Recently, however, the identification of a few, highly diverged hsp70 related sequences has significantly altered this long held view. Moreover, it has been shown that these proteins constitute a statistically distinct group, referred to as the hsp110/SSE subfamily, thus suggesting that all hsp70 and hsp110 related sequences can now be viewed as being members of either of one or the other subfamily (i.e. hsp110/SSE or dnaK). While the functional significance of the increased complexity of this important group of proteins is just beginning to be considered, the cloning of grp170 and the data presented here indicates that even this recently derived view is not inclusive of all hsp70 related sequences. Clearly, the relative similarities and differences in the functions of grp170, hsp110 and dnaK protein subfamilies and the physiological implications of these alterations presents a novel area of investigation in the field of stress protein biology which can be expected to provide new insights into the functions of this important family of proteins.

Acknowledgements: We greatly appreciate the expert assistance of Ms. Melanie Murawski. This work was supported by PHS Grant GM45994.

References

[1] Black, A.R. and Subjeck, J.R. (1991) Methods Achieve Exp. Pathol. 15, 126–166.
[2] Welch, W.J. (1992) Physiol. Rev. 72, 1063–81.
[3] Strudler, P. (1996) Radiat. Res. 145, 106–116.
[4] Liu, E.S. and Lee, A.S. (1991) Nucleic Acids Res. 19, 5425–5431.
[5] Cai, J.W., Henderson, B.W., Shen, J.W. and Subjeck, J.R. (1993). J. Cell. Physiol. 154, 229–237.
[6] Roy, B. And Lee, A.S. (1995) Mol. Cell. Biol. 15, 2263–2274.
[7] Kang, H.S. and Welch, W.J. (1991) J. Biol. Chem. 266, 5643–5649.
[8] Melnick, J., Dul, J.L. and Argon, Y. (1994) Nature 370, 373–375.
[9] Lin, H., Masso-Welch, P., Di. Y., Cai, J., Shen, J. and Subjeck R. (1993) Mol. Biol. Cell 4, 1109–1119.
[10] Lee-Yoon, D., Easton, D., Murawski, M., Burd, R. and Subjeck, J.R. (1995) J. Biol. Chem. 270, 15725–15733.
[11] Shen, J.W., Subjeck, J.R., Lock, R.B. and Ross, W.E. (1989) Mol. Biol. Cell 9, 3284–3291.
[12] Puissant, C., and Houdebine, L.-M. (1990) BioTechniques 8, 148–149.
[13] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (Eds.) (1995) Current Protocols in Molecular Biology, Greene Pub. and John Wiley and Sons.
[14] Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA 85, 8998–9002.
[15] Higgins, D.B. and Sharp, P.M. (1989) CABIOS 5, 1510153.
[16] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[17] Kozak, M. (1991) J. Biol. Chem. 266, 19867–19870.
[18] Birnstiel, M.L., Bussslinger, M. and Strub, K. (1985) Cell 41, 349–359.
[19] Bork, P., Sander, C. and Valencia , A. (1992) Proc. Natl. Acad. Sci USA 89, 7290–7294
[20] Ohlendieck, K., Partin, J.S. and Lennarz, W.J. (1994) J. Cell Biol. 125, 817–824.
[21] Becker, E. and Craig, E.A. (1994) Eur. J. Biochem. 219, 11–23.