The mammalian endoplasmic reticulum stress response element consists of an evolutionarily conserved tripartite structure and interacts with a novel stress-inducible complex

Binayak Roy and Amy S. Lee*

Department of Biochemistry and Molecular Biology and the USC/Norris Comprehensive Cancer Center, University of Southern California School of Medicine, Los Angeles, CA 90033, USA

ABSTRACT

When mammalian cells are subjected to calcium depletion stress or protein glycosylation block, the transcription of a family of glucose-regulated protein (GRP) genes encoding endoplasmic reticulum (ER) chaperones is induced to high levels. The consensus mammalian ER stress response element (ERSE) conserved among grp promoters consists of a tripartite structure CCAAT(N$_9$)CCACG, with N being a strikingly GC-rich region of 9 bp. The ERSE, in duplicate copies, can confer full stress inducibility to a heterologous promoter in a sequence-specific but orientation-independent manner. In addition to CBF/NF-Y and YY1 binding to the CCAAT and CCACG motifs, respectively, we further discovered that an ER stress-inducible complex (ERSF) from HeLa nuclear extract binds specifically to the ERSE. Strikingly, the interaction of the ERSF with the ERSE requires a conserved GGC motif within the 9 bp region. Since mutation of the GGC triplet sequence also results in loss of stress inducibility, specific sequence within the 9 bp region is an integral part of the tripartite structure. Finally, correlation of factor binding with stress inducibility reveals that ERSF binding to the ERSE alone is not sufficient; full stress inducibility requires integrity of the CCAAT, GGC and CCACG sequence motifs, as well as precise spacing among these sites.

INTRODUCTION

The signaling from intracellular compartments to the cell nucleus represents a fundamental regulatory process in maintaining cellular homeostasis. In *Escherichia coli*, the cell exhibits compartment-specific stress responses mediated by alternative σ factors, allowing it to respond separately to stress in the cytoplasm and periplasmic compartments (1). In eukaryotes, the endoplasmic reticulum (ER) is a major calcium signaling compartment as well as the cellular organelle where proteins destined for transport to the cell membrane or cell exterior are synthesized and processed. In *Saccharomyces cerevisiae*, accumulation of unfolded proteins in the ER activates transcriptional induction of molecular chaperones and folding enzymes in the ER. This novel signaling process, referred to as the unfolded protein response pathway, is mediated by a partial palindrome CAGCGTG referred to as the UPRE, which interacts with a yeast factor Hac1 (2–4). In mammalian cells, in response to overloading of protein in the ER, as distinct from the unfolded protein response, NF-κB DNA binding was induced, resulting in the activation of κB-dependent gene expression (5). When mammalian cells are depleted of sterol, an ER membrane-bound transcription factor SREBP was released by sterol-regulated proteolysis, resulting in the activation of the LDL receptor gene and other genes involved in cholesterol synthesis (6). Therefore, evidence is emerging that diverse and novel mechanisms have evolved from *E.coli* to human to respond to intracellular signaling.

The mammalian glucose-regulated protein (GRP) genes provide an interesting model for studying ER to nucleus signaling. Disruption of ER structure and function greatly enhances the nuclear transcription of the GRP genes, as exemplified by grp78, which encodes a 78 kDa ER luminal protein related to HSP70 and is also referred to as BiP (7,8). Other GRP genes include grp94, which encodes a 94 kDa ER glycoprotein related to HSP90 (9,10), and ERp72, which encodes a 72 kDa ER luminal protein related to protein disulfide isomerase (11). The most potent inducers of the ER stress response include the calcium ionophore A23187 and thapsigargin (Tg), the latter being a non-phorbol ester tumor promoter which inhibits specifically the ER calcium-ATPase (12,13). Both reagents activate grp transcription through depletion of calcium stores from the ER (14). Another inducer of grp transcription is brefeldin A, which causes absorption of Golgi membrane into the ER and blocks the transport of proteins from the ER to the Golgi apparatus (15). Further, over-expression of malfolded proteins which accumulate in the ER or treatment of cells with tunicamycin, which blocks N-linked glycosylation, also activate grp transcription (16–19).

To decipher the genetic code for the coordinate induction of the grp genes, we seek to identify common *cis*-regulatory elements

---

*To whom correspondence should be addressed. Tel: +1 323 865 0507; Fax: +1 323 865 0094; Email: amylee@zygote.hsc.usc.edu

on the grp gene promoters and the *trans*-acting factors which mediate the ER stress response. A unique feature shared among grp promoters is the presence of a large number of CCAAT elements flanked by GC-rich motifs (20,21). In the most well-characterized rat grp78 promoter, using a combination of deletion and site-directed mutagenesis approaches, it was discovered that the promoter is functionally redundant (21). Further, duplicate copies of at least three grp78 promoter subfragments (C1, C3 and core) each containing a CCAAT or a CCAAT-like motif and the flanking GC-rich sequence were able to confer stress inducibility to heterologous promoters (13). The CCAAT motifs were shown to be binding sites for the multimeric CCAAT binding factor NF-Y, also referred to as CBF (22,23), which is required for maximal grp promoter activity (24,25). Downstream of the CCAAT motif was an unusual binding site for YY1, a multifunctional transcription factor with DNA bending properties (26–29), and for the Y-box cold shock domain proteins (30). In NIH 3T3 cells, over-expression of YY1 resulted in specific activation of the grp78 promoter under ER stressed conditions (29), whereas Y-box proteins suppressed the grp78 promoter activity (30).

Since the CBF/NF-Y, YY1 and Y-box protein binding sites are present in many cellular promoters not inducible by ER stress, a mystery remains as to how these factor binding sites in combination can specifically mediate the ER stress response. We report here that in contrast to the yeast UPRE which consists of a simple palindromic sequence, the mammalian ER stress response element (ERSE) consists of an evolutionarily conserved tripartite structure CCAAT(N_9)CCACG, with N being a strikingly GC-rich region of 9 bp. The ERSE, in duplicate copies, can confer 12- and 6-fold stress inducibility to a heterologous promoter in response to ER calcium depletion and blockage in protein glycosylation, respectively. The enhancing activity of the ERSE acts in a sequence-specific but orientation-independent manner. In addition to NF-Y and YY1 binding to the ERSE, we further discovered that an ER stress-inducible complex (ERSF) from HeLa nuclear extract binds specifically to the ERSE. Strikingly, the interaction of the ERSF with the ERSE requires a conserved GGC motif within the 9 bp region. Since mutation of the GGC triplet sequence also results in loss of stress inducibility, specific sequence within the 9 bp region is an integral part of the tripartite structure. Finally, correlation of factor binding with stress inducibility reveals that ERSF binding to the ERSE alone is not sufficient; full stress inducibility requires integrity of the CCAAT, GGC and CCACG sequence motifs, as well as precise spacing among these sites. The identification of the unique features of the mammalian ERSE allows rapid identification of other ER stress-responsive genes. Since ERSF is a potential novel target for the ER to nucleus signaling machinery, the discovery of its binding site within the ERSE will greatly facilitate its isolation and characterization.

MATERIALS AND METHODS

Construction of plasmids

The general scheme for the construction of plasmids containing the rat grp78 promoter subfragments subcloned into CAT vector or synthetic ERSE oligomers subcloned into pMCAT with the minimal MMTV promoter driving the expression of the CAT reporter gene has been described (13). For the latter, oligomers corresponding to the wild-type or mutated sequence shown in Figure 3 were synthesized, annealed and ligated. Only the

oligomers linked in the sense–antisense orientation had *Hin*dIII sites at the termini and were inserted into the *Hin*dIII site of the pMCAT vector. The wild-type ERSE linked in the sense–sense orientation was constructed by PCR fill-in of overlapping oligomers, followed by cloning into the TA vector (Invitrogen, Carlsbad, CA) prior to insertion into the pMCAT vector. The sequence and orientation of all the promoter constructs were confirmed by DNA sequencing.

Cell culture and transfection conditions

The reporter gene constructs were transiently transfected into K12 cells as previously described (21). Briefly, 5 µg of the reporter plasmid was co-transfected with 2.5 µg of PCH110, an expression vector for β-galactosidase, along with 3 µg of HeLa genomic carrier DNA by the calcium phosphate precipitation method. Twenty-four hours after transfection, the cells were refed and treated with either 300 nM thapsigargin or 1.5 µg/ml tunicamycin for 16 h. The CAT assays were performed as described (21) and cell extracts corresponding to equal β-galactosidase units were used. The thin layer chromatography plates were quantitated with a PhosphorImager system (Molecular Dynamics, Sunnyvale, CA). Each experiment was repeated two to six times.

EMSAs

The sequence of the wild-type and mutated synthetic ERSE oligomers used in the gel shift assays are shown in Figures 3 and 4. The oligonucleotide strands were purified, annealed and labeled as described (23). Binding reactions were carried out using 4.5 µg of HeLa nuclear extract prepared as described (31) mixed with 100 ng of poly(dI·dC). After 10 min preincubation with poly(dI·dC), 1 ng of radiolabeled probe was added followed by 15 min incubation before loading the binding reaction onto 5% non-denaturing polyacrylamide gels. The gels were run at 190 V for 2 h as described (23). In case of competition, probe and competitor were added at the same time in the concentrations indicated followed by 15 min incubation. The rabbit polyclonal YY1 antibody used for the EMSAs was purchased from Santa Cruz Biotechnology and the rabbit polyclonal antibody against CBF was raised against the CBF A subunit and affinity purified.

RESULTS

Multiple copies of ERSE in mammalian grp promoters

Upon examination of the promoter sequences of ER stress genes grp78, grp94 and ERp72 from various species ranging from *Caenorhabditis elegans* to human, a sequence motif which occurs repetitively within the promoters of these genes was discovered (Fig. 1). Importantly, these sequences occupy regions of the grp promoters previously shown to be critical for the ER stress induction of these genes (13,23,32). The sequence motif, referred to as ERSE (ER stress element), consists of a 19 bp unit with a CCAAT element separated by exactly 9 bp from a CCACG or similar sequence motif (Fig. 2). For example, in the best characterized rat grp78 promoter, three ERSEs could be located. The most proximal ERSE occurs at ~90 bp upstream of the TATA element, with the first C base of the CCAAT motif located at –98. With the discovery of the common ERSE motif, this regulatory element, previously characterized as 78C1 (23), is now renamed as ERSE–98. Similarly, the second and third ERSE elements previously referred to C3 and core (13) are referred to below as
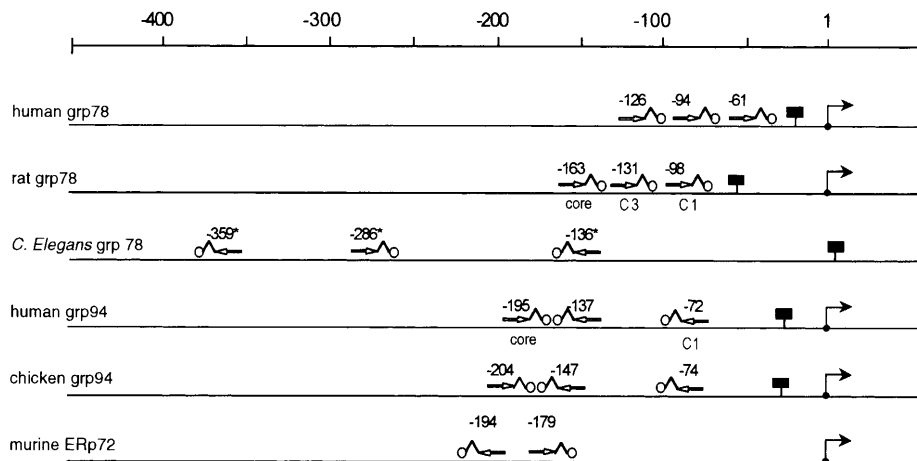
Figure 1. Schematic representation of the ERSE elements in promoters of ER genes. The occurrence of ERSE elements comprising a CCAAT (→) and a CCACG (○) element separated by a 9 bp (∧) repeat on the promoters of the grp78, grp94 and ERp72 genes are shown. The TATA boxes are indicated by ■ and the transcriptional initiation site by ↱. The ERSE elements are numbered according to the base count at the first C in the CCAAT element with the transcription initiation site set at +1, except in the case of the *C.elegans* ERSE where the elements are numbered with the putative TATA box set at 1 since the exact start of transcription is not known. A reversed orientation of the ERSE element denotes that the element is present on the non-coding strand.

```
                            1 2 3 4 5 6 7 8 9
human grp78
  ERSE-61       T C A | C C A A T | C G G C G G C C T | C C A C G | A C
  ERSE-94       G G G | C C A A T | G A A C G G C C T | C C A A C | G A
  ERSE-126      C T T | C G A A T | C G G C G G C G G | C C A G C | T T
rat grp78
  ERSE-98       T C A | C C A A T | C G G A G G C C T | C C A C G | A C
  ERSE-131      G A A | C C A A C | C A G C G G C C T | C C A A C | G A
  ERSE-163      C T T | C G A A T | C G G C A G C G G | C C A G C | T T
C.elegans grp78
  ERSE-136(L)   G C C | C G A A T | G C T C A A G G T | C G A C T | G T
  ERSE-286      T G T | C C T A T | C G T C C T A G G | C C A C G | T C
  ERSE-359(L)   C G G | C C A A T | C G G C G A C G G | C C G T G | T G
human grp94
  ERSE-72(L)    C C A | C C A A T | C G C G C C G C A | C C A C G | C A
  ERSE-137(L)   C C A | C C A A T | C G G A G C T G T | C C A G G | T G
  ERSE-195      T G A | C C A A T | C G G A A G G A G | C C A C G | C T
chicken grp94
  ERSE-74(L)    T G G | C C A A T | G G G A G C G C A | C C A C G | C C
  ERSE-147(L)   C T A | C C A A T | C G T G G C T T T | C C A T G | T G
  ERSE-204      C G G | C C A A T | C G A C G C C G G | C C A C G | C T
murine ERp72
  ERSE-179      C A G | C C A A T | C A G A A G G G G | G C A C C | T C
  ERSE-194(L)   C C A | C C A A T | C A C G G G C T G | C C A C T | C G

Consensus           | C C A A T | c g g c g g c g g | C C A C G |

A  0 2 8 0 0 16 17 0 0 4 2 5 4 2 1 1 2 0 0 16 2 0 2 3
C  9 6 1 17 14 0 0 1 1 41 2 9 2 5 9 6 0 16 16 0 105 5 5
G  3 6 5 0 3 0 0 0 3 12 10 3 11 9 5 8 8 1 1 1 3 10 3 4
T  5 3 3 0 0 1 0 16 0 0 3 0 0 1 2 2 7 0 0 0 2 2 7 5
```

Figure 2. Sequence and spatial conservation of ERSEs among ER gene promoters. Sequence alignment of putative ERSEs from the promoter regions of the ER protein genes human grp78 (40), rat grp78 (13), *C.elegans* grp78 (44), human grp94 (20), chicken grp94 (45) and murine ERp72 (32) shown in Figure 1. The conserved CCAAT and CCACG motifs, separated by precisely 9 bp, are boxed. The consensus ERSE and the frequency of occurrence of A, T, G and C at each position are shown.

ERSE–131 and ERSE–163, respectively. In each case, the numerical number denotes the location of the first C residue of the CCAAT or CCAAT-like motifs. Comparison of the organization of the ERSEs on the various grp promoters reveals that: (i) each promoter contains multiple copies of the ERSE; (ii) the spacing between individual ERSE units varies among the different grp promoters; (iii) the orientation of individual ERSE units varies such that both sense and antisense orientations are represented.

Unique features of the mammalian ERSE

To characterize further this conserved sequence motif, the nucleotide sequence of the individual ERSE from the promoters of the human, rat and *C.elegans* grp78 gene, the human and chicken grp94 gene and from the murine ERp72 gene were aligned (Fig. 2). A consensus for the ERSE emerges with the basic unit consisting of a tripartite structure with a 5 bp CCAAT sequence separated by an exactly 9 bp region from a 5 bp CCACG sequence. Closer examination of the shared ERSE reveals several unusual features within its tripartite structure. First, there is an exceptionally strong conservation of the CCAAT motif followed by a C residue. However, it is notable that the third ERSE of both the human and rat grp78 promoters (ERSE–126 and ERSE–163, respectively, previously referred to as the grp core), contain a C<u>G</u>AAT sequence motif instead. This same base change was also detected in ERSE–136 of the grp78 promoter of *C.elegans*. Second, the 9 bp sequence is far from random; rather, it is strikingly GC-rich. In fact, the consensus ERSE contains three tandem CGG sequences. Third, within the CCACG sequence, the CCA motif is strongly conserved but there are variations with the last two bases. For example, while the first ERSEs of both human and rat grp78 (ERSE–61 and ERSE–98, respectively) contain the consensus CCACG, their second ERSEs (ERSE–94 and ERSE–131, respectively) both contain a CCAAC motif, and their third ERSEs (ERSE–126 and ERSE–163, respectively) both contain a CCAGC motif. The conservation of these variations between the human and rat promoters could imply subtle functional differences among the ERSEs in response to different ER stress stimuli.

Mutational analysis of the ERSE

As shown previously for the rat grp78 promoter (21), ER stress inducibility does not require all three intact ERSEs, since deletion to –130 which only contains two ERSEs can still respond to the ER stress inducers Tg and tunicamycin; however, deletion to –104 with one intact ERSE (ERSE–98) resulted in near complete
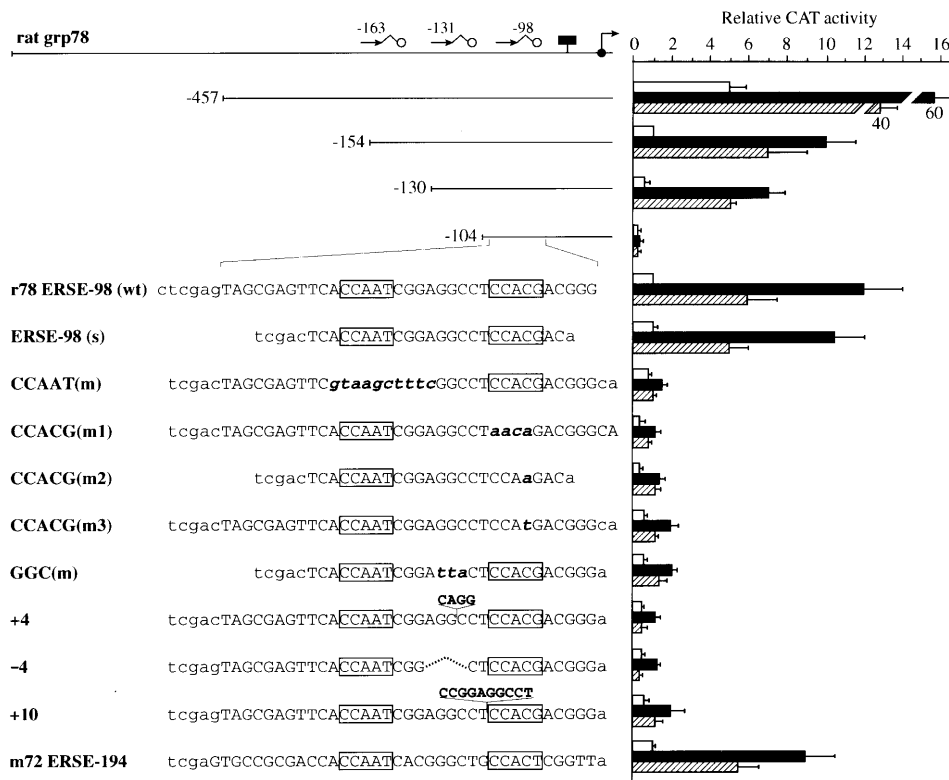
Figure 3. Stress inducibility of grp78–CAT or ERSE–pMCAT constructs. Effects of 5′ deletion of the rat grp78 promoter and of sequence mutation or disruption of the ERSE on the CAT promoter activity are summarized. The sequence of the synthetic oligomers corresponding to the wild-type (wt), truncated (s) and mutated (m) forms of rat grp78 ERSE–98 and that of the murine ERp72 ERSE–194 are shown. The conserved CCAAT and CCACG motifs are boxed. The mutated bases are in italics and are highlighted. The linker sequences added onto the 5′- or 3′-termini of the oligomers for cloning purposes are indicated by lower case alone. These synthetic oligomers were ligated in duplicate copies and subcloned into the minimal MMTV promoter driving the expression of a CAT reporter gene. For the 5′ deletion mutants of the native grp78 promoter, the basal CAT activity of the –154CAT construct was set at 1.0. For the ERSE constructs, the basal CAT activity of the wild-type ERSE was set at 1.0. The relative CAT activities are shown: open bar, control cells; black bar, cells treated with Tg; striped bar, tunicamycin. The standard deviations are indicated.

loss of ER stress inducibility (Fig. 3). Therefore, with the native grp78 promoter, the ER stress response requires at least two ERSE units, although they may not have identical ERSE sequence. Consistent with this, the promoter activity driven by one copy of the synthetic ERSE was minimally inducible by ER stress whereas the promoter activity driven by two copies of the synthetic ERSE–98 was elevated by 12- and 6-fold when the cells were treated with Tg and tunicamycin, respectively (Fig. 3 and data not shown). We further determined that the enhancing activity of ERSE–98 is independent of the orientation of the individual ERSE with respect to the TATA element since the stimulation conferred by the duplicate copies of ERSE–98 ligated in a sense–antisense orientation is identical to that of the sense–sense orientation (data not shown). In addition, two copies of ERSE–98, a 5′- and 3′-truncated version of ERSE–98, are fully stress-inducible (Fig. 3). Thus, all the sequence information required for induction by Tg or tunicamycin is contained within the 19 bp conserved sequence flanked by 3 and 2 bp on either end.

To test the requirement for the sequence integrity of the ERSE, selected bases within the tripartite structure of ERSE–98 were targeted for mutagenesis, the mutated forms were ligated and inserted into pMCAT vectors identically to the wild-type and the promoter activities were tested in both control and stressed cells.

The results are summarized in Figure 3. Mutation of the CCAAT motif and its immediate flanking sequence, as well as any mutation affecting the CCACG motif, greatly reduced Tg and tunicamycin inducibility. This includes mutation CCACG(m1), which changed the motif to AACAG, and the CCACG(m2) and CCACG(m3) mutations, which created only single base mutations to CCAAG and CCATG, respectively. Additionally, mutation of the GGC motif within the spacer region to TTA also resulted in substantial loss of stress inducibility. Collectively, these results show stringent sequence dependence for ERSE function not only within the CCAAT and CCACG motif, but also within the GC-rich 9 bp region.

The exact 9 bp spacing between the CCAAT and the CCACG sequence motif suggests that steric configuration within the ERSE may be a critical component for its function. To test this, 4 bp was either inserted (+4) or deleted (–4) within the spacer region (Fig. 3). Since a helical turn is 10.4 bp, these manipulations should disrupt the relative phasing of the transcription factors on the DNA helix. In both the +4 and –4 mutants, the GGC sequence motif was recreated within the spacer region and the CCAAT and CCACG sequences were kept intact. Despite the sequence integrity of these three elements, both the +4 and –4 mutants showed near complete loss of stress inducibility. We next tested
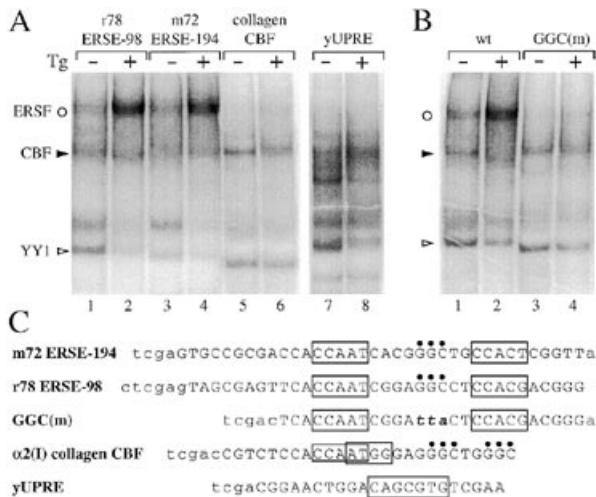
Figure 4. Stress-induced enhanced binding of ERSF to ERSE. (**A**) The EMSAs were performed with HeLa nuclear extracts prepared from either control (–) or Tg-treated cells (+), with different radiolabeled probes as indicated on top: lanes 1 and 2, rat grp78 ERSE–98; lanes 3 and 4, murine ERp72 ERSE–194; lanes 5 and 6, α2(I) collagen CBF binding site; lanes 7 and 8, yeast UPRE. The ERSF, CBF/NF-Y and YY1 complexes formed (lane 1) are indicated by an open circle and a closed and open arrow, respectively. (**B**) Effect of mutation of the GGC motif within the spacer region on factor binding. The probes used were: lanes 1 and 2, wild-type ERSE–98; lanes 3 and 4, GGC(m) oligomer. (**C**) Sequence alignment of the various oligomers used as probe. The core sequence motifs are boxed and the GGC sequence motifs are indicated.

the effect of inserting an extra 10 bp of spacer sequence into the ERSE by creating the +10 mutant (Fig. 3). This mutation should retain putative factor binding sites and maintain the same phasing but would double the distance between the CCAAT and CCACG motifs. Interestingly, while the effect of the +10 mutation was less severe than +4 or –4, inducibility by both Tg and tunicamycin was substantially reduced.

Finally, to examine the general applicability of ERSE function in promoters other than grp78, we tested the ability of the ERSE–194 located within the murine ERp72 promoter to serve as an ER stress-inducible element when linked to a heterologous promoter. As shown in Figure 3, the ERp72 element, in duplicate copies, is able to function as an ERSE by conferring a 9-fold induction to the MMTV promoter in Tg-treated cells and a 7-fold induction in tunicamycin-treated cells.

Stress-inducible complex interaction with the ERSE

Through binding site competitions and immuno-crossreactivity, we previously identified two major factors from HeLa nuclear extract that bind to rat grp78 ERSE–98 in electrophoretic mobility shift assays (EMSAs). The slower migrating complex contains the multimeric CCAAT binding factor CBF/NF-Y binding to the CCAAT element and its 3′ flanking sequence (23) and the faster migrating complex contains YY1 binding to the CCACG sequence (24). However, no major difference of complex binding to ERSE–98 was detected between nuclear extract prepared from control and stressed cells.

Through extensive re-investigation of the binding conditions used in EMSAs, we discovered that use of high concentrations of poly(dI·dC) as a competitor in previous assays, while helpful in eliminating non-specific binding, could have precluded

complexes binding to the ERSE through GC-rich sequence. In support, under the new conditions using a higher concentration of nuclear extract and minimal amount of poly(dI·dC), while CBF/NF-Y binding was weaker, enhanced binding of a novel complex, referred to as ERSF, was evident in the nuclear extract prepared from Tg-stressed cells (Fig. 4A). The binding of ERSF strongly prefers the ERSE since the stress-inducible complex was readily formed when r78ERSE–98 or m72ERSE–194 was used as probe (lanes 1–4), but not with the yeast UPRE sequence as probe (lanes 7 and 8). Interestingly, the α2(I) collagen CBF binding site contains an overlapping CBF/NF-Y and YY1 binding site within the CCAATGG motif (24) and is therefore capable of forming the CBF/NF-Y and the YY1 complexes as in the case of ERSE–98 and ERSE–194 (Fig. 4A, lanes 5 and 6). In addition, the sequence 3′ to its CCAAT motif is similar to that of ERSE–98, including the existence of GGC motifs (Fig. 4C). Nonetheless, despite these extensive similarities with ERSE–98, ERSF binding to the collagen CBF binding site is minimal.

To map the ERSF binding site within the ERSE and to verify the effect of base mutations on CBF/NF-Y and YY1 binding, we tested each of the mutated oligomers as radiolabeled probes for their ability to form the ERSF, CBF/NF-Y and YY1 complexes with nuclear extracts prepared from control and Tg-stressed cells. The results are summarized in Figure 5C. With CCAAT(m), where the mutation extends partially into the spacer region, the formation of the ERSF complex was slightly reduced. However, other mutations, such as CCACG(m1), CCACG(m2), CCACG(m3), +4, –4 and +10, were all able to form the ERSF complex (Fig. 5C). These results were independently confirmed by the ability of molar excess of the mutated oligomers to compete for the formation of ERSF with the wild-type ERSE–98 as probe in EMSAs (data not shown). Collectively, these results strongly suggest that the critical contact points for ERSF might lie within the 9 bp region. Since both the rat grp78 ERSE–98 and murine ERp72 ERSE–194 can form the ERSF complex, their 9 bp sequences were compared (Fig. 4C). Interestingly, embedded within their related but divergent sequences is a conserved GGC motif. This GGC motif was also retained in the ERSE mutants described above. Specific mutation of that GGC motif to TTA in form of synthetic oligomer or a longer 42 bp fragment greatly reduced ERSF binding while maintaining CBF/NF-Y and YY1 binding (Fig. 4B and data not shown). Therefore, stable association of the ERSF to the ERSE requires the GGC motif within the 9 bp region, correlating with the observation that mutation of the GGC motif [GGC(m)] results in substantial loss of stress inducibility (Fig. 3).

Several lines of evidence further suggest that the ERSF complex is a novel complex and may not contain CBF/NF-Y or YY1 as its component. First, specific antibodies against CBF/NF-Y and YY1 did not inhibit the formation of the ERSF complex, whereas they were effective in eliminating the respective CBF/NF-Y or YY1 complexes within the same sample (Fig. 5A). The addition of the antibody to the EMSA sometimes even resulted in slight enhancement of ERSF binding. This could be in part due to the stabilizing effect of additional protein from the serum since addition of BSA to the standard reaction mixture can cause the same effect (Fig. 5B, lanes 5 and 6). In agreement with the lack of effect of the antibodies, consensus CBF and YY1 binding sites were unable to inhibit the formation of the ERSF complex (data not shown). In addition, when either CCAAT(m) or CCACG(m1) were used as probe, ERSF was still able to form (Fig. 5B and data not shown). Finally, correlation of factor binding
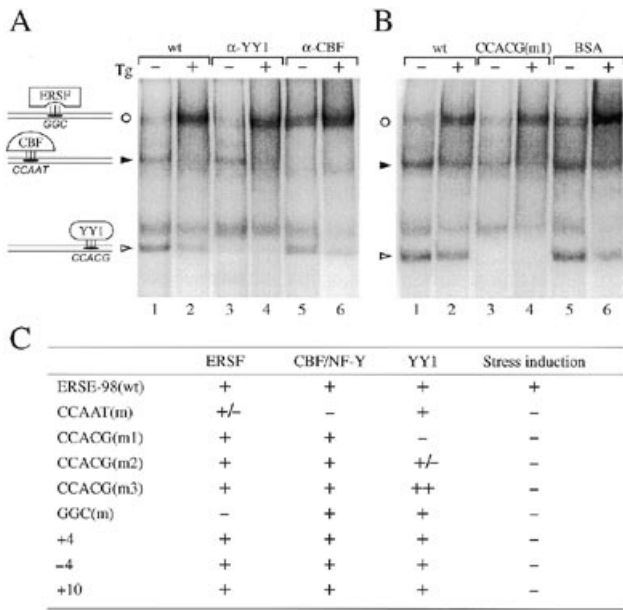
Figure 5. Relationship between factor binding and ERSE stress inducibility. (**A**) EMSAs were performed with HeLa nuclear extract from control (–) or Tg-treated cells (+), using ERSE–98 as probe. Lanes 1 and 2, no antibody was added; lanes 3 and 4, 1 µl of anti-YY1 antibody was added to the EMSA reactions; lanes 5 and 6, 1 µl of anti-CBF/NF-Y antibody was added. Schematic drawings of the critical contact bases within ERSE–98 for the ERSF, CBF/NF-Y and YY1 complexes are shown on the left. (**B**) The EMSAs were performed with the probes: wild-type ERSE–98 (lanes 1 and 2); CCACG(m1); ERSE–98 but with an additional 25 µg of BSA added in the reaction mixture. (**C**) Summary of the effect of ERSE mutations on ERSF, CBF/NF-Y and YY1 binding and stress inducibility. +, ability for factor to bind or for ERSE to confer strong stress induction; +/–, slightly negatively affected; –, loss of ability to bind factor or the stress inducibility is substantially reduced; ++, factor binds with stronger affinity.

with stress inducibility reveals that ERSF binding to the ERSE is not sufficient; full stress inducibility requires sequence integrity of the CCAAT, GGC and CCACG motifs representing CBF, ERSF and YY1 binding sites, as well as precise spacing among these sites within the tripartite structure of the ERSE (Fig. 5C).

DISCUSSION

Mammalian cells have the ability to induce the transcription of a group of genes encoding for ER proteins when the ER calcium store is depleted or when there is accumulation of malfolded proteins in the ER. In this report, we deciphered the genetic code mediating this ER stress signaling pathway, utilizing the grp genes as a model system. Our studies reveal that the consensus mammalian ERSE contains an unusual tripartite structure of CCAAT-(N9)CCACG. This basic ERSE sequence motif has recently been independently derived and appears to be evolutionarily conserved even among lower eukaryotes and plants (33). We further discovered that the 9 bp sequence strikingly rich in GC is an integral part of the tripartite structure needed for ERSE function as well as binding for a novel stress-inducible complex, not merely a spacer region with random sequence. Whereas one copy of the ERSE is insufficient to confer ER stress inducibility, tandem copies of ERSE are fully functional and can act in an orientation-independent manner. As the discovery of the heat

shock response element has made a tremendous impact in understanding the phenomenon of heat shock-mediated gene induction (34), the discovery of the novel features of the mammalian ERSE will allow rapid identification of other ER stress-responsive genes. It will also facilitate understanding the molecular control mechanisms of ER genes when the cells are subjected to ER stress.

What are the functional implications of the mammalian ERSE? The tripartite structure of the mammalian ERSE as opposed to the simple palindromic sequence which mediates the yeast unfolded protein response suggests that induction of the mammalian ER stress genes probably requires multiple regulatory proteins acting in synergy, as in the case of regulation of cholesterogenic genes by sterol mediated by CBF/NF-Y and SREBP (35). The strict requirement of precise spacing within the ERSE further implies that steric configurations of the interacting factors are critical to ERSE function. The unusually strong conservation of the CCAAT motif within the ERSE indicates the functional importance of the factor that occupies this site. For the grp78, grp94 and ERp72 promoters, this CCAAT binding factor has been identified as CBF/NF-Y (17,23,36). Since the mammalian cells contain multiple CCAAT binding proteins, the strong conservation of the bases flanking the CCAAT element is likely to help confer CBF/NF-Y binding specificity (22). The functional requirement of CBF/NF-Y as an activator of rat grp78 ERSE–98 has been established both *in vitro* (24) and *in vivo*, through the use of a dominant negative mutant (25). Within the ERSE, a high affinity CBF/NF-Y site is advantageous for mediating the ER stress response as conversion of the divergent CGAAT motif found in rat grp78 ERSE–163 to CCAAT increased its ER stress responsiveness (data not shown). How might CBF/NF-Y contribute to the transcriptional activation of ERSE? It is interesting to note that both the binding and transcriptional activation properties of CBF/NF-Y can be modulated by calcium (23,24). In addition to its ability to recruit and act synergistically with other transcription factors on the promoter (35,37), CBF/NF-Y has been shown to bind both the major and minor grooves of DNA and is able to bend DNA, particularly for CCAAT sequences embedded in GC-rich flanking sequences, as in the case of most ERSEs (22,38).

For the CCACG sequence motif, one binding factor to this site has been identified as YY1 (24). In contrast to the highly conserved CCAAT site, this sequence among the ERSEs exhibits some variations and the binding of YY1 to this particular sequence motif has been relatively weak compared with CCAGC in rat grp78 ERSE–163 (29). These subtle variations in the CCACG sequence motif could result in different relative contributions of the respective factors in stress response mediated by different ERSEs. Thus, while over-expression of YY1 stimulated ERSE–163-mediated ER stress induction in NIH 3T3 cells (29), the effect has not been observed for ERSE–98 (data not shown). In *in vitro* transcription assays, YY1 is not rate-limiting for ERSE–98-mediated transcription activation (24). Our mutational analysis here further revealed that any manipulation of the CCACG sequence, resulting in either eliminating [CCACG(m1)], weakening [CCACG(m2)] or strengthening [CCACG(m3)] of YY1 binding, all resulted in loss of stress inducibility (Fig. 5C). These results suggest that the binding of YY1 to this site could be fortuitous and not functionally relevant, opening up the possibility that another factor yet undefined is the functional interactive partner of this sequence. Alternatively, since YY1 has been shown to act both as a transcriptional activator and a repressor (26,28), any alteration of

its intermediate binding affinity to the ERSE target site could result in change of its regulatory property. Further, YY1 could act on the chromosomal configuration of the ERSE with its DNA bending properties (27,39), and such effects are difficult to detect in the assays used. Future investigations utilizing *in vivo* model systems with a loss of YY1 function will address this important issue.

A major new finding of this study is the discovery of the stress-inducible complex binding to the GGC motif within the 9 bp sequence of ERSE. In support of the functional importance of the ERSF binding site, mutation of the same GGC and its immediate flanking sequence in a ERp72 promoter subfragment also resulted in substantial loss of its inducibility by accumulation of μ heavy chain in the ER (36). Further, this same ERSF complex also exhibits affinity for ERSE–163 (data not shown). *In vivo* footprinting of the native grp78 promoter indicates that the identical GG residues within ERSE–163 exhibits specific stress-inducible changes (40), suggesting either a change in factor occupancy or conformation after ER stress. Currently, it is not known whether the binding activity changes of ERSF we observed in the EMSA are regulated at the transcriptional, translational or post-translational level in the stressed cell nuclei, nor could we rule out the possibility that the low but detectable ERSF binding activity in the control cells originated from slight contamination with stressed cells under tissue culture conditions. ERSF could be a single polypeptide which is not yet identified or a multiprotein complex. If it is a protein complex pre-formed in nuclei from stressed cells, our data suggest that it is at least in a form not recognizable by specific antibodies to the CBF-A subunit or YY1 nor does it exhibit affinity for the consensus CBF/NF-Y or YY1 binding sites. The binding property of ERSF to ERSE suggests that it could be a potential processing target for the newly discovered mammalian Ire1p (41,42). It could also be related to an A23187-inducible complex binding to the human grp78 promoter (43) or contain as a component ATF-6, which has recently been shown to undergo ER stress-induced proteolysis but is not able to bind ERSE on its own (33). Although ERSF binds to a GC-rich sequence resembling that of an AP-2 site, recombinant AP-2 does not bind ERSE–98 (data not shown). While the identity of ERSF remains to be determined, its enhanced binding to the ERSE could be observed in cells stressed by ER calcium depletion, as well as in cells where there is blockage in protein glycosylation (data not shown). Therefore, it is likely to be a common effector of the ER stress response and could be a functional counterpart of the yeast Hac1 protein which has yet to be identified. The discovery of the ERSF binding site within the ERSE will greatly facilitate its isolation and characterization.

REFERENCES

1 Connolly,L., De Las Penas,A., Alba,B.M. and Gross,C.A. (1997) *Genes Dev.*, **11**, 2012–2021.
2 Cox,J.S., Shamu,C.E. and Walter,P. (1993) *Cell*, **73**, 1197–1206.
3 Kawahara,T., Yanagi,H., Yura,T. and Mori,K. (1997) *Mol. Biol. Cell*, **8**, 1845–1862.
4 Mori,K., Sant,A., Kohno,K., Normington,K., Gething,M.J. and Sambrook,J.F. (1992) *EMBO J.*, **11**, 2583–2593.
5 Pahl,H.L. and Baeuerle,P.A. (1995) *EMBO J.*, **14**, 2580–2588.
6 Wang,X., Sato,R., Brown,M.S., Hua,X. and Goldstein,J.L. (1994) *Cell*, **77**, 53–62.
7 Lee,A.S. (1987) *Trends Biochem. Sci.*, **12**, 20–23.
8 Munro,S. and Pelham,H.R. (1986) *Cell*, **46**, 291–300.
9 Mazzarella,R.A. and Green,M. (1987) *J. Biol. Chem.*, **262**, 8875–8883.
10 Ramakrishnan,M., Schönthal,A.H. and Lee,A.S. (1997) *J. Cell. Physiol.*, **170**, 115–129.
11 Mazzarella,R.A., Srinivasan,M., Haugejorden,S.M. and Green,M. (1990) *J. Biol. Chem.*, **265**, 1094–1101.
12 Thastrup,O., Cullen,P.J., Drobak,B.K., Hanley,M.R. and Dawson,A.P. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 2466–2470.
13 Li,W.W., Alexandre,S., Cao,X. and Lee,A.S. (1993) *J. Biol. Chem.*, **268**, 12003–12009.
14 Drummond,I.A., Lee,A.S., Resendez,E.,Jr and Steinhardt,R.A. (1987) *J. Biol. Chem.*, **262**, 12801–12805.
15 Liu,E.S., Ou,J.H. and Lee,A.S. (1992) *J. Biol. Chem.*, **267**, 7128–7133.
16 Kim,Y.K., Kim,K.S. and Lee,A.S. (1987) *J. Cell. Physiol.*, **133**, 553–559.
17 Ramakrishnan,M., Tugizov,S., Pereira,L. and Lee,A.S. (1995) *DNA Cell Biol.*, **14**, 373–384.
18 Watowich,S.S., Morimoto,R.I. and Lamb,R.A. (1991) *J. Virol.*, **65**, 3590–3597.
19 Watowich,S.S. and Morimoto,R.I. (1988) *Mol. Cell. Biol.*, **8**, 393–405.
20 Chang,S.C., Erwin,A.E. and Lee,A.S. (1989) *Mol. Cell. Biol.*, **9**, 2153–2162.
21 Wooden,S.K., Li,L.J., Navarro,D., Qadri,I., Pereira,L. and Lee,A.S. (1991) *Mol. Cell. Biol.*, **11**, 5612–5623.
22 Maity,S.N. and de Crombrugghe,B. (1998) *Trends Biochem. Sci.*, **23**, 174–178.
23 Roy,B. and Lee,A.S. (1995) *Mol. Cell. Biol.*, **15**, 2263–2274.
24 Roy,B., Li,W.W. and Lee,A.S. (1996) *J. Biol. Chem.*, **271**, 28995–29002.
25 Zhou,Y. and Lee,A.S. (1998) *J. Natl Cancer Inst.*, **90**, 381–388.
26 Shi,Y., Seto,E., Chang,L.S. and Shenk,T. (1991) *Cell*, **67**, 377–388.
27 Natesan,S. and Gilman,M.Z. (1993) *Genes Dev.*, **7**, 2497–2509.
28 Shrivastava,A. and Calame,K. (1994) *Nucleic Acids Res.*, **22**, 5151–5155.
29 Li,W.W., Hsiung,Y., Zhou,Y., Roy,B. and Lee,A.S. (1997) *Mol. Cell. Biol.*, **17**, 54–60.
30 Li,W.W., Hsiung,Y., Wong,V., Galvin,K., Zhou,Y., Shi,Y. and Lee,A.S. (1997) *Mol. Cell. Biol.*, **17**, 61–68.
31 Shapiro,D.J., Sharp,P.A., Wahli,W.W. and Keller,M.J. (1988) *DNA*, **7**, 47–55.
32 Srinivasan,M., Lenny,N. and Green,M. (1993) *DNA Cell Biol.*, **12**, 807–822.
33 Yoshida,H., Haze,K., Yanagi,H., Yura,T. and Mori,K. (1998) *J. Biol. Chem.*, **273**, 33741–33749.
34 Lindquist,S. and Craig,E.A. (1988) *Annu. Rev. Genet.*, **22**, 631–677.
35 Jackson,S.M., Ericsson,J., Mantovani,R. and Edwards,P.A. (1998) *J. Lipid Res.*, **39**, 1712–1721.
36 Marcus,N. and Green,M. (1997) *DNA Cell Biol.*, **16**, 1123–1131.
37 Wright,K.L., Vilen,B.J., Itoh-Lindstrom,Y., Moore,T.L., Li,G., Criscitiello,M., Cogswell,P., Clarke,J.B. and Ting,J.P. (1994) *EMBO J.*, **13**, 4042–4053.
38 Ronchi,A., Bellorini,M., Mongelli,N. and Mantovani,R. (1995) *Nucleic Acids Res.*, **23**, 4565–4572.
39 Kim,J. and Shapiro,D.J. (1996) *Nucleic Acids Res.*, **24**, 4341–4348.
40 Li,W.W., Sistonen,L., Morimoto,R.I. and Lee,A.S. (1994) *Mol. Cell. Biol.*, **14**, 5533–5546.
41 Tirasophon,W., Welihinda,A.A. and Kaufman,R.J. (1998) *Genes Dev.*, **12**, 1812–1824.
42 Wang,X.Z., Harding,H.P., Zhang,Y., Jolicoeur,E.M., Kuroda,M. and Ron,D. (1998) *EMBO J.*, **17**, 5708–5717.
43 Chao,C.C.-K. and Lin-Chao,S. (1992) *Nucleic Acids Res.*, **20**, 6481–6485.
44 Heschl,M.F. and Baillie,D.L. (1990) *J. Mol. Evol.*, **31**, 3–9.
45 Kleinsek,D.A., Beattie,W.G., Tsai,M.J. and O'Malley,B.W. (1986) *Nucleic Acids Res.*, **14**, 10053–10069.