ELSEVIER

Gene 285 (2002) 127–139

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Genomic organization and promoter characterization of the gene encoding a putative endoplasmic reticulum chaperone, *ERp29*

Ernest Sargsyan[a], Mikhail Baryshev[a,b], Maria Backlund[a], Anatoly Sharipo[b], Souren Mkrtchian[a,*]

[a]*Division of Molecular Toxicology, Institute of Environmental Medicine, Karolinska Institute, 171 77 Stockholm, Sweden*
[b]*Biomedical Research and Study Center, University of Latvia, LV-1067, Riga, Latvia*

Received 21 September 2001; received in revised form 27 September 2001; accepted 15 January 2002
Received by R. Di Lauro

Abstract

ERp29 is a soluble protein localized in the endoplasmic reticulum (ER) of eukaryotic cells, which is conserved in all mammalian species. The N-terminal domain of ERp29 displays sequence and structural similarity to the protein disulfide isomerase despite the lack of the characteristic double cysteine motif. Although the exact function of ERp29 is not yet known, it was hypothesized that it may facilitate folding and/or export of secretory proteins in/from the ER. ERp29 is induced by ER stress, i.e. accumulation of unfolded proteins in the ER. To gain an insight into the mechanisms regulating *ERp29* expression we have cloned and characterized the rat *ERp29* gene and studied in details its distribution in human tissues. Comparison with the murine and human genes and phylogenetic analysis demonstrated common origin and close ortholog relationships of these genes. Additionally, we have cloned ~3 kb of the 5′-flanking region of the *ERp29* gene and functionally characterized its promoter. Such characteristics of the promoter as GC-rich sequence, absence of TATA-box, multiple transcription start sites taken together with the ubiquitous gene expression, reaching maximum levels in the specialized secretory tissues, indicate that *ERp29* belongs to the group of the constitutively expressed housekeeping genes. A 337 bp fragment of the 5′ flank was identified as a core promoter sufficient for the transcriptional activation of the gene. Gel mobility shift assay indicated interaction of the predicted GC and E box elements within the core promoter with Sp1/Sp3 and USF1/USF2 transcription factors, respectively, suggesting their key role in the basal expression of the gene.

Keywords: Chaperone; Housekeeping gene; Promoter; Transcriptional regulation

1. Introduction

ERp29 is a 24.5 kDa soluble endoplasmic reticulum (ER) protein postulated to function within a network of molecular chaperones and foldases that facilitate folding, posttranslational modifications and transport of the secretory proteins. ERp29 DNA complementary to RNA (cDNA) was cloned from the rat liver (Fang et al., 1997; Mkrtchian et al., 1998) and enamel cells (Demmer et al., 1997) and gene expression was demonstrated in all studied mammalian cells (Mkrtchian et al., 1998; Hubbard et al., 2000). The cDNA of the protein termed ERp28, which is 87% identical to the cDNA of ERp29 was cloned from the human liver (Ferrari et al., 1998).

The ERp29 polypeptide consists of two domains with the N-terminal domain resembling the ***a*** domain of human protein disulfide isomerase (PDI), similar thioredoxin domains of the P5-like (or PDI-Dα) PDIs of plant and amoebal origin and Windbeutel, ERp29-like protein from *Drosophila melanogaster*. Sequence similarity is supported by the structural homology demonstrated by NMR spectroscopy of the N-terminal domain of ERp29, which adopts a typical thioredoxin fold characteristic for the proteins with the redox function (Liepinsh et al., 2001). However, the active double cysteine motif of PDI is absent in ERp29, apparently ruling out the PDI-like redox function. The C-terminal domain represents a novel all-helical fold, which is absent in human PDI but found in the P5-like PDIs and also in Windbeutel. The latter may provide interesting insights into the physiological function of ERp29 as it was suggested to play a chaperone-like role in the ovarian follicle cells, facilitating the specific Golgi targeting of Pipe, oligosac-

Abbreviations: A, adenosine; bp, base pairs; cDNA, DNA complementary to RNA; CHO, Chinese hamster ovary cells; PDI, protein disulfide isomerase; EMSA, electro mobility shift assay; ER, endoplasmic reticulum; kb, kilobase; RACE, rapid amplification of cDNA ends; SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis

* Corresponding author. Tel.: +46-8-728-7711; fax: +46-8-337-327.
E-mail address: souren.mkrtchian@imm.ki.se (S. Mkrtchian).

0378-1119/02/$ - see front matter
PII: S0378-1119(02)00417-1

charide-modifying enzyme, essential for the embryonic development (Konsolaki and Schupbach, 1998; Sen et al., 2000).

ERp29 is induced under the conditions collectively termed as ER stress, i.e. accumulation of unfolded proteins in the ER (Mkrtchian et al., 1998). Recently the transcriptional activation of *ERp29* gene was demonstrated in rat thyrocytes upon the induction by the thyroid-stimulating hormone (TSH) (Kwon et al., 2000), which implicitly indicates ERp29 involvement in the thyroglobulin (Tg) maturation and/or secretion. However, genetic elements responsible for the basal and inducible expression of *ERp29* are currently unknown. To facilitate the identification of *cis*- and *trans*-regulatory factors important for the transcriptional regulation of *ERp29* we have cloned and characterized the rat *ERp29* gene including ~3 kb of the 5′-flanking region. GC-rich 5′ flank and the absence of TATA box in combination with the multiple transcription start sites and ubiquitous gene expression demonstrated in this study are indicative of a typical housekeeping gene properties of *ERp29*. Functional study of the 5′-flanking region identified a core promoter region and Sp1/Sp3 and USF transcription factors as key regulators of the *ERp29* gene expression.

2. Materials and methods

2.1. Genomic library screening

To isolate the *ERp29* gene, P1 rat genomic library was screened in the contract study with *Genome Systems* (St. Louis, MO) using a pair of ERp29 cDNA-specific polymerase chain reaction (PCR) primers (5′-ATCGAGGCCTCCAGCAGAGA, forward and 5′-CAAGGATCCTCACAGCTCCTCCTTCTC, reverse) that have been previously shown to amplify a 277 bp band from the rat liver genomic DNA. The presence of ERp29-specific DNA in the genomic clone was confirmed by PCR using the same pair of primers and Southern blotting (results not shown). P1 recombinant plasmid (pAd10SacBII) was isolated as to *Genome Systems* recommendations. The plasmid was digested with *Hin*dIII and subjected to Southern blot using as a probe 252 bp fragment of the ERp29 cDNA, encompassing translation start of the gene. A 4 kb fragment, which hybridized with the probe, was isolated and subcloned into pUC18 vector to facilitate sequencing of the 5′-flanking region of the gene.

2.2. DNA sequencing

Sequencing was performed using BigDye terminator cycle sequencing kit (PE Biosystems) and ABI Prism 377 sequencer. The sequence of the introns was determined by direct sequencing of P1 plasmid utilizing gene-walking approach whereas 5′-flanking region was sequenced using 4 kb gene fragment in pUC18 as a template. The complete sequence of the rat *ERp29* gene was deposited in GenBank with accession number AY004254.

2.3. Determination of the transcription start site

To identify the transcription start site we employed a modified procedure of the 5′-RACE technique, SMART 5′-RACE (Clontech) using the rat liver and brain poly(-A)$^+$RNA (Ambion) and gene-specific (5′-ACCAAGAGATCATCGCTGGAGGCTGAGTT) reverse primer for the first-strand cDNA synthesis. Obtained cDNA fragments were PCR amplified with the reverse gene-specific nested primer (5′-GTACTGGGTGTCGAACTTCACCAAGACGAA) and nested universal primer from the kit. PCR fragments were subsequently cloned using TOPO TA Cloning kit (Invitrogen) and sequenced.

2.4. Reporter plasmid constructs

The pGL3 expression vector system (Promega) with a firefly luciferase as a reporter gene was used to evaluate the transcriptional activity of the fragments of the 5′-flanking region of the *ERp29* gene. The 5′ deletion fragments were generated by the digestion of the pUC18 recombinant vector containing 5′-flanking segment of the *ERp29* gene ($\approx$ 3 kb) with corresponding endonucleases (Fig. 5A). These constructs share 3′ boundary at the position +1 (the A of the ATG codon) (an engineered *Bgl*II site) while 5′ boundary varies between −44 and −3040. Additionally, 3′ deletion fragments were constructed with a constant 5′ boundary at −337 (*Bse*RI site) and variable 3′ boundary (*Bss*HII, *Fsp*I, *Dra*III). All fragments were cloned into the pGL3-Basic promoterless vector and used for transfections.

2.5. Cell culture and transient transfections

FAO (a rat hepatoma cell line obtained from M. Weiss, Pasteur Institute, Paris, France) were maintained in Ham's F12 Coon's modified medium (Sigma, St. Louis, MA, USA), HeLa (human epithelial cells, American Type Culture Collection (ATCC), Manassas, VO, USA), NIH 3T3 (mouse fibroblasts, ATCC), CHO (Chinese hamster ovary cells, ATCC) were grown in Dulbeccos's modified essential medium and FRTL-5 Fisher rat thyrocytes (ATCC) were cultivated in the Coon's modified Ham's F-12 medium containing 6 hormone (6H) mixture: TSH (1 mU/ml), transferrin (5 μg/ml), somatostatin (10 ng/ml), glycyl-1-histidyl-1-lysine (10 ng/ml), hydrocortisone (10 nM), insulin (10 μg/ml). All hormones were purchased from Sigma. All culture media except Ham's F12 were obtained from Life Technologies and supplemented with 5% fetal bovine serum (FBS) and 100 U/ml penicillin and streptomycin from the same source. A total of 80–90% confluent cells were plated 24 h before transfection into 24-well tissue culture plates. Recombinant pGL3 reporter vectors (0.65 μg/well) containing the firefly luciferase gene and a pRL-TK vector (0.065 μg/well) containing *Renilla* luciferase reporter gene used as

an internal control of the transfection efficiency, were transfected into FAO cells using Lipofectin reagent (Gibco-BRL) as to manufacturer's recommendations. Firefly and *Renilla* luciferase activities were determined using Dual-Luciferase Reporter Assay System (Promega). Within each experiment transfections were done in quadruplicate or triplicate.

2.6. Electrophoretic mobility shift assay (EMSA)

Nuclear extracts from the confluent FAO cells were prepared as it is described elsewhere (Backlund et al., 1997). Electrophoretic mobility shift assays (EMSA) were carried out in a total volume of 20 μl with 0.6 ng labeled probe ($\sim 25 \times 10^4$ cpm), 3.8 μg of the nuclear extract protein in the 20 mM Hepes, 10% glycerol, 1.5 mM $MgCl_2$, 60 mM KCl, 1 mM EDTA, 0.12 mM PMSF, 1 mM DTT, and 1 μg/μl poly(dI-dC)-poly(dI-dC) (Amersham Pharmacia Biotech, Uppsala, Sweden), pH-7.9. Double-stranded probes (Fig. 8) were labeled with γ-[^{32}P]ATP using T4 polynucleotide kinase (Life Technologies). For identification of Sp1, Sp3 and USF protein/DNA complexes 1–2 μg of corresponding antibodies (Santa Cruz Biotechnology, Santa Cruz, CA) were added to the binding reaction. The DNA-protein complexes were separated on a 4% non-denaturing polyacrylamide gel, in the Tris-glycine buffer (50 mM Tris, 380 mM glycine, 2 mM EDTA) at 4°C. The dried gels were analyzed by autoradiography using BAS 1800 phosphoimager (Fuji film).

2.7. Profiling of ERp29 gene expression

A ^{32}P-labeled ERp29 probe (783 bp fragment of the ERp29 cDNA containing the entire coding sequence) was hybridized for overnight at 65°C with the RNA Master blot nylon membrane (Clontech) containing poly(A)$^+$RNA from various human tissues immobilized in separate dots. These multiple tissue expression arrays allow fast profiling of the gene expression and at the same time accurate estimation of the relative levels of mRNA abundance as poly(A)$^+$ RNA loading on the membrane was normalized using eight different housekeeping genes. Membrane was washed as to manufacturer's recommendations and analyzed by autoradiography using BAS 1800 phosphoimager (Fuji film).

3. Results and discussion

3.1. Genomic organization of the rat, mouse and human ERp29 genes

The rat *ERp29* genomic clone selected by the PCR screening of the P1 phage genomic library with *ERp29*-specific primers was used for the sequencing of the introns utilizing gene-walking approach. Sequencing of the 5′-flanking region using the same approach was unsuccessful probably due to the high GC content and complex secondary structure of the region approximately 200 bp upstream of the translation start (Fig. 4). This obstacle was circumvented by cloning the 5′-flanking region with a part of the first exon into the pUC18 vector and subsequent sequencing using vector primers. The 6.4 kb-long gene contains three relatively small exons separated by the large (4.7 kb) and small (0.38 kb) introns with the polyadenylation signal located in the 3′-flanking region 268 bp downstream of the stop codon TGA (Fig. 1). The 3′-end of the *ERp29* transcript indicating the end of the gene sequence was found immediately downstream of the polyadenylation signal (Mkrtchian et al., 1998). With the exception of the exon-intron boundary 2, all other exon-intron junctions demonstrated conserved 5′- and 3′-splice consensus sequences (Table 1) (Shapiro and Senapathy, 1987). First intron is class 0 (located between codons) and the second belongs to class 1 (interrupts first and second bases of the codon).

As it was already reported (Hubbard and McHugh, 2000), the homology search in GenBank identifies almost complete sequence of the human *ERp29* gene. Nearly identical sequence was obtained from the *Celera* human genome database (www.celera.com). The cDNA of the human gene was cloned (Ferrari et al., 1998) and the protein termed ERp28. UNIGENE database search mapped the human gene to the chromosome 12. Additionally, an intronless pseudogene was found on the chromosome 20. The homology search in the *Celera* mouse genome database identified the complete sequence of the mouse *ERp29* gene. The exon/intron structures of *ERp29* genes of all three species are identical with the nearly same lengths of all three exons and the first intron (Fig. 1). Intron 2 is substantially longer in the mouse and human genes. Conserved AATAAA polyadenylation signals were found at the same positions ~300 bp downstream of the stop codon in all three genes which taken together with the high degree of identity of all 3′-UTRs (>70%) may indicate similar pattern of regulation of the processing and stability of mRNA. Coding sequences of all *ERp29* genes are also interrupted at the identical conserved positions (Table 1). The search for the alternative splice sites did not reveal any such motifs in neither of these genes, which is consistent with the previously reported results indicating a single transcript band in the Northern blot (Mkrtchian et al., 1998).

Phylogenetic analysis of ERp29s from different mammalian species and related PDIs (Fig. 2) shows clustering of mammalian ERp29 sequences where rodent and murine species form a distinct subgroup. Windbeutel may be tentatively included in the *ERp29* family as a distant ortholog member although the comparison of the promoters did not reveal significant similarity suggesting the diverse regulation of these genes. High amino acid (>90%) sequence identity, analogous gene organization (Fig. 1) shared by rat, mouse and human *ERp29s* and phylogenetic analysis allow their classification as orthologs and confirm the strong conservation in mammals. Similar conclusions were reached recently (Hubbard and McHugh, 2000) in the

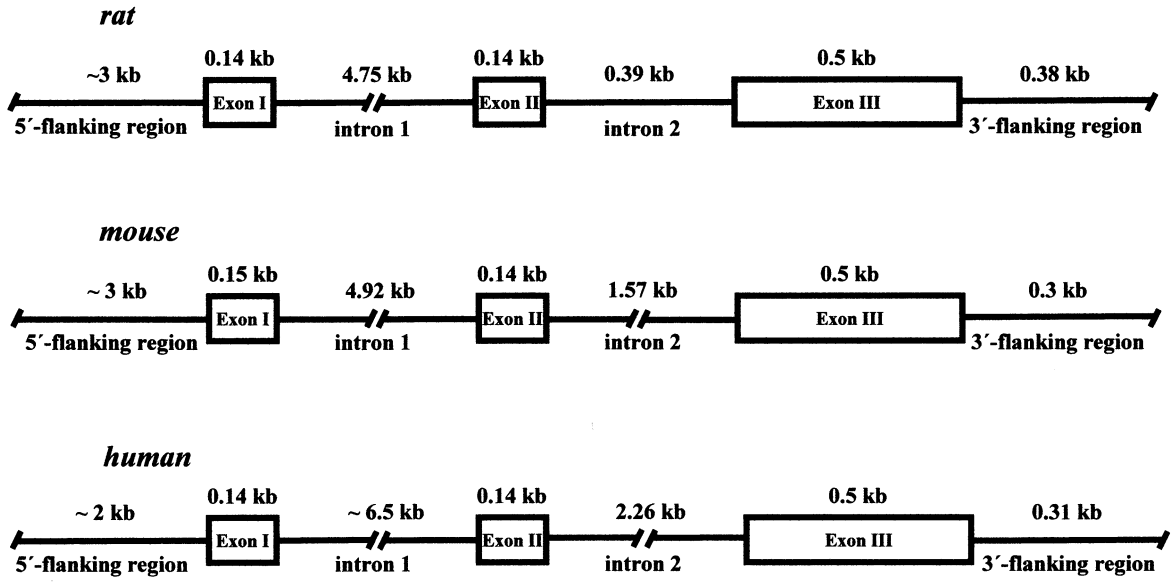

Fig. 1. Schematic representation of the rat, mouse and human *ERp29* genes. 5′-flanking region of the rat *ERp29* gene is represented by the ~3 kb fragment which was sequenced as it is described in the Section 2.2. 5′-flanking regions of the mouse and human genes are represented by the identical in length fragments derived from the gene sequences found in GenBank and *Celera* databases. 3′-UTR of the rat gene was identified earlier (Mkrtchian et al., 1998) whereas the sizes of the mouse and human 3′-UTRs were deduced taking into account the positions of the polyadenylation signals.

comparative study of the physico-chemical characteristics of the human and rat proteins. Additionally, the phylogenetic tree clearly indicates a common origin of the ERp29 family and a group of PDIs from various organisms. Despite the absence of direct functional link between these two groups due to the lack of the active redox motif in ERp29, it is speculated (Liepinsh et al., 2001) that ERp29 may have similar to PDI general chaperone function (Puig and Gilbert, 1994; Wang, 1998) within the folding/secretion machinery of the ER (see Section 3.5). The search in all available completed genome databases revealed an intriguing pattern of the evolutionary expression of ERp29, which is found exclusively in multicellular organisms, such as vertebrates and some invertebrates (*Drosophila*) and absent from the monocellular life forms. This lends additional support to the hypothetical 'secretory' role of ERp29 assuming that the protein export function is most extensively developed in the multicellular organisms.

3.2. Characterization of the ERp29 transcription start site

Transcription start site determination was accomplished by the modified 5′-RACE procedure, SMART 5′-RACE, which allows isolation of complete 5′-sequences more consistently than previous versions of the same method. First-strand cDNA was reverse transcribed from the rat

Table 1
Exon-intron junctions of the rat, mouse and human *ERp29* genes

Exon number	bp	5′ donor sequence		3′ acceptor sequence		Intron number	bp
		Exon	Intron	Intron	Exon		
Rat							
1	144	TTCTACA**AG**	**gt**aatgaga……gaccctc**ag**		**GT**CATTCCC	1	4747
2	139	GGATCTC**AG**	**gt**atttaca……tgttcac**ag**		**AC**TATGGTG	2	388
3	500	To the polyadenylation signal					
Mouse							
1	150	TTCTACA**AG**	**gt**aatgagg……cgccctc**ag**		**GT**CATTCCC	1	4917
2	139	GGATCTC**AG**	**gt**atggaca……tgtccac**ag**		**AC**TATGGCG	2	1568
3	500	To the polyadenylation signal					
Human							
1	144	TTCTACA**AG**	**gt**aaccggg……cgccctc**ag**		**GT**CATTCCC	1	6500
2	139	GGATCTC**AG**	**gt**atggaca……tgctcac**ag**		**AT**TATGGTG	2	2255
3	503	To the polyadenylation signal					

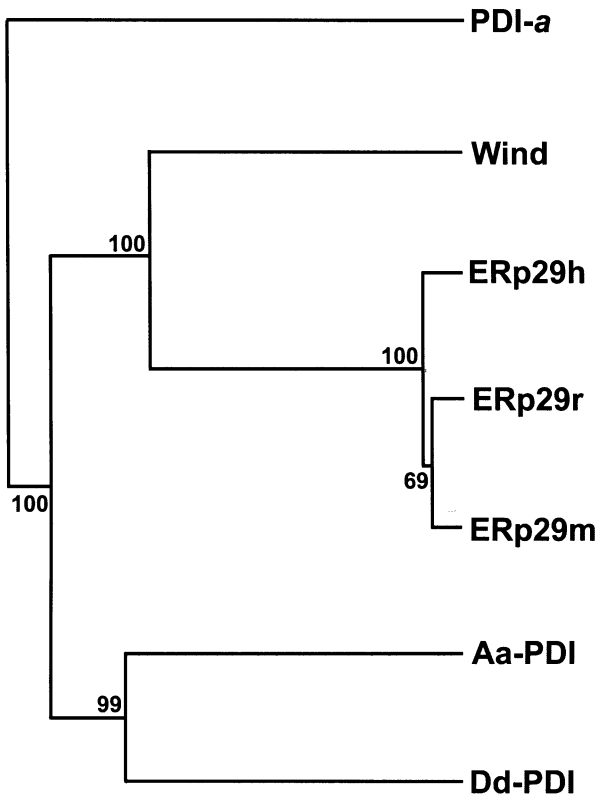

Fig. 2. Phylogenetic tree of ERp29s and related PDIs. The amino acid sequences of the different mammalian ERp29s (r, rat, m, mouse, h, human), Windbeutel (Wind) protein from *Drosophila melanogaster* (GenBank, AF025408) and related proteins from PDI family (PDI-***a***, ***a*** domain of the human PDI (P07237), PDI-aa, PDI from the plant alfaalfa (P38661), PDI-dd, PDI from the amoebae *Dictyostelium discoideum* (AAB86685) were aligned as in (Liepinsh et al., 2001). Phylogenies were inferred by distance matrix analysis using the PHYLIP program (ver. 3.5) (Felsenstein, 1989). The values supporting each node are derived from 100 resamplings. The image based on the PHYLIP output file was developed by the TreeView software. The primary structures of PDI-***a*** and Aa-PDI and Dd-PDI were used as outgroup.

liver and brain poly(A)$^+$RNA using gene-specific primer and PCR-amplified with the nested reverse gene-specific primer and nested universal primer. First PCR reaction did not produce any discrete products (results not shown) whereas secondary PCR using the same primers resulted in the amplification of two identical fragments (~420 and 300 bp) from both tissue samples and additionally ~380 and ~250 bp products from the brain sample and ~1200 bp fragment from the liver cDNA (Fig. 3). All these PCR products were isolated and cloned into the pcDNA3.1 vector with subsequent sequencing using nested gene-specific primer. Sequencing of the ~420 and ~300 bp products from both mRNA sources (Fig. 3, lanes 1 and 3) identified oligonucleotides with 5′-termini corresponding to the positions −148 and −46 (Fig. 4). The ~380 bp product (Fig. 3, lane 3) had 5′-end at the position −87. The ~1200 bp fragment from the liver and ~250 bp product from the brain cDNA turned to be unrelated sequences. The length of the 5′-UTR of the previously cloned ERp29 cDNA is 43 bp (Mkrtchian et al., 1998), which is very close to the first transcription start identified in current work.

3.3. Structural analysis of the 5′-flanking region of ERp29 gene

The 5′-flanking region of the rat *ERp29* gene was found within the *ERp29* genomic clone by Southern blot mapping, cloned into pUC18 vector and ~3 kb of the insert was sequenced. Analysis of the sequence revealed the presence of a GC-rich region spanning from the nucleotide −469 to −2 with the observed/expected CpG ratio about 1.0 and GC content >70% which are the characteristic features of CpG islands frequently found close to the transcription initiation sites of housekeeping genes (Larsen et al., 1992). Nearly identical in size and other parameters CpG islands were found also in the 5′-flanking regions of the mouse and human genes. In general, the first 600 bp of rat and human 5′ flanks exhibit more than 60% sequence identity whereas similarity between rat and mouse genes is even higher, 80% (Fig. 4).

Screening 5′ flanks of all three genes with MathInspector V2.2 (transfac.gbf.de/TRANSFAC) based on the TRANSFAC database of transcriptional factors (Wingender et al., 2000) failed to locate canonical TATA-like sequences and CAAT boxes, the elements that commonly specify the transcription start site in most of the genes. However, two GC boxes were mapped immediately upstream of one of the putative transcription start sites, −148, identified in this study. Similar TATA-less, GC-rich promoters are frequently found in the constitutively active genes (Dynan et al., 1986). Sp1 binding to GC boxes in such promoters is critical for the transcription initiation (Pugh and Tjian, 1991; Purnell and Gilmour, 1993), which is often directed from the multiple sites. A functional analog of TATA box,

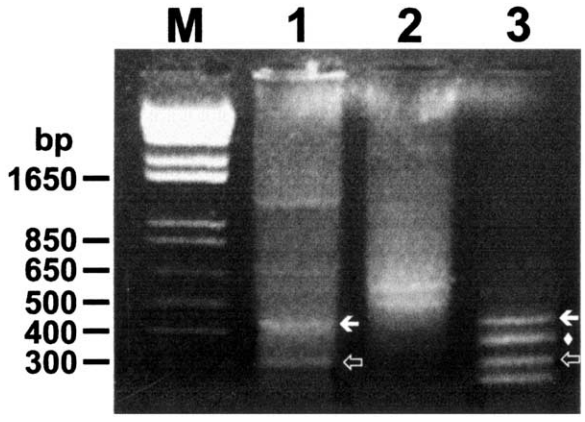

Fig. 3. Transcription start site mapping by SMART-RACE method. The second round of the PCR amplification of ERp29 cDNA from the rat liver and brain using nested universal and gene-specific primers. M, 100-bp marker, lane 1, liver cDNA, lane 2, negative control, brain cDNA amplified with one primer (nested gene-specific primer), lane 3, brain cDNA. Closed and open arrows identify ~420 and ~300 bp PCR products correspondingly. Diamond indicates ~380 bp PCR product.

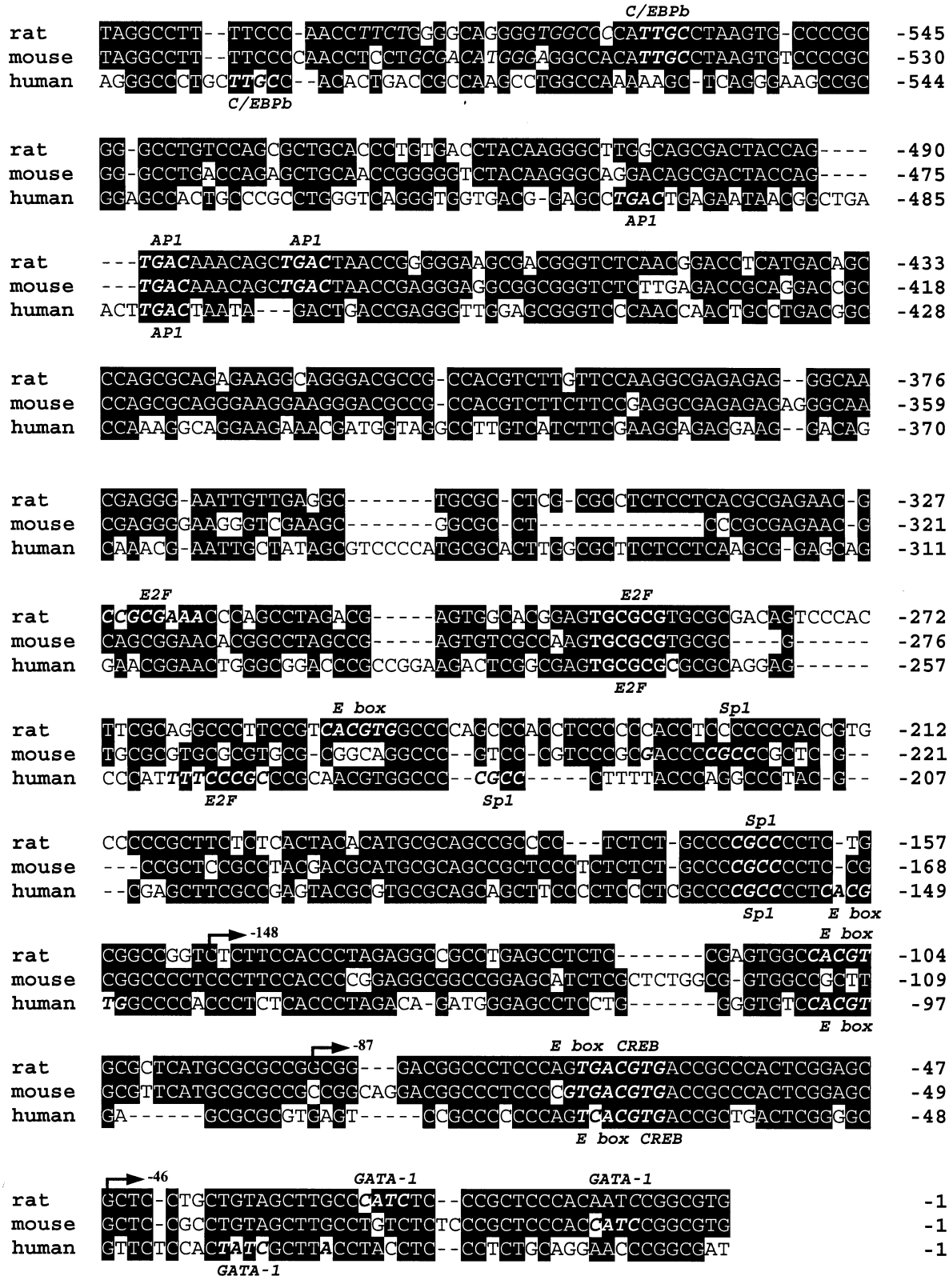

Fig. 4. Comparison of the nucleotide sequences of the rat, mouse and human 5′-flanking region of *ERp29*. Identical nucleotides are indicated by dark shading. The nucleotide preceding the translation start codon (ATG) is designated as −1. Putative transcription start sites are indicated by bent arrows on the rat sequence. Potential *cis*-elements (cores) are shown in bold with their names indicated above the alignment for the rat and mouse sequences and below for the human sequence.

an initiator element (Inr) (Lo and Smale, 1996) that usually encompasses the transcription start site was also absent in the proximal promoter lending additional support to the existence of multiple sites of transcription.

Within the 600 bp proximal fragment of the 5′ flank of the *ERp29* gene several AP-1, GATA-1, palindromic B-type E boxes, E2F and CREB binding sites were predicted, as well as a putative C/EBP element. Most of these motifs are conserved between all three sequences (Fig. 4), suggesting their importance for the transcriptional regulation of the *ERp29* gene.

We have previously shown that accumulation of unfolded proteins in the ER (ER stress) elevates the level of ERp29 in the FAO cells (Mkrtchian et al., 1998). ER stress triggers the unfolded protein response (UPR) signal transduction mechanism (Kaufman, 1999), activating ER stress response elements (ERSE) (Yoshida et al., 1998; Roy and Lee, 1999) on the promoters of the genes of the major ER chaperones. Thorough examination of the rat, mouse and human 5′-flanking regions of the *ERp29* gene did not reveal exact matches to ERSE. However, as it was shown earlier, some ER stress-inducible genes may lack or contain only weak ERSE-like sequences (Yoshida et al., 1998). For instance, the promoter of the *PDI* gene lacks exact matches to full ERSE ($CCAATN_9CCAAG$) despite the presence of the multiple CCAAT boxes, a characteristic core of the ERSE (Tasanen et al., 1988). Interestingly, PDI levels were seldom induced more than 3–4-fold by the ER stress whereas the induction of typical ERSE-containing genes, such as BiP or GRP94, is substantially stronger (Macer and Koch, 1988; Dorner et al., 1990). Similar to PDI levels of induction were observed also for ERp29 (Mkrtchian et al., 1998). Additionally, it may be speculated that the functional regulation of ERp29 is accomplished not (or not only) on the transcriptional but also on the translational or post-translational levels. The other possibility is that *ERp29* gene contains yet unknown regulatory elements that may act as *cis*-elements in UPR. Work is in progress to identify such putative stress responsive elements in the *ERp29* gene.

3.4. Functional analysis of the ERp29 promoter

To delineate the DNA elements responsible for the basal expression of the *ERp29* gene, a series of 5′-deletion constructs were generated by restriction digestion of the 5′-flanking region (Fig. 5A). The constructs were inserted into a firefly luciferase reporter gene plasmid and transiently transfected into the rat hepatoma FAO cells. As shown in

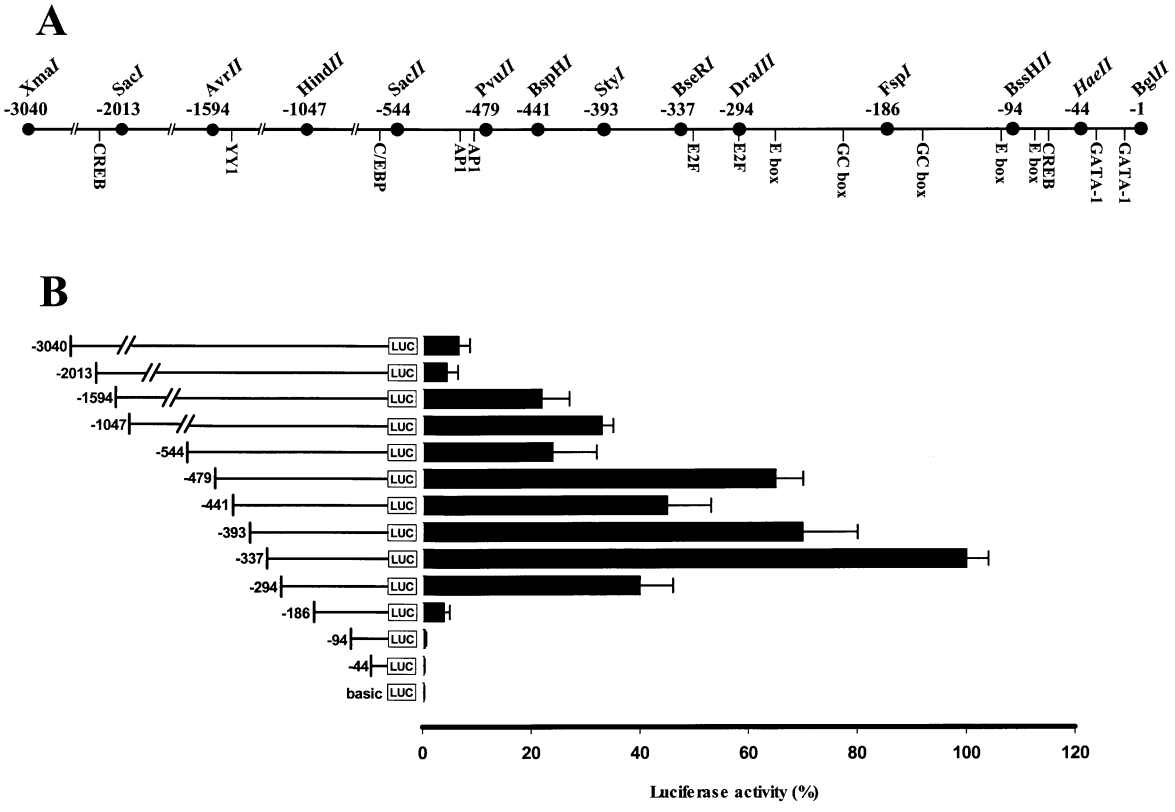

Fig. 5. 5′-deletion analysis of the 5′-flanking region of the rat *ERp29* gene. (A) The map of the restriction sites used for the construction of the 5′ deletion fragments of the 5′-flanking region and putative transcription factor binding sites. (B) Constructs (shown on the left) were produced by the restriction digestion of the genomic clone and then inserted into the promoterless pGL3-Basic vector containing firefly luciferase gene and transiently co-transfected with the internal control, pRL-TK vector carrying *Renilla* luciferase gene, into the FAO cells. Luciferase activity of each construct (shown on the right) was expressed relative to that of the −337/−1 construct. Data are means ± SE of eight to ten independent experiments.

Fig. 5B luciferase expression driven by the gene reporter constructs up to the position −186 was essentially not discernible from that of the promoterless construct despite the presence of GC and E boxes (Fig. 5A) that are typical binding sites for transcriptional initiators such as Sp1 and USF in TATA-less promoters. Further extension of the fragment to the position −294 incorporating additional GC and E boxes drastically increased the luciferase activity of the reporter gene. The maximum strength promoter reached in the −337/−1 fragment, indicating the presence of the additional positively acting regulatory element(s) in the short sequence between −294 and −337 bp. The putative E2F-binding sites found upstream of the Sp1 element (Fig. 5A) may be suggested as potential candidates. Further 5′ extension of the constructs had in general negative effect on the luciferase activity and revealed potential transcription enhancer (−479/−441) and silencer (−544/−479 and −2013/−1594) regions (Fig. 5B).

In order to localize the 3′ boundary of the core promoter region, the gene reporter construct with the highest activity (−337/−1) was chosen as a basis for 3′ deletions Three 3′ deletion constructs were inserted into the pGL3 vector and analyzed in the FAO cells by transient transfection. The shortening of the −337/−1 reporter construct by approximately 100, 200 and 300 bp led to the linear rather than abrupt decrease of the activity, which apparently reflects the gradual deletion of elements essential for the transcription initiation (Fig. 6). Deletion of the putative E box, CREB and GATA-1 sites (Fig. 5A) and two potential transcription start sites (fragment −337/−94) reduced the activity down to 65% of that of the original −337/−1 fragment. Further deletion of GC and E boxes and the −148 start site (fragment −337/−186) led to the loss of nearly 80% of the initial activity. The remaining 20% of activity may reflect the important role of the last GC box, E box and E2F elements. These data is consistent with the results of the 5′ deletion analysis and delineate the 337 bp proximal fragment of the 5′ flanking region as a core promoter of the gene.

Binding of Sp1 to GC boxes of TATA-less promoters has been shown to be critical for transcriptional initiation of numerous constitutively active genes (Dynan et al., 1986; Pugh and Tjian, 1991; Purnell and Gilmour, 1993). In addition, E box-specific binding proteins, such as USF1 and USF2 belonging to the basic helix-loop-helix family, have also been shown to be important for the basal activity of a number of TATA-less, Inr-less promoters (Sirito et al., 1994). Furthermore, it was demonstrated that the cooperation of Sp1 and USF factors is crucial for the transcriptional activation of the human transcobalamin II TATA-less promoter (Li and Seetharam, 1998). Additionally, E2F and Sp1 cooperation was shown to be essential for the full promoter activity of several TATA-less genes (Huang et al., 2001). Taken together, these data allow suggesting transcription factors, such as Sp1, USF and possibly E2F as key regulators of the basal activation of the *ERp29* gene.

Based on the structural and functional analysis of the *ERp29* promoter, as described above, we sought to identify the *trans*-acting nuclear factors that might be responsible for the constitutive activation of the *ERp29* gene. Oligonucleotides derived from the *ERp29* promoter that encompass two potential GC boxes (GC-1 and GC-2), one E-box (USF) and two putative E2F sites (E-1 and E-2) (Fig. 7B) were used in EMSA analysis with nuclear extracts isolated from FAO cells. Four DNA/protein complexes were revealed when the GC-1 oligonucleotide, containing a canonical Sp1 binding site, was used as a probe (Fig. 7A, left panel). Addition of antibody against Sp1 to the binding reaction caused a supershift of the band with lowest mobility, whereas two additional bands were supershifted by the Sp3 antibody, verifying the involvement of Sp1/Sp3 proteins in the

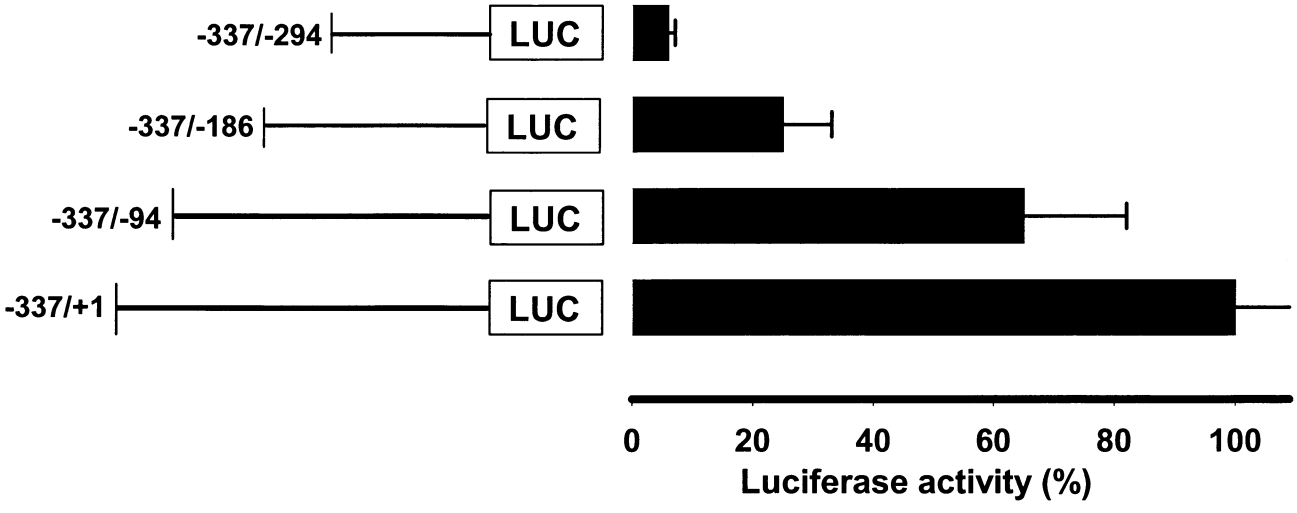

Fig. 6. 3′-deletion analysis of the 5′-flanking region of the rat *ERp29* gene. 3′ deletion constructs were generated by the restriction digestion of the ′337/′1 *ERp29* promoter fragment, transfected to FAO cells and analyzed as it is described in the legend to Fig. 5.

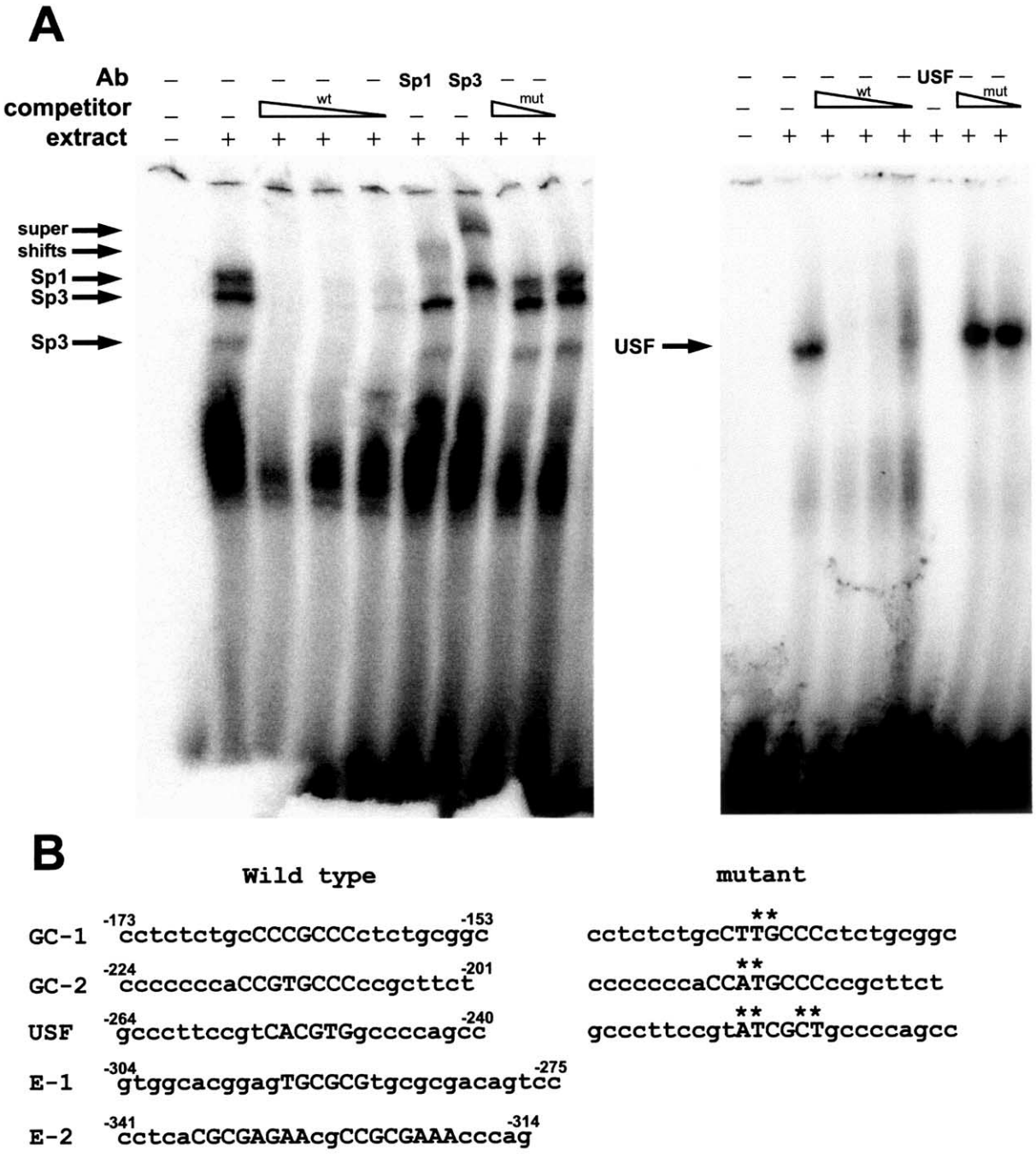

Fig. 7. Electrophoretic mobility shift assay. (A) Nuclear extracts from FAO cells were incubated with labeled wild type (wt) GC-1 (left panel) and USF (right panel) oligonucleotides and protein/DNA complexes (indicated by arrows) were resolved on the non-denaturing gel. A total of 1 μg of Sp1 or Sp3 and 2 μg of USF-1 antibodies were used in supershift assay. Wild type oligonucleotides binding were competed with decreasing concentrations of non-radioactive (50-5 × for GC-1 and 25-2 × for USF) and mutant oligonucleotides (100-50 ×). (B) Core consensus sequences of corresponding *cis*-elements are capitalized. Start and end positions of the wild type oligonucleotides are indicated above each sequence. Mutated bases are indicated with asterisks.

complex formation. Additionally, competition with non-labeled probe and absence of competition by an oligonucleotide containing point mutations in the core sequence of the GC box confirmed specificity of the Sp1/Sp3 complexes. Computer predictions indicate existence of a second, though less conserved GC box located in the fragment −186/−294 that showed high activity in the gene reporter assay. However, when the oligonucleotide GC-2 containing this element (Fig. 7B) was examined by EMSA, no binding to the nuclear proteins was observed (results not shown). To understand better which DNA element(s) may account for the high transcriptional activity of the promoter fragment −186/−294, an oligonucleotide derived from that fragment with a putative E box sequence was analyzed by EMSA.

Results presented on the Fig. 7A (right panel) indicate formation of the DNA/protein complex as evidenced by the USF oligonucleotide mobility shift. Identity of the complex-forming nuclear protein was revealed by complete elimination of the mobility shift by adding USF-1 antibody to the binding reaction. Again, binding was specific as shown by the competition by non-radioactive oligonucleotide and the absence of inhibition by the mutant probe.

Transcriptional activation of TATA-less and Inr-less promoters is often accomplished by Sp1 binding to tandem GC boxes. However, gel shift analysis failed to demonstrate binding to the predicted GC element adjacent to the functional GC box in the −186/−1 construct. Consequently, the low luciferase activity of this fragment suggests that Sp1 binding to the canonical GC box is not sufficient to trigger the basal gene expression. It has been suggested earlier that USF binding to E boxes of TATA-less promoters may function as a major factor directing basal transcription and moreover, the interplay between Sp1 and USF may further increase the transcriptional activity of the promoter (Li and Seetharam, 1998). The combined results of the reporter gene assays and the EMSA shown in this study would indicate that a similar interaction between Sp1 and USF transcription factors is important for the basal transcriptional activity of the *ERp29* gene. As already indicated, extension of the promoter fragment to the position −337 further enhances the activity of the luciferase reporter, suggesting involvement of additional *trans*-acting factors. Our preliminary results show that one of the E2F elements may contribute to such high activity as we observed binding of nuclear protein(s) to the oligonucleotides E-1 containing E2F-binding site (results not shown). Concluding this section, it may be argued that basal expression of the *ERp29* gene is most likely determined by the combined

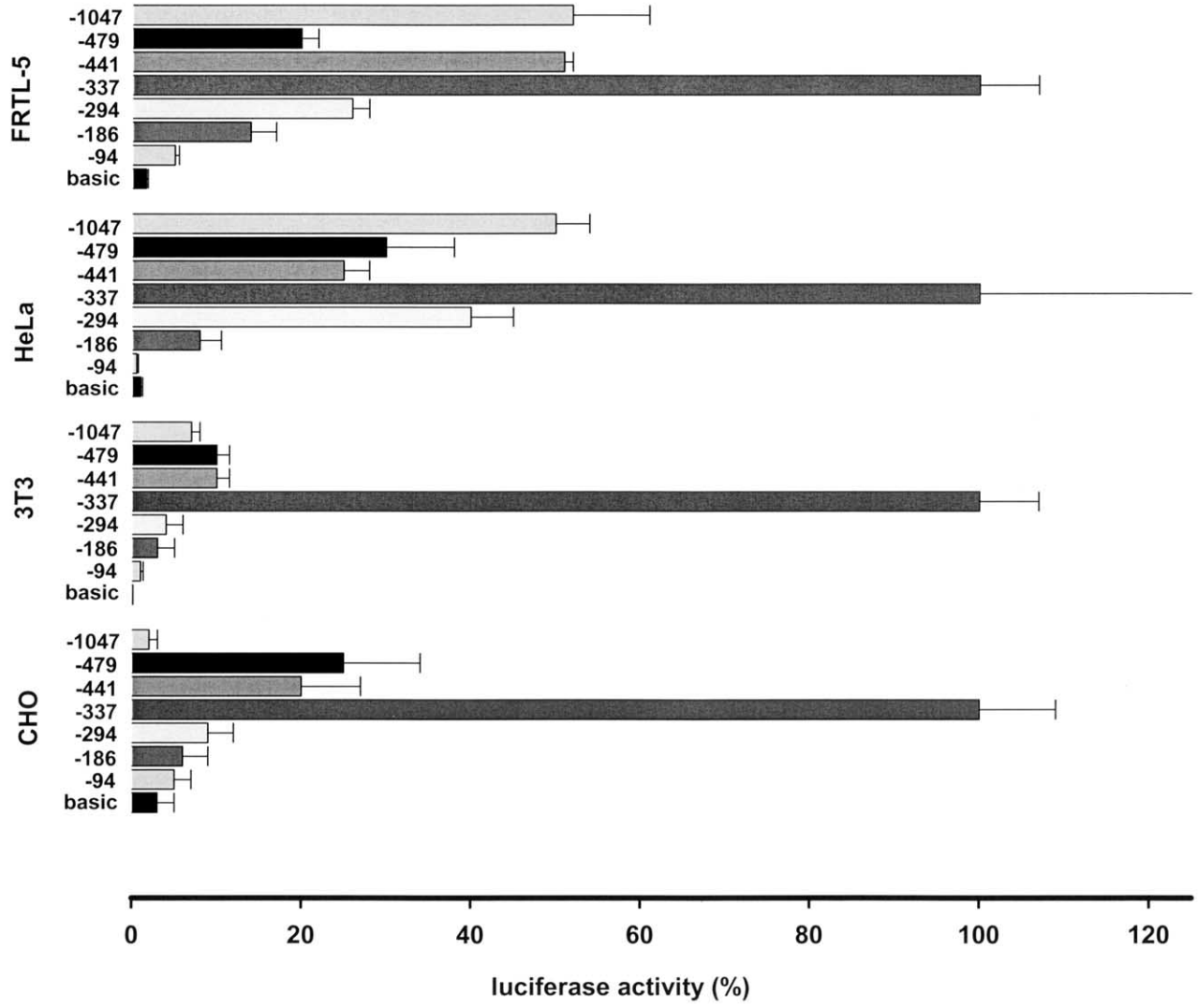

Fig. 8. 5′-deletion analysis of the 5′-flanking region of the *ERp29* gene in the different cell lines. 5′ deletion constructs of the 5′-flanking region of the rat *ERp29* gene were produced, cloned, transfected into FRTL-5, 3T3, CHO and HeLa cells as indicated in the legend to Fig. 5 and luciferase activity was expressed relative to that of the −337/−1 construct.. Data are means ± SE of 3–4 independent experiments.

action of several transcription factors including Sp1, USF and possibly E2F.

3.5. ERp29 promoter activity in various cell lines

Activities of the same gene reporter constructs were examined also in murine 3T3, hamster CHO, rat FRTL-5 and human HeLa cells (Fig. 8), which all express ERp29, as detected by Western blot (results not shown). The fragment that displayed the highest activity in FAO cells (−337/−1) was found strongest also in all tested cells. Moreover, despite some differences in the fold activation or repression, the general patterns of luciferase activities of the fragments encompassing the basal promoter region (−1 to −337) were similar to that of the FAO cells (compare Figs. 5B and 6). At the same time, we have observed certain differences between tested cells, which points to the specie- and tissue-dependent diversity in the *ERp29* regulation. For instance, the activity of the fragment −294/−1 was significantly higher in human HeLa and rat FRTL-5 cells as compared to 3T3 and CHO cells (Fig. 8) and comparable with that of FAO cells (Fig. 5B). This discrepancy could probably be attributed to the different extent of transcription factors present in the different cell lines. Similar discrepancy is observed in the fragment −1047/−1, containing putative enhancer. Its activity was significantly higher in HeLa and FRTL-5 cells as compared to the 3T3 and CHO cells (Fig. 8). Since this fragment contains a predicted C/EBP site (Fig. 5A), this element might be suggested as a putative enhancer that is active only in HeLa and FRTL-5 cells. The latter cell lines display rather similar pattern of the activity of the promoter fragments implying comparable mechanisms of the constitutive transcriptional regulation. One can speculate that tissues, such as thyroid gland that have substantially higher levels of expression of *ERp29* (see Section 3.6), would need different mode of the basal gene regulation as compared to the tissues with a low expression of ERp29. Immunoblotting (results not shown) indicate that HeLa cells express ERp29 in the amounts comparable with hepatocytes and thyrocytes, which may be suggested as a possible explanation of the similar pattern of the promoter activity.

3.6. ERp29 gene expression profile

The high level of the identity between the rat and human cDNAs (87%) allowed us to utilize the fragment of rat *ERp29* cDNA as a specific probe for the hybridization with the human RNA dot blot array. This method offers significantly broader range of tissues as compared with the previous studies of rat and human tissue expression utilizing Northern blot approach (Demmer et al., 1997; Ferrari et al., 1998; Mkrtchian et al., 1998).

As shown on Fig. 9, the ERp29 transcript was detected in all tissues represented on the array. The highest level of expression is observed in the tissues with high secretory activity such as pituitary, adrenal, thyroid and salivary glands, prostate, pancreas, which are specialized in secreting proteins and additionally, liver and kidney. Interestingly, the level of *ERp29* expression is substantially lower in testis and ovary, also active secretory tissues, which however, export mostly low molecular weight molecules such as testosterone and estrogen. This observation implies that ERp29 might be involved primarily in the processing of protein substrates. Apparently, *ERp29* expression commences at the early stages of development, as mRNA is detected in all tested fetal tissues. The expression profile is similar to that of the adult tissues with the highest expression observed in the lung and liver.

Although such ubiquitous expression of ERp29 is consistent with the structural characteristics of the 5′ flank of the

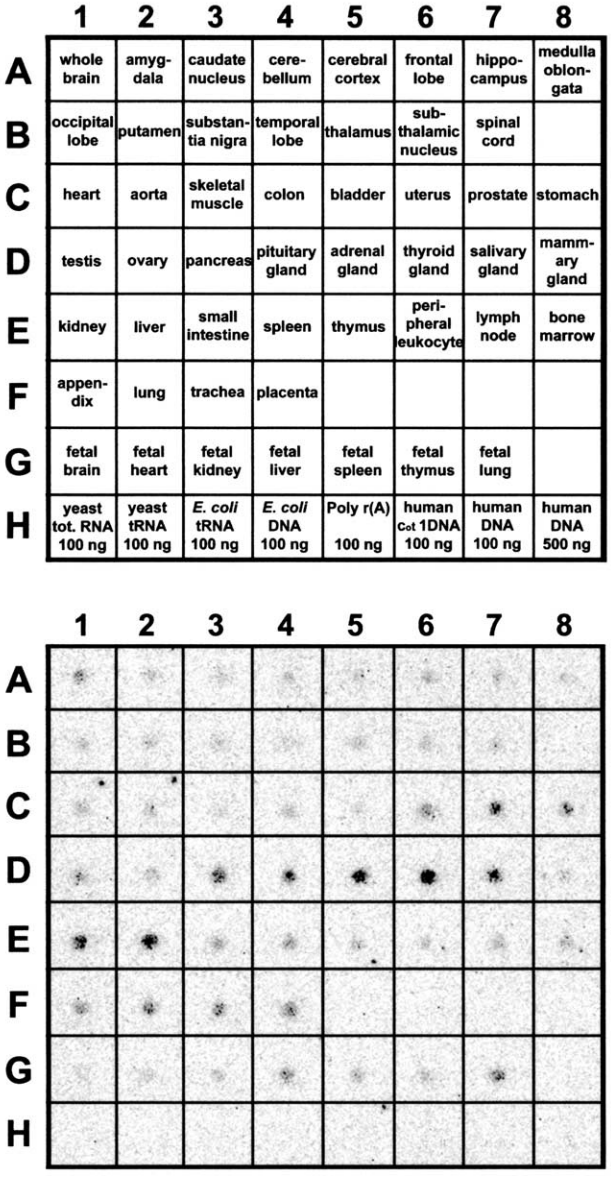

Fig. 9. Tissue-specific expression of the *ERp29* gene. Nylon membrane loaded with the normalized amounts of poly(A^+)RNA from human tissues (indicated on the upper panel) was hybridized with the radiolabeled ERp29 probe and analyzed by autoradiography.

ERp29 gene predicting the housekeeping properties of the protein, it displays rather uneven distribution pattern. However, even though generally considered to have low expression fluctuations, the expression of a number of housekeeping genes varies significantly in different tissues (Thellin et al., 1999; Warrington et al., 2000). Apparently, this may be caused by quite variable demand for the basic cellular functions in different cell types and hence lead to the variable expression of the housekeeping genes responsible for the maintenance of such functions.

Secretory function is an essential feature of all mammalian cells and is strongly accentuated in the cells specialized solely in the export of proteins or low molecular weight molecules. It is reasonable to assume that in the tissues with a low secretory activity, the expression of secretion factors, such as ERp29, would be kept at relatively low levels, whereas higher amounts of these proteins would be required in the specialized secretory tissues. For example, despite the identical widespread expression, a number of ER chaperones and foldases such as calreticulin, PDI and ERp72, were also distributed unevenly in different tissues being more abundant in tissues with high secretory activity (Bjelland, 1987; Tharin et al., 1992).

In conclusion, we have cloned the rat *ERp29* gene, characterized its organization and identified the core promoter and transcriptional factors that may contribute to the basal expression of this gene. Such characteristics of the 5′ flank as CpG island, the absence of TATA-box, multiple transcription start sites and also ubiquitous gene expression indicate that *ERp29* belongs to the group of constitutively expressed housekeeping genes. Identical gene organization, conservation of the rat, mouse and human nucleotide sequences and phylogenetic analysis provide enough basis for suggesting *ERp29* as a single name for all mammalian genes, thus avoiding current ambiguity in nomenclature of human (*ERp28*) and rodent (*ERp29*) genes. The studies described here provide a starting point for examination of the regulation of *ERp29* gene in more details.

Acknowledgements

Mikhail Baryshev and Ernest Sargsyan received scholarships from the Royal Swedish Academy and Swedish Institute. We thank Dr Inger Johansson for helpful comments on the manuscript and Magnus Ingelman-Sundberg for the constant support and encouragement. This work was supported by the Swedish Medical Research Council and Swedish Society for Medical Research.

References

Backlund, M., Johansson, I., Mkrtchian, S., Ingelman-Sundberg, M., 1997. Signal transduction-mediated activation of the aryl hydrocarbon receptor in rat hepatoma H4IIE cells. J. Biol. Chem. 272, 31755–31763.

Bjelland, S., 1987. Tissue distribution and molecular heterogeneity of bovine thiol: protein-disulphide oxidoreductase (disulphide interchange enzyme). Comp. Biochem. Physiol. B 87, 907–914.

Demmer, J., Zhou, C., Hubbard, M.J., 1997. Molecular cloning of ERp29, a novel and widely expressed resident of the endoplasmic reticulum. FEBS Lett. 402, 145–150.

Dorner, A.J., Wasley, L.C., Raney, P., Haugejorden, S., Green, M., Kaufman, R.J., 1990. The stress response in Chinese hamster ovary cells. Regulation of ERp72 and protein disulfide isomerase expression and secretion. J. Biol. Chem. 265, 22029–22034.

Dynan, W.S., Sazer, S., Tjian, R., Schimke, R.T., 1986. Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. Nature 319, 246–248.

Fang, C., Mkrtchian, S., Ingelman-Sundberg, M., 1997. Combination of direct DNA sequencing with degenerate primer-mediated PCR and 5′-/3′-RACE to screen novel cDNA sequences. Biotechniques 23, 52–56.

Felsenstein, J., 1989. PHYLIP – Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.

Ferrari, D.M., Nguyen Van, P., Kratzin, H.D., Soling, H.D., 1998. ERp28, a human endoplasmic-reticulum-lumenal protein, is a member of the protein disulfide isomerase family but lacks a CXXC thioredoxin-box motif. Eur. J. Biochem. 255, 570–579.

Huang, D., Jokela, M., Tuusa, J., Skog, S., Poikonen, K., Syvaoja, J.E., 2001. E2F mediates induction of the Sp1-controlled promoter of the human DNA polymerase epsilon B-subunit gene POLE2. Nucleic Acids Res. 29, 2810–2821.

Hubbard, M.J., McHugh, N.J., 2000. Human ERp29: isolation, primary structural characterization and two-dimensional gel mapping. Electrophoresis 21, 3785–3796.

Hubbard, M.J., McHugh, N.J., Carne, D.L., 2000. Isolation of ERp29, a novel endoplasmic reticulum protein, from rat enamel cells evidence for a unique role in secretory-protein synthesis. Eur. J. Biochem. 267, 1945–1957.

Kaufman, R.J., 1999. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. Genes Dev. 13, 1211–1233.

Konsolaki, M., Schupbach, T., 1998. Windbeutel, a gene required for dorsoventral patterning in *Drosophila*, encodes a protein that has homologies to vertebrate proteins of the endoplasmic reticulum. Genes Dev. 12, 120–131.

Kwon, O.Y., Park, S., Lee, W., You, K.H., Kim, H., Shong, M., 2000. TSH regulates a gene expression encoding ERp29, an endoplasmic reticulum stress protein, in the thyrocytes of FRTL-5 cells. FEBS Lett. 475, 27–30.

Larsen, F., Gundersen, G., Lopez, R., Prydz, H., 1992. CpG islands as gene markers in the human genome. Genomics 13, 1095–1107.

Li, N., Seetharam, B., 1998. A 69-base pair fragment derived from human transcobalamin II promoter is sufficient for high bidirectional activity in the absence of a TATA box and an initiator element in transfected cells. Role of an E box in transcriptional activity. J. Biol. Chem. 273, 28170–28177.

Liepinsh, E., Baryshev, M., Sharipo, A., Ingelman-Sundberg, M., Otting, G., Mkrtchian, S., 2001. Thioredoxin fold as homodimerization module in the putative chaperone ERp29. NMR structures of the domains and experimental model of the 51 kDa dimer. Structure 9, 457–471.

Lo, K., Smale, S.T., 1996. Generality of a functional initiator consensus sequence. Gene 182, 13–22.

Macer, D.R., Koch, G.L., 1988. Identification of a set of calcium-binding proteins in reticuloplasm, the luminal content of the endoplasmic reticulum. J. Cell Sci. 91, 61–70.

Mkrtchian, S., Fang, C., Hellman, U., Ingelman-Sundberg, M., 1998. A stress-inducible rat liver endoplasmic reticulum protein, ERp29. Eur. J. Biochem. 251, 304–313.

Pugh, B.F., Tjian, R., 1991. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev. 5, 1935–1945.

Puig, A., Gilbert, H.F., 1994. Protein disulfide isomerase exhibits chaper-

one and anti-chaperone activity in the oxidative refolding of lysozyme. J. Biol. Chem. 269, 7764–7771.

Purnell, B.A., Gilmour, D.S., 1993. Contribution of sequences downstream of the TATA element to a protein-DNA complex containing the TATA-binding protein. Mol. Cell. Biol. 13, 2593–2603.

Roy, B., Lee, A.S., 1999. The mammalian endoplasmic reticulum stress response element consists of an evolutionarily conserved tripartite structure and interacts with a novel stress-inducible complex. Nucleic Acids Res. 27, 1437–1443.

Sen, J., Goltz, J.S., Konsolaki, M., Schupbach, T., Stein, D., 2000. Windbeutel is required for function and correct subcellular localization of the *Drosophila* patterning protein Pipe. Development 127, 5541–5550.

Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15, 7155–7174.

Sirito, M., Lin, Q., Maity, T., Sawadogo, M., 1994. Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells. Nucleic Acids Res. 22, 427–433.

Tasanen, K., Parkkonen, T., Chow, L.T., Kivirikko, K.I., Pihlajaniemi, T., 1988. Characterization of the human gene for a polypeptide that acts both as the beta subunit of prolyl 4-hydroxylase and as protein disulfide isomerase. J. Biol. Chem. 263, 16218–16224.

Tharin, S., Dziak, E., Michalak, M., Opas, M., 1992. Widespread tissue distribution of rabbit calreticulin, a non-muscle functional analogue of calsequestrin. Cell Tissue Res. 269, 29–37.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., Heinen, E., 1999. Housekeeping genes as internal standards: use and limits. J. Biotechnol. 75, 291–295.

Wang, C.C., 1998. Protein disulfide isomerase assists protein folding as both an isomerase and a chaperone. Ann. N. Y. Acad. Sci. 864, 9–13.

Warrington, J.A., Nair, A., Mahadevappa, M., Tsyganskaya, M., 2000. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol. Genomics 2, 143–147.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., Schacherer, F., 2000. TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. 28, 316–319.

Yoshida, H., Haze, K., Yanagi, H., Yura, T., Mori, K., 1998. Identification of the cis-acting endoplasmic reticulum stress response element responsible for transcriptional induction of mammalian glucose-regulated proteins. Involvement of basic leucine zipper transcription factors. J. Biol. Chem. 273, 33741–33749.