PERGAMON

Journal of Steroid Biochemistry & Molecular Biology 72 (2000) 89–102

The Journal of Steroid Biochemistry & Molecular Biology

www.elsevier.com/locate/jsbmb

Identification of human estrogen-inducible transcripts that potentially mediate the apoptotic response in breast cancer

Jozsef Szelei, Ana M. Soto, Peter Geck, Marcelle Desronvil, Nancy V. Prechtl, Beau C. Weill, Carlos Sonnenschein*

Department of Anatomy and Cell Biology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA

Received 3 September 1999; accepted 2 November 1999

Abstract

Hormone manipulation has been used for several decades with the purpose of inducing breast cancer regression. On the one hand, hormone ablation and antiestrogen administration were used on the rationale that estrogens induce proliferation of their target cells. Before the advent of the antiestrogen tamoxifen, on the other hand, the estrogen agonist DES was used to obtain clinical remissions. The rationale for the use of diethylstilbestrol (DES) was totally empirical. In fact, the efficacy of both treatments was comparable. A mechanistic explanation for estrogen-induced regression is urgently needed in order to provide a rationale for its use in therapeutic fields, and to develop markers to identify this phenotype in order to recognize responsive tumors. In this report, we use E8CASS cells (a MCF7 variant) as a model to study estrogen-mediated regression. The proliferation rate of E8CASS cells is decreased by estrogens. In order to isolate mRNA sequences induced by estradiol, a subtracted library was prepared from E8CASS cells grown in the presence and absence of estrogens. Twenty nine differentially expressed unique sequences were found. Seven of them were homologous to known genes, 12 of them were homologous to expressed sequence tags (EST), and 10 sequences had no homologues in the databases. The two sequences showing the highest induction by estradiol (E9 and E43) were chosen for further analysis. The sequence of the E43 coding region has 96% homology to the bovine actin2 gene and 100% identity to bovine actin2 protein, and it is homologous to the human actin-related protein 3 (Arp3). It has been suggested that Arp3 is involved in actin nucleation. The phenotype of E8CASS cells is clearly affected by estrogen treatment. It is likely that E43 may be involved in these morphological changes. The E9 cDNA is a putative zinc-finger protein of the PHD family of transcriptional transactivators. A member of this family, Requiem, is involved in apoptosis. The E9 mRNA is highly expressed in E8CASS cells treated with estrogens, a treatment which results in decreased proliferation rate and increased DNA degradation. This correlation suggests that E9 may be a mediator of estrogen-induced regression of breast cancer.

Keywords: Breast carcinoma; Tumor regression; Estrogens; Transcription factor

1. Introduction

Despite significant progress made in the detection and identification of prognostic factors in breast cancer during the last 40 years, decisions about treatment continue to be a statistical game for the affected individual [1]. Endocrine treatment is mostly given as adjuvant therapy or in advanced disease [2]. Before the advent of tamoxifen, endocrine therapy consisted of estrogen ablation (ovariectomy, adrenalectomy) or massive estrogen administration (diethylstilbestrol, DES). Although counterintuitive, DES treatment resulted in spectacular regressions; in randomized trials, tamoxifen had a similar response rate to DES [3,4]. However, DES had undesirable side effects. As a result, the use of estrogens to induce regression was

* Corresponding author. Tel.: +1-617-636-6581; fax: +1-617-636-6536.

E-mail address: csonnenschein@infonet.tufts.edu (C. Sonnenschein).

0960-0760/00/$ - see front matter
PII: S0960-0760(00)00025-X

practically abandoned. Recently, however, DES administration was reported to induce regression in tumors that initially responded to tamoxifen and later relapsed [5]. A mechanistic explanation for estrogen-induced regression is urgently needed to provide a rationale for its use in therapeutic fields, and to develop markers to identify this phenotype to recognize responsive tumors.

In order to address these issues, we have identified MCF7 breast cancer cell variants (e.g. E8CASS) that respond to estrogens by decreasing their proliferation rate and undergoing cell death [6]. In order to identify the genes that may be involved in estrogen-induced regression, we conducted a search for genes that turned on in response to estradiol treatment in E8CASS cells. Sensitive selection techniques are now available to identify differentially expressed genes. The Wang and Brown differential amplification method [7] was chosen to study the effect of estradiol on gene expression in E8CASS cells. A subtracted library was prepared in order to identify estradiol-inducible genes. In this article we report the identification of estrogen-induced sequences that, due to their pattern of expression in these cells, are candidates for mediators of estrogen-induced regression.

2. Materials and methods

2.1. Reagents

17 β-estradiol was purchased from Calbiochem (Richmond, CA). ^{32}P α- and γ-dATP and 8,1 α-dCTP were obtained from ICN (Irvine, CA).

2.2. Tissue culture

MCF7 cells were kindly supplied to us by Dr. Charles McGrath, Michigan Cancer Foundation, Detroit, Michigan, in 1983 [8]. Since then, they have been cloned repeatedly and routinely grown in DMEM supplemented with 5% heat-inactivated (56°C, 30 min) fetal calf serum (FBS) (Hyclone, Logan, Utah). The parental MCF7 cells are inhibited from proliferating when exposed to 10% charcoal-dextran stripped human serum (CDHuS) and express a monophasic proliferative response to estrogens [9]. The serum inhibitor was identified as albumin [10]. E8CASS cells were selected by continuous exposure to medium supplemented with 10% CDHuS. These cells proliferate maximally in 10% CDHuS, and estradiol significantly reduces their proliferation rate [6].

2.3. Cell proliferation studies

Proliferation rate experiments were carried out in 12-well plates (Falcon, Lincoln Park, NJ). Forty thousand cells/well were seeded in 1 ml 5% CDFBS. After 24 h, this medium was changed to the experimental media being tested (10% CDHuS with or without estradiol at different concentrations). Cells were lysed and their nuclei were counted daily by a Coulter Counter Model ZM (Hialeah, FL) to measure their proliferation rate.

2.4. Apoptotic DNA fragmentation analysis

The protocol described by Tilly and Hsueh [11] was followed. Briefly, DNA was prepared from E8CASS cells, grown in T75 flasks in 10% CDHuS medium with and without 1 nM estradiol which were harvested at several time intervals. Then, 3′ end-labeling was performed by incubating the DNA aliquots with terminal deoxynucleotidyl transferase (Gibco BRL Life Technologies, Gaithersburg, MD) and ^{32}α-P-dATP (3000 Ci/nmol; Amersham, Arlington Heights, IL). The incorporated radioactivity was monitored by Cherenkov counts. The labeled samples were loaded on 2.5% agarose gels and separated by electrophoresis for 3–4 h at 40–50 V using 1× TAE solution (40 mM Tris-acetate buffer, pH 8.0–1 mM EDTA). The gels were dried, sealed in plastic wrap and exposed to X-ray film.

2.5. The Wang–Brown differential amplification

Poly-A$^+$ mRNA was isolated from E8CASS cells after 40 h of exposure to 1 nM estradiol-supplemented 10% CDHuS (E-mRNA), and from E8CASS cells harvested after 40 h of exposure to CDHuS (CD-mRNA) by using a FastTrack Kit (InVitrogen, San Diego, CA). Double-stranded cDNA pools from both RNA preparations were synthesized using the Copy Kit (InVitrogen), with oligo-dT priming. Two micrograms aliquots of cDNA from both, the CD and E preparations were digested with Sau3A, and the cDNAs were ligated with specific double-stranded adaptors, CD-adaptor (5′-**GATC**GGACGGTGATGTCTCGAGAGTG-3′) and E-adaptor (5′-**GAT**CTGCGGTGAGAGGCTG-3′). Each adaptor was designed to have a four-base, 5′ protruding end (shown in bold), complementary to the Sau3A-generated termini of the cDNA [12]. The adaptors were generated by the annealing of the appropriate primer pairs at 45°C for 10 min. The unligated adaptors were removed from the ligation mixtures by DNA precipitation and electrophoresis in a 2% low-melting agarose gel. The cDNAs were PCR-amplified using the appropriate primers specific to the adaptors (5′-GACACTCTCGAGACATCACCGTCC-3′, and 5′-AGCACTCTCCAGCCTCTCACCGCA-3′) and using primers of the GeneAmp Kit (Perkin Elmer, Foster

City, CA). Fifty micrograms aliquots of amplified cDNAs were photobiotinylated (driver cDNA) with Photoprobe biotin (Vector Laboratories, Burlingame, CA), and hybridized to a 2 μg aliquot of non-biotinylated cDNA from the other pool. Each subtraction cycle consisted of one long and one short hybridization steps; they were performed in order to increase subtraction efficiency for low and high abundance transcripts [7]. After three cycles of subtraction, the final amplified pools of sequences specifically expressed in the presence of estrogen (E) and absence of estrogen (CD) were ligated into the pCRII vector using the TA Cloning Kit (In Vitrogen). Colonies were transferred to nylon filters and differentially screened using ^{32}P α-ATP labeled subtracted E and CD final amplified pools.

2.6. DNA sequence analysis

PCR sequencing reactions were performed by the DyeTerminator Cycle Sequencing Kit and the 373A Automated DNA Sequencer (Applied Biosystems, Foster City, CA). The results were analyzed with the FASTA [13], the TRANSLATE (Wisconsin Package, Version 9.0, Genetics Computer Group, Madison, WI) and the BLAST (National Center for Biotechnology Information, Bethesda, MD) programs. Most of the protein motif and profile predictions were done by the corresponding programs of the Wisconsin Package. We also used remote servers providing protein sequence analysis programs through the Internet. The Pfam program, Release 3.3 (http://pfam.wustl.edu/) is a database of protein domain family alignments maintained by the Pfam Consortium through the Sanger Center (UK), the Washington University, St. Louis (US) and the Karolinska Institute (Sweden). The ExPaSy Molecular Biology Server (http://expasy.hcuge.ch/) is dedicated to the analysis of protein sequences and structures as well as two-dimensional protein pattern analysis and maintained by the Swiss Institute of Bioinformatics (Geneva, Switzerland).

2.7. Northern blot analysis

Total RNA was isolated by the RNeasy kit (Qiagen, Chatsworth, CA) from E8CASS and MCF7 cells grown in the presence and absence of estradiol. Twenty micrograms purified RNA aliquots were resolved by electrophoresis in 1.2% agarose–2.2 M formaldehyde gels and transferred to Nytran membranes using the TurboBlot system (Schleicher and Schuell, Keene, NH). The membranes were baked under vacuum (80°C) and hybridized to the random primed ^{32}P-labeled cDNA probes according to standard protocols [14]. Membranes were analyzed by PhosphorImager (Molecular Dynamics, Stanton, CA) using the ImageQuant program. Multiple Tissue Northern blots containing poly-A$^+$ RNA preparations from an array of normal human tissues were purchased from Clontech (Palo Alto, CA). RNA loading was normalized to either glyceraldehyde 3-phosphate dehydrogenase (GAPDH) or β-actin signals in the same blot.

2.8. 5′- and 3′-RACE

The Marathon-Ready cDNA preparation from the human mammary gland was used as a substrate of the 5′- and 3′-RACE reactions (Clontech Marathon-Ready cDNA kit). The amplified products were cloned into the pCR2.1 vector using T4 ligase (In Vitrogen). In every RACE PCR amplification, internal (nested) primers were used to screen the extended cDNA clones (E9 5′ RACE primers: 5′-GCCTTGACTGACTGGCACCTACTAC (positions 3921–3897), 5′-CTGCCTCAGATTCAAAGGTGACAC (positions 3566–3543), and 5′-TGAGTGTGTCCCCAACAAGCAAG (positions 3147–3125); E9 3′ RACE primers: 5′-GGGGTGGATAAAAACATGTCTGCT (positions 3682–3705), and 5′-CTCAGCTCTCTTTTCTTCTTAGCC (positions 3764–3788); and E43 RACE primers: 5′-CTGGTCCAACACTCTTGTCTTTAC (positions 2118–2135) and 5′-CAGGGCACAGAAA-

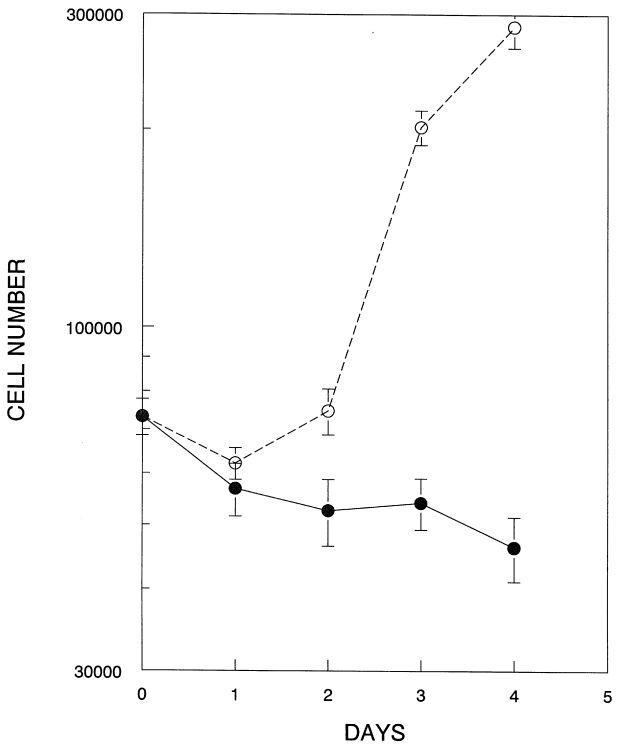

Fig. 1. Proliferation curves of E8CASS cells cultured in 10% CDHuS (○- - - - -○) alone and 10% CDHuS plus 1 nM estradiol (●———●). Cell numbers are expressed as the mean ± SD of triplicate wells.

GAACTAAAACC (positions 1735–1712). DNAs coding for the open reading frames (ORF) were synthesized from freshly isolated RNA obtained from E8CASS cells treated with 1 nM estradiol, using 5′- and 3′-end primers with the Advantage cDNA Kit (Clontech).

3. Results

3.1. Proliferation profile of E8CASS cells

E8CASS cells were cultured in 10% CDHuS and 10% CDHuS plus 1 nM estradiol. Cells were counted daily to determine their proliferation rate. One nanomolar estradiol inhibited the proliferation of E8CASS cells after 1–2 days (Fig. 1); therefore, the search for genes that mediate this inhibitory response was performed using RNA from cells exposed to estradiol for 40 h. These cell population kinetics data do not discriminate between estradiol-triggered inhibition of cell proliferation and estradiol-induced apoptosis. However, DNA degradation, considered as a biochemical marker for apoptosis, was observed in E8CASS cells starting 24 h after being exposed to estradiol at 1 nM (Fig. 2).

3.2. Characterization of the estradiol-specific subtracted library

One hundred and eight clones from the E preparation were plated and screened using the ^{32}P α-dCTP-labeled CD- and E-subtracted, PCR amplified cDNA fragments as probes. Cross-hybridizations with both probes showed that all the isolated clones were present exclusively in the E set; they represented 44 different sequences. Comparative slot blot analysis of total RNA from estradiol-treated and -untreated E8CASS cells revealed that these 44 sequences were specifically expressed by estradiol-treated E8CASS cells. Further analysis by sequencing revealed that among these cDNA clones there were seven clones that represented known human repetitive sequences (Alu and L1 family), and 37 clones that represented different portions of 29 unique sequences. The average insert-size was approximately 172 bp. A summary of the clones and sequences identified by the Wang and Brown method is shown in Table 1. Ten sequences had no matches and the remaining 19 had homologies in the GenBank database. The accession numbers, the lengths of the sequences, and the degrees of homologies are shown in Table 2.

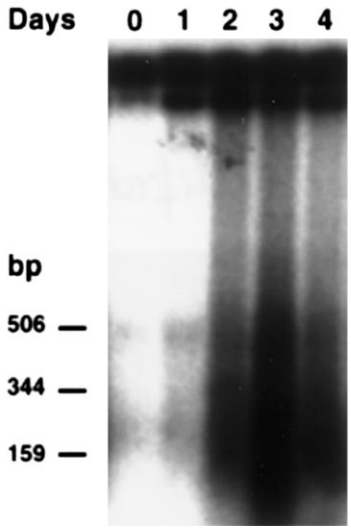

Fig. 2. DNA degradation in E8CASS cells. The 3′ termini of the DNA fragments were labeled by ^{32}P incorporation using terminal deoxynucleotidyl transferase and separated by agarose electrophoresis. After estradiol treatment at day 0, samples were taken at days indicated on the top.

3.3. Analysis of differential gene expression

Eleven of the 29 sequences were randomly chosen for further analysis by Northern blots to identify the extent of the differential gene expression using RNA from E8CASS cells treated with and without estradiol. Five estradiol-induced cDNAs did not give a signal on Northern blots; however, RT-PCR analysis revealed that these sequences were significantly induced (see Table 2). The lack of detection of these estradiol-

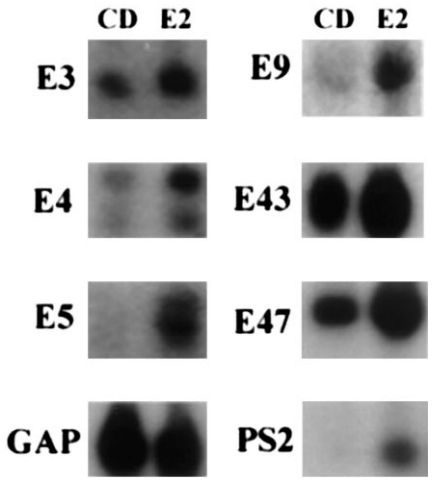

Fig. 3. Northern analysis of differential gene expression. RNA was prepared from E8CASS cells grown in 10% CDHuS (CD) and 10% CDHuS plus 1 nM estradiol (E2) for two days; 20 μg of total RNA were applied to each lane. The filters were hybridized with gene-specific probes indicated at the left of the autoradiograms. High level expression of E43 was detected in the autoradiogram, where phosphorimager analysis indicated fourfold induction by E2. The estrogen-inducible PS2 gene was used as a positive control. Glyceraldehyde phosphate dehydrogenase (GAP) was used as the normalization control.

Table 1
Characterization of the estradiol-specific subtracted library[a]

E library clones	Differential expression by slot-blot	Unique sequences	Sequences with GenBank homology	Newly identified sequence	Differential expression by Northern	Differential expression by RT-PCR
108 clones	44 clones	29 clones	19 E1 (E4, E46) E2 (E26, E41) E3 (E25) E9 (E32) E10 E12 E15 (E16) E19 E24 E28 E30 E33 E34 E35 E37 E39 E40 E42 E43	10 E5 E6 E8 E22 E23 E27 (E31) E29 E36 E38 E47	6 E1 (E4, E46) E3 (E25) E5 E9 (E32) E43 E47	5 E6 E8 E15 (E16) E24 E34

[a] The original clones in the subtracted library were named E followed by a number (E1–E47). When a unique sequence is represented by a single clone it is identified by the name of such clone. When a unique sequence is represented by more than one clone, it is identified by the clone with the lowest number followed by the other clones enclosed in parenthesis.

induced sequences by Northern blots may have been due to their low copy number or short size of the probe. The rest of the sequences, however, were readily detectable on the blots.

The sequences that showed high differential expression were chosen for further analysis. Quantitative phosphorimager analysis revealed that E43 was induced four-fold and E9 was induced seven- to nine-

Table 2
Sequence homologies of the isolated clones[a]

Clones	Insert (bp)	Homology	Identity (%)	Species
E1, E4, E46	189	T26006 (EST)	97	Human
E2, E26, E41	191, 198, 204	AB020864	98	Human
E3, E25	197	X89962	79	Rat
E9, E32	124, 197	D86969	100	Human
E10	197	U42390	98	Human
E12	142	W23314 (EST)	98	Human
E15, E16	176	H47788 (EST)	98	Human
E19	181	AA464480 (EST)	88	Human
E24	168	W79040 (EST)	100	Human
E28	185	AA604400 (EST)	95	Human
E30	150	L19711	93	Human
E33	154	AA128322 (EST)	97	Human
E34	179	D87117	92	Mouse
E35	177	AA477849 (EST)	95	Human
E37	174	AA384816 (EST)	96	Human
E39	192	AA505904 (EST)	98	Human
E40	168	H84303 (EST)	96	Human
E42	164	AA594805 (EST)	98	Human
E43	105	D12816	97	Bovine

[a] The Insert column indicates the cloned sequence tag sites in base pairs from the estradiol-specific library. EST homologue sequences are indicated in the homology column. Only the sequence with the highest score is shown among the EST homologue sequences. Accession # D86969 is the KIAA0215 gene. Accession # U42390 is the human Trio mRNA. Accession # L19711 is the human dystroglycan (DAG1) mRNA. Accession # U50147 is the Rattus norvegicus synapse-associated protein 102 mRNA. Accession # D12816 is the Bovine mRNA for actin2.

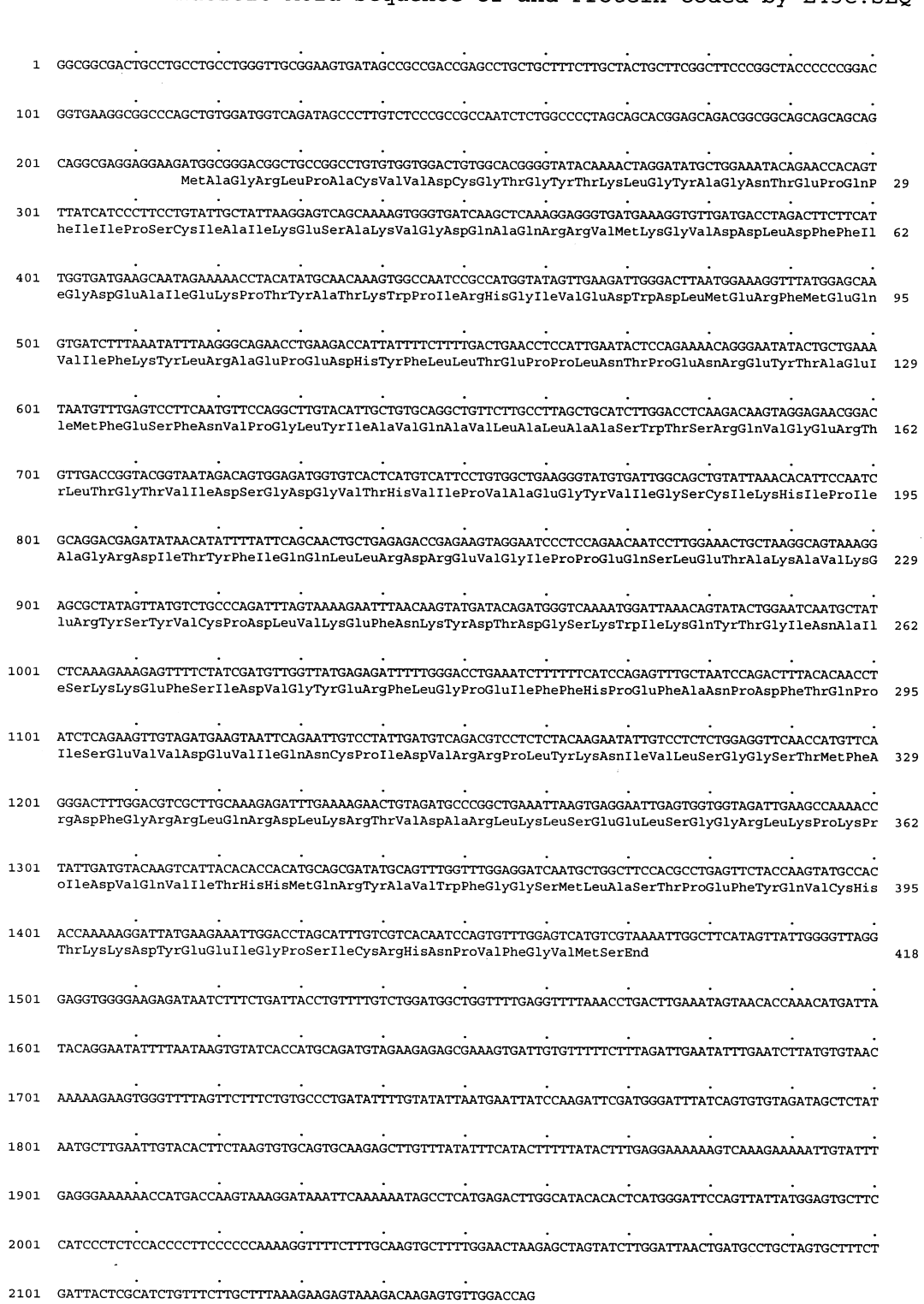

Fig. 4. E43 cDNA sequence and open reading frame. The amino acid sequence is under the coding strand. Numbers on the left indicate the positions of the base pairs; numbers on the right indicate the positions of the amino acids.

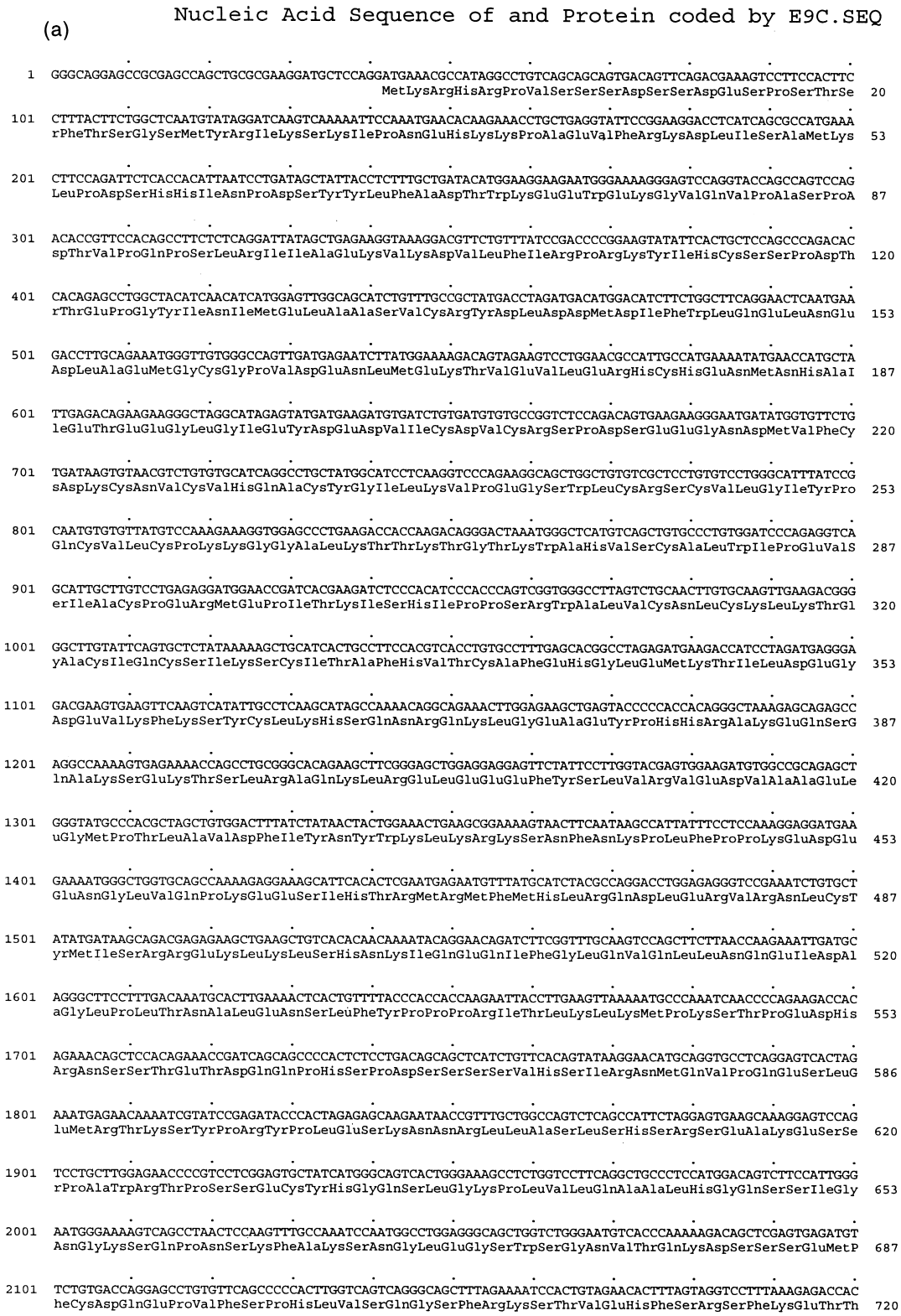

Fig. 5. E9 cDNA sequence and open reading frame. The amino acid sequence is under the coding strand. Numbers on the left indicate the positions of the base pairs; numbers on the right indicate the positions of the amino acids. Destabilizing signals are underlined.

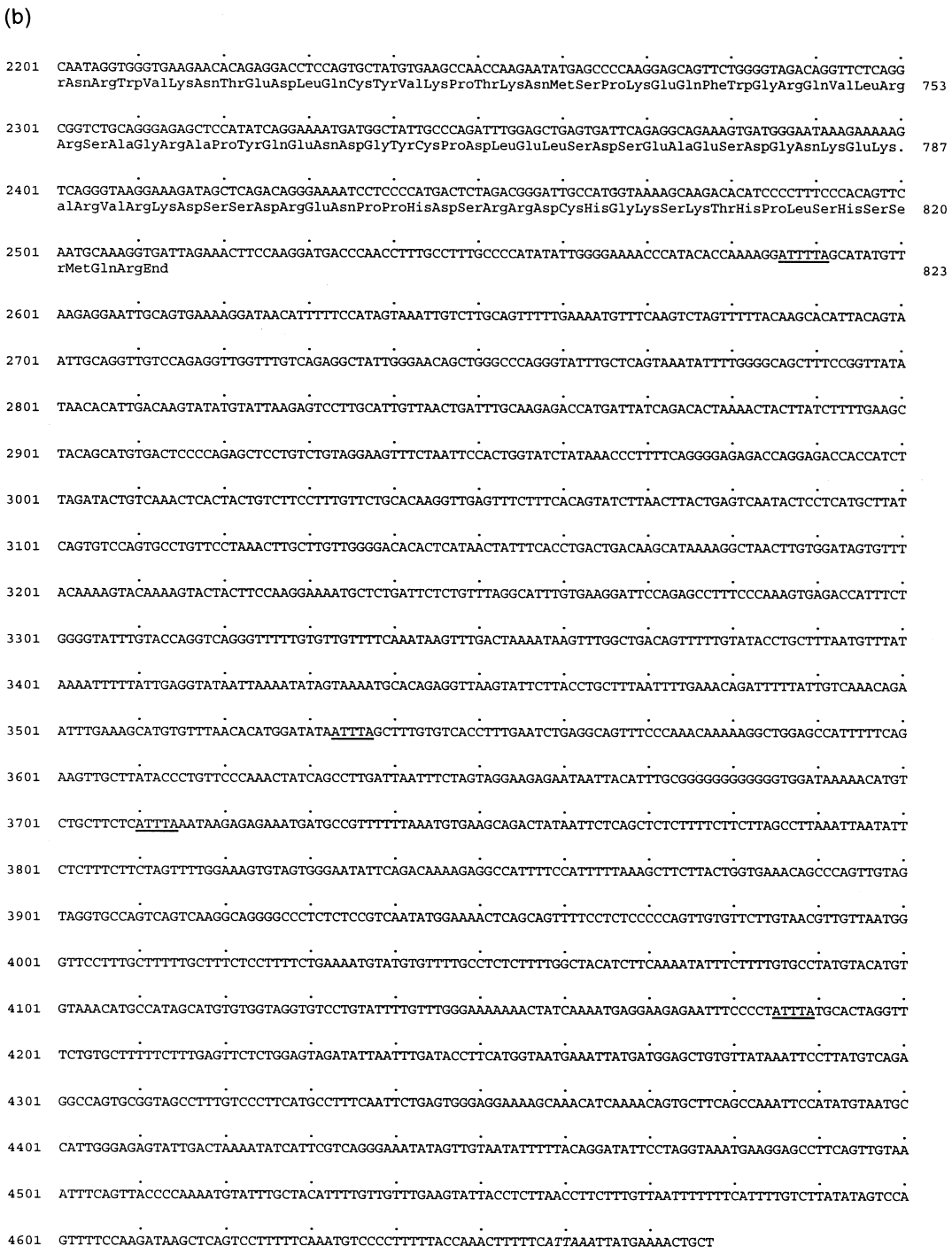

Fig. 5 (*continued*)

fold by estradiol in E8CASS cells (Fig. 3). Although the E5 sequence was also highly induced, the isolated tag sequence was insufficient for successful RACE extensions and has not been analyzed further. However, with the E43 and E9 genes the RACE and Marathon PCR extensions were successful and the complete open reading frames of both these genes have been established.

3.4. Characterization of the E43 cDNA

The E43 sequence is homologous to bovine actin2

[15]. The analysis of the E43 gene was of interest to us not only because of its high inducibility but also because E8CASS cells changed their shape after the addition of estradiol to the medium, a process where actin may be involved. Northern analysis revealed two transcripts, a 2.2 and a 2.5 kb mRNAs. The 2158 base long cDNA represents the shorter transcript; it contains a 1254 base long ORF (Fig. 4), which has 96% homology to the bovine actin2 gene coding region and 47% homology to the ORF of the β-actin gene [15]. The cDNA sequence is 98% homologous to the recently cloned human actin-related protein 3 (Arp3) gene [16]. The AUG codon is in a strong Kozak's initiator codon context [17]. The 418 amino acid long protein sequence is identical to the bovine actin2 and the recently described human Arp3 protein. It has extensive homology to other Arp3 proteins: *Drosophila melanogaster* actr66B (80%), *Acanthamoeba castellanii* Arp3 (71%), *Dictyostelium discoideum* ACLA (70%), *Neurospora crassa* Arp3 (63%), *Saccharomyces cerevisiae* ACT4 (61%), and *Schizosaccharomyces pombe* act2 (59%) [18,19].

3.5. Characterization of the E9 cDNA

The E9 and E32 clones represent fragments of the same gene; hereafter this gene will be called E9. These sequences were extended by 5′- and 3′-RACE (Table 1). While our study was in progress, a sequence similar to E9 (KIAA0215 protein) was submitted to the DDB/EMBL/GenBank databases [20]. The E9 cDNA (4677 bp) contains a long ORF (Fig. 5), and this region has a 99.9% homology to the KIAA0215 protein coding cDNA [20,21]. There is no published information on the function of this sequence. The initiator codon is at position 32, and the stop codon is at position 2511. The 3′ non-coding region has four AT-rich elements (ARE, AUUUA) [22] (Fig. 6). These destabilizing sequences may contribute to the regulation of the E9 transcript and protein levels. The polypeptide consists of 823 amino acids, and has the same sequence as the KIAA215 protein. The salient features of this sequence are (Fig. 6): (i) serine-rich regions at positions 8, 566, 608 and 650, (ii) five ATP/GTP-binding site motifs (P-loops at positions 24, 263, 353, 650, and 804), (iii) the C-terminal region of the protein contains a domain of charged and polar amino acids (the basic amino acids are mainly at position 780–823, while the acidic residues clustered at position 770–800), (iv) two putative nuclear localization signals, at positions 259 (PKKGGALKTTKTGTK) and 435 (KLKRKSNFNKPLFPPK), and (v) PHD (plant homology domain) zinc fingers. A highly conserved PHD Domain I is found between positions 200–250 and a less conserved PHD Domain II is localized between positions 309–368 (Fig. 7). PHD fingers have been recently recognized as a subgroup within the zinc-finger family [23]. PHD fingers are characterized by a special arrangement of the metal binding amino acids (Cys4HisCys3). The zinc-finger containing region (residues 130–367) has 42% sequence identity and 77% similarity with a human zinc-finger protein, BR140 [24].

An extensive homology to a P1 artificial chromo-

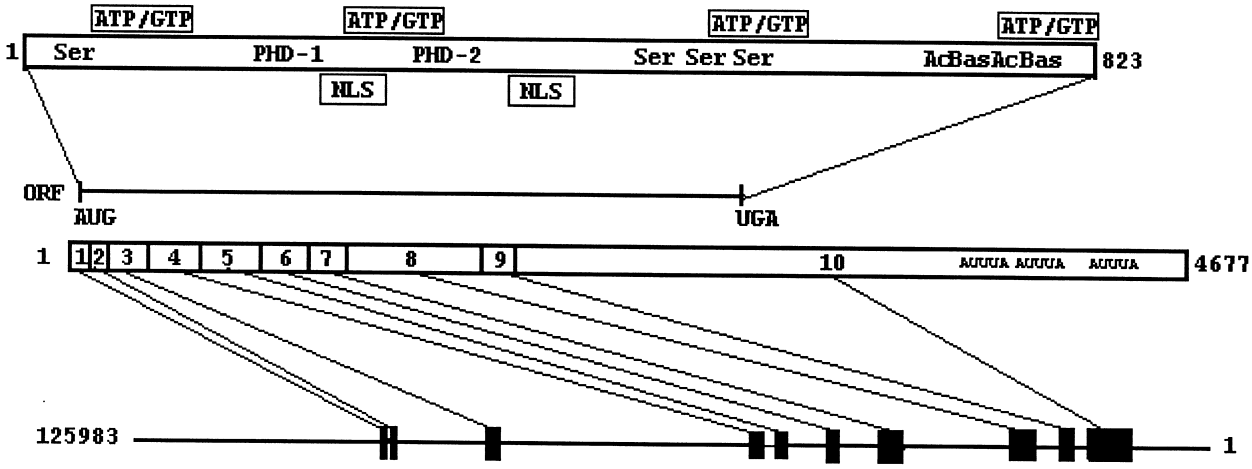

Fig. 6. Diagram of the E9 polypeptide domains (top), exon map (middle) and the exon–intron structure of the genomic coding region of the PAC306D1 sequence (bottom). The boxes labeled "ATP/GTP" placed above the open reading frame represent the putative ATP/GTP binding domains. The boxes labeled "NLS" below the polypeptide represent nuclear localization sequences. Domains within the open reading frame box are labeled as Ser (serine-rich domains), PHD-1, PHD-2 (PHD zinc-fingers), and AcBas (acid–basic domains). The exons are numbered within the transcript box. The AUUUA regulatory sequences are also shown. The full boxes in the PAC306D1 sequence represent the exons. The numbers at the beginning and end of the polypeptide and nucleic acid sequences indicate amino acid and nucleotide positions, respectively.

Consensus PHD sequence versus E9 PHD domain 1

```
PHD Consensus*->dnfCsvCgstdddldgdmvqCdGggCdewfHqkClnpplelqsskeslnppdgkWlCpkCkrk<-*
                d++C vC+s d++ ++dmv+Cd  +C++++Hq C+++ +           p+g WlC+ C
      E9  200   DVICDVCRSPDSEEGNDMVFCD--KCNVCVHQACYGILKV----------PEGSWLCRSCVLG  250
```

Consensus PHD sequence versus E9 PHD domain 2

```
PHD Consensus*->dnfCsvCgstdddldgdmvqCdGggCdewfHqkC.....lnpplelqsskeslnppdgkWlCpkCkrk<-*
                 +C +C+ +     g ++qC   +C ++fH++C  +++l+ + +    +++ +  ++k +C k+++
      E9  309   ALVCNLCKLK----TGACIQCSIKSCITAFHVTCafehgLEMKTI----LDEGDEVKFKSYCLKHSQN  368
```

Fig. 7. Comparison between the consensus PHD sequence and the PHD domains of E9. The upper case characters in the consensus sequence indicate the amino acids directly involved in zinc binding. The numbers before and after the E9 sequences show the positions of the PHD domains in the E9 sequence. The "+" sign indicates functionally conservative substitution.

some, PAC 306D1, which was mapped to the X chromosome, was found through a BLAST search [25]. The comparison of the genomic sequence to our cDNA sequence by the BLAST program resolved the exon–intron structure of the E9 gene (Fig. 8). A summary of the predicted functional domains of the E9 polypeptide and the exon–intron structure of the genomic coding region are shown in Fig. 6.

3.6. Time course and specificity of the expression of the E9 and E43 genes

The 5 kb mRNA for the E9 gene increased 1.5-fold after 14 h of exposure to 1 nM estradiol, and continued increasing steadily throughout the period studied; a three-fold increase was found at 24 h and an eight-fold increase at 48 h. The E43 gene showed a similar

```
                1                                  87  (90113)
                gggcaggagc ... Exon 1  ... tcagacgaaa   gtgagtagaa

    (89342)  88                                      167  (89261)
tctttcacag * gtccttccac ... Exon 2  ... acctgctgag * gtgagatcgt

    (76934)  168                                     325  (76775)
atctttccag * gtattccgga ... Exon 3  ... cttctctcag * gtatttgcat

    (50330)  326                                     516  (50138)
gcttttctag * gattatagct ... Exon 4  ... gcagaaatgg * gtaagttctt

    (47162)  517                                     728  (46949)
ttctgtgcag * gttgtgggcc ... Exon 5  ... tgtgcatcag * gttagtgaga

    (41433)  729                                     896  (41264)
atacctccag * gcctgctatg ... Exon 6  ... gatcccagag * gtaagaattc

    (36105)  897                                    1013  (35987)
ttatttgtag * gtcagcattg ... Exon 7  ... ttgtattcag * gtaagtcccc

    (20896)  1014                                   1484  (202424)
cttcctacag * tgctctataa ... Exon 8  ... cctggagagg * gtaaggtgac

    (18972)  1485                                   1602  (18853)
gggctttcag * gtccgaaatc ... Exon 9  ... attgatgcag * gtaacctgta

    (16886)  1603                                   4677  (13814)
tttatctcag * ggcttccttt ... Exon 10 ... gaaaactgct * poly-A
```

Fig. 8. Exon–intron structure of E9 genomic DNA. Asterisks represent the exon–intron boundaries. The area between asterisks represents the exons, the numbers indicate the position of exon base pairs, and the numbers in parenthesis refer to the position of base pairs of PAC306D1.

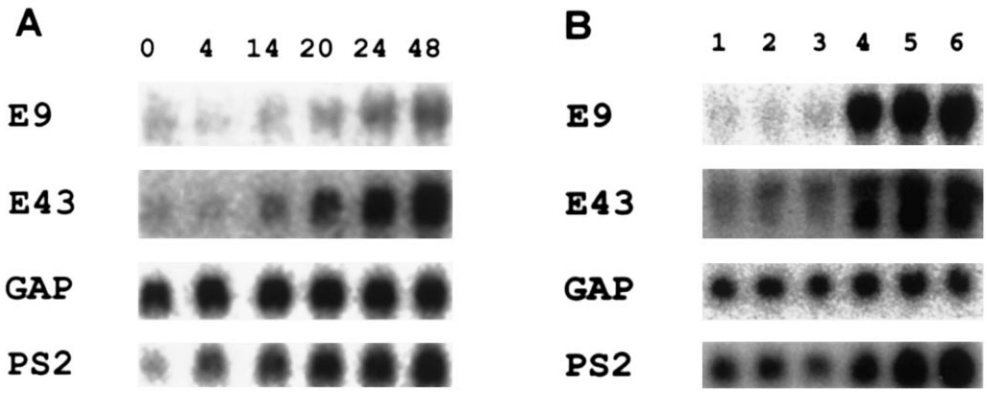

Fig. 9. Northern analysis of the time-course induction and the effect of estradiol concentration on E9 and E43 gene expression. Panel A: the time-course of induction of E9 and E43 expression by 1 nM estradiol. The time of exposure to estradiol (hours) is indicated by the numbers above the lanes. Panel B: the effect of estradiol concentration on the induction of the E9 and E43 genes in E8CASS cells. Lane 1: 10% CDHuS; Lanes 2–6: 10% CDHuS plus estradiol (2, 0.1 pM; 3, 1 pM; 4, 10 pM; 5, 100 pM; 6, 1 nM). Cells were harvested after 48 h of exposure to estradiol.

pattern; the first significant increase (2.7-fold) was observed at 14 h after exposure to 1 nM estradiol, it increased to four-fold at 24 h, and 4.7-fold at 48 h (Fig. 9, panel A).

E9 was induced by estradiol exclusively in E8CASS cells. The induction of E43 by estradiol was also observed in MCF7 cells; however, the induction levels in MCF7 cells were modest (not shown) in comparison to those observed in E8CASS cells. The induction of the E9 gene occurred at the estradiol concentrations required for inhibition of E8CASS cell proliferation [6], namely, 10 pM and higher estradiol concentrations

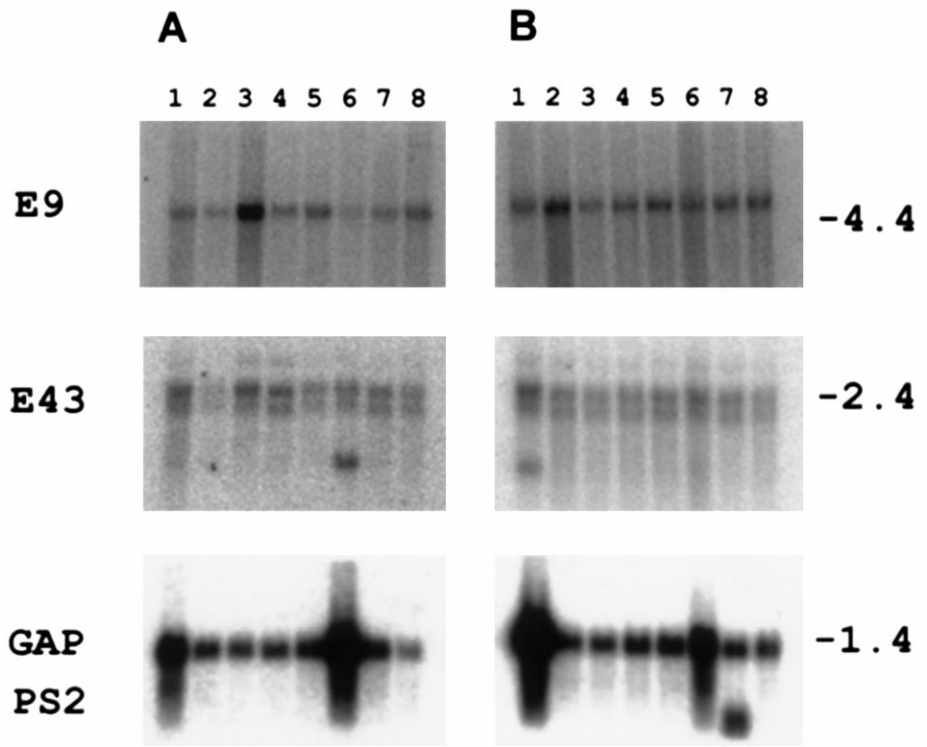

Fig. 10. Northern filters containing normalized amounts of poly-A selected mRNA preparations from various normal human tissues (CLONTECH Laboratories) were hybridized with probes derived from E9, E43, GAPDH and PS2 genes. Panel A: 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas. Panel B: 1, skeletal muscle; 2, uterus; 3, colon; 4, small intestine; 5, bladder; 6, heart; 7, stomach; 8, prostate.

(Fig. 9, panel B). The two isoforms of the E43 mRNA displayed a similar pattern of induction by estradiol (Fig. 9, panel B).

3.7. Expression of the E9 and E43 genes in human tissues

Northern blots of mRNA preparations from 14 normal human tissues were tested, using the Multiple Tissue Northern blots (Fig. 10). The 5 kb mRNA of E9 was expressed in all tissues; however, it was most abundantly expressed in the placenta and the uterus. Double bands were detected for E43 mRNA (2.5 and 2.2 kb) in all tissues, and a small (1 kb) band was seen in skeletal muscle and the heart. Northern blots were slightly more dense in skeletal muscle, heart, placenta and the lung. The E9 and E43 genes are also expressed at high levels in the mammary gland and uterus (Fig. 11).

4. Discussion

The mechanisms underlying estrogen-induced regression in breast cancer are poorly understood. We have previously reported the identification of a MCF7 breast cancer cell variant, E8CASS, that was selected after nine months of continuous exposure to 10% CDHuS [6]. The parental MCF7 cells are inhibited from proliferating in the presence of 10% CD-stripped serum, and proliferate maximally when estradiol is added to this medium. Instead, the proliferation rate of the E8CASS variant is significantly decreased by estrogen treatment (Fig. 1). Moreover, estradiol induced genomic DNA degradation, a marker of apoptosis, in these E8CASS cells. Hence, we used these cells to study the expression of genes which may be involved in estrogen-induced proliferation arrest and apoptosis in breast cancer cells, as a model for breast cancer regression. A subtracted library was made in order to identify these sequences. We have identified several candidate sequences that may mediate estradiol-induced regression, two of which have been chosen for analysis in this paper.

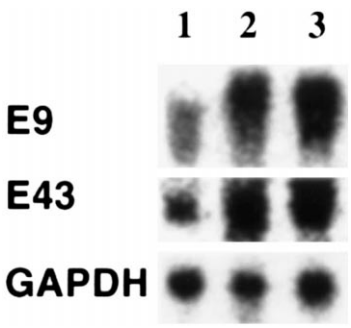

Fig. 11. Analysis of E9 and E43 gene expression in human placenta (1), uterus (2) and mammary gland (3). Twenty grams of total RNA (CLONTECH Laboratories) were loaded in each lane.

4.1. Analysis of the subtracted library

The Wang and Brown method [7] was used to identify up-regulated genes following estradiol exposure with the modification [12] of using two different primers for the amplification of the driver (CD) and the tracer (E) cDNA fragments. As a result, cross-hybridization of clones on colony filters and slot blots with subtracted cDNA probes (CD and E) showed greatly reduced cross-contamination between the two sets of cDNAs.

Sequence analysis of the differentially expressed, unique cDNA sequences in estradiol-treated E8CASS cells revealed 12 sequences that were homologous to EST sequences (Table 2). E1 is a 5′-end mRNA sequence expressed in Burkitt's lymphoma. E12 is homologous to a sequence involved in renal carcinoma, and has been isolated from a putative tumor suppressor region, 3p104.2 [26]. E15 and E24 are 5′-end mRNA sequences from a human fetal liver–spleen library. E19 and E35 are partial 5′-end transcripts from a human ovary tumor library. E28, E39 and E42 are 3′-end partial cDNAs from the Cancer Genome Anatomy Project at NCI. E33, E37 and E40 are partial 5′-end sequences from cDNA libraries representing pregnant uterus, thyroid and retina transcripts, respectively.

Seven cDNAs were found to be homologous to known sequences containing ORF regions: (i) E3 is homologous to the rat PIPPin, a putative RNA-binding protein expressed in the brain [27], (ii) E10 is the human Trio mRNA, which codes a multidomain protein (protein kinase domain and separate rac-specific and rho-specific guanine nucleotide exchange factor domain), and binds to the transmembrane protein tyrosine phosphatase [28], (iii) E30 is the human dystroglycan (DAG1) mRNA, (iv) E34 is the rat SAP102 mRNA coding for a postsynaptic protein and is homologous to the Drosophila disc-large tumor suppressor protein DlgA [29], and (v) E2, E26 and E41 clones were identical to each other and they were homologous to a human genomic sequence at 8p21.3-p22 that codes for a putative suppressor gene in hepatocellular, colorectal and non-small cell lung cancer (Table 2). Finally, 10 sequences had no homologues in the data bases. Further analysis (5′- and 3′-RACE) of these novel sequences is currently under way.

4.2. The E43 cDNA characterization and relevance

The complete homology of the Arp3 protein to the putative E43 polypeptide and the 98% similarity of the

Arp3 cDNA to the E43 sequence show that E43 is the human Arp3. E43 is also homologous to the Arp3 sequences from other species. The Arp2/3 complex seems to play a role in actin filament nucleation [30]. The Arp3 gene is essential for the viability of *Schizosaccharomyces pombe* [19] because of its function in the reorganization of the actin cytoskeleton during the cell cycle [31]; remarkably, overexpression of Arp3 protein results in the lysis of the cell. Northern analysis revealed two (2.5 and 2.2 kb) mRNA isoforms in E8CASS cells, possibly the result of alternative polyadenylation; a comparable phenomenon has been described about bovine actin2 [15]. E43 mRNA is expressed at higher levels in mammary gland, skeletal muscle, heart, placenta, lung and kidney, than in the other organs tested. E43 mRNA expression was first increased at 14 h, with maximal induction at 48 h after estradiol exposure in E8CASS cells (Fig. 9). The change in the shape of E8CASS cells after estradiol exposure [6], together with an increasing detachment of the cells from the growing surface, signal the initial phase of apoptosis. The underlying mechanism appears to be cytoskeletal reorganization of these cells; the E43/Arp3 protein may play an important role in this process.

4.3. The E9 cDNA: characterization and relevance

The protein coded by the E9 gene has several features suggestive of a regulatory role. The five putative ATP/GTP-binding site motifs indicate that the E9 protein may participate in signal transduction, phosphorylation, nucleotide transport, and/or may be regulated by nucleotide derivatives. Serine-rich regions are also found in receptors and transcription factors. The C-terminal region of transcriptional transactivators frequently features acidic domains. The pattern and localization of charged and polar amino acids suggests that E9 is a transcriptional transactivator. A most prominent feature of the E9 protein is the presence of two PHD zinc fingers. Members of the zinc-finger family are mostly DNA or RNA binding proteins, and factors that participate in protein–protein interactions. Among the members of the PHD family are the Drosophila proteins PcG [23], TrxG [32] and Pcl [33]. These proteins, as well as their identified yeast homologues, are parts of a multicomponent transcriptional activator that modulates the chromatin structure of the regulated genes, and maintains stable expression patterns for clustered genes (e.g. homeotic genes). Similar genes were found in plants; they are known as plant homeodomain proteins (hence the name PHD). Examples of human genes with PHD domains are HuHRX, a gene involved in chromosomal translocations in acute leukemia [34], HuXE169, a gene that escapes X chromosome inactivation [35], and transcription factors such as HuND4 [36].

The most relevant homology regarding the putative function of E9 is that observed with a novel member of the PHD family, called Requiem (REQ). This protein appears to be involved in the regulation of programmed cell death [37,38] It also has two PHD fingers. Functional expression and antisense studies suggest that its expression is a prerequisite for the activation of the apoptotic response. The E9 protein also has two putative nuclear localization signals. Altogether, the presence of nuclear localization signals in a protein with zinc-finger and transactivator domains strongly suggests that the E9 protein may have a regulatory function in the nucleus.

The time-course and dose-response pattern of E9 induction correlated with the estrogen-induced decrease of the cell proliferation rate and induction of genomic DNA degradation; the increase of E9 mRNA levels precedes the inhibition of cell proliferation by at least 10 h (Fig. 9).

In summary, the expression pattern of the E9 gene strongly suggests that it may be involved in mediating the estrogen-induced decrease of the proliferation rate and induction of apoptosis in E8CASS cells. From these data, one may speculate that E9 mRNA may play a central role in estrogen-mediated breast cancer regression.

Acknowledgements

This work was supported in part by PHS NIH grants AG-13807, CA-13410, CA-55574, and NSF DCB-8711746, and the Breast Cancer Research Program form the Massachusetts Department of Public Health. We are thankful to Dr. Nazneen Aziz for editorial suggestions, and to Cheryl Michaelson and Lisa Stagon for their excellent technical help. The assistance of The Center for Reproductive Research at Tufts University (P30 HD 28897) is gratefully acknowledged.

References

[1] Anonymous, Carcinoma of the breast, ACOG Technical Bulletin Number 158, August 1991, International Journal of Gynaecology and Obstetrics 39 (1992) 237–243.
[2] W.L. McGuire, A.K. Tandon, D.C. Allred, G.C. Chamness, P.M. Ravdin, G.M. Clark, Prognosis and treatment decisions in patients with breast cancer without axillary node involvement, Cancer 70 (1992) 1775–1781.
[3] Anonymous, Cooperative breast cancer group: results of studies of the cooperative breast cancer group 1961–1963, Cancer Chemother. Rep 41 (1964) 1–24.
[4] J.N. Ingle, D.L. Ahman, S.J. Green, Randomized clinical trial of DES versus tamoxifen in post-menopausal women with advanced breast cancer, N. Engl. J. Med. 304 (1981) 16–21.

[5] M.J. Boyer, M.H.N. Tattersall, Diethylstilbestrol revisited in advanced breast cancer management, Med. Pediat. Oncol. 18 (1990) 317–320.
[6] C. Sonnenschein, J. Szelei, T.L. Nye, A.M. Soto, Control of cell proliferation of human breast MCF7 cells serum and estrogen resistant variants, Oncology Res. 6 (1994) 373–381.
[7] Z. Wang, D.D. Brown, A gene expression screen, Proc. Natl. Acad. Sci. USA 88 (1991) 11505–11509.
[8] H.D. Soule, C.M. McGrath, Estrogen responsive proliferation of clonal human breast carcinoma cells in athymic mice, Cancer Letters 10 (1980) 177–189.
[9] A.M. Soto, C. Sonnenschein, The role of estrogens on the proliferation of human breast tumor cells (MCF-7), J. Steroid Biochem. 23 (1985) 87–94.
[10] C. Sonnenschein, A.M. Soto, C.L. Michaelson, Human serum albumin shares the properties of estrocolyone-I, the inhibitor of the proliferation of estrogen-target cells, J. Steroid Biochem. Molec. Biol. 59 (1996) 147–154.
[11] J.L. Tilly, J.W. Hsueh, Microscale autoradiographic method for the qualitative and quantative analysis of apoptotic DNA fragmentation, J. Cell Physiol. 154 (1993) 519–526.
[12] M. Nurminskaya, T.F. Linsenmayer, Identification and characterization of up-regulated genes during chondrocyte hypertrophy, Developmental Dynamics 206 (1996) 260–271.
[13] W.R. Pearson, D.J. Lipman, Improved tools for biological sequence comparison, Proc. Natl. Acad. Sci. USA 85 (1988) 2444–2448.
[14] J. Sambrook, E.F. Fritsch, T. Maniatis. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[15] T. Tanaka, F. Shibasaki, M. Ishikawa, N. Hirano, R. Sakai, J. Nishida, T. Takenawa, H. Hirai, Molecular cloning of bovine actin-like protein, actin2, Biochem. Biophys. Res. Comm. 187 (1992) 1022–1028.
[16] M.D. Welch, A.H. dePace, S. Verma, A. Iwamatsu, T.J. Mitchison, Arp2/3 complex is composed of evolutionarily conserved subunits and is localized to cellular regions of dynamic actin filament assembly, J. Cell Biol. 138 (1997) 375–384.
[17] M. Kozak, An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs, Nucleic Acids Res. 15 (1987) 8125–8148.
[18] C. Fyrberg, E. Fyrberg, A Drosophila homologue of the *Schizosaccharomyces pombe* act2 gene, Biochem. Genet. 31 (1993) 329–341.
[19] J.P. Lees-Miller, G. Henry, D.M. Helfman, Identification of act2, as essential gene in the fission yeast *Schizosaccharomyces pombe* that encodes a protein related to actin, Proc. Natl. Acad. Sci. USA 89 (1992) 80–83.
[20] T. Nagase, N. Seki, K. Ishikawa, M. Ohira, Y. Kawarabayasi, O. Ohara, A. Tanaka, H. Kotani, N. Miyajima, N. Nomura, Prediction of the coding sequences of unidentified human genes. Part VI: The coding sequences of 80 new genes (KIAA0201–KIAA0280) deduced from analysis of cDNA clones from cell line KG-1 and brain, DNA Res. 3 (1996) 321–329.
[21] T. Nagase, N. Seki, K. Ishikawa, M. Ohira, Y. Kawarabayasi, O. Ohara, A. Tanaka, H. Kotani, N. Miyajima, N. Nomura, Prediction of coding sequences of unidentified human genes. Part VI: The coding of 80 new genes (KIAA0201–KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain (Supplement), DNA Res. 3 (1996) 341–354.
[22] G. Shaw, R. Kamen, A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degredation, Cell 46 (1986) 659–667.
[23] R. Aasland, T.J. Gibson, A.F. Stewart, The PHD finger: implications for chromatin-mediated transcriptional regulation (Review, 29 refs), Trends Biochem. Sci. 20 (1995) 56–59.
[24] K.A. Thompson, B. Wang, W.S. Argraves, F.G. Giancotti, D.P. Schranck, E. Ruoslahti, BR140 a novel zinc-finger protein with homology to the TAF250 subunit of TFIID, Biochem. Biophys. Res. Comm. 198 (1994) 1143–1152.
[25] P.A. Ioannou, C.T. Amemiya, J. Garnes, P.M. Kroisel, H. Shizuya, C. Chen, M.A. Batzer, P.J. de Jong, A new bacteriophage P1-derived vector for the propagation of large human DNA fragments, Nature Genet. 6 (1994) 84–89.
[26] A. Lux, W. Bardenheuer, D. Michael, F. Brocker, K. Julicher, L. Vieten, S. Michaelis, S. Seeber, B. Opalka, J. Schutte, Identification of novel 'expressed sequence tags' within the FHIT gene locus in human chromosome region 3p142, Hum. Genet. 100 (1997) 90–95.
[27] D. Castiglia, M. Scaturro, T. Nastasi, A. Cestelli, I. DiLiegro, PIPPin, a putative RNA-binding protein specifically expressed in the rat brain, Biochem. Biophys. Res. Comm. 218 (1996) 390–394.
[28] A. Debant, C. Serra-Pages, K. Seipel, S. O'Brien, M. Tang, S.-H. Park, M. Streuli, The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase contains a protein kinase domain and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains, Proc. Natl. Acad. Sci. USA 93 (1996) 5466–5471.
[29] K. Langnaese, C. Seidenbecher, H. Wex, B. Seidel, K. Hartung, U. Appeltauer, A. Garner, B. Voss, B. Mueller, C.C. Garner, E.D. Gundelfinger, Protein components of a rat brain synaptic junctional protein preparation, Molecular Brain Research 42 (1996) 118–122.
[30] M.D. Welch, J. Rosenblatt, J. Skoble, D.A. Portnoy, T.J. Mitchison, Interaction of human Arp2/3 complex and the Listeria monocytogenes ActA protein in actin filament nucleation, Science 281 (1998) 105–108.
[31] D. McCollum, A. Feoktistova, M. Morphew, M. Balasubramanian, K.L. Gould, The Schizosaccharomyces pombe actin-related protein Arp3 is a component of the cortical actin cytoskeleton and interacts with profilin, EMBO J. 15 (1996) 6438–6446.
[32] J.A. Kennison, Transcriptional activation of Drosophila homeotic genes from distant regulatory elements (Review, 35 refs), Trends in Genetics 9 (1993) 75–79.
[33] A. Lonie, R. D'Andrea, R. Paro, R. Saint, Molecular characterisation of the Polycomblike gene of Drosophila melanogaster, a trans-acting negative regulator of homeotic gene expression, Development 120 (1994) 2629–2636.
[34] Y. Gu, T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C.M. Croce, The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene related to Drosophila trithorax, to the AF-4 gene, Cell 71 (1992) 701–708.
[35] J. Wu, J. Ellison, E. Salido, P. Yen, T. Mohandas, L.J. Shapiro, Isolation and characterization of XE169, a novel human gene that escapes X-inactivation, Human Molecular Genetics 3 (1994) 153–160.
[36] A.V. Chestkov, I.D. Baka, M.V. Kost, G.P. Georgiev, V.L. Buchman, The d4 gene family in the human genome, Genomics 36 (1996) 174–177.
[37] T.G. Gabig, P.L. Mantel, R. Rosli, C.D. Crean, Requiem — a novel zinc finger gene essential for apoptosis in myeloid cells, J. Biol. Chem. 269 (1994) 29515–29519.
[38] T.G. Gabig, C.D. Crean, A. Klenk, H. Long, N.G. Copeland, D.J. Gilbert, D. Quincey, F. Parente, F. Lespinasse, G.F. Carle, P. Gaudray, A. Calender, J. Hoeppener, K. Kas, R.V. Thakker, F. Farnebo, B.T. Teh, C. Larsson, F. Piehl, J. Lagercrantz, S. Khodaei, E. Carson, G. Weber, Expression and chromosomal localization of the Requiem gene, Mammalian Genome 9 (1998) 660–665.