PERGAMON

Journal of Steroid Biochemistry & Molecular Biology 86 (2003) 219–224

The Journal of Steroid Biochemistry & Molecular Biology

www.elsevier.com/locate/jsbmb

The human *CYP19* (aromatase P450) gene: update on physiologic roles and genomic organization of promoters☆

Serdar E. Bulun[a,*], Siby Sebastian[a], Kazuto Takayama[b], Takashi Suzuki[c], Hironobu Sasano[a], Makio Shozu[d]

[a] *Department of Obstetrics and Gynecology and Molecular Genetics, University of Illinois at Chicago, 820 S. Wood St. M/C 808, Chicago, IL 60612, USA*
[b] *Department of Obstetrics and Gynecology, Tohoku University, Sendai, Japan*
[c] *Department of Pathology, Tohoku University, Sendai, Japan*
[d] *Department of Obstetrics and Gynecology, Kanazawa University, Kanazawa, Japan*

Abstract

The human *CYP19* (P450arom) gene is located in the chromosome 15q21.2 region and is comprised of a 30 kb coding region and a 93 kb regulatory region. The Internet-based Human Genome Project data enabled us to elucidate its complex organization. The unusually large regulatory region contains 10 tissue-specific promoters that are alternatively used in various cell types. Each promoter is regulated by a distinct set of regulatory sequences in DNA and transcription factors that bind to these specific sequences. In most mammals, P450arom expression is under the control of gonad- and brain-specific promoters. In the human, however, there are at least eight additional promoters that seemed to have been recruited throughout the evolution possibly via alterations in DNA. One of the key mechanisms that permit the recruitment of such a large number of promoters seems to be the extremely promiscuous nature of the common splice acceptor site, since activation of each promoter gives rise splicing of an untranslated first exon onto this common junction immediately upstream of the translation start site in the coding region. These partially tissue-specific promoters are used in the gonads, bone, brain, vascular tissue, adipose tissue, skin, fetal liver and placenta for physiologic estrogen biosynthesis.

The most recently characterized promoter (I.7) was cloned by analyzing P450arom mRNA in breast cancer tissue. This TATA-less promoter accounts for the transcription of 29–54% of P450arom mRNAs in breast cancer tissues and contains endothelial-type *cis*-acting elements that interact with endothelial-type transcription factors, e.g. GATA-2. We hypothesize that this promoter may upregulate aromatase expression in vascular endothelial cells. The in vivo cellular distribution and physiologic roles of promoter I.7 in healthy tissues, however, are not known.

The gonads use the proximally located promoter II. The normal breast adipose tissue, on the other hand, maintains low levels of aromatase expression primarily via promoter I.4 that lies 73 kb upstream of the common coding region. Promoters I.3 and II are used only minimally in normal breast adipose tissue. Promoters II and I.3 activities in the breast cancer, however, are strikingly increased. Additionally, the endothelial-type promoter I.7 is also upregulated in breast cancer. Thus, it appears that the prototype estrogen-dependent malignancy breast cancer takes advantage of four promoters (II, I.3, I.7 and I.4) for aromatase expression. The sum of P450arom mRNA species arising from these four promoters markedly increase total P450arom mRNA levels in breast cancer compared with the normal breast.

Keywords: Aromatase; *CYP19*; Estrogen; Alternative transcription; Tissue-specific promoters; Chromosome 15q21.2; Human genome project; Ovary; Adipose tissue; Placenta; Breast cancer; Endometriosis; Skin; Brain; Vascular endothelial cell

1. Introduction

The *CYP19* gene localized at chromosome 15q21.2[1] encodes aromatase that is the key enzyme for estrogen biosynthesis [1,2]. In most mammals, aromatase is expressed only in gonads and the brain, whereas primates express this gene in additional extragonadal sites [1]. Aromatase expression and estrogen production continuously increase, as evolutionary tree progresses and reach the maximum in the human. This is achieved by the more efficient use of existing promoters and recruitment of additional novel tissue-specific promoters in fat, skin, placenta and the bone [1]. Estrogen is essential in females for the development of reproductive organs, and in both sexes for bone mineralization and gonadal function [3,4]. Moreover, in estrogen-dependent pathologic

☆ Presented at the VI International Aromatase conference: AROMATASE 2002, Kyoto, Japan, 26–30 October 2002.

* Corresponding author. Tel.: +1-312-996-8197; fax: +1-312-996-4238.
E-mail address: sbulun@uic.edu (S.E. Bulun).

[1] The chromosome location of the *CYP19* gene was originally reported as 15q21.1 [2]. The current information based on the Human Genome Project, however, identifies the precise location as 15q21.2.

0960-0760/$ – see front matter © 2003 Published by Elsevier Ltd.
doi:10.1016/S0960-0760(03)00359-5

tissues such as breast cancer and endometriosis, aromatase is upregulated via inappropriate activation of aberrant promoters [5,6]. Alternative use of multiple promoters, which regulate mature aromatase mRNA levels by splicing of each first exon or 5′-untranslated region (5′-UTR) onto a common splice junction immediately upstream of the coding region, is the key molecular mechanism conferring tissue-specific expression of the *CYP19* gene. For instance, the proximal ovary-specific promoter (PII) gives rise to a 5′-UTR contiguous with the first coding exon (exon II), whereas, a constitutively active distal promoter (I.1) in placenta is the basis of strikingly elevated levels of circulating estrogen (100–1000 times normal) in pregnant women [7,8]. Since all mRNA species contain the identical open reading frame (exons II–X), the encoded protein is the same regardless of the promoter used. Some of these promoters do not contain canonical TATA and CAAT elements, and each promoter is regulated in response to a distinct set of hormones or cytokines [1].

2. The genomic organization of the *CYP19* (P450arom) gene

The size of the *CYP19* (aromatase) gene has been estimated to be larger than 80 kb [1]. Characterization of the entire genomic organization and the accurate size of this large gene, however, had not been possible to date using conventional phage or cosmid genomic libraries [1]. Recently published Human Genome Project Data allowed us for the first time to precisely locate all known promoters and elucidate the extraordinarily complex organization of the entire human *CYP19* gene (Fig. 1).

Nucleotide sequence information of various 5′-UTRs and the coding region of the *CYP19* gene were subjected to the Basic Local Alignment Search Tool (BLAST) homology search against the High Throughput Genomic Sequence (HTGS) database of the National Center for Biotechnology Information [9]. GenBank Accession numbers of the previously characterized unique 5′-UTRs and corresponding

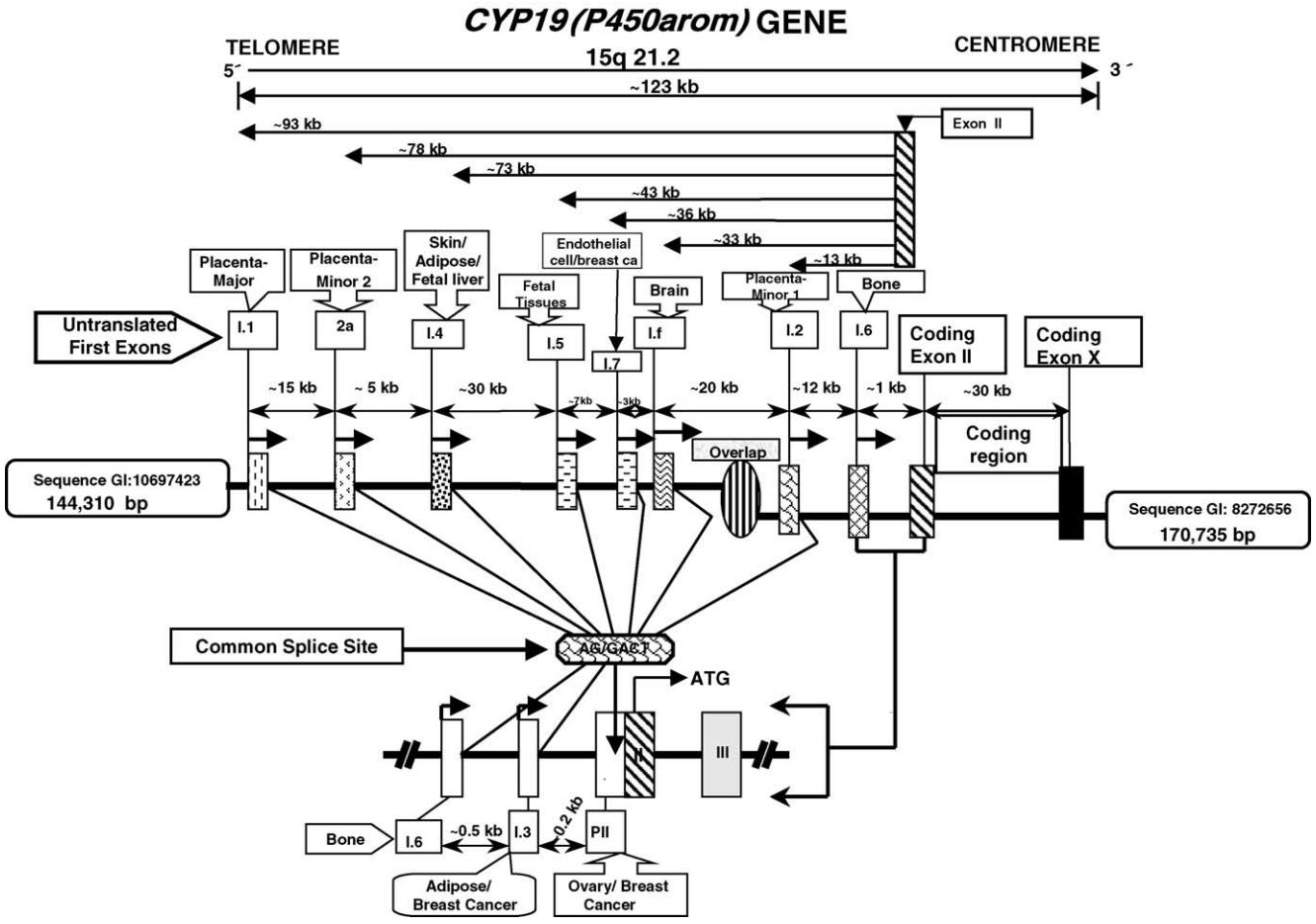


Fig. 1. Mapping the promoters of the *CYP19* (P450arom) gene to the human genome. Model depicting location of each tissue-specific promoter of the P450arom gene in the chromosome 15q21.2 region. BLAST searches of various promoters and the coding region revealed alignment to distinct locations in two overlapping BAC clones (RP11-522G20 and RP11-108K3) of the chromosome 15q21.2 region. The distance of each promoter with respect to the first coding exon (exon II) was also determined. The major placental promoter I.1 is the most distally located (approximately 93-kb). Even though each tissue expresses a unique first exon or 5′-untranslated region (5′-UTR), by splicing into a highly promiscuous splice acceptor site (AG/GACT) of the exon II, the coding region and the translated protein product are identical in all the tissues.

promoters of the human *CYP19* gene utilized for pairwise local sequence alignments were: M22246 (human full-length aromatase cDNA), S52794 (ovary-specific promoter PII), D21241 (adipose/breast cancer, I.3), S96437 (placenta-minor, I.2), D29757 (brain, I.f), S71536 (fetal tissues, I.5), L21982 (skin & adipose, I.4), D14473 (placenta-minor, 2a), and X55983 (placental-major, I.1). Sequences of bone-specific 5′-UTR I.6 and endothelial-specific I.7 were obtained from the original publication [10,11]. BLAST search permitted us to align the *CYP19* gene coding region (exons II–X) and promoters II, I.3, I.6 and I.2 to a 178,762 bp bacterial artificial chromosome (BAC) clone (RP11-522G20) mapped to chromosome 15q21.2 (GenBank accession no: AC012169). Similarly, the remaining promoters were precisely aligned within another 144,714 bp BAC clone (RP11-108K3) mapped to chromosome15q21.2 (GenBank accession no: AC020891). Next, using BLAST 2 search program (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html), we set out to examine the alignment between these two clones and results showed an overlap of a 6442-bp region, demonstrating continuity of the gene sequence in these BAC clones (Fig. 1).

Based on an analysis of these two BAC clones, we demonstrated that the entire gene spans over approximately 123 kb of DNA. Additionally, the *CYP19* gene is found between markers stSG12786 and stSG47530 with the 3′-end of the gene centromeric to the 5′-end of the gene, showing the direction of transcription as from telomere to centromere. Only the 30 kb 3′-region encodes aromatase, whereas a large 93 kb 5′-flanking region serves as the regulatory unit of the gene. The most proximal gonadal-specific promoter II and two other proximal promoters, I.3 (expressed in adipose tissue and breast cancer) and I.6 (expressed in bone) are found to be located within the 1 kb region upstream of the ATG translation start site in exon II, as expected (Fig. 1). Promoter I.2, the minor placenta-specific promoter, is located approximately 13 kb upstream of the ATG site in exon II. The promoters specific for the brain (I.f), endothelial cells (I.6), fetal tissues (I.5), adipose tissue (I.4) and placenta (2a and I.1) are localized in tandem order at ∼33, 36, 43, 73, 78 and 93 kb, respectively, upstream of the first coding exon, the exon II. Each promoter gives rise to a P450arom mRNA with a specific 5′-untranslated end for that particular promoter but with an identical coding region (Fig. 2).

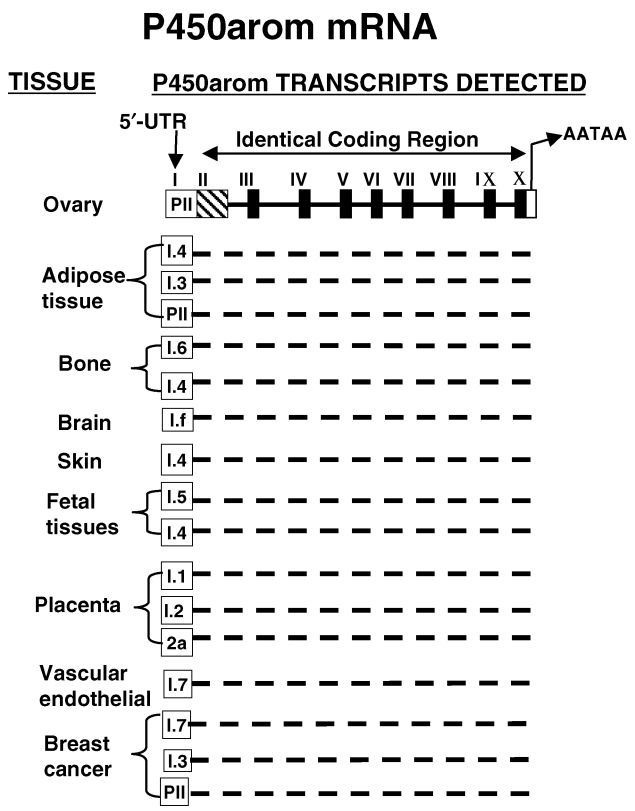


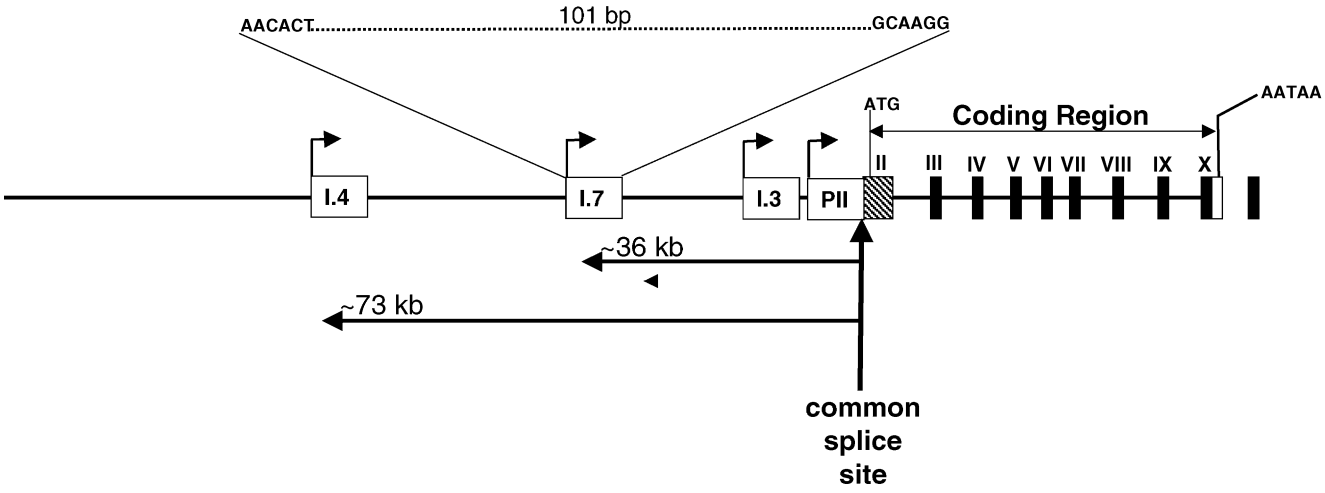
Fig. 2. Diagrammatic sketch of the structure of P450arom mRNA species found in various human tissues. The untranslated first exon (5′-UTR) of mRNA species may be viewed as a signature of the alternatively used promoter in that particular tissue.

3. A genomic mechanism for recruitment of multiple promoters throughout evolution

The placental promoter I.1 located approximately 93 kb upstream of the coding region is the most distally located promoter, which gave rise to splicing of a 103 bp first exon onto the common splice junction immediately (38-bp) upstream of the ATG translational start site (Fig. 3). The activity of promoter I.1 is the basis for 100–1000 times elevated levels of circulating estrogen in pregnant women [12,13]. Thus, recruitment of this most distal promoter may have an evolutionary impact, since, of all species, humans are unique to acquire and maintain extraordinarily high levels of aromatase expression in placenta.

A transgene containing the human genomic region upstream of the placental exon I.1 was found to exhibit readily detectable promoter activity in the mouse placenta [14]. Thus, although mouse placental tissue does not express aromatase endogenously, it contains the necessary transcriptional factors for human promoter I.1 expression. This is suggestive that the recruitment of aromatase expression in human placenta occurred via alterations in the mammalian genome throughout the evolution. One of the key mechanisms that permit the recruitment of such a large number of promoters seems to be the extremely promiscuous nature of the common splice acceptor site, since activation of each promoter gives rise to splicing of an untranslated first exon onto this common junction immediately upstream of the translation start site in the coding region.

Definition of the complete structure and organization of the human *CYP19* gene in its fine details will facilitate further characterization of various molecular mechanisms by which the tissue-specific and temporal expression of this gene is regulated, in normal tissues and pathological conditions such as breast cancer and endometriosis.



DISTRIBUTION OF P450arom PROMOTERS IN BREAST CANCER SAMPLES (n=5):

Number of 5′-RACE Clones sequenced	I.7	PII	I.3	I.4	Total
Breast Cancer 1	4	3	4	3	14
Normal Breast				10	10

5′-RACE/ Colony hybridization	Percentage of I.7-specific clones
Breast Cancer 2	39%
Breast Cancer 3	42%
Breast Cancer 4	42%
Breast Cancer 5	54%

Fig. 3. Exon I.7. A novel first exon (5′-UTR) of P450arom mRNA species was originally cloned from human breast cancer tissues. This 5′-UTR was mapped to 36 kb upstream of the translation start site. We determined the distribution of I.7-specific mRNA in 5′-rapid amplification of cDNA ends (5′-RACE) libraries from five cancer specimens. This suggested that 29–54% of P450arom mRNA species in breast cancer were directed by promoter I.7.

This review also highlights a potential use of the Human Genome Project data and powerful bioinformatics tools (especially the freely available web based programs) to disseminate information about the structure and organization of very large and complex genes. In absence of this freely available data base and other resources, this type of work is not only technically challenging to many laboratories, but also require significant amount of time and effort devoted to traditional molecular biology techniques such as cloning, library screening and sequencing.

4. Endothelial-type promoter I.7 is upregulated in breast cancer tissue

We recently cloned a novel 101 bp untranslated first exon (I.7) that comprises the 5′-end of 29–54% of P450arom mRNA species isolated from breast cancer tissues [11] (Fig. 3). The levels P450arom mRNA with exon I.7 were significantly increased in breast cancer tissues and adipose tissue adjacent to tumors (Fig. 4). We identified a promoter immediately upstream of exon I.7 and mapped this to about 36 kb upstream of ATG translation start site of the P450arom gene [11] (Fig. 3). Sequence analysis of I.7 revealed a TATA-less promoter containing an initiator, two consensus GATA sites and *cis*-regulatory elements found in megakaryocytic and endothelial-type promoters (Fig. 5). Promoter activity directed by a −299/+81 sequence was four-fold greater than a minimum length promoter sequence (−35/+81) in human microvascular endothelial cells, but only two-fold greater in MCF-7 breast cancer cell line (Fig. 5). There was no promoter activity in primary breast adipose fibroblasts. Site-directed mutations and DNAse I protection demonstrated the maximal basal promoter activity required GATA, *Ets* and E47 motifs in this critical −299/−35 region (Fig. 5). Gel shift demonstrated the binding of GATA-2 protein to the −196/−191 bp GATA motif (Fig. 5). Overexpression of GATA-2 in endothelial cells increased promoter I.7 activity by five-fold. In conclusion, promoter I.7 is GATA-2-regulated endothelial-type promoter of the human P450arom gene and may increase estrogen biosynthesis in vascular endothelial cells of breast cancer. The activity of this promoter may also be important for intracrine and paracrine effects of estrogen on blood vessel physiology.

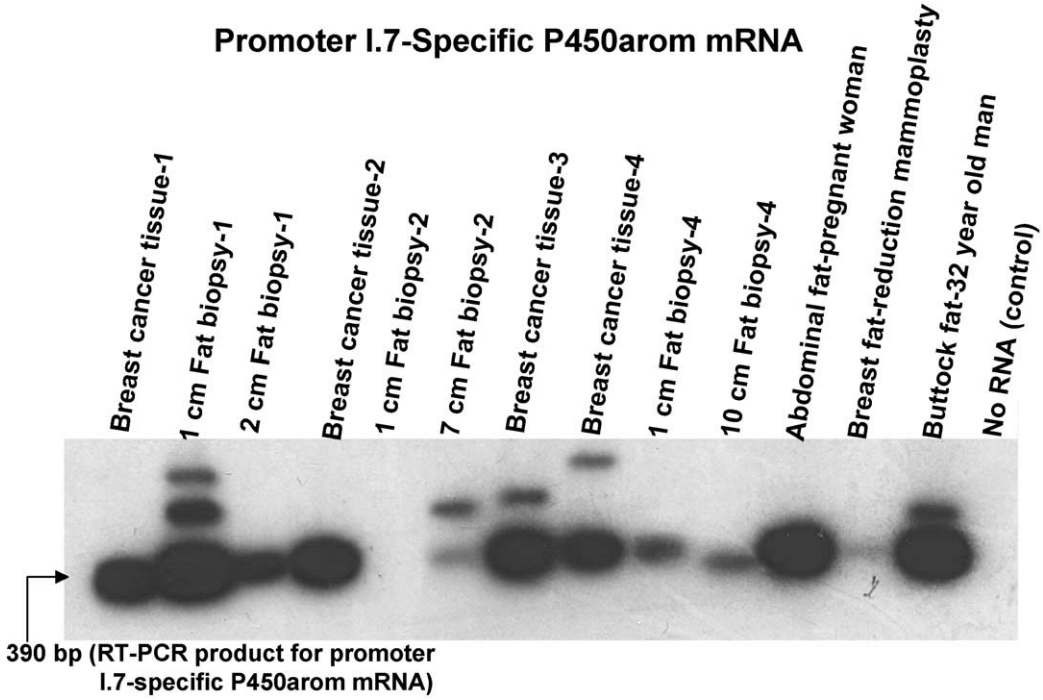


Fig. 4. RT-PCR to amplify P450arom mRNA species with the exon I.7-specific sequence was conducted using total RNA from breast cancer tissue, breast cancer tissue adjacent to the carcinoma, breast adipose tissue from a cancer-free reduction mammoplasty specimen, abdominal adipose tissue from a pregnant woman and buttock adipose tissue from a normal man. A gradient of I.7-specific mRNA in mastectomy specimens was observed. The levels increased towards the cancer tissue in three mastectomy specimens. Breast fat from a disease-free woman contained very low I.7-specific mRNA levels. The significance of this promoter in male buttock and pregnant female abdominal fat tissues is not known. Equal loading was confirmed by RT-PCR amplification of glyceraldehyde-3-phosphate dehydrogenase mRNA (data not shown).

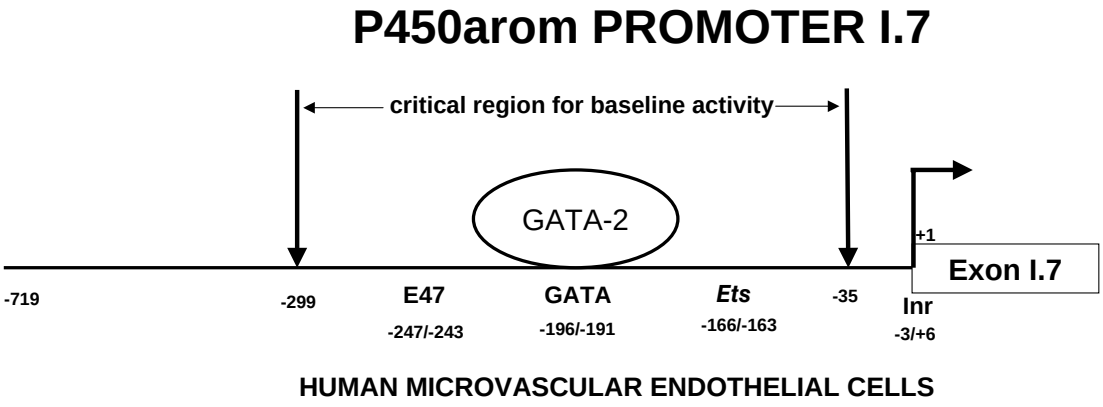


Fig. 5. P450arom promoter I.7 is a TATA-less promoter and contains an initiator (Inr) or cap signal. The −299/−35 region confers maximum basal activity in endothelial cells. This regulatory region contains at least three critical endothelial-type motifs including *Ets*, GATA and E47. The binding of GATA-2 protein to the −196/−191 bp site is important for baseline promoter activity.

5. A genetic mechanism for increased aromatase expression in breast cancer: additive effects of multiple promoters

The normal breast adipose tissue maintains low levels of aromatase expression primarily via promoter I.4 that lies 73 kb upstream of the common coding region (Fig. 6). Promoters I.3 and II are used only minimally in normal breast adipose tissue. Promoters II and I.3 activities in the breast cancer, however, are strikingly increased [6]. Additionally, the endothelial-type promoter I.7 is also upregulated in breast cancer [11]. Thus, it appears that the prototype estrogen-dependent malignancy breast cancer takes advantage of four promoters (II, I.3, I.7 and I.4) for aromatase expression (Fig. 6). The sum of P450arom mRNA species arising from these four promoters markedly increase total

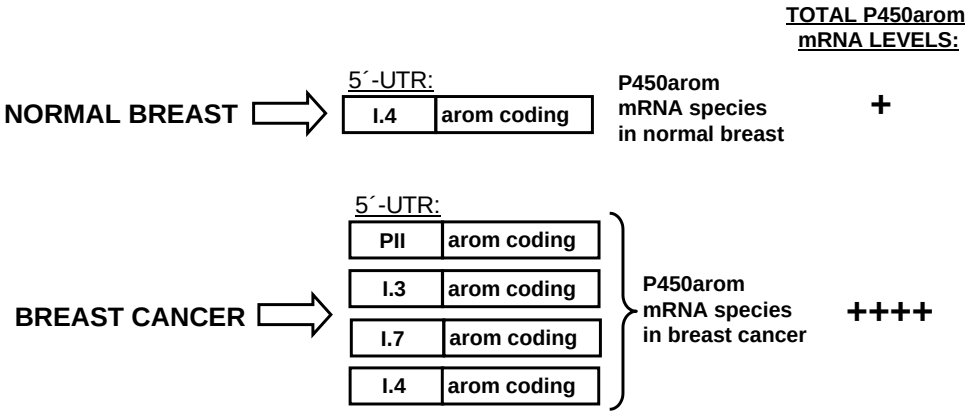


Fig. 6. Distribution of promoter-specific P450arom mRNA species found in normal and malignant breast tissues. The levels of total P450arom mRNA levels in breast cancer tissue are strikingly higher than normal breast tissue. The normal breast adipose tissue maintains low levels of aromatase expression primarily via promoter I.4. Promoters I.3 and II are used only minimally in normal breast adipose tissue. Promoters II and I.3 activities in the breast cancer, however, are strikingly increased. Additionally, the endothelial-type promoter I.7 is also upregulated in breast cancer. Thus, it appears that the prototype estrogen-dependent malignancy breast cancer takes advantage of four promoters (II, I.3, I.7 and I.4) for aromatase expression. The sum of P450arom mRNA species arising from these four promoters markedly increase total P450arom mRNA levels in breast cancer compared with the normal breast.

P450arom mRNA levels in breast cancer compared with the normal breast that uses almost exclusively promoter I.4 (Fig. 6).

Acknowledgements

Grant Support: NIH grants CA67167 and HD38691 to Serdar E. Bulun.

References

[1] E.R. Simpson, M.S. Mahendroo, G.D. Means, M.W. Kilgore, M.M. Hinshelwood, S. Graham-Lorence, B. Amarneh, Y. Ito, C.R. Fisher, M.D. Michael, C.R. Mendelson, S.E. Bulun, Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis, Endocr. Rev. 15 (1994) 342–355.

[2] S.A. Chen, M.J. Besman, R.S. Sparkes, S. Zollman, I. Klisak, T. Mohandas, P.F. Hall, J.E. Shively, Human aromatase: cDNA cloning, Southern blot analysis, and assignment of the gene to chromosome 15, DNA 7 (1988) 27–38.

[3] E.P. Smith, J. Boyd, G.R. Frank, H. Takahashi, R.M. Cohen, B. Specker, T.C. Williams, D.B. Lubahn, K.S. Korach, Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man, N. Engl. J. Med. 331 (1994) 1056–1061.

[4] A. Morishima, M.M. Grumbach, E.R. Simpson, C. Fisher, K. Qin, Aromatase deficiency in male and female siblings caused by a novel mutation and the physiological role of estrogens, J. Clin. Endocrinol. Metab. 80 (1995) 3689–3698.

[5] K. Zeitoun, K. Takayama, M.D. Michael, S.E. Bulun, Stimulation of aromatase P450 promoter (II) activity in endometriosis and its inhibition in endometrium are regulated by competitive binding of steroidogenic factor-1 and chicken ovalbumin upstream promoter transcription factor to the same *cis*-acting element, Mol. Endocrinol. 13 (1999) 239–253.

[6] V.R. Agarwal, S.E. Bulun, M. Leitch, R. Rohrich, E.R. Simpson, Use of alternative promoters to express the aromatase cytochrome P450 (*CYP19*) gene in breast adipose tissues of cancer-free and breast cancer patients, J. Clin. Endocrinol. Metab. 81 (1996) 3843–3849.

[7] M. Shozu, K. Akasofu, T. Harada, Y.J. Kubota, A new cause of female pseudohermaphroditism: placental aromatase deficiency, J. Clin. Endocrinol. Metab. 72 (1991) 560–566.

[8] S.E. Bulun, Clinical review 78, aromatase deficiency in women and men: would you have predicted the phenotypes? J. Clin. Endocrinol. Metab. 81 (1996) 867–871.

[9] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.

[10] M. Shozu, Y. Zhao, S.E. Bulun, E.R. Simpson, Multiple splicing events involved in regulation of human aromatase expression by a novel promoter, I.6, Endocrinology 139 (1998) 1610–1617.

[11] S. Sebastian, K. Takayama, M. Shozu, S.E. Bulun, Cloning and characterization of a novel endothelial promoter of the human *CYP19* (aromatase P450) gene that is upregulated in breast cancer tissue, Mol. Endocrinol. 16 (2002) 2243–2254.

[12] K. Toda, L.X. Yang, S.J. Shizuta, Transcriptional regulation of the human aromatase cytochrome P450 gene expression in human placental cells, Steroid Biochem. Mol. Biol. 53 (1995) 181–190.

[13] K.J. Ryan, Biological aromatization steroids, J. Biol. Chem. 234 (1959) 268–272.

[14] A. Kamat, K.H. Graves, M.E. Smith, J.A. Richardson, C.R. Mendelson, A 500-bp region, approximately 40 kb upstream of the human *CYP19* (aromatase) gene mediates placenta-specific expression in transgenic mice, Proc. Natl. Acad. SCI. U.S.A. 96 (1999) 4575–4580.