Control of vitellogenin genes expression by sequences derived from transposable elements in rainbow trout.

Anthony Bouter, Nicolas Buisine, Adélaïde Le Grand, Nathalie Mouchel, Franck Chesnel, Catherine Le Goff, Véronique Le Tilly, Jacques Wolff, Olivier Sire

Control of vitellogenin genes expression by sequences derived from transposable elements in rainbow trout

Anthony Bouter[a], Nicolas Buisine[b], Adélaïde Le Grand, Nathalie Mouchel[c], Franck Chesnel[d], Catherine Le Goff[d], Véronique Le Tilly*, Jacques Wolff and Olivier Sire

Laboratoire d'Ingénierie des Matériaux de Bretagne, Université de Bretagne-Sud, CER Yves Coppens, BP573, 56017 Vannes CEDEX, France

[a] *present address: Molecular Imaging and NanoBioTechnology, UMR 5248 CBMN, CNRS-Université Bordeaux 1-ENITAB, IECB, 2 rue Robert Escarpit, 33607 Pessac, France.*

[b] *present address: Museum National d'Histoire Naturelle, 57 Rue Cuvier, 75231 Paris Cedex 05, France.*

[c] *present address: Paediatric Molecular Genetics, Institute of Molecular Medicine, Oxford University, John Radcliffe Hospital, Oxford OX3 9DS, UK.*

[d]*CNRS UMR 606, Institut Génétique et Développement de Rennes, IFR140, Campus de médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes CEDEX, France.*

* *Corresponding author:* Tel. 33 297 017 135; Fax. 33 297 017 071

E-mail address: letilly@univ-ubs.fr (V. Le Tilly)

Keywords: estrogen response element, estrogen receptor, vitellogenin, rainbow trout, transfection assays, fluorescence anisotropy.

Abbreviations: TE, transposable element; TSS, transcriptional start site; cyc, cytochrome-c oxidase; DO, drop-out supplement; E2, 17β-estradiol; E_S, estrogens; ER, estrogen receptor; ERE(s), estrogen response element(s); ERE_{cs}, consensus estrogen response element; hER,

human estrogen receptor; OD, optical density; rtER, rainbow trout estrogen receptor; rtVTG, rainbow trout vitellogenin.

Abstract

In most of oviparous animals, vitellogenins (VTG) are the major egg yolk precursors. They are produced in the liver under the control of estrogens. In rainbow trout (*Oncorhynchus mykiss*), the *vtg* genes cluster contains an unusually large number of almost identical gene copies. In order to identify the regulatory elements in their promoters, we used a combination of reporter plasmids containing genomic sequences including putative estrogen response elements (EREs) and we performed transient transfection assays in MCF-7 and yeast cells. We found a functional ERE corresponding to the sequence GGGGCAnnnTAACCT (*rtvtg*ERE), which differs from the consensus ERE (ERE_{cs}) by three base pairs. This non-palindromic ERE is located in the *env* gene of a retrotransposon relic, 180 base pairs upstream of the transcriptional start site. Fluorescence anisotropy experiments confirmed that the purified human estrogen receptor α (hERα) can specifically bind to *rtvtg*ERE. Furthermore, we observe that the stability of hERα-ERE_{cs} and hERα-*rtvtg*ERE complexes is similar with equilibrium dissociation constants of 3.0 nM and 6.2nM respectively, under our experimental conditions. Additionally, this *rtvtg*ERE sequence displays a high E2-responsiveness through ER activation *in cellulo*.

In the rainbow trout, the functional ERE (*rtvtg*ERE) lies within promoter sequences which are mostly composed of sequences derived from transposable elements (TEs), which therefore may have acted as an evolutionary buffer to secure the proper expression of these genes.

1. Introduction

Estrogens (E_S) signaling pathways are key components of many biological processes such as differentiation, growth and embryogenesis, and are thus critical for many life traits. E_S biological activity is mediated by the estrogen receptor (ER), a member of the superfamily of nuclear receptors. In the presence of E_S, ER binds to specific DNA sequences called Estrogen Response Elements (EREs) [1]. These short sequences are usually located in the promoter, although some have been identified in introns or exons [2]. Recently, genome-wide analysis found many functional ER binding sites located at large distances from the transcriptional start site [3-4]. It is well established that ER has the highest affinity for a 15 bp sequence composed of two 6 bp inverted repeats separated by a 3 bp spacer [2, 5]. This sequence, **AGGTCA**nnn**TGACCT**, is designated as the consensus ERE sequence (ERE_{cs}). The ERE_{cs} sequence is rarely found in natural promoters of E_S-regulated genes; in fact, a multitude of imperfect palindromic-like ERE sequences has been identified as functional EREs. Additionally, by using natural and synthetic imperfect EREs, it has been shown that single nucleotide alteration in each half-site of the ERE palindrome affects more the ER binding and its transcriptional activity than if two mutations occur in only one half-site of the ERE [2]. Besides, the spacer size between the two half-sites also affects the binding affinity and the conformation of the receptor in human estrogen receptor-ERE complex: human estrogen receptors bind strongly to ERE_{cs} exhibiting no spacer or with a spacer size of 3 bp between half-sites. In comparison, the association is much lower with a spacer size of 1 or 2 bp [6].

Vertebrate genomes usually encode two distinct ERs, ERα and ERβ, which significantly differ in their biological activities [7]. In rainbow trout, ERα is present as two isoforms generated by alternative splicing ($rtER_S$ and $rtER_L$) [8]. The $rtER_S$ expression is restricted to the liver where it is the dominant isoform, whereas the $rtER_L$ expression pattern is more ubiquitous, suggesting a specific role of $rtER_S$ in vitellogenesis [9]. Human ERα (hERα) and $rtER_S$ are well-conserved (92% and 60% similarities for the DNA-binding domain and the

ligand-dependent transactivation domain, respectively), except for the ligand-independent transactivation domain, which is poorly conserved (20% similarity). hERα and $rtER_S$ also exhibit important functional differences [10]: (i)- $rtER_S$ has a marked transcriptional activity in the absence of estrogens; (ii)- $rtER_S$ needs a 10-fold higher estradiol (E2) concentration to achieve maximal transactivation compared to hERα; (iii)- $rtER_S$ displays a weaker transactivation activity compared to hERα, in yeast assays with a reporter gene containing one, two or three copies of ERE_{cs}. It is important to note that the binding of E2 to $rtER_S$, but not to hERα, was shown to be temperature sensitive [11].

Vitellogenin (VTG) is the major precursor of egg yolk proteins which are essential for the early development of non-mammalian vertebrates. VTG, produced by the liver of the mature female, is mainly under the control of estrogens. Besides inducing *vtg* genes transcription, E_S also increase the stability of the corresponding messenger RNA [12-13]. VTG is then secreted into the bloodstream and selectively incorporated into the growing oocytes [14]. Because of their remarkable E2-mediated stimulation, *vtg* genes have for long been the model of choice to decipher the molecular mechanisms of transcriptional regulation by ERs. Most of our knowledge of the basic mechanisms of transcriptional regulation is based on *Xenopus laevis* and chicken *vtg* genes [15-18]. Besides, the *tilapia vtg* gene promoter [19] was shown to contain several regions exhibiting more than 70% similarity with the *X. laevis vtg*A2 gene promoter. This, together with additional functional analysis [20], suggest that many features of *vtg* gene expression have been conserved through evolution between teleosts and tetrapods.

VTG are usually encoded by small multigene families which mostly form a *vtg* gene cluster in a conserved syntenic group [21]. The promoter structure of *vtg* genes is generally quite simple [15-16, 19, 22-23], consisting of a consensus or imperfect EREs plus additional enhancers located close to the transcriptional start site (TSS). In salmonids, two paralogous vtg gene clusters arose from an ancestral tetraploïdization, at the base of salmonid radiation. In the course of evolution, *Oncorhynchus* species have retained only one cluster [24]. In the

rainbow trout (*Oncorhynchus mykiss*), this cluster contains about 20 highly conserved genes, and functional genes, plus 10 truncated pseudogenes for which the truncation breakpoint corresponds to a putative retrotransposon located in intronic sequences [12]. Genes and pseudogenes are arranged in a head-to-tail orientation, typical of tandemly arrayed genes subject to concerted evolution [25-26]. Repeated units are separated from each other by a highly conserved 4.6 kb intergenic region which is mostly composed of transposable elements (TEs)-related sequences. Strikingly, these sequences are found very close to the TSS, raising the question whether they participate functionally or interfere with *vtg* genes expression. They show no sequence similarity with other known *vtg* promoter sequences, *ie tilapia, Xenopus* and chicken.

In this paper, we describe the identification and the physicochemical characterization of a functional imperfect ERE which drives *rtvtg* genes transcription. This *rtvtg*ERE is located in TE-related sequences that compose almost all the promoters of *rtvtg* gene, suggesting that these sequences participate in the regulation of the expression of *vtg* genes and were co-opted during the course of evolution. Therefore, the recent reshaping of the structure and organization of *rtvtg* genes provides us with a unique opportunity to decode the interplay between the evolution of gene structure and the regulation of gene expression.

2. Materials and methods

2.1. Vector construction

DNA fragments containing putative EREs were obtained by PCR from clone S5 of the *rtvtg* gene [27] and cloned in a luciferase reporter plasmid (pGL2-b, Promega). *Mlu*I and *Bgl*II restriction sites were included in primers to facilitate cloning. DNA was amplified between positions -480 to +22 and -140 to +22, with respect to the transcription initiation site. The PCR products were cloned into the *Mlu*I/*Bgl*II digested pGL2-b plasmid, giving the $p_{-480/+22}$ and $p_{-140/+22}$ vectors. The resulting plasmids contain the first 21 nucleotides of the *rtvtg* exon 1, in which the ATG (position +18) was mutated (ATA) to prevent interferences with reporter gene. All constructs were controlled by sequencing. pGL2-b (promoterless plasmid) and pGL2-p (SV40 promoter) were used as negative and positive controls respectively.

β-galactosidase reporter vectors are derivatives of the YRPE2 vector, in which 2 ERE_{cs} are located upstream of a minimal *cyc* promoter [28]. A unique *Xho*I site upstream of the 2 ERE_{cs} and a unique *Bam*HI in the *lac*Z gene permitted the replacement of this region by appropriate sequences prepared in pBK plasmid (Stratagene). The *cyc*-*lac*Z sequence was amplified from YRPE2 with primers containing *Xho*I-*Bgl*II sites (forward) and *Bam*HI site (reverse) and subcloned into pBK at *Xho*I and *Bam*HI sites (pBKcyc). Genomic fragments encompassing the putative EREs were amplified using primers containing *Xho*I (forward) and *Bgl*II (reverse) sites. These sequences were subcloned into the *Xho*I-*Bgl*II sites of pBKcyc and *Xho*I-*Bam*HI fragment was used to replace the corresponding fragment in YRPE2. The resulting plasmids were designated $Y_{-450/-130}$, $Y_{-260/-130}$, $Y_{-210/-130}$, $Y_{-450/-190}$ and $Y_{-450/-310}$. The same strategy was used to construct vectors with a single or no consensus ERE (Y_{EREcs} and Y_{cyc}, respectively). Site mutations were introduced in $Y_{-450/-130}$ and the oligonucleotides used for mutation of the -180 bp fragment were GCTAAATGGCAGTG**CC**G**A**Aggt**A**AACCTAACCTTTAT and

ATAAAGGTTAGGTT**T**accT**TCGG**CACTGCCATTTAGC (bold represents mutated nucleotides).

Yeast $rtER_S$ and hERα expression plasmids (pY60$rtER_S$ and pY60hERα, respectively) are derivatives of pYeDP60-AQPcic [29] from which the cDNA encoding AQPcic was excised by *Eco*RI and *Sac*I and replaced by the complete coding sequence of the $rtER_S$ cDNA amplified from pCMV5/$rtER_S$ [30], or the human ERα cDNA amplified from YEPE15 [28]. The plasmid has the yeast ura marker and the bacterial amp^R selection markers. $rtER_S$ and hERα expression is under the control of a GAL10-cyc promoter and can be induced by 2% galactose.

2.2. Cell culture and transient expression assays

MCF-7 human breast cancer cells were seeded at 10^6 cells per 60 mm dish in phenol red free Dubelcco's modified medium and Ham's nutrient mix F12 (DMEM Ham F12) supplemented with 10% Fetal Calf Serum (FCS), 10 U/ml penicillin, 10 µg/ml streptomycin, 0.0025 µg/ml amphotericin, 25 mM Hepes and 4.8 mM of bicarbonate. After 24 h of incubation at 37 °C in this medium (DMEM Ham F12 + 10 % FCS), cells were kept for 12 h in DMEM Ham F12 plus 5 % dextran coated charcoal stripped FCS (DCC-FCS). MCF-7 cells were then transfected by the calcium phosphate procedure using 5 µg of reporter plasmids (pGL2-b, pGL2-p, $p_{-140/+22}$ or $p_{-480/+22}$). The plasmid pCH110 (Pharmacia) containing the β-galactosidase reporter gene was used as an internal control. To improve transfection efficiency, a glycerol shock was performed 6 h after transfection. Cells were then incubated with DMEM Ham F12 supplemented with 1 % DCC-FCS with or without E2 (10^{-8} M). Cells were harvested 36 h after transfection. Luciferase activity was measured by scintillation counting with the luciferase assay system (Promega), and counts were normalized to levels of β-galactosidase activity. Displayed values are the means of 3 independent experiments.

Yeast *Saccharomyces cerevisiae* cells W303.1B (α, leu2, his3, trp1, ura3, ade2-1, can^R, and cyr+) were grown in a rich YPRE medium (2% tryptone, 1% yeast extract, 0.5% raffinose, pH 7 and 3% ethanol) or selective SD medium (0.67% nitrogen base without amino acids, 1% raffinose or 2% glucose, pH 5.8, plus drop-out supplements: DO-Ura or DO-Ura-Leu). Yeast cells were transformed with YRPE2-derivative reporter plasmids and/or expression plasmids by the lithium acetate chemical method (Yeast protocols handbook, Clontech). Transformed yeast cells were grown in the selective medium at 170 rpm and 30°C up to $OD_{600nm,l=1cm}$ = 1 and diluted 10 times into the YPRE medium. They were grown again to $OD_{600nm,l=1cm}$ = 0.6 before induction with 2% galactose, and/or hormonal stimulation with 17β-estradiol (E2), for 16 h. Activity determined by β-galactosidase assays (Yeast protocols handbook, Clontech) was expressed in Miller units according to:

$$Activity = \frac{1000 \times OD_{420nm,l=1cm}}{t \times V \times OD_{600nm,l=1cm}} \quad (1)$$

with t, the time of reaction (min) at 30°C and V the culture volume used for the assay (0.05 mL). Displayed values are the means of at least 3 independent experiments. The data of the dose-response curves have been fitted according to the Hill equation from which the recovered value of EC_{50} was extracted:

$$Activity = Activity_{[L]\to 0} + \frac{a \times [L]^n}{EC_{50} + [L]^n} \quad (2)$$

2.3. In vitro binding assays

Double stranded oligonucleotide solutions were prepared from complementary 21 bp-long oligonucleotides synthesized (with labeling or not), purified and adjusted to 100 µM by the Proligo company (Paris, France). The sense strand of each 21 bp oligonucleotide sequence was labeled with fluorescein at the 5' end. These oligonucleotide sequences are GTCAGGTCACAGTGACCTGAT (ERE_{cs}), AGTGGGGCAGGTTAACCTAAC ($pERE_2$),

AGTG**CC**GA**A**AGGT**A**AACCTAAC (mutated pERE2), CCCACGTAAAACTGACCATCC ($pERE_1$) and CCATTAGACCGTTAGG (mim, a negative control corresponding to a Myb response element). Double stranded oligonucleotides were used at a working concentration of 1 nM. Purified human ERα expressed from baculovirus was provided by PanVera Corporation at a concentration of 3 µM. One hour before experiments, protein samples were diluted four times in buffer B (10 mM Tris-HCl, 140 mM KCl, pH 7.5, 1 mM DTT, 0.1 mM EDTA and 10% glycerol) and kept in ice for the duration of the experiments. The experiment could not be carried out with rtERs because we could not produce a functional protein. We note that functional estrogen receptors are notoriously difficult to over-express and purify. Instead of using electro-mobility gel shift assay (EMSA), K_D values have been inferred from anisotropy fluorescence experiments. As opposed to the limited window of experimental conditions available with EMSA (especially pH), fluorescence anisotropy has the considerable advantage of being able probe the physical properties of protein-DNA interaction in a wide range of experimental conditions, and thus providing us with biologically relevant parameters.

Fluorescein fluorescence anisotropy values were monitored with a 480 nm linear polarized excitation light (vertical or horizontal) using an SLM 8100 spectrofluorometer. Fluorescein emission anisotropy (A) was obtained from parallel, $F_{//}$, and perpendicular, $F_{\perp}$, emission components:

$$A = \frac{F_{//} - F_{\perp}}{F_{//} + 2 \times F_{\perp}} \quad (3)$$

Parallel and perpendicular fluorescence intensities were monitored through a 520 nm cut-off Oriel filter. For each displayed anisotropy value, the solvent contribution was subtracted. Fluorescence anisotropy measurements were carried out with an integration time of 5 s on each emission component. Each binding curve was performed at 10°C and repeated at least 3 times. The values of the dissociation equilibrium constant (K_D) were inferred by nonlinear

least-squares curve fitting using commercial Peakfit software and by assuming that only one equilibrium is observed within the range of protein concentration being used.

3. Results

3.1. The promoter of rtvtg genes is composed almost entirely of sequences derived from TEs

In the rainbow trout genome, *vtg* genes are arranged in tandem arrays. They are separated by a 4.6 kb highly conserved intergenic region (Fig. 1). Sequence analysis between -1090 and +410 of a *vtg* gene did not reveal blocks of regulatory elements such as those described for the *Xenopus* and chicken genes [15, 18]. No obvious similarity in the organization and sequence of the promoter region was observed between the *rtvtg* and other vertebrate *vtg* genes. This is not surprising considering that most of the 5' region of *rtvtg* genes corresponds to sequences derived from retroelements [31]: the region between positions -1380 and -290 encodes for a truncated reverse transcriptase of a LINE-type retrotransposon (50% similarity with a *jockey* element) whereas the sequence from positions -1495 to -1380 displays 70% similarities with an *env* coding sequence (position 41 to 69, EMBL accession number: U56288) of a HIV-type retrovirus [31]. Additional sequence comparisons carried out with BLASTX against NCBI's databases and REPEATMASKER against a collection of salmonid specific TE sequences (http://lucy.ceh.uvic.ca/repeatmasker/cbr_repeatmasker.py) revealed that the sequence between positions -191 and -107 also shares significant similarities (64%) with the HIV *env* coding sequence (position 176 to 196, EMBL accession number: U56288), strongly suggesting that the *jockey*-like element has been inserted into a retrovirus-like element already present in the *rtvtg* promoter region. In addition, we found that the sequence between -3987 and -2851 is related to a *Salmo salar* LINE-like element (SsaL2.1). Altogether, these data show that almost all the entire intergenic region up to position -107 (which contains the promoter) unambiguously originates from TE-like sequences, leaving a particularly short proximal sequence, putatively corresponding to the ancestral promoter. We note that the TE dynamics and the extent of genomic rearrangement in the intergenic region were much greater than initially reported [24, 31].

3.2. Putative imperfect ERE located in TE sequences within intergenic region

Sequence analysis of the *rtvtg* promoter revealed no ERE_{cs} but two motifs corresponding to putative imperfect EREs ($pERE_1$ and $pERE_2$) within the first ~ 450 bp proximal sequences (Fig. 1). An additional motif located, at -220, is composed of two half EREs separated by a single nucleotide spacer and almost certainly does not correspond to a structural and functional ERE [5-6, 32]. Compared to the consensus sequence, the two putative identified EREs, denoted as $pERE_1$ (A<u>C</u>GT<u>A</u>AaacTGACC<u>A</u>) and $pERE_2$ (<u>G</u>GG<u>G</u>CAggtT<u>A</u>ACCT), exhibit 3 nucleotide differences distributed in the two half-sites. Importantly, both putative EREs are located in sequences derived from TEs. Nonetheless, given that half EREs may act in synergism, we cannot rule-out that they functionally participate to the regulation of *rtvtg* genes expression. Therefore, we undertook *in cellulo* and *in vitro* experiments to probe the biological activity of these putative EREs, acting alone or synergistically. To this end, we first tested their ability to induce reporter gene expression in an E2-dependent manner. We then characterized the physical properties of the interactions between putative EREs and ER, with *in vitro* binding experiments.

3.3. A putative ERE exhibits an E2-dependent activity in MCF-7 cells

In order to assess the ability of the region containing the putative EREs to mediate E2-dependent gene induction, the fragments, -480 to +22 and -140 to +22, were inserted in the promoterless plasmid pGL2-b, upstream of the luciferase reporter gene. The resulting plasmids, $p_{-480/+22}$ and $p_{-140/+22}$, were used to transfect breast cancer MCF-7 cells, which naturally express hER. Luciferase activity was measured with or without E2 treatment (10^{-8} M). Control experiments were carried out in similar conditions with MCF-7 cells transfected with pGL2-p vector, containing the SV40 promoter, and showed no E2-dependent activity (Fig. 2). Cells transfected with $p_{-140/+22}$, devoid of putative ERE, exhibit no significant luciferase activity with and without E2 treatment. Luciferase activities measured with this

construct were similar to those of the promoterless pGL2-b vector and thus represent the background level of the assay. Cells transformed with the construct $p_{-480/+22}$, in the absence of E2, also showed a background level of luciferase activity, whereas an increased luciferase activity (X ~12) is observed upon E2 stimulation. This result shows that the sequence between -480 and -140 contains a transcriptional enhancer whose activity is dependent on E2.

3.4. An imperfect ERE is located in TE-like sequences within 480 bp upstream of the transcription start site of rtvtg

The functional characterization of the putative EREs was performed by transfection assay in yeast expressing either hERα or $rtER_S$. We used this simple cell system over human breast cancer cells because it allows to study, in the same cellular context, the properties of two ERs originating from different species (namely human -hERα- and rainbow trout -$rtER_S$-), towards consensus and imperfect EREs. This system also has the advantage of being appropriate to monitor ligand-dependent transactivation mediated by both ERs [28, 33-34], without the confounding effects of endogenous steroid factors.

Yeast cells were co-transformed with a plasmid expressing ER (pY60ER) under the control of a galactose-inducible promoter and with a β-galactosidase reporter plasmid under the control of one ERE_{cs} (Y_{EREcs}) or various genomic sequences located between -450 and -130 positions from *rtvtg* gene. Five constructs were tested: $Y_{-450/-130}$ containing both pEREs, $Y_{-450/-190}$ and $Y_{-450/-310}$ containing only $pERE_1$, $Y_{-260/-130}$ and $Y_{-210/-130}$ containing only $pERE_2$.
Cells containing only the reporter construct (*i.e.* without ER expression vector or non-induced expression of ER) exhibited only a basal level of β-galactosidase activity (< 1 Miller unit) with or without E2 treatment (data not shown).

In order to compare the ability of hERα or $rtER_S$ to induce reporter gene expression, cells were co-transformed with pY60hERα or pY60$rtER_S$ together with the Y_{EREcs} plasmid. ER expression was induced with addition of galactose and activities were measured in the

absence or in the presence of 10^{-7} M of E2 (Fig. 3). Preliminary experiments showed that this concentration induced maximal β-galactosidase activity. In the absence of E2, only cells expressing $rtER_S$ exhibited a significant activity. Following E2-stimulation, β-galactosidase activity was strongly increased in hERα-expressing cells whereas $rtER_S$-expressing cells exhibited an activity increased by twice in comparison to without E2. This control experiment clearly indicates that hERα transcriptional activity is strictly E2-dependent, whereas $rtER_S$ exhibits a marked E2-independent activity when acting towards ERE_{cs}, as previously described [8, 10].

The functional analysis of the two putative ERE-like sequences was performed in yeast cells expressing hERα, or $rtER_S$, and transfected with reporter plasmids (Fig. 4A and 4B), which contain a single or a combination of the putative ERE sequences identified *in silico*. Yeast cells co-transformed with the expression plasmid and a reporter plasmid containing either no ERE (Y_{cyc}; lane 1) or one ERE_{cs} (Y_{EREcs}; lane 7) were used as negative and positive controls, respectively.

Cells expressing ER (hERα or $rtER_S$) transformed with $Y_{-450/-190}$ (lane 5) or $Y_{-450/-310}$ (lane 6) exhibited only background level of β-galactosidase activity. This result suggests that there is no E2-dependent regulatory sequence within positions -450 and -190 and that $pERE_1$ plays no direct role in ERs E2-dependent transactivation. In contrast, cells transformed with any of the three constructs containing $pERE_2$ ($Y_{-450/-130}$, $Y_{-260/-130}$ or $Y_{-210/-130}$, lanes 2-4), displayed a strong E2-induced β-galactosidase activity, which can be mediated by both hERα and $rtER_S$.

This result indicates that the region between positions -190 and -130 is responsible for E2-stimulation, strongly suggesting that $pERE_2$ sequence is the active ERE.

3.5. E2 responsiveness of pERE2-containing sequence

We examined the E2 responsiveness of pERE2-containing sequence with dose-response curves carried out with yeast cells transformed with the reporter plasmids, Y_{EREcs} (positive

control) or $Y_{-210/-130}$ (containing pERE2) or $Y_{-450/-310}$ (containing the non E2-responsive $pERE_1$ as a negative control), together with hERα or $rtER_S$ expression plasmid (Fig. 5 and Table 1). As expected, the $Y_{-450/-310}$ construct did not yield significant levels of β-galactosidase activity, at whatever E2 concentration. In contrast, the Y_{EREcs} and $Y_{-210/-130}$ constructs responded strongly to E2, with hERα or $rtER_S$. Indeed, the EC_{50} (that is, the concentration of E2 corresponding to 50% of the maximal response) is very similar for both ERs, although the EC_{50} measured with $Y_{-210/-130}$ is consistently 4.5 times higher than that of Y_{EREcs} (3.1 and 14.2 *vs* 4.1 and 18.5 with Y_{EREcs} and $Y_{-210/-130}$ constructs, respectively).

In agreement with previous reports [10-11], we found marked functional differences between hERα and $rtER_S$: (i)- the maximal level of reporter activity of the $Y_{-210/-130}$ construct with $rtER_S$ is only half that of hERα (~ 63 *vs* ~ 135 Miller units), and (ii)- $rtER_S$ (but not hERα) displays a strong level of background expression (~ 100 Miller units) without E2 treatment, implying that it has a high E2-independent biological activity. Likewise, the induction factor mediated by $rtER_S$ is much lower than that of hERα (hERα : 17.0 and 10.3 *vs* $rtER_S$: 2.4 and 4.2, with Y_{EREcs} and $Y_{-210/-130}$ constructs, respectively), despite the fact that maximal induction levels measured with the control construct (Y_{EREcs}) are comparable for hERα and $rtER_S$ and they are reached at similar E2 concentration. These differences have important physiological implications which will be discussed below. Nonetheless, these results show that despite functional differences between hERα and $rtER_S$, the genomic region between -210 and -130 consistently drives an E2-dependent transcriptional induction. Given that this genomic region contains a single ERE-like motif, we speculate that $pERE_2$ corresponds to a functional ERE.

To check whether this putative ERE_2 is indeed functional, a site-directed mutagenesis to change the sequence of $pERE_2$ from GGGGCAggtTAACCT to GCCGAAggtAAACCT was performed from the $Y_{-450/-130}$ construct; the mutated bases have been previously demonstrated as making ERE sequence to be non-functional [35]. As shown in Fig. 5, the mutation of

$pERE_2$ leads to a loss of ERs transactivation in response to E2. Thus, $pERE_2$ sequence is responsible for the E2-responsiveness of *vtg* gene expression in rainbow trout.

3.6. ER specifically binds to the pERE2

The physical properties of the interactions between ER and the different putative EREs ($pERE_2$, $pERE_1$, ERE_{cs} or mim, a nonspecific sequence) were inferred from fluorescence anisotropy DNA-binding curves. This technology is remarkably sensitive and robust for measuring protein-DNA equilibrium constants [36-39]. We carried out titration curves with purified recombinant hERα but not with $rtER_S$, since we could not produce biologically active proteins, despite the use of several over-production and purification systems. These fluorescence anisotropy measurements were performed by using a fixed concentration of fluorescein-labeled oligonucleotides (1.0 nM) and various amounts of hERα, at 140 mM KCl. Different oligonucleotide sequences have been used: ERE_{cs}, $pERE_2$, mutated $pERE_2$, $pERE_1$ and mim. Data were fitted to recover the equilibrium dissociation constant (K_D). The fluorescence anisotropy binding curve of the hERα–ERE_{cs} complex provides an estimated K_D value of 3.0 ± 0.4 nM (Fig. 6A). Strikingly, a K_D value 2-fold higher (6.2 ± 0.3 nM) was observed with $pERE_2$ (Fig. 6B). Thus, at 140 mM KCl, hERα exhibits a similar affinity for the consensus ERE and for the natural $pERE_2$ sequence. Furthermore, with ERE_{cs} and $pERE_2$, anisotropy values increased twice following hERα-ERE complex formation, implying that the stoechiometry of the complex is the same in both cases. In contrast, the binding curve obtained with $pERE_1$ or mutated $pERE_2$ sequence was similar to that obtained with mim, indicating that hERα is unable to bind specifically to this sequence, thus confirming the fact that it is not a functional ERE.

These results unambiguously show that hERα specifically binds to the $pERE_2$ sequence with an affinity close to the value displayed for ERE_{cs}, confirming that E2-regulatory element within the genomic region between -210 and -190 is pERE2. As $pERE_2$ appears as the main

and unique ERE sequence inside *rtvtg* promoter, this sequence will be referred to as *rtvt*gERE from now on.

4. Discussion

4.1. A functional ERE resides in transposable element sequences

Sequence analysis of the *rtvtg* promoter revealed two putative imperfect EREs in the region between -480 and -140, relative to the TSS. This *in silico* analysis is based on well known ERE sequences: the canonical ERE inverted repeat AGGTCAnnnTGACCT, the direct repeat AGGTCAn$_{(>10bp)}$AGGTCA, the inverted repeat TGACCTnnnAGGTCA, a half ERE flanked by an AT rich region and the ERE variants found in Alu sequences (an imperfect ERE palindrome 9 bp upstream of a perfect ERE half-site) [2 and ref. therein]. There are two putative ERE inverted repeat sequences in the proximal promoter: ACGTAAaacTGACCA and GGGGCAggtTAACCT. The TGGACAtTGATCT sequence, found in the region, is composed of two imperfect half-sites separated by a single base pair. Therefore, it cannot correspond to a functional ERE [5-6, 32]. Since *in silico* sequence analysis tends to predict large numbers of putative EREs, of which only a fraction displays some biological activity, it is necessary to validate their biological activity experimentally. To this end, we probed the biological properties of ERE candidates *in cellulo* and *in vitro.*

The E2-dependent enhancer activity of the sequence spanning positions -480 to +22 of the *rtvtg* gene was first tested with a reporter assay in MCF-7 cells, a breast cancer cell line, that stably express hER. This is a well-established model system which has been used to characterize the regulation of gene expression by E$_S$, *e.g.* the identification of EREs controlling the expression of *Xenopus*, chicken and *tilapia vtg* genes [17, 38]. Our constructs include the natural proximal 5'-flanking sequence with the TATA box (TTAAAA) and 22 bp of 5' untranslated sequences of the *rtvtg* gene. The transient transfection assays indicate that a functional element driving E2-induced transcription is located between positions -480 and -140.

We next assessed the enhancer activity of this genomic region in yeast, with pY60ER and YRPE2-derivative plasmids. This convenient cellular system has long proved instrumental for the characterization of the rtER, the promoter of several E_S-inducible genes and to screen the estrogenic-potency of xenobiotics [6, 8, 31, 39]. This analysis restricted the position of the E2-dependent enhancer to a short region located between positions -130 and -190 from the TSS. This is an indication that the $pERE_2$ motif, located at position -180, may correspond to the enhancer. In contrast, the $pERE_1$ motif exhibits no biological activity, thus confirming that it is not a functional ERE. Strikingly, the $pERE_2$ sequence (GGGGCAggtTAACCT) differs from the ERE_{cs} (AGGTCAnnnTGACCT) by two and one mutations, in the first and second half-sites, respectively. This motif has partially lost its palindromy and features one of the most degenerated functional ERE motifs ever identified. Interestingly, recent genomic data suggest that binding of ER to degenerated ERE motifs is in fact quite common [40].

We used *in vitro* DNA-binding assay to address whether hERα specifically binds to the $pERE_2$ sequence. We show that, in our experimental conditions, hERα specifically binds to the $pERE_2$ sequence with a very high affinity, but not to the $pERE_1$ sequence. In agreement with previous work *in vitro*, we found that hERα has a slightly lower affinity for a degenerated ERE ($pERE_2$) compared to ERE_{cs} [2, 40], although in the case of $pERE_2$, it is only two fold lower. At higher salt concentration, hERα is known to display a much stronger preference for ERE_{cs} over degenerated EREs [36, 41], ensuring dynamic interactions and the fast biological responses. This should have strong physiological implication and might explain the frequent occurrence of degenerated EREs in promoter sequences.

As a whole, our results show that the $pERE_2$ sequence, now designated *rtvtg*ERE, is the functional element driving the E2-dependent *rtvtg* gene transcription.

4.2. Functional impact of ERE_{cs} and rtvtgERE: The driving force of rtvtg genes evolution?

Paradoxically, even though the ERE_{cs} is the optimal ER DNA-binding site and mediates the

strongest transcriptional induction, it is found in less than 10% of E_S-responsive genes [2]. Instead, imperfect motifs, rather than ERE_{cs}, seem to have been selected in the course of evolution. Thus, comparing the biological activity of the various ERE sequences might provide some clues about the mechanisms of transcriptional regulation mediated by E_s and their physiological relevance. In the absence of E2, a promoter construct containing an ERE_{cs} separated by 20 bp is sufficient to drive the transcription of the reporter gene with $rtER_S$. This strong E2-independent basal activity has already been observed in various cell systems (including yeast) and is assigned to the absence of the N-terminal domain [8, 10, 42]. In the presence of E2, transcription is further increased by 2.4 fold with construct containing one ERE_{cs}. $rtER_S$ is therefore an efficient basal transcriptional activator which weakly responds to E2 when bound to an ERE_{cs}. In sharp contrast, $rtER_S$ displays a much lower basal transcriptional activity but a high responsiveness to E2, when it is bound to *rtvtg*ERE. This point has important physiological consequences because E2-responsive genes will be properly expressed in the liver, where the $rtER_S$ isoform dominates, only if they are under the control of imperfect EREs (such as *rtvtg*ERE), not ERE_{cs}.

One might argue that the level of transcriptional induction driven by $rtER_S$ is too weak to account for the dramatic induction of *rtvtg* genes [43]. However, given that rainbow trout VTGs are encoded by ~20 (almost) identical genes, even a modest induction could be amplified by the large number of *rtvtg* genes. This functional hypothesis fits remarkably well with the current model of *rtvtg* genes evolution [24], in which the weak expression of an ancestral *rtvtg* gene would have been compensated by series of tandem amplification (see below). Therefore, the work presented here provides key molecular and functional clues for our understanding of *rtvtg* gene evolution and the interplay between the transposable elements and the evolution of gene regulation (see below).

4.3. Co-opted transposable elements and (non)evolution of gene regulation

Vtg genes structure and organization in rainbow trout are the result of a complex evolutionary process [24]. Our new structural and functional data support well the current model of *rtvtg* genes evolution, which proposes that the promoter of an ancestral *rtvtg* gene was disrupted by the insertion of two retrotransposons, resulting in reduced transcription levels. Subsequently, this weak expression level would have been rescued by local gene amplification, promoting the rise of the present-day gene organization. Given that VTG synthesis is restricted to the liver and controlled by E2 in the rainbow trout, this model implies that the functional impact of the TE sequences was limited to a reduction of the *rtvtg* transcription level, with little or no effect on its expression pattern and timing. Indeed, VTG synthesis mobilizes considerable resources and improper expression would have been severely counter-selected; no evidence of ectopic expression of VTG has been found so far in the rainbow trout. Perhaps surprisingly, we found that a key regulatory element resides in the remnants of TE sequences. This indicates that despite initiating the disruption of the ancestral promoter, they also contributed to restore the proper regulation of the *rtvtg* gene expression. Thus, TE sequences provided both the plague and the cure.

The fact that gene expression may be influenced by EREs located nearby TE sequences is not new, since it has already been described for the androgen receptor gene [44], the apolipoprotein A gene [45] and the BRCA-1 gene [46]. In these cases, TEs are likely to have significantly contributed to the evolution of the gene expression profiles by providing novel regulatory sequences.

In the case of *rtvtg* genes, however, they did not provide additional enhancers responding to new regulation pathways, but rather they provided regulatory elements helping to keep a similar expression profile. Thus, although they are usually regarded as a source of genomic and functional innovations, in the case of *rtvtg* genes, they may have acted instead as some kind of regulatory buffers, in order to preserve the proper expression of a protein of vital importance for egg survival and reproduction.

The fate of TEs in genomes is dictated by a delicate balance between opposite evolutionary forces: On one hand, they can be maintained by genetic drift, high transposition rate or because they provide some advantageous character. On the other hand, they can be purged by stochastic loss, their genetic burden (by selection) or self inactivation, to name a few [47]. Although their propensity to alter gene regulation plays an important role in genome expression dynamics, it also increases their genetic burden since they can drive the cell into aberrant transcriptional programs. Thus, the potential to buffer alterations of the transcriptional networks might balance the genetic burden associated with their biological activity and may represent a powerful strategy to secure their fate in genomes.

Acknowledgements

This work was supported by Le Ministère de l'Enseignement Supérieur et de la Recherche and Le Comité du Morbihan de la Ligue Nationale Contre le Cancer. The authors are grateful to J. Bouma and J. Nagler (University of Idaho, USA) for their technical help. The authors also thank Pr. O'Malley for the gift of the YRPE2 vector.

References

[1] R.M. Evans, The steroid and thyroid hormone receptor superfamily, Science 240 (1988) 889-895.

[2] C.M. Klinge, Estrogen receptor interaction with estrogen response elements, Nucleic Acids Res 29 (2001) 2905-2919.

[3] Y.F. Pan, K.D. Wansa, M.H. Liu, B. Zhao, S.Z. Hong, P.Y. Tan, K.S. Lim, G. Bourque, E.T. Liu, E. Cheung, Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements, J Biol Chem 283 (2008) 32977-32988.

[4] F. Parisi, B. Sonderegger, P. Wirapati, M. Delorenzi, F. Naef, Relationship between estrogen receptor alpha location and gene induction reveals the importance of downstream sites and cofactors, BMC Genomics 10 (2009) 381.

[5] C.J. Gruber, D.M. Gruber, I.M. Gruber, F. Wieser, J.C. Huber, Anatomy of the estrogen response element, Trends Endocrinol Metab 15 (2004) 73-78.

[6] S. El Marzouk, R. Gahattamaneni, S.R. Joshi, W.M. Scovell, The plasticity of estrogen receptor-DNA complexes: binding affinity and specificity of estrogen receptors to estrogen response element half-sites separated by variant spacers, J Steroid Biochem Mol Biol 110 (2008) 186-195.

[7] H.A. Harris, Estrogen receptor-beta: recent lessons from in vivo studies, Mol Endocrinol 21 (2007) 1-13.

[8] F. Pakdel, R. Metivier, G. Flouriot, Y. Valotaire, Two estrogen receptor (ER) isoforms with different estrogen dependencies are generated from the trout ER gene, Endocrinology 141 (2000) 571-580.

[9] A. Menuet, I. Anglade, G. Flouriot, F. Pakdel, O. Kah, Tissue-specific expression of two structurally different estrogen receptor alpha isoforms along the female

reproductive axis of an oviparous species, the rainbow trout, Biol Reprod 65 (2001) 1548-1557.

[10] F. Petit, Y. Valotaire, F. Pakdel, Differential functional activities of rainbow trout and human estrogen receptors expressed in the yeast Saccharomyces cerevisiae, Eur J Biochem 233 (1995) 584-592.

[11] J.B. Matthews, J.H. Clemons, T.R. Zacharewski, Reciprocal mutagenesis between human alpha (L349, M528) and rainbow trout (M317, I496) estrogen receptor residues demonstrates their importance in ligand binding and gene expression at different temperatures, Mol Cell Endocrinol 183 (2001) 127-139.

[12] G. Flouriot, H. Brand, S. Denger, R. Metivier, M. Kos, G. Reid, V. Sonntag-Buck, F. Gannon, Identification of a new isoform of the human estrogen receptor-alpha (hER-alpha) that is encoded by distinct transcripts and that is able to repress hER-alpha activation function 1, EMBO J 19 (2000) 4688-4700.

[13] M. Saceda, R.K. Lindsey, H. Solomon, S.V. Angeloni, M.B. Martin, Estradiol regulates estrogen receptor mRNA stability, J Steroid Biochem Mol Biol 66 (1998) 113-120.

[14] E.H. Lim, J.L. Ding, T.J. Lam, Estradiol-induced vitellogenin gene expression in a teleost fish, *Oreochromis aureus*, General and Comparative Endocrinology 82 (1991) 206-214.

[15] J.B. Burch, M.I. Evans, T.M. Friedman, P.J. O'Malley, Two functional estrogen response elements are located upstream of the major chicken vitellogenin gene, Mol Cell Biol 8 (1988) 1123-1131.

[16] J.B. Burch, A.H. Fischer, Chromatin studies reveal that an ERE is located far upstream of a vitellogenin gene and that a distal tissue-specific hypersensitive site is conserved for two coordinately regulated vitellogenin genes, Nucleic Acids Res 18 (1990) 4157-4165.

[17] D.L. Davis, J.B. Burch, The chicken vitellogenin II gene is flanked by a GATA factor-dependent estrogen response unit, Mol Endocrinol 10 (1996) 937-944.

[18] L. Klein-Hitpass, M. Schorpp, U. Wagner, G.U. Ryffel, An estrogen-responsive element derived from the 5' flanking region of the Xenopus vitellogenin A2 gene functions in transfected human cells, Cell 46 (1986) 1053-1061.

[19] B.Y. Teo, N.S. Tan, E.H. Lim, T.J. Lam, J.L. Ding, A novel piscine vitellogenin gene: structural and functional analyses of estrogen-inducible promoter, Mol Cell Endocrinol 146 (1998) 103-120.

[20] B.Y. Teo, N.S. Tan, T.J. Lam, J.L. Ding, Synergistic effects of nuclear factors--GATA, VBP and ER in potentiating vitellogenin gene transcription, FEBS Lett 459 (1999) 57-63.

[21] P.J. Babin, Conservation of a vitellogenin gene cluster in oviparous vertebrates and identification of its traces in the platypus genome, Gene 413 (2008) 76-82.

[22] L. Klein-Hitpass, M. Kaling, G.U. Ryffel, Synergism of closely adjacent estrogen-responsive elements increases their regulatory potential, J Mol Biol 201 (1988) 537-544.

[23] E. Martinez, W. Wahli, Cooperative binding of estrogen receptor to imperfect estrogen-responsive DNA elements correlates with their synergistic hormone-dependent enhancer activity, EMBO J 8 (1989) 3781-3791.

[24] N. Buisine, V. Trichet, J. Wolff, Complex evolution of vitellogenin genes in salmonid fishes, Mol Genet Genomics 268 (2002) 535-542.

[25] D. Liao, Concerted evolution: molecular mechanism and biological implications, Am J Hum Genet 64 (1999) 24-30.

[26] D. Liao, T. Pavelitz, J.R. Kidd, K.K. Kidd, A.M. Weiner, Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves

rapid intrachromosomal homogenization and rare interchromosomal gene conversion, EMBO J 16 (1997) 588-598.

[27] N. Mouchel, V. Trichet, B.Y. Naimi, J.P. Le Pennec, J. Wolff, Structure of a fish (Oncorhynchus mykiss) vitellogenin gene and its evolutionary implication, Gene 197 (1997) 147-152.

[28] D.P. McDonnell, Z. Nawaz, C. Densmore, N.L. Weigel, T.A. Pham, J.H. Clark, B.W. O'Malley, High level expression of biologically active estrogen receptor in Saccharomyces cerevisiae, J Steroid Biochem Mol Biol 39 (1991) 291-297.

[29] V. Lagree, I. Pellerin, J.F. Hubert, F. Tacnet, F. Le Caherec, N. Roudier, D. Thomas, J. Gouranton, S. Deschamps, A yeast recombinant aquaporin mutant that is not expressed or mistargeted in Xenopus oocyte can be functionally analyzed in reconstituted proteoliposomes, J Biol Chem 273 (1998) 12422-12426.

[30] F. Pakdel, F. Petit, I. Anglade, O. Kah, F. Delaunay, T. Bailhache, Y. Valotaire, Overexpression of rainbow trout estrogen receptor domains in Escherichia coli: characterization and utilization in the production of antibodies for immunoblotting and immunocytochemistry, Mol Cell Endocrinol 104 (1994) 81-93.

[31] V. Trichet, N. Buisine, N. Mouchel, P. Moran, A.M. Pendas, J.P. Le Pennec, J. Wolff, Genomic analysis of the vitellogenin locus in rainbow trout (Oncorhynchus mykiss) reveals a complex history of gene amplification and retroposon activity, Mol Gen Genet 263 (2000) 828-837.

[32] K. Pettersson, K. Grandien, G.G. Kuiper, J.A. Gustafsson, Mouse estrogen receptor beta forms estrogen response element-binding heterodimers with estrogen receptor alpha, Mol Endocrinol 11 (1997) 1486-1496.

[33] D. Metzger, J.H. White, P. Chambon, The human oestrogen receptor functions in yeast, Nature 334 (1988) 31-36.

[34] T.A. Pham, Y.P. Hwung, D. Santiso-Mere, D.P. McDonnell, B.W. O'Malley, Ligand-dependent and -independent function of the transactivation regions of the human estrogen receptor in yeast, Mol Endocrinol 6 (1992) 1043-1050.

[35] A.D. Adamson, S. Friedrichsen, S. Semprini, C.V. Harper, J.J. Mullins, M.R. White, J.R. Davis, Human prolactin gene promoter regulation by estrogen: convergence with tumor necrosis factor-alpha signaling, Endocrinology 149(2008) 687-694.

[36] M. Boyer, N. Poujol, E. Margeat, C.A. Royer, Quantitative characterization of the interaction between purified human estrogen receptor alpha and DNA using fluorescence anisotropy, Nucleic Acids Res 28 (2000) 2494-2502.

[37] V. Le Tilly, C. Royer, Fluorescence anisotropy assays implicate protein-protein interactions in regulating trp repressor DNA binding, Biochemistry 32 (1993) 7753-7758.

[38] M.S. Ozers, J.J. Hill, K. Ervin, J.R. Wood, A.M. Nardulli, C.A. Royer, J. Gorski, Equilibrium binding of estrogen receptor with DNA using fluorescence anisotropy, J Biol Chem 272 (1997) 30405-30411.

[39] L. Zargarian, V. Le Tilly, N. Jamin, A. Chaffotte, O.S. Gabrielsen, F. Toma, B. Alpert, Myb-DNA recognition: role of tryptophan residues and structural changes of the minimal DNA binding domain of c-Myb, Biochemistry 38 (1999) 1921-1929.

[40] A.S. Cheng, V.X. Jin, M. Fan, L.T. Smith, S. Liyanarachchi, P.S. Yan, Y.W. Leu, M.W. Chan, C. Plass, K.P. Nephew, R.V. Davuluri, T.H. Huang, Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters, Mol Cell 21 (2006) 393-404.

[41] A. Bouter, V. Le Tilly, O. Sire, Interplay of flexibility and stability in the control of estrogen receptor activity, Biochemistry 44 (2005) 790-798.

[42] R. Metivier, F.G. Petit, Y. Valotaire, F. Pakdel, Function of N-terminal transactivation domain of the estrogen receptor requires a potential alpha-helical structure and is negatively regulated by the A domain, Mol Endocrinol 14 (2000) 1849-1871.

[43] C. Vaillant, C. Le Guellec, F. Pakdel, Y. Valotaire, Vitellogenin gene expression in primary culture of male rainbow trout hepatocytes, Gen Comp Endocrinol 70 (1988) 284-290.

[44] C. Ramakrishnan, D.M. Robins, Steroid hormone responsiveness of a family of closely related mouse proviral elements, Mamm Genome 8 (1997) 811-817.

[45] Z. Yang, D. Boffelli, N. Boonmark, K. Schwartz, R. Lawn, Apolipoprotein(a) gene enhancer resides within a LINE element, J Biol Chem 273 (1998) 891-897.

[46] J. Norris, D. Fan, C. Aleman, J.R. Marks, P.A. Futreal, R.W. Wiseman, J.D. Iglehart, P.L. Deininger, D.P. McDonnell, Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers, J Biol Chem 270 (1995) 22777-22782.

[47] C. Feschotte, E.J. Pritham, DNA transposons and the evolution of eukaryotic genomes, Annu Rev Genet 41 (2007) 331-368.

Figure legends

Fig. 1. The intergenic region separating the tandemly arrayed *rtvtg* genes is mostly composed of transposable elements. The 3'-end of a gene (exons 33-35) and the 5'-end of the next gene (exons 1-3) are shown with *black boxes*. The intergenic region is 4.6 kb long and numbered relative to the transcription start site of the downstream *rtvtg* gene (+1). Sequences shown with *stippled boxes* are remnants of retrotransposable elements (*Gypsy*-like*, LINE*-like and *Jockey*-like*).* A short sequence analogous to a HIV *env* gene (*H*) has been disrupted by the insertion of the *Jockey-like* sequence. The *Jockey*-like sequence corresponds to a reverse transcriptase domain and the *Gypsy-like* sequence shows typical features of protease and reverse transcriptase domains of Pol polyproteins. Putative EREs (pERE1 and $pERE_2$) and half ERE motifs are shown.

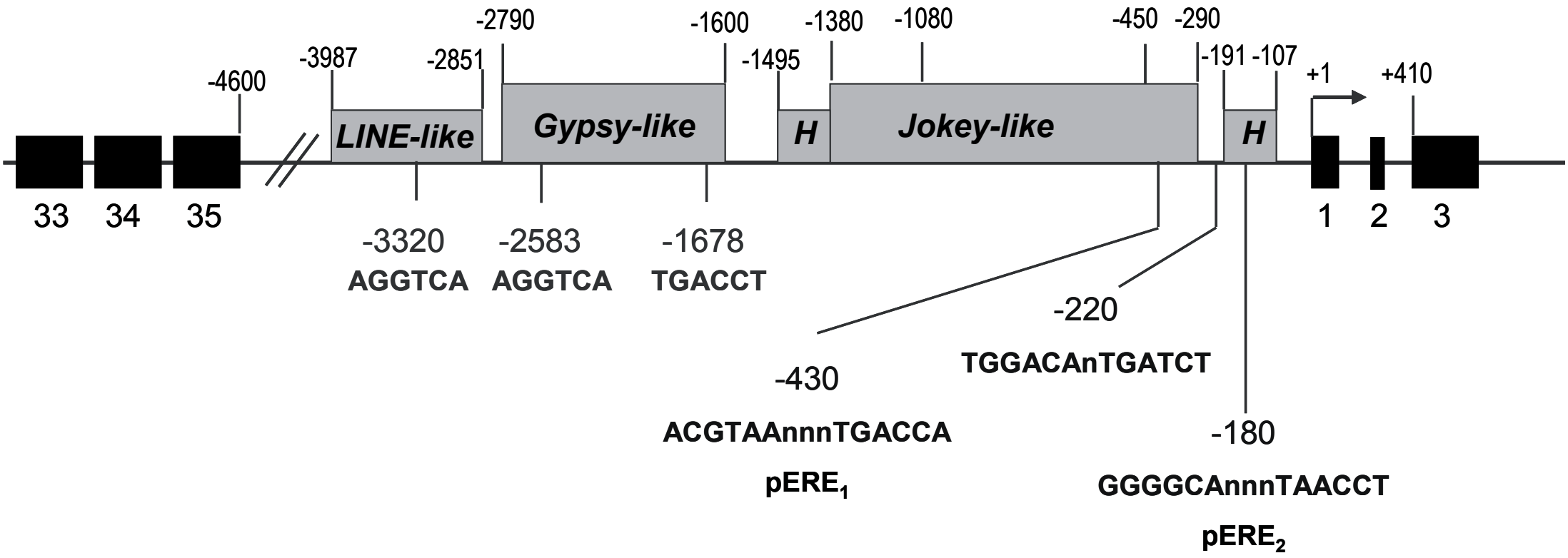
Fig. 2. *rtvtg* gene promoter activity in MCF-7 cells. Cells were transfected with plasmids containing the luciferase reporter under no promoter (pGL2-b), the SV40 promoter (pGL2-p) or different regions of the 5'-end of the *rtvtg* gene. Constructs are shown on the left side. On top, a schematic representation of the *rtvtg* 5'-region indicates the positions of putative EREs. The respective promoter activities in the absence (-E2) and presence (+E2) of 17β-estradiol are shown in the histogram on the right. Values are the average of three biological replicates.

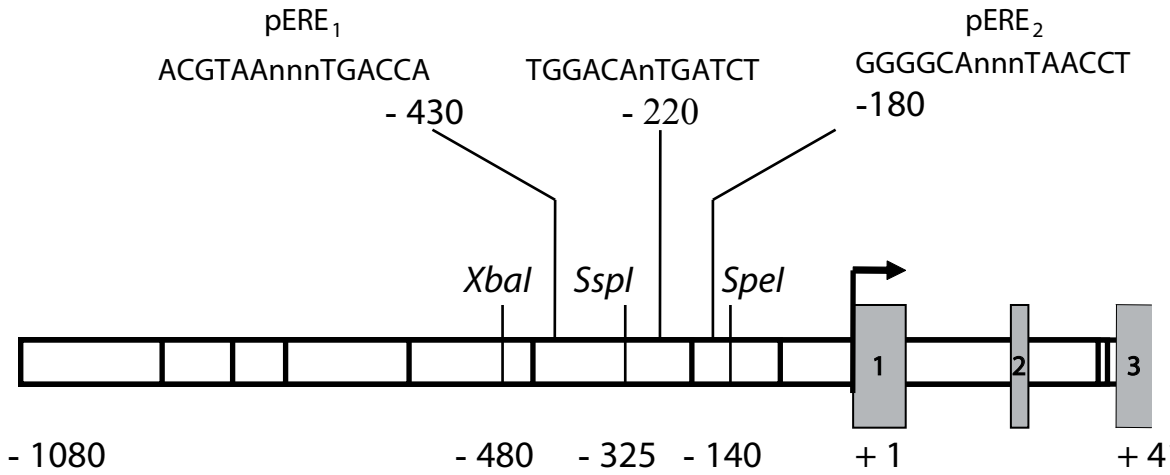
Fig. 3. E2-dependent enhancer activity of hERα and $rtER_S$ in *Saccharomyces cerevisiae*. Cells were co-transformed with pY60hERα (black bar charts) or pY60$rtER_S$ (grey bar charts) expression vector and the β-galactosidase reporter vector Y_{EREcs}. The recombinant cells were grown in a YPRE medium until $OD_{600nm,\ l=1cm}$ = 0.6 and then induced with 2% of galactose and stimulated or not with 10^{-7} M of E2. After 16h of incubation, β-galactosidase activity was measured. Values represented a mean of at least three independent experiments.

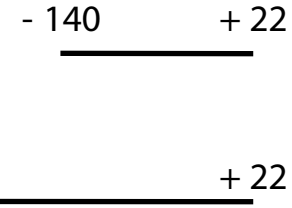
Fig.4. An ERE is located in the *rtvtg* gene promoter: functional analysis using hERα (A) and $rtER_S$ (B)**.** Yeast cells were co-transformed with the pY60ER expression vector and one of the YRPE2-derivative reporter vectors: Y_{cyc} (lane 1), $Y_{-450/-130}$ (lane 2), $Y_{-260/-130}$ (lane 3), $Y_{-210/-130}$ (lane 4), $Y_{-450/-190}$ (lane 5), $Y_{-450/-310}$ (lane 6) and Y_{EREcs} (lane 7). The recombinant cells were grown in a YPRE medium until $OD_{600nm,\ l=1cm} = 0.6$ and then induced with 2% of galactose and stimulated (+E2) with 10^{-7} M of E2 or not (-E2). After 16h of incubation, β-galactosidase activity was measured. Values represented a mean of at least three independent experiments.

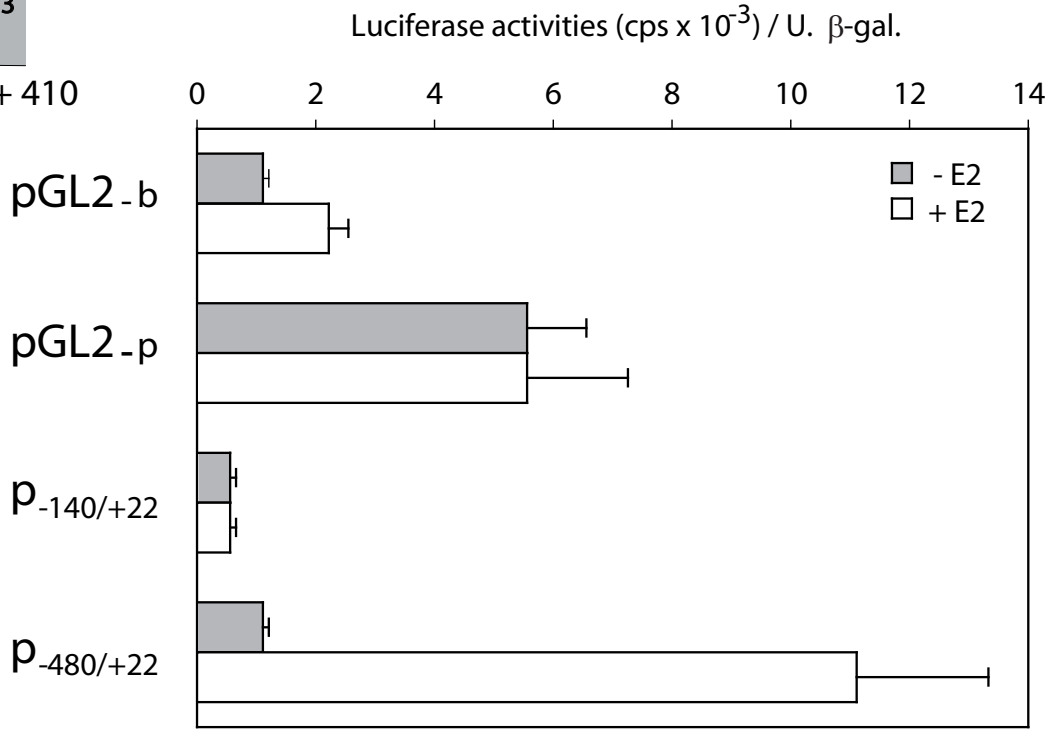
Fig. 5. Functional analysis of the putative EREs and functional differences between hERα and $rtER_S$. Dose-response curves were carried out with yeast cells co-transformed with ERE reporter plasmids (Y_{EREcs} (●), $Y_{-210/-130}$ (Δ),$Y_{-450/-310}$ (■),$Y_{-450/-130}$ (∇) or mutated $Y_{-450/-130}$ (▼) and an ER expression vector (pY60hERα or $pY60rtER_S$). Recombinant cells were grown in YPRE medium until $OD_{600nm,\ l=1cm} = 0.6$ and ER expression was induced with 2% of galactose and treated with different E2 concentrations. β-galactosidase activity was measured after 16h induction. Values represented a mean of at least three independent experiments.

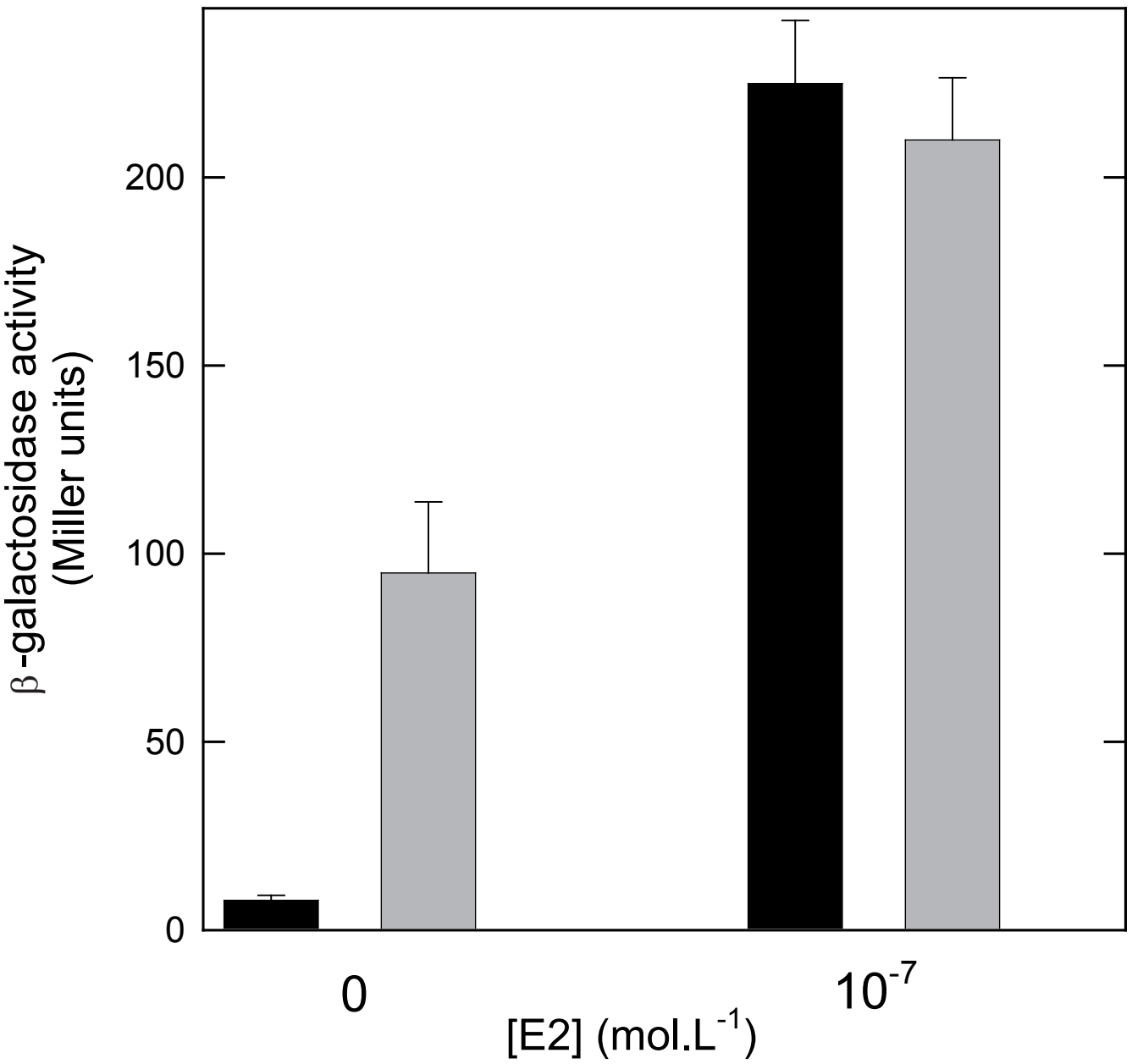
Fig. 6. ER has a strong affinity for *rtvtg*ERE. Fluorescence anisotropy was measured at low salt concentration. Titration of 1 nM fluorescein-labeled 21-mer with hERα. Lines through the data are the result of the best fit (simple binding model) through non-linear regression. Panel **A**: ERE_{cs} (•) and non-specific oligonucleotide targets, mim (□); Panel **B**: $pERE_2$ (Δ), mutated $pERE_2$ (▼) and $pERE_1$ (■). The inferred K_D values are 3.0 ± 0.4 nM, 6.2 ± 0.3 nM and > 40 nM for ERE_{cs}, $pERE_2$ and non-specific sequences (mutated $pERE_2$, $pERE_1$ and mim), respectively. Experiments are performed at 10 ± 2 °C in 10 mM Tris-HCl, 140 mM KCl, pH 7, 0.1 mM EDTA, 1 mM DTT, and 10% glycerol.

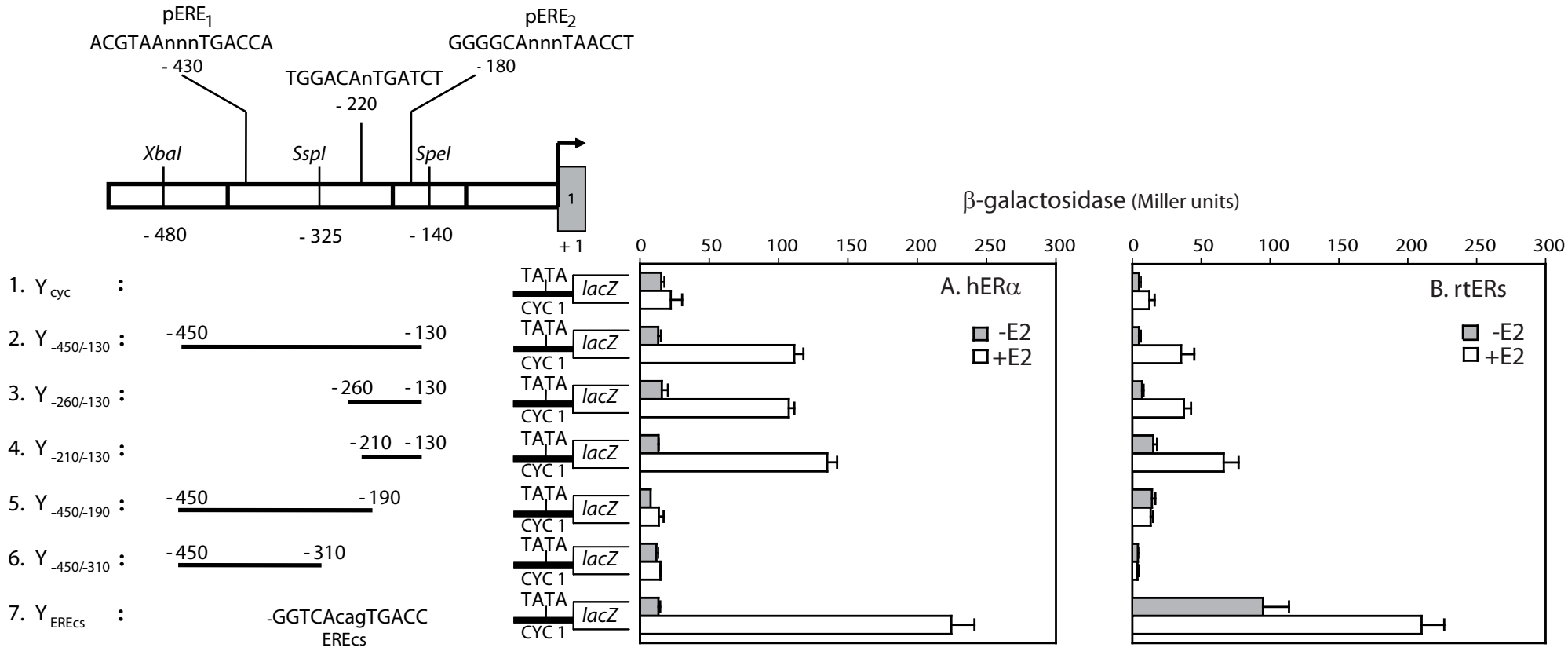

-4600
-3987
-2851
-2790
-1600
-1495
-1380
-1080
-450
-290
-191
-107
+1
+410
LINE-like
Gypsy-like
H
Jokey-like
H
33
34
35
1
2
3
-3320
AGGTCA
-2583
AGGTCA
-1678
TGACCT
-220
TGGACAnTGATCT
-430
ACGTAAnnnTGACCA
pERE1
-180
GGGGCAnnnTAACCT
pERE2

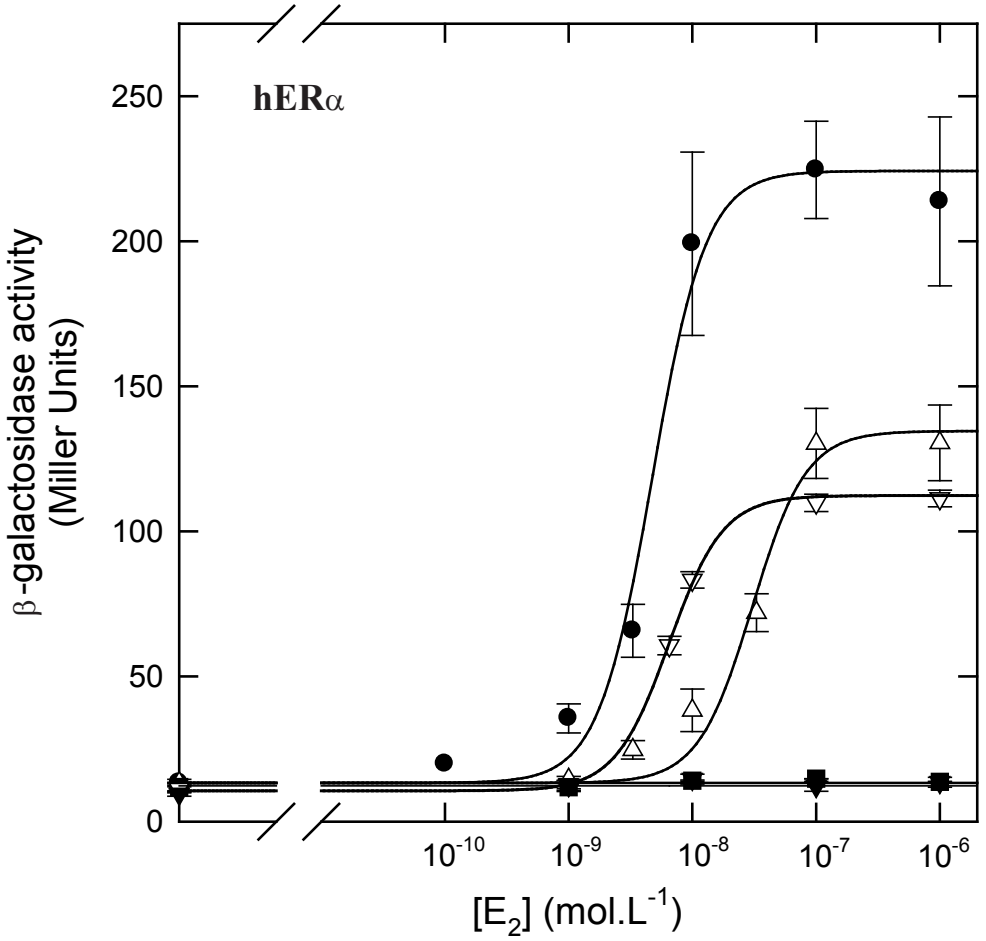

pERE1
ACGTAAnnnTGACCA
- 430
TGGACAnTGATCT
- 220
pERE2
GGGGCAnnnTAACCT
-180
XbaI
SspI
SpeI
1
2
3
- 1080
- 480
- 325
- 140
+ 1
+ 410

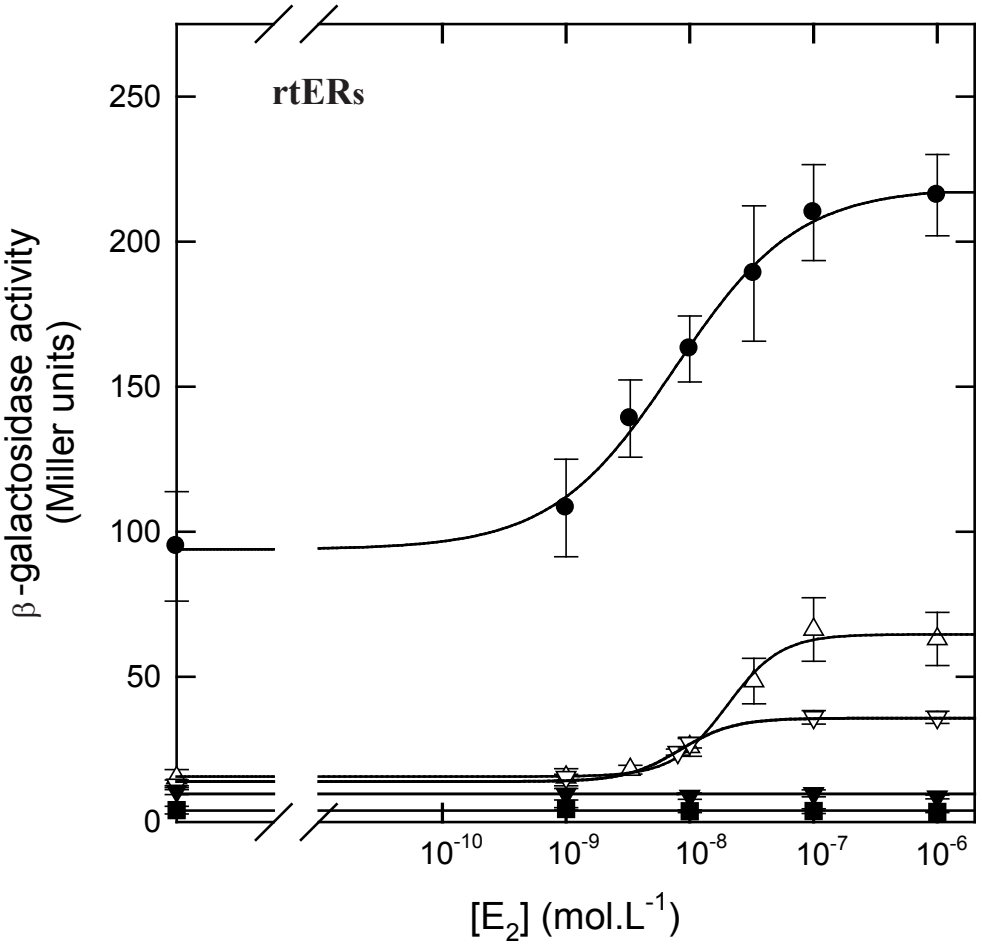

- 140
+ 22
- 480
+ 22

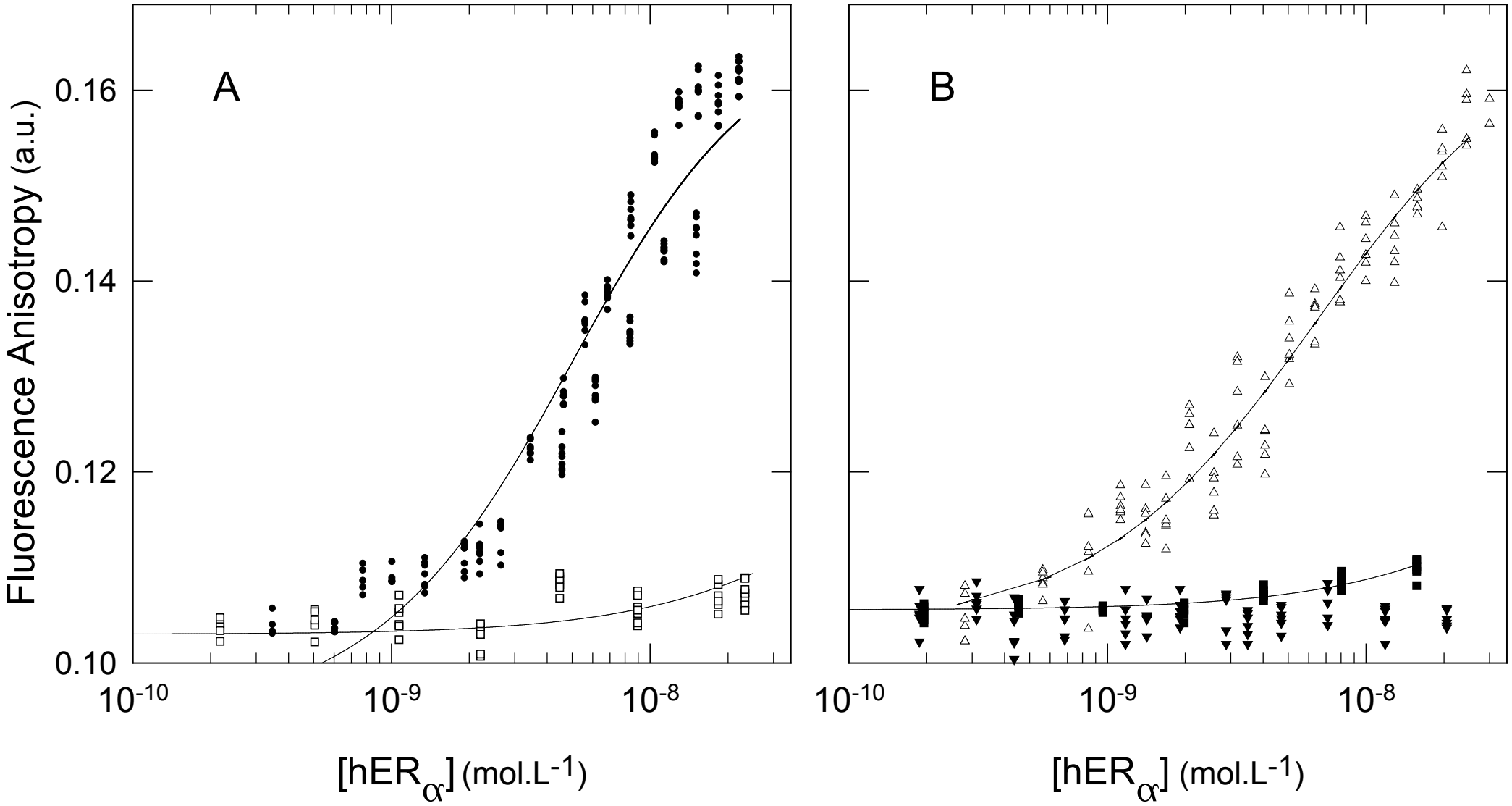

Luciferase activities (cps x 10-3) / U. β-gal.
0
2
4
6
8
10
12
14
- E2
+ E2
pGL2-b
pGL2-p
p-140/+22
p-480/+22


β-galactosidase activity
(Miller units)
200
150
100
50
0
0
10-7
[E2] (mol.L-1)


pERE1
ACGTAAnnnTGACCA
- 430
TGGACAnTGATCT
- 220
pERE2
GGGGCAnnnTAACCT
- 180
XbaI
SspI
SpeI
1
- 480
- 325
- 140
+ 1
β-galactosidase (Miller units)
0
50
100
150
200
250
300
1. Y cyc :
2. Y -450/-130 : -450 -130
3. Y -260/-130 : -260 -130
4. Y -210/-130 : -210 -130
5. Y -450/-190 : -450 -190
6. Y -450/-310 : -450 -310
7. Y EREcs : -GGTCAcagTGACC
EREcs
TATA
lacZ
CYC 1
A. hERα
B. rtERs
-E2
+E2


hERα
β-galactosidase activity
(Miller Units)
250
200
150
100
50
0
10-10
10-9
10-8
10-7
10-6
[E2] (mol.L-1)


rtERs
β-galactosidase activity
(Miller units)
250
200
150
100
50
0
10-10
10-9
10-8
10-7
10-6
[E2] (mol.L-1)


A
B
Fluorescence Anisotropy (a.u.)
0.16
0.14
0.12
0.10
10-10
10-9
10-8
[hERα] (mol.L-1)

Table

Table 1. EC50 values and induction factor following E2 treatment. The EC_{50} values were determined from the *in cellulo* dose-response assays using non linear regression with a sigmoid curve fitting (see Material and Methods section). The Activity $_{[L]\rightarrow max}$ is the β-galactosidase activity measured with the maximal E2 concentration used (10^{-7} M). E2 stimulation factor was calculated as ratio of β-galactosidase activities measured in the presence of 10^{-7} M of E2 and in the absence of hormonal stimulation.

		EC_{50} (10^{-9} mol.L^{-1})	**Activity $_{[L]\rightarrow max}$ (Miller units)**	**E2 Induction Factor**
hERα	**Y_{EREcs}**	3.1 ± 1.2	224.6 ± 16.7	17.0 ± 2.4
	$Y_{-210/-130}$	14.2 ± 2.8	135.1 ± 7.1	10.3 ± 0.9
$rtER_S$	**Y_{EREcs}**	4.1 ± 1.0	216.8 ± 14.0	2.4 ± 0.4
	$Y_{-210/-130}$	18.5 ± 4.5	66.3 ± 10.9	4.2 ± 0.9