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

To cite this version:
Control of vitellogenin genes expression
by sequences derived from transposable elements in rainbow trout

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

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

\textsuperscript{a} present address: Molecular Imaging and NanoBioTechnology, UMR 5248 CBMN, CNRS-Université Bordeaux 1-ENITAB, IECB, 2 rue Robert Escarpit, 33607 Pessac, France.
\textsuperscript{b} present address: Museum National d'Histoire Naturelle, 57 Rue Cuvier, 75231 Paris Cedex 05, France.
\textsuperscript{c} present address: Paediatric Molecular Genetics, Institute of Molecular Medicine, Oxford University, John Radcliffe Hospital, Oxford OX3 9DS, UK.
\textsuperscript{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\textsubscript{S}, estrogens; ER, estrogen receptor; ERE(s), estrogen response element(s); ERE\textsubscript{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 $\text{GGGGCAnnnTAACCT}$ (*rtvtg*ERE), which differs from the consensus ERE (EREcs) 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 $\alpha$ (hER$\alpha$) can specifically bind to *rtvtg*ERE. Furthermore, we observe that the stability of hER$\alpha$-EREcs and hER$\alpha$-*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$^\text{nnn}$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, ER$\alpha$ and ER$\beta$, which significantly differ in their biological activities [7]. In rainbow trout, ER$\alpha$ 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 ER$\alpha$ (hER$\alpha$) 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). hER\(\alpha\) 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 hER\(\alpha\); (iii) rtER\(S\) displays a weaker transactivation activity compared to hER\(\alpha\), 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 hER\(\alpha\), 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\(_3\) 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 vtgA2 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, *i.e.* 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 rtvtgERE 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 \textit{rtvtg} gene [27] and cloned in a luciferase reporter plasmid (pGL2-b, Promega). \textit{Mlu}I and \textit{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 \textit{Mlu}I/\textit{Bgl}II digested pGL2-b plasmid, giving the \textit{p-480/+22} and \textit{p-140/+22} vectors. The resulting plasmids contain the first 21 nucleotides of the \textit{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.

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

Yeast rtER₅ and hERα expression plasmids (pY60rtER₅ and pY60hERα, respectively) are derivatives of pYeDP60-AQPcic [29] from which the cDNA encoding AQPcic was excised by EcoRI and SacI and replaced by the complete coding sequence of the rtER₅ cDNA amplified from pCMV5/rtER₅ [30], or the human ERα cDNA amplified from YEPE15 [28]. The plasmid has the yeast ura marker and the bacterial ampR selection markers. rtER₅ 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 10⁶ 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₋₁₄₀/₀₂₂ or p₋₄₈₀/₀₂₂). 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:

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

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$_{cs}$), AGTGGGCGAGTGAACCTAAC (pERE$_2$),
AGTGCCGAAGGTAAACCTAAC (mutated pERE2), CCCACGTAAAACGTACCATCC (pERE1) 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\textsubscript{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, \( F_\parallel \), and perpendicular, \( F_\perp \), emission components:

\[
A = \frac{F_\parallel - F_\perp}{F_\parallel + 2 \times F_\perp} \tag{3}
\]

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\textsubscript{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 EREs but two motifs corresponding to putative imperfect EREs (pERE1 and pERE2) 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 pERE1 (ACGTAaacTGACCA) and pERE2 (GGGGCAggtTACCT), 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<sub>-480/+22</sub> and p<sub>-140/+22</sub>, were used to transfsect breast cancer MCF-7 cells, which naturally express hER. Luciferase activity was measured with or without E2 treatment (10<sup>-8</sup> 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<sub>-140/+22</sub>, 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 rtERs. 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 -rtERs-), 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 EREcs (YEREcs) 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 pERE1, Y-260/-130 and Y-210/-130 containing only pERE2. 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 rtERs to induce reporter gene expression, cells were co-transformed with pY60hERα or pY60rtERs together with the YEREcs plasmid. ER expression was induced with addition of galactose and activities were measured in the
absence or in the presence of 10^{-7} \text{M} \text{ of E2 (Fig. 3). Preliminary experiments showed that this concentration induced maximal $\beta$-galactosidase activity. In the absence of E2, only cells expressing rtER_S exhibited a significant activity. Following E2-stimulation, $\beta$-galactosidase activity was strongly increased in hER$\alpha$-expressing cells whereas rtER_S-expressing cells exhibited an activity increased by twice in comparison to without E2. This control experiment clearly indicates that hER$\alpha$ 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 hER$\alpha$, 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 (hER$\alpha$ or rtER_S) transformed with Y_{-450/-190} (lane 5) or Y_{-450/-310} (lane 6) exhibited only background level of $\beta$-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 $\beta$-galactosidase activity, which can be mediated by both hER$\alpha$ 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\textsubscript{-210/-130} (containing pERE2) or Y\textsubscript{-450/-310} (containing the non E2-responsive pERE1 as a negative control), together with hER\textalpha{} or rtER\textsubscript{S} expression plasmid (Fig. 5 and Table 1). As expected, the Y\textsubscript{-450/-310} construct did not yield significant levels of β-galactosidase activity, at whatever E2 concentration. In contrast, the Y\textsubscript{EREcs} and Y\textsubscript{-210/-130} constructs responded strongly to E2, with hER\textalpha{} or rtER\textsubscript{S}. Indeed, the EC\textsubscript{50} (that is, the concentration of E2 corresponding to 50% of the maximal response) is very similar for both ERs, although the EC\textsubscript{50} measured with Y\textsubscript{-210/-130} is consistently 4.5 times higher than that of Y\textsubscript{EREcs} (3.1 and 14.2 vs 4.1 and 18.5 with Y\textsubscript{EREcs} and Y\textsubscript{-210/-130} constructs, respectively).

In agreement with previous reports [10-11], we found marked functional differences between hER\textalpha{} and rtER\textsubscript{S}: (i)- the maximal level of reporter activity of the Y\textsubscript{-210/-130} construct with rtER\textsubscript{S} is only half that of hER\textalpha{} (~ 63 vs ~ 135 Miller units), and (ii)- rtER\textsubscript{S} (but not hER\textalpha{}) 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\textsubscript{S} is much lower than that of hER\textalpha{} (hER\textalpha{}: 17.0 and 10.3 vs rtER\textsubscript{S}: 2.4 and 4.2, with Y\textsubscript{EREcs} and Y\textsubscript{-210/-130} constructs, respectively), despite the fact that maximal induction levels measured with the control construct (Y\textsubscript{EREcs}) are comparable for hER\textalpha{} and rtER\textsubscript{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 hER\textalpha{} and rtER\textsubscript{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\textsubscript{2} corresponds to a functional ERE.

To check whether this putative ERE\textsubscript{2} is indeed functional, a site-directed mutagenesis to change the sequence of pERE\textsubscript{2} from GGGGCAggtTAACCT to GCCGAAggtAAACCT was performed from the Y\textsubscript{-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
pERE2 leads to a loss of ERs transactivation in response to E2. Thus, pERE2 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 (pERE2, pERE1, EREcs 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 rtERβ, 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: EREcs, pERE2, mutated pERE2, pERE1 and mim. Data were fitted to recover the equilibrium dissociation constant (K_D). The fluorescence anisotropy binding curve of the hERα–EREcs 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 pERE2 (Fig. 6B). Thus, at 140 mM KCl, hERα exhibits a similar affinity for the consensus ERE and for the natural pERE2 sequence. Furthermore, with EREcs and pERE2, 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 pERE1 or mutated pERE2 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 pERE2 sequence with an affinity close to the value displayed for EREcs, confirming that E2-regulatory element within the genomic region between -210 and -190 is pERE2. As pERE2 appears as the main
and unique ERE sequence inside \textit{rtvtg} promoter, this sequence will be referred to as \textit{rtvtgERE} 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 GGGGCAaggTAACCT. 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 Eσ-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 pERE2 motif, located at position -180, may correspond to the enhancer. In contrast, the pERE1 motif exhibits no biological activity, thus confirming that it is not a functional ERE. Strikingly, the pERE2 sequence (GGGGCAggtTAACCT) differs from the EREcs (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 pERE2 sequence. We show that, in our experimental conditions, hERα specifically binds to the pERE2 sequence with a very high affinity, but not to the pERE1 sequence. In agreement with previous work in vitro, we found that hERα has a slightly lower affinity for a degenerated ERE (pERE2) compared to EREcs [2, 40], although in the case of pERE2, it is only two fold lower. At higher salt concentration, hERα is known to display a much stronger preference for EREcs 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 pERE2 sequence, now designated rtvtgERE, is the functional element driving the E2-dependent rtvtg gene transcription.

4.2. Functional impact of EREcs and rtvtgERE: The driving force of rtvtg genes evolution?

Paradoxically, even though the EREcs is the optimal ER DNA-binding site and mediates the
strongest transcriptional induction, it is found in less than 10% of E\textsubscript{S}-responsive genes [2]. Instead, imperfect motifs, rather than ERE\textsubscript{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\textsubscript{S} and their physiological relevance. In the absence of E2, a promoter construct containing an ERE\textsubscript{cs} separated by 20 bp is sufficient to drive the transcription of the reporter gene with rtER\textsubscript{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\textsubscript{cs}. rtER\textsubscript{S} is therefore an efficient basal transcriptional activator which weakly responds to E2 when bound to an ERE\textsubscript{cs}. In sharp contrast, rtER\textsubscript{S} displays a much lower basal transcriptional activity but a high responsiveness to E2, when it is bound to rtvtgERE. This point has important physiological consequences because E2-responsive genes will be properly expressed in the liver, where the rtER\textsubscript{S} isoform dominates, only if they are under the control of imperfect EREs (such as rtvtgERE), not ERE\textsubscript{cs}.

One might argue that the level of transcriptional induction driven by rtER\textsubscript{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


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


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.


**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 pERE2) 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 in *Saccharomyces cerevisiae*. Cells were co-transformed with pY60hERα (black bar charts) or pY60rtER (grey bar charts) expression vector and the β-galactosidase reporter vector Y_{EREcs}. The recombinant cells were grown in a YPRE medium until $OD_{600nm, \, i=1cm} = 0.6$ 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 rttvg gene promoter: functional analysis using hERα (A) and rtERs (B). Yeast cells were co-transformed with the pY60ER expression vector and one of the YRPE2-derivative reporter vectors: Ycyc (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 YEREcs (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 rtERs. Dose-response curves were carried out with yeast cells co-transformed with ERE reporter plasmids (YEREcs (●), Y-210/-130 (Δ), Y-450/-310 (■), Y-450/-130 (▼)) or mutated Y-450/-130 (▼) and an ER expression vector (pY60hERα or pY60rtERs). 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 rttvgERE. 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: EREcs (●) and non-specific oligonucleotide targets, mim (□); Panel B: pERE2 (Δ), mutated pERE2 (▼) and pERE1 (■). The inferred $K_D$ values are 3.0 ± 0.4 nM, 6.2 ± 0.3 nM and > 40 nM for EREcs, pERE2 and non-specific sequences (mutated pERE2, pERE1 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.
Gypsy-like Jokey-like H HLINE-like

-3320 AGGTCA -2583 AGGTCA -1678 TGACCT

pERE$_1$

-1495 -1380 -1080 -220 -450 -290 -191 -107 +410 +1

ACGTAAnnnTGACCA

GGGGCAnnnTAACCT

pERE$_2$
Luciferase activities (cps x 10^{-3}) / U. \( \beta \)-gal.

- \( pERE_1 \)
  - ACGTAAnnnTGACCA
  - TGGACAnTGATCT
  - XbaI
  - SspI
  - Spel
  - -430
  - -325
  - -140

- \( pERE_2 \)
  - GGGGCAnnnTAACCT
  - -180

- Lucerase activities (cps x 10^{-3}) / U. \( \beta \)-gal.
  - \( pGL2 \) -b
  - \( pGL2 \) -p
  - \( p-140/22 \)
  - \( p-480/22 \)

- Graph showing luciferase activities for different constructs under control of E2 and +E2.
β-galactosidase activity (Miller units)

[E2] (mol.L⁻¹)

0 10⁻⁷
1. Y\textsubscript{cyc} : 

2. Y\textsubscript{-450/-130} : -450 -130

3. Y\textsubscript{-260/-130} : -260 -130

4. Y\textsubscript{-210/-130} : -210 -130

5. Y\textsubscript{-450/-190} : -450 -190

6. Y\textsubscript{-450/-310} : -450 -310

7. Y\textsubscript{EREcs} : -GGTCAcagTGACC

EREcs

\begin{itemize}
\item A. hER\textalpha
\item B. rtERs
\end{itemize}
Table 1. EC50 values and induction factor following E2 treatment. The EC50 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 $[L]_{\text{→max}}$ 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.

<table>
<thead>
<tr>
<th></th>
<th>EC50 ($10^{-9}$ mol.L$^{-1}$)</th>
<th>Activity $[L]_{\text{→max}}$ (Miller units)</th>
<th>E2 Induction Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>hERα</td>
<td>Y$_{EREcs}$</td>
<td>3.1 ± 1.2</td>
<td>224.6 ± 16.7</td>
</tr>
<tr>
<td></td>
<td>Y$_{210/-130}$</td>
<td>14.2 ± 2.8</td>
<td>135.1 ± 7.1</td>
</tr>
<tr>
<td>rtER$_S$</td>
<td>Y$_{EREcs}$</td>
<td>4.1 ± 1.0</td>
<td>216.8 ± 14.0</td>
</tr>
<tr>
<td></td>
<td>Y$_{210/-130}$</td>
<td>18.5 ± 4.5</td>
<td>66.3 ± 10.9</td>
</tr>
</tbody>
</table>