Quantitative effects on gene silencing by allelic variation at a tetranucleotide microsatellite

Véronique Albanèse, Nicole Faucon Biguet, Hélène Kiefer, Elisa Bayard, Jacques Mallet* and Rolando Meloni

Laboratoire de Génétique de la Neurotransmission, CNRS UMR C9923, Hôpital de la Pitié Salpêtrière, 83 Boulevard de l’hôpital, 75013 Paris, France

Received April 11, 2001; Revised and Accepted June 18, 2001

Microsatellites are common repeated sequences, which are useful as genetic markers and lack any clearly established function. In a previous study we suggested that an intronic polymorphic TCAT repeat in the tyrosine hydroxylase (TH) gene, the microsatellite HUMTH01, may regulate transcription. The TH gene encodes the rate-limiting enzyme in the biosynthesis of catecholamines, and the microsatellite HUMTH01 has been used in genetic studies of neuropsychiatric and cardiovascular diseases, in which disturbances of catecholaminergic neurotransmission have been implicated. HUMTH01 alleles associated with these diseases act as transcriptional enhancers when linked to a minimal promoter and are recognized by specific nuclear factors. Here we show that allelic variations of HUMTH01 commonly found in humans have a quantitative silencing effect on TH gene expression. Two specific proteins, ZNF191, a zinc finger protein, and HBP1, an HMG box transcription factor, which bind the TCAT motif, were then cloned. Finally, allelic variations of HUMTH01 correlate with quantitative and qualitative changes in the binding by ZNF191. Thus, this repeated sequence may contribute to the control of expression of quantitative genetic traits. As the HUMTH01 core motif is ubiquitous in the genome, this phenomenon may be relevant to the quantitative expression of many genes in addition to TH.

INTRODUCTION

Microsatellites, consisting of di-, tri- or tetranucleotide repeats, are frequent, polymorphic and randomly distributed across the genome, characteristics that make them useful as DNA markers for gene mapping and linkage analysis (1). These sequences have long been considered as essentially neutral elements devoid of biological effect. However, several studies suggest that repeated sequences may have a function in recombination (2), in the generation of nucleosome positioning signals (3) and in transcription (4). For example, abnormal expansions of trinucleotide repeats in non-coding sequences interfere with normal transcriptional activity and are responsible for several human neurological diseases. In fragile X syndrome, the expansion of a CGG repeat located in the 5′-untranslated region (5′-UTR) of the FMR-1 gene abolishes transcription of the gene (5). Similarly, in myotonic dystrophy, a CTG expansion in the 3′-UTR of the DMPK gene affects the transcription of a neighboring gene coding for a homeodomain protein (6).

Repeated sequences may not be only associated with pathological expansions of unstable DNA stretches causing Mendelian diseases with complete penetrance. They may also have more subtle effects on gene expression (4). We recently demonstrated that a tetranucleotide repeat, the HUMTH01 microsatellite, in the first intron of the tyrosine hydroxylase (TH) gene, acts as a transcriptional enhancer in vitro (7). TH is the rate limiting enzyme in the biosynthesis of catecholamines, a class of neurotransmitters involved in a wide range of neuro-physiological functions, such as stress response, motor control, cognitive and behavioral functions (8). HUMTH01 has been widely used as a marker in genetic studies, and is significantly associated with diseases in which abnormal catecholamine transmission is implicated: essential hypertension (9), bipolar disorder (10–12) and schizophrenia (13,14). The HUMTH01 sequence consists of 5 to 10 repetitions of a (TCAT) motif. The perfect (TCAT)_10p variant is very rare (<1%). The imperfect (TCAT)_10p(CAT(TCAT)₅) variant, called the (TCAT)_10i repeat, is the most common allele in Caucasians (~30%) (15). Both (TCAT)_10p and (TCAT)_10i enhance transcription when placed upstream from a minimal promoter and a luciferase reporter gene. Moreover, these repeated sequences interact specifically and with high affinity with nuclear proteins (7).

In the present study, we have investigated the contribution of the HUMTH01 microsatellite to the regulation of TH gene expression. We show that this microsatellite participates in the transcriptional regulation of the TH gene by a silencing effect that correlates with the number of repetitions. The cloning of two specific proteins, ZNF191, a zinc finger protein, and HBP1, an HMG box transcription factor, which bind the TCAT motif, has allowed us to further investigate the molecular properties of the TCAT repeated sequence. In fact, the binding of the specific transcription factor ZNF191 is also correlated in a quantitative fashion to the number of repeats. Thus, the HUMTH01 microsatellite may participate in the transcriptional regulation of the TH gene and of other genes

*To whom correspondence should be addressed. Tel: +33 1 42177563; Fax: +33 1 42177533; Email: mallet@infobiogen.fr
containing such repeated sequences. The characteristics of the interaction of this repeated sequence may provide a molecular basis for the modulation of gene expression relevant for the genetics of quantitative traits.

RESULTS

Role of the HUMTH01 microsatellite in the transcriptional regulation of the TH gene

In a preliminary study, the (TCAT)$_{10i}$ and (TCAT)$_{10p}$ alleles of the HUMTH01 microsatellite located in the first intron of the TH gene were found to regulate transcription when placed upstream of the thymidine kinase minimal promoter (7). The transcriptional effects of this repeat were then tested in its natural intronic context. The sequence of the human TH gene from –2187 to +1300, including the distal promoter, three exons and two introns, was fused to the luciferase gene generating the TH-Luc construct (Fig. 1A). To test how the transcriptional activity relates to the number of repeats, constructs (TCAT)$_{\delta}$, (TCAT)$_{3}$, (TCAT)$_{5}$, (TCAT)$_{8}$, (TCAT)$_{10i}$ and (TCAT)$_{10p}$ were generated. The constructs were introduced by transient transfection into two catecholaminergic cell lines, PC12 and CHP212.

In PC12 cells, the (TCAT) repeats inhibited transcriptional activity in a stepwise mode as compared to the control (TCAT)$_{\delta}$ construct (Fig. 1B; P < 0.01 for all constructs): 50% inhibition for three repeats, 60% for five repeats and 70% inhibition for eight repeats or more. This effect was not dependent on the orientation of the repeat, since the antisense (ATGA)$_{10}$ sequence also resulted in 70% inhibition of the transcriptional activity (Fig. 1B; P < 0.01). The unrelated (CCAC)$_{10}$ repeat had no effect on transcriptional activity indicating that this transcriptional repression was specific to (TCAT) repeats.

In CHP212 cells, (TCAT) repeats also had similar inhibitory effects (Fig. 1B; P < 0.01 for all constructs). (TCAT)$_{3}$ and (TCAT)$_{5}$ were the least powerful inhibitors, ~30%; (TCAT)$_{8}$ caused 65% inhibition, and the longer repeats, (TCAT)$_{10i}$ and (TCAT)$_{10p}$, resulted in only 55 and 40% transcriptional inhibition, respectively. However, the differences between the effects of (TCAT)$_{5}$ and (TCAT)$_{10}$ repeats were not significant.

As shown previously, (TCAT)$_{10p}$ acts as an enhancer when placed upstream from the minimal promoter of the thymidine kinase gene (TCAT-ik-LUC construct) (7). In contrast, (TCAT)$_{10i}$ in the first intron of the TH gene-inhibited transcription (Fig. 1). The different effects are possibly due to the different positions of the (TCAT) repeat with respect to the promoter. To address this issue, the fusion construct was modified to move the (TCAT)$_{10i}$ repeat from the first intron to the AarII restriction site at position –44 in the proximal promoter (Fig. 2A). The transcription level from this construct was the same as that from the construct with the repeats in the first intron (Fig. 2B). Thus, the inhibitory effect of the tetranucleotide repeat was not strictly dependent on position.

Additionally, when two (TCAT)$_{10i}$ repeats were present, one in the promoter and one in the first intron, no further inhibition was observed.

We then investigated whether the (TCAT) repeat interacts with other unidentified regulating sequences. We constructed fusion genes in which the 2.2 kb of the TH promoter was completely removed and the rest of the TH-Luc minigene was linked to a heterologous promoter, the minimal collagenase promoter (Fig. 3A). As shown in Figure 3B, the results were very similar to those for constructs containing the entire 2.2 kb promoter, both in terms of amplitude and transcriptional profile, with a stepwise inhibitory effect up to a maximum of 70% for eight repeats. No further inhibition was detected, but a reduction in the extent of inhibition, which is more marked for (TCAT)$_{10i}$, was observed for longer repetitions. These results suggest that the difference between the inhibitory effects of the (TCAT) repeat observed here and the activating effects observed in the synthetic TCAT-ik-LUC construct used previously (7) is due to elements outside the TH proximal promoter. Moreover, the lack of further inhibition or the partial loss of the inhibitory effect for 10 repetitions suggest that the lengthening of the HUMTH01 microsatellite alleles counteracts the silencing effect on gene transcription of its shorter alleles.

Identification of proteins interacting with the tetranucleotide repeated sequence using the yeast one hybrid system

To isolate the protein(s) interacting with the (TCAT) repeat, we set up a yeast one-hybrid system (16) using the (TCAT)$_{10}$ as bait for screening a human brain cDNA library. 100 clones were isolated based on their ability to induce HIS3 transcription placed under the control of two (TCAT)$_{10}$ motifs. The clones were also able to activate HIS3 transcription in yeast
containing only one copy of the binding site, while they failed to rescue the His^{-} phenotype of strains with the HIS3 gene under the control of unrelated binding sites (TRE element and E box), demonstrating that the activation was specific. The most abundant cDNA (70\%) also had the largest effect in the one-hybrid system. It encoded an in-frame fusion between the Gal4 activation domain and the zinc finger protein, ZNF191, lacking the first 32 amino acids (17). ZNF191 is a putative transcription factor belonging to the Kruppel family and containing four C2H2 zinc fingers in its C-terminus. The conserved amino acid knuckle or (H/C) link between zinc fingers typified by the amino acid sequence TGE(R/K)P(F/Y)X is also present in ZNF191. A SCAN box element like those in the zinc finger proteins SRE-ZBP, Ctfin51, Znf165 and Znf174 was present upstream of the zinc finger domain. This highly conserved region may be an interaction domain allowing homologous and heterologous oligomerization of the members of this family of transcription factors (18). The region between the SCAN box and the zinc finger domains does not share any sequence similarity with other transcription factors.

ZNF191 was further investigated since it was encoded by the most abundant cDNA identified by the one-hybrid system, and had a stronger effect in this system. The full-length ZNF191 cDNA was used as a probe to screen for this zinc finger transcript in a variety of human organs. The transcript was abundant in heart, lung and spleen. It was also found in urogenital tissues (kidney, testis and prostate) and lymphatic tissues (lymph node, thymus and tonsil). Low levels of ZNF191 transcript were also detected in brain, which may reflect dilution of specific expressing regions in total tissue. In all organs containing ZNF191 mRNA, two forms were found (10 and 3.9 kb), probably as a result of alternative splicing of the pre-messenger (data not shown).

We next investigated the distribution of ZNF191 mRNA in rat brain structures using the human sequence as a probe. Transcript levels in total rat brain were low. However, it was present at high levels in catecholaminergic tissues including the substantia nigra, the hypothalamus and the olfactory bulb. It was less abundant in the striatum, the cortex and the cerebellum. Interestingly, the transcript was also abundant in the adrenal medulla, a peripheral catecholaminergic tissue. In the rat embryo, the mRNA was present at high levels in total embryonic head, which may reflect a role in brain development. The transcript was abundant in the mesencephalon, a structure that contains the major embryonic catecholaminergic tissues (data not shown).

Transcriptional activity of ZNF191

Next, we tested the transcriptional activity of ZNF191 using a functional assay. Previous work and present results had shown that the (TCAT) repeats exhibit stimulatory effects when
placed upstream of the minimal thymidine kinase promoter, or inhibitory effects when located in the first intron of a TH-Luc fusion gene. These two kinds of reporter constructs were thus tested in ZNF191 cotransfection experiments in PC12 cells. Co-transfection with ZNF191 and (TCAT)$_{10}$ in its native intronic context resulted in transcription indistinguishable from that in the absence of ZNF191. However, the cotransfection with ZNF191 significantly increased the enhancing effects of (TCAT)$_{10}$ on the expression of a minimal thymidine kinase-Luc construct, evidence of a functional interaction between ZNF191 and the (TCAT)$_{10}$ repeat (Fig. 4). These results suggest that in the context of the TH gene either the effect of ZNF191 is already maximal or its interaction with the HUMTH01 sequence is competed by other factors.

**In vitro binding of ZNF191 to the TCAT repeat sequence**

To ascertain the specificity of interaction between ZNF191 and the (TCAT) repeated sequence, the protein was purified to homogeneity. The purified protein was resolved on an SDS–polyacrylamide gel as a 45 kDa band, in accordance with the estimated molecular weight, and was tested by EMSA. To determine the minimal sequence necessary for interaction with ZNF191, (TCAT)$_{5}$, (TCAT)$_{6}$, (TCAT)$_{7}$, (TCAT)$_{8}$, (TCAT)$_{9}$, (TCAT)$_{10}$i, and (TCAT)$_{10}$p probes were used in EMSA. So that the various probes were of comparable lengths, the shorter probes were made up to 40 bp with the flanking intronic sequences of the TH gene. As illustrated in Figure 5A, all the probes interacted with ZNF191. As shown in Figure 5A (lanes 2 and 4), the protein bound the (TCAT)$_{10}$ and (TCAT)$_{10}$p probes, giving two complexes, denoted I and II. The complex with higher mobility (complex II), corresponding to the binding of a second ZNF191, was not observed with the shortest probes (TCAT)$_{5}$ and (TCAT)$_{7}$ (Fig. 5A). This experiment indicated that ZNF191 can interact with a repeat as short as three copies of the motif TCAT. Moreover, interaction with two proteins was only possible if there were at least six copies, indicating that the minimal binding motif is a (TCAT)$_{3}$ repeat.

An excess of unlabeled homologous probe, but not an excess of the heterologous probe (CCAC)$_{10}$ eliminated the DNA–protein complexes (Fig. 5B, lanes 3 and 4). Additionally, no complex was observed with a labeled (CCAC)$_{10}$ probe (data not shown). The flanking sequences were not involved in this interaction as: (i) equivalent probes with (CCAC) repeats instead of (TCAT) repeats did not result in any detectable complex (data not shown), and (ii) footprint experiments show that ZNF191 does not interact with these flanking sequences (see below). The binding was completely abolished in the presence of the zinc chelator, orthophenanthroline, demonstrating the role of the zinc fingers (Fig. 5B, lane 5). The addition of anti-histidine antibody to the reaction resulted in a supershifted complex, indicating that the observed complexes I and II were due to the recombinant His-ZNF191 and not a contaminant protein (Fig. 5B, lane 6). Finally, as shown in Figure 5C, UV cross-linking experiments demonstrated that complexes I and II corresponded to the probe bound to one or two molecules, respectively, of ZNF191.

This was further confirmed by footprint experiments used to determine the binding site of ZNF191 within the (TCAT) repeated sequences. Different DNA probes were tested, consisting of 5, 8 or 10 TCAT repeats together with 40 bp flanking intronic sequences on each side (Fig. 6). The (TCAT)$_{3}$ probe had one binding site for ZNF191, whereas the probes (TCAT)$_{6}$, (TCAT)$_{10}$i and (TCAT)$_{10}$p had two binding sites. In all cases, the protected region covered approximately three to four repeats showing that the minimal binding site is a (TCAT)$_{3}$ motif. As ZNF191 has four zinc fingers, it is possible that each finger binds one (TCAT) motif. For (TCAT)$_{10}$i and (TCAT)$_{10}$p Probes, a central region between the two sites was not protected. The differences in the protection profiles of the (TCAT)$_{5}$ and (TCAT)$_{10}$p probes suggest that the presence of an imperfect motif renders these sequences non-equivalent for the ZNF191 binding. Interestingly, the two sites in these longer probes were also not equally protected, showing that the affinities of ZNF191 for these two sites were not identical. The affinities were thus determined more accurately for probes (TCAT)$_{10}$i, (TCAT)$_{10}$p and (TCAT)$_{10}$ using Scatchard analysis. Free probe and bound probes, corresponding to complexes I and II, were quantified after incubation of a fixed amount of ZNF191 with various amounts of DNA probes (see Materials and Methods). All probes had one identical binding site at the 3′ site of the (TCAT)$_{10}$ sequence with an affinity of ∼2 μM. The probes with longer repeats also exhibited a second binding site in 5′ with an affinity of ∼20 μM.

**DISCUSSION**

Our findings suggest that the HUMTH01 microsatellite regulates TH gene transcription and acts as a quantitative genomic effector modulating the activity of the TH gene. In their orthologous position in the first intron of the TH gene, the HUMTH01 alleles inhibited transcription, and the inhibition was proportional to the number of repeats from three to eight. However, increasing the number of repeats from 8 to 10
perfect or imperfect repetitions did not produce any further inhibitory effect, but rather resulted in no changes or in a relative enhancement of transcription depending on the cell line used. This phenomenon suggests that the silencing effect on gene transcription of the HUMTH01 shorter alleles is counteracted by the lengthening of its alleles above eight repetitions of its core motif. A different repeated sequence with a tetranucleotide repeat motif did not affect the reporter gene expression, demonstrating that the effect on transcription was specifically due to the (TCAT)n sequence and was not due to a 'spacer' effect. (TCAT)_{10}i and (TCAT)_{10}p alleles were previously shown to act as enhancers when placed upstream of the minimal promoter of the thymidine kinase gene, whereas these sequences acted as repressors when placed in the first intron of the TH gene. Thus, the effects of the HUMTH01 microsatellite depend on the presence of other sequences in the TH gene. However, the substitution of the 5′ regulatory region of the TH construct with a collagenase gene minimal promoter did not affect the profile of transcriptional regulation exerted by the (TCAT)_{n} sequence. Furthermore, the activity of the (TCAT)_{10}p sequence is independent of both its orientation and its position relative to the flanking sequences (Fig. 2A). These results strongly suggest that the repeated sequence acts independently of 5′ promoter regulatory elements and is not dependent on orientation or physical continuity with adjacent flanking sequences. They do not exclude the involvement of elements in the first intron or its flanking exons in the HUMTH01 microsatellite activity. Interestingly, not only the (TCAT)n tetranucleotide repeat but also its flanking sequences are conserved at their orthologous position in the first intron of the TH gene in several non-human primate species, consistent with there being evolutionary constraints maintaining its integrity (20). Moreover, the longer sequence encountered in non-human primates corresponds to eight repeats. The presence of longer alleles and the fact that the most frequent and longer allele in humans presents a point mutation that prevents further expansion would strengthen the fact that the HUMTH01 microsatellite is endowed with a functional role.

The molecular basis of the HUMTH01 action may be a modular interaction of this sequence with factors such as ZNF191 or HBP1. ZNF191 illustrates this putative interaction, which is based both on quantitative and qualitative differences associated with the length of the (TCAT)_{n} sequence. This may

![Figure 5](https://via.placeholder.com/150)

**Figure 5.** Interaction of recombinant ZNF191 with the (TCAT) repeated sequence. (A) EMSA with the various (TCAT)n probes, either alone (lanes 1, 3, 5, 7, 9, 11 and 13) or incubated with 500 ng recombinant His-ZNF191 (lanes 2, 4, 6, 8, 10, 12 and 14). Arrows indicate free probe and complexes I and II. (B) EMSA. ^32P-labeled (TCAT)_{10}p probe (lane 1), the labeled (TCAT)_{10}p Probe incubated with 500 ng recombinant His-ZNF191, either alone (lane 2) or with homologous unlabeled (TCAT)_{10}p Probe (lane 3), heterologous unlabeled (CCAC)_{10}p probe (lane 4), orthophenantroline (lane 5) or anti-histidine tag antibody (lane 6). Arrows indicate free probe and complexes I and II, and the arrowhead indicates supershifted complex in the presence of antibody. (C) SDS–PAGE analysis after UV cross-linking of the (TCAT)_{10}p probe incubated with 500 ng recombinant His-ZNF191. Arrows indicate complexes I and II.

![Figure 6](https://via.placeholder.com/150)

**Figure 6.** Footprinting experiments. Footprinting experiments with single-end labeled intronic fragments containing the (TCAT)_{10}i, (TCAT)_{10}p, (TCAT)_{6} or (TCAT)_{3} motifs, incubated with various amounts of recombinant His-ZNF191. Bars indicate protected regions.
be the first example of a new type of control of transcriptional activity involving a complex interplay between allelic variations in the consensus sequence length, the number of binding factor molecules acting on the same DNA stretch and the differences in the affinity of the same factors depending on their relative position on the same sequence. The effects mediated by ZNF191 may account for the transcriptional regulation of the many other human genes containing (TCAT) repeats. ZNF191 expression is not limited to catecholaminergic structures. It was previously isolated from bone marrow and promyelocytic leukemia cell lines using homologous PCR amplification with primers based on conserved sequences of the Kruppel family of transcription factors. This first identification had suggested that ZNF191 has a role in hematopoiesis (17).

Other binding factors including HBP1 may also participate in the transcriptional regulation and may account for further modulation of the effects of the HUMTH01 microsatellite either in an alternative or complementary fashion to ZNF191. The HMG box transcription factor HBP1 clones were the second most frequent (20%) among those identified by our yeast one-hybrid screening. This factor has a chromatin remodeling role in the LCR of the CD2 gene (19) where it specifically binds a (TTCA)₃ motif. This motif is embedded in the (TCAT)ₙ sequence. Preliminary EMSAs indicate that it binds to the HUMTH01 microsatellite (data not shown). Previous EMSAs using nuclear extracts show that other trans-acting factors of the fos–jun family, that bind the API consensus sequence TGATTCA in the TH gene (21), are also able to bind specifically the tetranucleotide repeat (TCAT)ₙ sequence (7). Moreover, a TCATT sequence, which is not tandemly present in other genes, the associated control mechanisms are likely to be relevant to phenotypic manifestations of a variety of quantitative genetic traits.

MATERIALS AND METHODS

Reporter constructs

The pGL3 basic vector for firefly luciferase expression (Promega) was modified as follows: (i) the NotI restriction site was disrupted by digesting and ligating refilled ends; (ii) the HindIII–NcoI fragment was shortened to AAGCTGGAT-AAGCTTCATG (a new HindIII site is underlined); and (iii) the ATG initiation codon was replaced with GTG, which abolished production of wild-type luciferase (data not shown). The human TH gene sequence from −2187 to +1300 was then inserted in-frame with the luciferase gene: the modified pGI3 vector was digested with KpnI and blunted, then digested with HindIII and ligated to the TH gene obtained from the digestion of pJ4.7 with EcoRV and HindIII. The resulting construct was designated (TCAT)₁₀₃.

To facilitate modification of the HUMTH01 microsatellite, the KpnI–NotI fragment of the TH gene was then inserted into the pBluescript vector. Using PCR, the sequence CCC(TCAT)₉CAT(TCAT)₂TCACCATGG, including the (TCAT)₁₀₃ motif and ending at the unique Ncol restriction site (underlined) was replaced either with CCCCATGG or CCCGGGCATGG. In the first construct, (TCAT)₁₀₃, the microsatellite was eliminated. The second construct resulted in the creation of a Smal restriction site, an enzyme that gives blunt ends. Various sequences were thus easily inserted into the Smal–Ncol cassette without modification of the sequences surrounding the HUMTH01 microsatellite. The KpnI–NotI fragment could then be transferred from pBluescript back to the original (TCAT)₁₀₃ construct. (TCAT)₁₀₃ (n = 3, 5, 8 and 10) and (CCAC)₁₀₃ plasmids were thereby constructed. To generate (TCAT)₁₀₃A and (TCAT)₁₀₃A⁺₁₀₃ constructs, in which a (TCAT)₁₀₃ motif is present in the proximal promoter, the GACGTC(TCAT)₁₀₃GACGTC sequence was inserted into the AatII site in both the (TCAT)₁₀₃ and (TCAT)₁₀₃ constructs. The reporter constructs with the thymidine kinase minimal promoter were described previously (7).

Transcriptional activity measurement of reporter constructs

PC12 cells (rat pheochromocytoma) were grown and transfected as described (7), using 2 µg of the reporter construct, 10 µg of pBluescript and 3 µg of the Renilla luciferase expression plasmid to assess transfection efficiency (pRL-SV40; Promega). In co-transfection experiments, pBluescript was replaced with the empty pcDNA3.1 expression plasmid (Invitrogen) or with the ZNF191 cDNA inserted into pcDNA3.1 vector. CHP212 cells (human neuroblastoma) were grown in F-15 DMEM (Sigma) supplemented with 15% heat-inactivated fetal calf serum and 1% glutamine. The cells were seeded at 1.5 × 10⁴ cells/well in a six-well plate and transfected 48 h later with 20 µg of reporter plasmid and 3 µg of pRL-SV40 by the calcium phosphate co-precipitation method (23). The medium was replaced after 24 h.

Firefly and Renilla luciferase activities were measured 48 h after transfection using the Dual-Luciferase Reporter Assay System (Promega) according to manufacturer’s instructions. Light emission was measured with a Lumat LB9501 (Berthold) luminometer. Preliminary studies indicated that the luciferase and Renilla activities were both in the linear range. Experiments were repeated with at least two DNA preparations of each construct.

One-hybrid system

Two copies of the (TCAT)₁₀₃ motif were inserted into the blunted XhoI restriction site upstream from the HHS3 gene in the yeast expression shuttle vector pYi (24). The yeast strain
formants were grown to an A 595 nm of 0.4, then induced with M15 overproducing the Lac repressor protein. The transfor six histidines to the N-terminus of the encoded ZNF191 restriction sites of pQE30 (Qiagen) thereby fusing a tag coding The full-length cDNA was inserted between the sequence was obtained by PCR using human genomic DNA. × phosphate buffer pH 7.0, 0.8 M NaCl, 2.5 55 Blots I and III; Invitrogen) were pre-hybridized overnight at blots, commercial filters (Northern Territory Human Normal Riboprobe Combination System Sp6/T7 according to the ZNF191 Northern blot analysis expression of ZNF191 protein in bacteria and purification. The recombinant protein was purified from cell extracts by column chromatography using Ni-NTA resin under denaturing conditions (8 M urea), according to the manufacturer’s recom mendations. The protein was renatured slowly by successive dialysis (Slide-A-Lyzer dialysis cassettes; Pierce) in buffers containing 7, 6, 5, 4, 3, 2, 1, 0.5, 0.25 and 0 M urea (25 mM HEPES pH 7.6, 100 mM KCl, 0.1 mM EDTA, 12.5 mM MgCl2, 10 µM ZnSO4, 10% glycerol, 0.1% Nonidet P-40, 0.5 mM PMSF, 1 mM DTT). Approximately 500 µg of eluted affinity-purified protein was recovered as estimated by Bradford protein assay (Bio-Rad) and was stored in aliquots at –80°C. The purified protein was resolved by SDS–PAGE as a 45 kDa band, consistent with the estimated molecular weight.

EMSA
So that the various probes were of comparable lengths, the shorter probes were made up to 40 bp with the flanking intrinsic sequences of the TH gene. The complementary oligonucleotides encoding the (TCAT) motifs were annealed and labeled by incubating with T4 polynucleotide kinase and [γ-32P]ATP (250 Ci/mmol). The probes were purified with Biogel P-10 resin (Bio-Rad) and isolated from a 6% poly acrylamide gel by elution and ethanol precipitation. Each binding reaction mix contained 1 µg of purified ZNF protein and 0.1 pmol of labeled DNA in 10 µl buffer [15 mM HEPES pH 7.6, 60 mM KCl, 6 µM EDTA, 7.5 mM MgCl2, 7.5 mM ZnCl2, 6% glycerol, 0.6% Nonidet P-40, 0.3 mM PMSF, 0.6 mM DTT, 1 µg poly(dI)/poly(dC)]. Protein, poly(dI)/poly(dC) and unlabeled competitors were first incubated for 15 min at room temperature. The radiolabeled probe was then added and the samples incubated for another 15 min. The DNA–protein complexes were analyzed by electrophoresis through a 6% polyacrylamide gel, which was then dried for autoradiography.

To determine the affinity of ZNF191 for (TCAT)n probes, a fixed amount of ZNF191 was incubated with various amounts of probe. The amounts of free probe, F, and bound probe, B1 and B2, corresponding to the binding of one or two proteins (complexes I and II) were determined. According to Scatchard analysis for one site binding: B/F = (P0 – B)/K, where K is the affinity. In our experiments, there was a molar excess of ZNF191 protein such that P0 >> B, leading to the simplified linear relation B = (P0/K)F. If the probe has two binding sites (a) and (b) with affinities Ka and Kb and there is no cooperativity, the amount of probe with a protein bound to site (a) is B(a) = (P0/Ka)F and that with a protein at site (b) is B(b) = (P0/Kb)F. B(a) and B(b) are not accessible by these experiments, as EMSA experiments measure the total amount of probe with one protein bound to site (a) is B(a) = (P0/Ka)F + (P0/Kb)F. The binding of a second ZNF191 molecule to site (b) of a probe with site (a) already occupied corresponds to equation B(a) + B(b) = (P0/Ka)F + (P0/Kb)F. The binding of a second ZNF191 molecule to site (b) of a probe with site (a) already occupied corresponds to equation B(a) + B(b) = (P0/Ka)F + (P0/Kb)F. Therefore, B1 = (P0/Ka)F and B2 = (P0/Kb)F. Thus, B1 and B2 are expected to be linearly related to free probe F, and Ka and Kb are obtained from the slopes of the plots.

The affinities were thus determined for probes (TCAT)n, (TCAT)n, and (TCAT)10, free probe, F, and bound probe, B1 and B2, corresponding to complexes I and II were quantified after incubation of a fixed amount of ZNF191 with various amounts of DNA probes. Models with one or two sites were used to determine the affinities of the recombinant ZNF191 protein. In both models, B1 and B2 were assumed to be linearly

Expression of ZNF191 protein in bacteria and purification
The cDNA isolated lacked 90 bp of the ZNF191 gene. This sequence was obtained by PCR using human genomic DNA. The full-length cDNA was inserted between the PstI–HindIII restriction sites of pQE30 (Qiagen) thereby fusing a tag coding for six histidines to the N-terminus of the encoded ZNF191 protein. The vector was introduced into Escherichia coli strain M15 overproducing the Lac repressor protein. The transfor mants were grown to an A 595 nm of 0.4, then induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside for 4 h. The

EMSA
So that the various probes were of comparable lengths, the shorter probes were made up to 40 bp with the flanking intrinsic sequences of the TH gene. The complementary oligonucleotides encoding the (TCAT)n motifs were annealed and labeled by incubating with T4 polynucleotide kinase and [γ-32P]ATP (250 Ci/mmol). The probes were purified with Biogel P-10 resin (Bio-Rad) and isolated from a 6% polyacrylamide gel by elution and ethanol precipitation. Each binding reaction mix contained 1 µg of purified ZNF protein and 0.1 pmol of labeled DNA in 10 µl buffer [15 mM HEPES pH 7.6, 60 mM KCl, 6 µM EDTA, 7.5 mM MgCl2, 7.5 mM ZnCl2, 6% glycerol, 0.6% Nonidet P-40, 0.3 mM PMSF, 0.6 mM DTT, 1 µg poly(dI)/poly(dC)]. Protein, poly(dI)/poly(dC) and unlabeled competitors were first incubated for 15 min at room temperature. The radiolabeled probe was then added and the samples incubated for another 15 min. The DNA–protein complexes were analyzed by electrophoresis through a 6% polyacrylamide gel, which was then dried for autoradiography.

To determine the affinity of ZNF191 for (TCAT)n probes, a fixed amount of ZNF191 was incubated with various amounts of probe. The amounts of free probe, F, and bound probe, B1 and B2, corresponding to the binding of one or two proteins (complexes I and II) were determined. According to Scatchard analysis for one site binding: B/F = (P0 – B)/K, where K is the affinity. In our experiments, there was a molar excess of ZNF191 protein such that P0 >> B, leading to the simplified linear relation B = (P0/K)F. If the probe has two binding sites (a) and (b) with affinities Ka and Kb and there is no cooperativity, the amount of probe with a protein bound to site (a) is B(a) = (P0/Ka)F and that with a protein at site (b) is B(b) = (P0/Kb)F. B(a) and B(b) are not accessible by these experiments, as EMSA experiments measure the total amount of probe with one protein bound to site (a) is B(a) = (P0/Ka)F + (P0/Kb)F. The binding of a second ZNF191 molecule to site (b) of a probe with site (a) already occupied corresponds to equation B(a) + B(b) = (P0/Ka)F + (P0/Kb)F. Therefore, B1 = (P0/Ka)F and B2 = (P0/Kb)F. Thus, B1 and B2 are expected to be linearly related to free probe F, and Ka and Kb are obtained from the slopes of the plots.

The affinities were thus determined for probes (TCAT)n, (TCAT)n, and (TCAT)10, free probe, F, and bound probe, B1 and B2, corresponding to complexes I and II were quantified after incubation of a fixed amount of ZNF191 with various amounts of DNA probes. Models with one or two sites were used to determine the affinities of the recombinant ZNF191 protein. In both models, B1 and B2 were assumed to be linearly
dependent on F, and affinity values were deduced from the slopes of the plots and the amount of protein, $P_0$.

**DNAse I footprinting analysis**

PCR amplification of a 140 bp fragment from the human TH intron containing the (TCAT) repeats generated the templates for footprint assay. The 5' primer contained a NotI restriction site. The PCR fragment was purified on a Concert column (Gibco), digested with NotI and labeled with [$\alpha$-$^32$P]dCTP using Klenow fragment. Labeled products were purified using Biogel P-10 resin. Binding reactions were as described for EMSA, using 20 000 c.p.m. of the labeled DNA and 0–5 µg of purified protein. After 30 min incubation, 1 µl of Mg–Ca buffer (10 mM MgCl$_2$ and 5 mM CaCl$_2$) and 1 U of DNAse I was added to the reaction which was incubated for an additional 30 s. The reaction was then stopped by adding 100 µl of stop buffer (0.15% SDS, 30 mM EDTA, 20 mM sodium acetate, 3 µg/ml yeast tRNA). The samples were extracted with phenol–chloroform and precipitated with ethanol and the precipitated DNA was dissolved in 3 µl loading buffer (80% formamide, 45 mM Tris base, 45 mM boric acid, 1 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol) and analyzed on a 6% sequence gel.

**ACKNOWLEDGEMENTS**

We thank Dr Fabienne Chatail in our laboratory for help with the one-hybrid system. V.A. is a recipient of a fellowship from the Association pour la Recherche contre le Cancer. This work was supported by grants from the Association Française contre les Myopathies, Aventis Pharma and the Centre National de la Recherche Scientifique.

**REFERENCES**