Nucleotide sequence conservation of novel and established *cis*-regulatory sites within the tyrosine hydroxylase gene promoter

Meng Wang[1], **Kasturi Banerjee**[1,*], **Harriet Baker**[1,2], and **John W. Cave**[*,1,2]

[1]Burke Medical Research Institute, White Plains, NY 10605, USA

[2]Weill Cornell Medical College Brain and Mind Research Institute, NY 10065, USA

Abstract

Tyrosine hydroxylase (TH) is the rate-limiting enzyme in catecholamine biosynthesis and its gene proximal promoter (< 1 kb upstream from the transcription start site) is essential for regulating transcription in both the developing and adult nervous systems. Several putative regulatory elements within the *TH* proximal promoter have been reported, but evolutionary conservation of these elements has not been thoroughly investigated. Since many vertebrate species are used to model development, function and disorders of human catecholaminergic neurons, identifying evolutionarily conserved transcription regulatory mechanisms is a high priority. In this study, we align *TH* proximal promoter nucleotide sequences from several vertebrate species to identify evolutionarily conserved motifs. This analysis identified three elements (a TATA box, cyclic AMP response element (CRE) and a 5′-GGTGG-3′ site) that constitute the core of an ancient vertebrate *TH* promoter. Focusing on only eutherian mammals, two regions of high conservation within the proximal promoter were identified: a ~250 bp region adjacent to the transcription start site and a ~85 bp region located approximately 350 bp further upstream. Within both regions, conservation of previously reported *cis*-regulatory motifs and human single nucleotide variants was evaluated. Transcription reporter assays in a *TH* -expressing cell line demonstrated the functionality of highly conserved motifs in the proximal promoter regions and electromobility shift assays showed that brain-region specific complexes assemble on these motifs. These studies also identified a non-canonical CRE binding (CREB) protein recognition element in the proximal promoter. Together, these studies provide a detailed analysis of evolutionary conservation within the *TH* promoter and identify potential *cis*-regulatory motifs that underlie a core set of regulatory mechanisms in mammals.

Keywords

catecholamine; dopamine; transcription; evolution

Correspondence: John W. CAVE, joc2042@med.cornell.edu.
*Current address: The Gene and Linda Voiland School of Chemical Engineering and Bioengineering, Washington State University, Pullman, WA 99164-6515, USA

Introduction

Tyrosine hydroxylase (TH), the rate-limiting enzyme for catecholamine neurotransmitter biosynthesis, is essential for proper neurological function. In mice, homozygous disruption of the *TH* gene is embryonic lethal due to altered cardiac and/or cardiovascular development (Kobayashi et al., 1995; Zhou et al., 1995). By contrast, *TH* heterozygous mutant mice are viable, fertile with a normal physical appearance, but they have reduced noradrenaline levels in multiple brain regions (Kobayashi et al., 2000). These reduced noradrenaline levels are associated with impaired associative and latent learning, although hippocampus-mediated spatial learning and long-term potentiation are normal. In humans, single nucleotide variations in the human *TH* coding region are more frequent than deletion of the locus. Individuals with mutations in both *TH* alleles are associated with TH deficiency, which is a spectrum of movement disorders that typically first present in infants (reviewed in Willemsen et al., 2010). The majority of pathogenic single nucleotide variations in the human *TH* gene are expected to produce mutant TH proteins with either reduced stability or enzyme activity.

Because of its essential neurological role, the spatial and temporal regulation of *TH* transcription in the nervous system has been extensively studied. The majority of studies have concentrated on the role of the promoter region (< 1 kb upstream) in the human and rodent *TH* gene and identified several *cis*-regulatory elements that modulate expression (Schimmel et al., 1999; Yang et al., 1998; Kessler et al., 2003; Kim et al., 2003b). The conservation of these sites across species, however, has not been well investigated. Given the extensive use of rodents and other vertebrates to model human catecholaminergic neuronal function, identifying conserved proximal promoter *cis*-regulatory elements is a high priority. The shared brain region-specific expression of TH in mammals suggests that there should be strong conservation of *cis*-regulatory elements, but some promoter regulatory elements have been previously found to lack conservation. For example, the transcription factor ER81 (ETV1) directly binds the mouse *TH* proximal promoter to activate transcription in the olfactory bulb, but this mechanism is not conserved in other mammals, including humans (Cave et al., 2010).

In this study, the *TH* promoter from several vertebrate species was aligned to identify conserved nucleotide motifs. This analysis identified three elements (the TATA box, cyclic AMP response element (CRE) and a 5′-GGTGG-3′ site) that constitute an ancient vertebrate *TH* core proximal promoter. In eutherian mammals, conservation was highest within the proximal ~250 bp promoter and a separate ~85 bp region centered approximately 580 bp upstream of the translation start site. Conservation of previously reported *cis*-regulatory motifs and human single nucleotide variants within both of these promoter regions was evaluated. Transcription reporter assays with a *TH*-expressing cell line demonstrated the functionality of these proximal promoter regions, and electromobility shift assays (EMSAs) with tissue lysate showed that brain region-specific complexes assemble on these motifs. These studies also identified a non-canonical CRE binding (CREB) protein recognition element in the mammalian proximal promoter. Together, these studies provide a detailed analysis of evolutionary conservation within the *TH* promoter and identify potential *cis*-regulatory motifs that underlie a set of conserved regulatory mechanisms in mammals.

Materials and methods

Nucleotide sequence alignments

Genomic sequences containing the 1 kb *TH* promoter region (relative to the translation start) were downloaded from Ensembl (http://www.ensembl.org). The species used for the alignments were: anolis (*Anolis carolinensis*), cat (*Felis catus*), chicken (*Gallus gallus*), cod (*Gadus morhua*), coelacanth (*Latimera chalumnae*), cow (*Bos taurus*), dolphin (*Tursiops truncatus*), duck (*Anas platyrhynchos*), ferret (*Mustela putorius furo*), flycatcher (*Ficedula albicollis*), frog (*Xenopus tropicalis*), fugu (*Takifugu rubripes*), gorilla (*Gorilla gorilla*), guinea-pig (*Cavia porcellus*), Horse (*Equus caballus*), human (*Homo sapiens*), Medaka (*Oryzias latipes*), mouse (*Mus musculus*), opossum (*Monodelphis domestica*), orangutan (*Pongo abelii*), rat (*Rattus norvegicus*), squirrel (*Ictidomys tridecemlineatus*), Stickleback (*Gasterosteus aculeatus*), turkey (*Meleagris gallapavo*), turtle (*Pelodiscus sinensis*), zebra finch (*Taeniopygia guttata*), and zebrafish (*Danio rerio*). Sequence alignments and visualizations were performed using Multi-LAGAN and mVista web-based services (http://genome.lbl.gov/vista/mvista/submit.shtml) (Brudno et al., 2003; Frazer et al., 2004).

Luciferase activity assays and cell culture

The olfactory bulb (OB) *TH*-expressing clonal cell line (OB4671) was derived from transgenic mice expressing the SV40 large T-antigen under the control of the rat 9kb *TH* upstream region (Son et al., 1999). The line was generated using olfactory bulb tissue dissected from post-natal day 2 animals and maintained at 33°C with 95% air and 5% CO_2 on primaria coated culture dishes in DMEM cell culture medium supplemented with 10% FBS and 1% penicillin/streptomycin.

Luciferase activity assays in the OB cell line were conducted with stably transfected cells. For each line, 2.5×10^6 cells in 60 mm primaria coated culture dishes were transfected using Fugene HD transfection reagent (Roche) with 2 μg of pGL4.20 firefly luciferase reporter plasmid (Promega). The firefly luciferase was under the control of the rat 4.5 kb *TH* upstream region containing either the wild-type sequence or mutations to highly conserved motifs. The pGL4.20 plasmid also constitutively expressed a puromycin resistance gene to enable selection of transfected cells. 48 h after transfection, culture media was supplemented with 25 μg/mL puromycin to select transfected cells. The puromycin concentration was reduced to 8.3 μg/mL after one subculture and this reduced level was used to maintain the lines.

To measure luciferase activity in stably transfected cell lines, sub-confluent cultures in log phase of growth were dissociated with trypsin, washed and suspended in 1ml of PBS. The density of the isolated cells was measured at a wavelength of 600 nm. Cells were then pelleted and treated with reporter lysis buffer (Promega). Using a firefly luciferase assay kit (Promega), luminescence was measured with a LMaxII luminometer (Molecular Devices). Luminescence readings were normalized to the respective cell densities. Luciferase activity was reported as the mean of at least three independent transfection experiments with error bars representing the standard deviation. Significant changes in activity ($p < 0.05$) were assessed by ANOVA with appropriate post hoc tests.

Electromobility shift assays (EMSAs)

Oligonucleotides were end-labeled with γ-P^{32} ATP (Perkin Elmer) using T4 polynucleotide kinase (New England Biolabs) and unincorporated nucleotides were removed using Bio-30 spin columns (Bio-Rad). All protein/DNA binding reactions were conducted in 20 μL with 5 mM HEPES pH 7.9, 25 mM KCl, 0.05 mM EDTA, 0.125mM phenylmethysulfonyl fluoride, 1mg/mL of non-specific herring sperm carrier DNA, 0.0l% NP40 and 2 μL of radiolabeled DNA probe. Nuclear lysate was prepared from the OB and midbrain (MB) of adult C57BL6 mice using NEPER Nuclear Extraction kit (Pierce/Thermo Scientific) according the manufacturer's instructions. Briefly, fresh OB and MB tissue obtained from adult mice was rinsed with PBS and weighed before being immersed in a Dounce homogenizer containing ice-cold CER I buffer (Pierce/Thermo Scientific) supplemented with a protease inhibitor cocktail (Pierce/Thermo Scientific). Following homogenization, the tissue lysate was incubated on ice for 10min before addition of CERII buffer (Pierce/Thermo Scientific) and incubated on ice for additional 1 min. The tissue homogenate was spun down at 16000×g for 5 min at 4°C. The supernatant (containing the cytoplasmic fraction) was removed from the pellet (containing the cell nuclei) and the pellet was re-suspended with NER buffer (Pierce/Thermo Scientific) containing a protease inhibitor cocktail (Pierce/Thermo Scientific). The suspended nuclear fraction was incubated on ice for 40 min, with a 15 s of vortexing every 10 min, before being spun at Centrifuge at 16000×g for 10 min at 4°C. After which, the supernatant (containing the nuclear extract) was transferred to a clean pre-chilled tube and protein content was estimated by A$_{280}$ absorbance. Purified CREB was a generous gift from Dr. Jennifer K. Nyborg (Colorado State University). Binding reactions were carried out at 4°C for 1 h before being resolved on 3% 29:1 acrylamide/bis-acrylamide gels with 1/4×TBE buffer at 215 V for 2 h. Gels were dried and then imaged using a BAS2500 phosphoimager (GE Healthcare). Oligonucleotide sequences used for EMSAs are shown in the figures.

Results

Vertebrate *TH* proximal promoter nucleotide sequence alignments

To identify regions of evolutionary conservation in the vertebrate *TH* promoter, we aligned orthologous nucleotide sequences (1kb upstream relative to the translation start site) from 28 vertebrate species representing mammals, avians, reptiles, amphibians and bony fish (osteichthyans). Using the human *TH* gene as a reference, this alignment revealed that only ~100 bp upstream of the translation start site is highly conserved in all five vertebrate orders (Fig. 1). Within this region, both the TATA box and CRE are highly conserved (Fig. 2). There is also a 5′-GGTGG-3′ motif (GGTGG box) upstream of the CRE that is conserved in all tetrapods (mammals, avians, reptiles and amphibians) and partially conserved in fish (Fig. 2). The two fish species examined lacking the GGTGG box (stickleback and puffer fish) are the most recently divergent species of the ray-finned fish examined in this study (which also included cod, medaka and zebrafish) (Yamanoue et al., 2006). The conservation of the three motifs (GGTGG box, CRE and TATA box) in the older ray-finned fish species (cod, medaka and zebrafish) as well as in lobe-finned fish (coelacanth) and tetrapods suggests that these three elements represent the core of a vertebrate *TH* proximal promoter that evolved at least 450 million years ago.

Except for eutherian (placental) mammals, vertebrates also contain a *TH* gene paralogue (*TH2*) that may have arisen from whole-genome duplication in early vertebrate evolution (Yamamoto et al., 2010). *In situ* hybridization studies in both developing and adult fish have revealed that these two *TH* paralogues are largely expressed in complementary patterns, suggesting that the *cis*-regulatory regions for these paralogues have substantially diverged (Chen et al., 2009; Filippi et al., 2010; Yamamoto et al., 2010). Consistent with this interpretation, we did not find conservation of either the CRE or GGTGG box in the *TH2* proximal promoter of several vertebrate species (data not shown).

When fish were excluded from the alignment, the remaining tetrapod species showed conservation extending approximately 200 bp upstream of the transcription start site. This extended conservation included two G:C-rich regions upstream from the CRE (G:C rich regions 1 and 2; Fig. 3). G: C Region 1 contained the GGTGG box described in the analysis above that included the fish species. G:C-rich Region 2 also contained GGTGG box on the complementary strand, but this motif was conserved only in mammals (Fig. 3). In both mammals and sauropsids (avians and reptiles), an additional region was conserved upstream from G:C-rich Region 2 (Fig. 3). This additional region contained a pair of canonical homeodomain recognition motifs (5′-TAAT-3′). One motif (HD-1) was conserved in all mammalian and sauropsid species. The other motif (HD-2) was also conserved in both mammals and sauropsids, but absent specifically in primates (Fig. 3). Upstream from these homeodomain motifs, sauropsid sequences diverge from the mammalian orthologs and there is no further significant homology between these two groups of vertebrates.

Conservation of cis-regulatory elements in the mammalian *TH* proximal promoter

When analyzing only eutherian mammals, two distinct promoter regions of conservation were evident (Fig. 4A). The more distal region is about 85 bp long and is centered about 580 bp upstream of the translation start site (Fig. 4A and 4B). To the best of our knowledge, this is a previously unrecognized region of high conservation in the mammalian *TH* proximal promoter. Within this distal promoter region is a butyrate response element (BuRE; Fig. 4B) (Patel et al., 2005). This response element was originally identified in rodents and our alignment revealed that this site is partially conserved in mammals.

The proximally conserved region covers about 250 bp, with the distal edge defined by an AP-1 binding site (Fig. 4C) (Cambi et al., 1989). The AP-1 site is conserved in all mammals and recruits several Fos and Jun-related basic-leucine zipper (bZIP) transcription factors (Nagamoto-Combs et al., 1997; Liu et al., 1999). The AP-1 site is encompassed within a dyad motif identified in the rat promoter (Yoon and Chikaraishi, 1992). This dyad also contains an E-box (5′-CANNTG-3′), which is the canonical recognition site for basic-helix–loop–helix transcription factors. Except for the AP-1 site, the dyad sequence is poorly conserved in non-rodent species.

The region containing the homeodomain motifs overlaps with previously reported Octamer and Heptamer sites as well as a neural restrictive silencing element (NRSE)(Dawson et al., 1994; Kim et al., 2006). The Octamer and Heptamer sites (which overlap HD-1 and HD-2, respectively) were originally recognized for their similarity to binding sites for POU-domain transcription factors (Dawson et al., 1994), which are a subclass of homeodomain proteins.

With the exception of the Heptamer in primates, both the Octamer and Heptamer motifs are conserved across mammals. By contrast, the putative NRSE was originally identified in the human *TH* promoter (Kim et al., 2006) and our analysis showed this motif is poorly conserved outside of primates. The NRSE mediates repression by recruiting the neural repressive silencing factor (NRSF), and the poor conservation of this element suggests that regulation of *TH* by NRSF directly targeting the proximal promoter may be unique to primates.

The G:C-rich regions also contain several previously reported regulatory motifs. G:C-rich region 2 contains partially overlapping binding sites for EGR-1 and SP-1 (Kobayashi et al., 1988; Papanikolaou and Sabban, 1999, 2000). The EGR-1 site also overlaps with the distal GGTGG box. The nucleotide sequences for both the EGR-1 and SP-1 sites have some species variability, but they are generally conserved across mammals. By contrast, the ER81 and CRE2 sites in G:C Region 1 display poor conservation in mammals. The lack of conservation for the ER81 site was expected since previous studies showed direct regulation of *TH* transcription by this protein at this site is a rodent-specific mechanism (Cave et al., 2010). The current analysis confirms the poor conservation of the ER81 site across a wide range of mammalian species. The CRE2 site was reported to enhance the induction of *TH* transcription in response to cAMP (Best et al., 1995). Unlike the CRE, however, the CRE2 sequence identified in the rat *TH* promoter is not conserved outside of rodents.

The *TH* core proximal promoter is centered on the highly conserved TATA box. This site is bound by the TATA binding protein (TBP) subunit of the TFIID general transcription factor. On canonical TATA box-dependent promoters, TFIIB straddles TBP by binding to recognition elements, uBRE and dBRE, that are adjacent to the TATA box. In mammalian *TH* promoters, there is modest conservation of the uBRE, but conservation of the dBRE consensus sequence was limited to primates. Given that there is little tolerance for variable spacing between the TATA box and either the uBRE or dBRE (Venters and Pugh, 2013), the alignment data indicate that TFIIB interacts with the *TH* promoter through sites that substantially deviate from consensus recognition sequences. The dBRE position also overlaps a partial dyad sequence identified in the rat *TH* promoter (Patankar et al., 1997). This partial dyad is likely involved in the assembly or activation of the basal transcription complex, but displays only modest conservation throughout mammals.

The core promoter region has also been reported to recruit tissue-specific transcription factors. Like the upstream AP-1 site, the CRE in the core promoter is highly conserved throughout mammals and is recognized by several bZIP transcription factors (Sabban, 1997; Liu et al., 1999; Yukimasa et al., 1999; Suzuki et al., 2002). A putative PITX3 binding site partially overlaps the CRE (Cazorla et al., 2000; Kim et al., 2003b). PITX3 is a member of the bicoid class of paired-homeodomain transcription factors and is essential for development of a subset of midbrain dopami-nergic neurons (Hwang et al., 2003; Nunes et al., 2003; Smidt et al., 2004). The putative binding site is highly conserved in the mammalian *TH* promoter, but substantially deviates from the canonical bicoid paired-homeodomain recognition sequence. The rodent core promoter is also reported to recruit the orphan nuclear receptor NURR1 to a site that overlaps the uBRE and TATA box (Iwawaki et al., 2000; Kim et al., 2003a). NURR1 is also critical for development of midbrain

dopaminergic neurons (Saucedo-Cardenas et al., 1998; Zetterström et al., 1997), but the putative binding site in the *TH* promoter is not well conserved outside of rodents.

Overlap of single nucleotide variants with conserved regions of the human *TH* promoter

The regions of high conservation also overlapped several single nucleotide variants identified in the human *TH* promoter (Fig. 4B and 4C). Most polymorphisms in the *TH* promoter are not associated with a phenotype, but known variants in the CRE are considered to be causal for TH deficiency for individuals lacking mutations in the TH coding region (Verbeek et al., 2007). These mutations presumably impair recognition of the CRE by bZIP transcription factors and diminish *TH* expression levels. In the distally conserved portion of the *TH* promoter, there are two adjacent variants in proximity to the BuRE (Fig. 4B). One of these variants (rs10770140) is associated with elevated norepinephrine secretions and blood pressure during stress (Rao et al., 2007). Unlike the polymorphisms in either the CRE or proximal GGTGG box, however, the rs10770140 position is not highly conserved outside of primates, suggesting that regulatory mechanisms affected by this variant are primate-specific.

Modulation of *TH* proximal promoter activity by mutation of highly conserved motifs

To establish whether mutation of the most highly conserved motifs in the mammalian *TH* proximal promoter could modulate gene expression levels, luciferase activity assays were performed in a *TH* -expressing cell line derived from the mouse olfactory bulb. These assays used luciferase reporters under the control of the 4.5kb rat *TH* promoter, which is the minimal sequence necessary to activate gene expression in all catecholaminergic brain regions (Schimmel et al., 1999). The luciferase activity assays tested promoters with mutations to either the AP-1, both HD, the overlapping EGR-1/SP-1, GGTGG, PITX-3 or CRE motifs (Fig. 5A). As a comparison, a region of lower conservation was also mutated (labeled "control"; Fig. 5A). The assays revealed that mutations to all of the highly conserved motifs substantially reduced promoter activity (Fig. 5B). Disruption of the CRE had the strongest effect, but all mutations reduced reporter expression by at least 50%. By contrast, mutating the region of lower conservation did not significantly reduce reporter activity (Fig. 5B). Together, these results indicate that regions of highest conservation are functional in modulating *TH* promoter activity.

Recruitment of brain region-specific complexes

Region-specific expression of transcription factors is critical for establishing the spatial distribution of catecholaminergic neuronal groups within the nervous system (Goridis and Rohrer, 2002; Burbach et al., 2003; Cave and Baker, 2009). To test whether brain region-specific protein complexes assemble on the highly conserved motifs within the *TH* proximal promoter, EMSAs were performed using nuclear lysate from either the mouse olfactory bulb (OB) or midbrain (MB). These lysates were used with five oligonucleotides that encompassed either the AP-1, HD-1/HD-2, EGR-1/SP-1, GGTGG or CRE motifs (Fig. 6A). Although potential differences in the proportion of dopaminergic neurons within the OB and MB lysates may account for some of the differences in band intensities, brain region-dependent and independent complexes were detected on each target sequence (Fig. 6B). The

oligonucleotides with either the HD-1/HD-2 or EGR-1/SP-1 motifs produced the greatest number of regionally distinct complexes, suggesting that these motifs are important for integrating regional input to direct *TH* transcription. By contrast, the AP-1, CRE and GGTGG motifs only contained either 1 or 2 regionally distinct complexes, suggesting that these motifs function primarily as part of a region-independent mechanism for regulating *TH* expression.

Non-canonical CREB binding sites in the *TH* proximal promoter

In an EMSA experiment initially designed to probe transcription factor protein interactions on an oligonucleotide containing both the CRE and GGTGG sites, CREB was capable of binding a target sequence lacking the CRE (Fig. 7A). Since this EMSA used purified CREB protein and not nuclear lysate, the observed shift was not the result of other, unknown proteins associating with the target oligonucleotide. Inspection of the nucleotide sequence upstream of the CRE revealed three overlapping motifs that resembled the CRE2 site (Fig. 7A) (Best et al., 1995). Based on this resemblance, a CRE2-like consensus sequence of 5′-AGGNNGGG-3′ was derived and used to search the rat *TH* proximal promoter (between the AP-1 and CRE) for additional CRE2-like sites. The search criteria allowed for variation within the three adjacent G nucleotides at the 3′ end of the motif, as long as two of the three positions remained a G nucleotide. This search identified a total of 8 sites matching these criteria (including the original CRE2 site; Fig. 7B).

Since the CRE2 site functions in combination with the CRE to induce *TH* transcription in response to cAMP (Best et al., 1995), EMSAs tested whether purified CREB could also bind the CRE2. These experiments showed that CREB bound the CRE2 site, but with a lower affinity than the canonical CRE motif (Fig. 7C). CREB also bound all of the CRE2-like sites, but with different affinities that were also less than the CRE (Fig. 7C). Thus, the CRE2-like sites appear to be low affinity, non-canonical CREB binding sites. CREB also bound the AP-1 site, but also with lower affinity than the CRE, which is consistent with previously published studies (Barrett and Vedeckis, 1996; Moll et al., 2002). The inability of CREB to bind target oligonucleotides containing the HD-1 site confirmed that the observed protein-DNA complexes formed in these EMSAs required specific interactions with either the CRE2-like or canonical CRE sites (Fig. 7C).

To establish whether CRE2-like sites were conserved, alignments of the first ∼250bp of the mammalian *TH* proximal promoters were re-examined and modified to minimize gaps in CRE2-like motifs. This re-examination revealed that the number of CRE2-like sites varied for each species, but all were clustered in the G:C-rich regions or adjacent to the AP-1 site (Fig. 8). These findings indicated that the general location of sites within the promoter were conserved, but the precise number and location of individual sites were species-specific. The only exception was the almost complete conservation of the CRE2-like site that overlapped the proximal GGTGG box.

Discussion

Vertebrate conservation of the core *TH* proximal promoter

The present study identified a core mammalian *TH* promoter, containing a TATA-box, CRE and GGTGG box, conserved in vertebrate evolution. The conservation of this core promoter in mammals, avians, reptiles, amphibians and bony fish indicates that it evolved at least 450 million years ago. Whether these core promoter elements are also present in the common ancestor for all seven classes of vertebrates requires analyses of promoter sequences from cartilaginous fish (chondrichthyans) and jawless vertebrates (agnathostomes). Current assemblies for representative species from these two additional vertebrate classes, however, are not sufficient for the required analysis.

Within tetrapod species, the mammalian *TH* promoter showed the strongest conservation with sauropsids. Given the greater evolutionary divergence between mammals and amphibians, the greater conservation between mammals and sauropsids was not unexpected. Amphibians, however, were under-represented in the analysis when compared to other vertebrate orders examined. Sequencing of additional amphibian genomes will further refine our phylogenetic comparison. The current analysis showed that G:C rich regions and the core promoter elements were also present in mammals, sauropsids and amphibians, but the homeodomain motif region was noticeably absent from amphibians. This differential conservation of homeodomain motifs correlates with the division between amniotic and anamniotic species. Phylogenetic analyses of brain catecholaminergic regions have shown distinct differences in the organization of midbrain dopaminergic cell groups between these two sets of species (Smeets and González, 2000; Yamamoto and Vernier, 2011). Given the role of homeodomain transcription factors in regional specification during development, it is tempting to speculate that these conserved homeodomain motifs integrate regional developmental input that spatially organizes the dopaminergic midbrain in amniotes. Further studies are required to both clearly define the functional role of these specific sites in the regulation of *TH* expression and identify the transcription factors that bind this region (discussed further below).

Conservation within the mammalian *TH* proximal promoter

This study identified two distinct, conserved regions in the mammalian *TH* promoter that contained several previously reported *cis*-regulatory elements (summarized in Table 1). The shorter, distal region is a previously unrecognized area of conservation. The BuRE within this distal region augments activation of *TH* expression in response to butyrate by recruiting CREB (Patel et al., 2005). Inspection of the BuRE and its 5′ flanking nucleotides reveals a CRE2-like sequence (5′-AGGNNGGG-3′), but only in rodents. The reported recruitment of CREB by the BuRE is consistent with the current findings that CREB targets CRE2-like sites. Even though the BuRE shows reasonably modest conservation in mammals, variation exists at positions previously shown to diminish the butyrate-mediated induction of *TH* expression (Patel et al., 2005). Together with the observation that the CRE2-like consensus sequence is present in only rats and mice, these findings suggest that regulation of *TH* expression by the BuRE is a species-specific mechanism.

In addition to the BuRE, the human single nucleotide variant (rs10770140) is at a position in a primate-specific motif within the distally conserved region. It is interesting that this polymorphism in humans and heterozygous deletion of *TH* in mice are both associated with altered norepinephrine production (Kobayashi et al., 2000; Rao et al., 2007), but whether this distally conserved region in the *TH* promoter is essential for proper norepinephrine production remains to be established.

The proximal *TH* promoter contained five principal areas of high conservation: an AP-1 site, a putative homeodomain region, two G:C-rich regions and the core promoter. This AP-1 site recruits bZIP transcription factors that, in combination with factors bound to the CRE, enhance *TH* transcription in response to stimuli, such as depolarization (Nagamoto-Combs et al., 1997; Liu et al., 1999; Lewis-Tuffin et al., 2004). In both rodents and zebrafish, *TH* expression in the olfactory bulb is induced by odorant-mediated synaptic activity-dependent depolarization (Baker et al., 1983; Baker et al., 1993; Byrd, 2000; Paskin et al., 2011), and mutation of the AP-1 site strongly reduces *TH* expression in the rodent olfactory bulb (Baker et al., 2001). Conservation of the AP-1 site, however, was only observed in mammals. Thus, inducible *TH* transcription in non-mammalian species may be mediated primarily by the CRE.

Downstream from the AP-1 site, two canonical homeodomain recognition motifs were conserved in all mammals, except for primates, which only had one. The significance of this primate-specific variation is not clear because the transcription factors that bind this region are not known. Previous studies suggest that the POU/homeodomain transcription factor OCT-2 targets this region (Dawson et al., 1994), but in vivo evidence for regulation of *TH* by OCT-2 remains to be established. By contrast, members of the BRN family of POU-domain transcription factors are reported to modulate dopaminergic differentiation (Ichikawa et al., 2005; Tan et al., 2014) and may regulate *TH* through this region. EMSAs with olfactory bulb and midbrain tissue lysate suggest that this homeodomain motif region integrates region-specific regulatory input. Thus, brain region-specific homeodomain factors (including PAX6 and PHOX2A/B), which are established regulators of catecholaminergic differentiation, may also target this region (Lo et al., 1999; Hack et al., 2005; Kohwi et al., 2005). Although homeodomain proteins are logical candidates for binding to this region, other classes of transcription factors may bind instead of, or in cooperation with, homeodomain proteins. Elucidating the regulatory factors recruited by this region will likely provide critical insight into cell-type specific mechanisms that control *TH* expression.

The two G:C-rich regions are distinct from other areas within the proximal promoter because they are defined by the conservation of high G:C base content rather than specific nucleotide motifs. The only specific motifs within these regions that showed strong conservation were the GGTGG boxes. The proteins that recognize the GGTGG boxes are not known, but the distal site overlaps, and is adjacent to, reported binding sites for EGR-1 and SP-1, respectively. The role of SP-1 in *TH* regulation, however, is uncertain. Previous EMSAs with rat adrenal medulla extracts detected SP-1 binding to this region (Papanikolaou and Sabban, 1999), but studies have not demonstrated a functional role for SP-1 in *TH* transcription. EMSAs with rat adrenal medulla extracts also showed EGR-1 binding to this region (Papanikolaou and Sabban, 1999), but unlike SP-1, *in vitro* transcription assays have

shown a functional role for EGR-1 in upregulating *TH* promoter activity (Papanikolaou and Sabban, 2000; Nakashima et al., 2003; Stefano et al., 2006). These findings support a role for EGR-1 in inducing *TH* expression within the adrenal medulla in response to stress, but the lack of significant co-expression between EGR1 and TH within the brain suggests that EGR-1 does contribute to *TH* regulation within the CNS (Hebert et al., 2005; Akiba et al., 2009).

One function of the G:C-rich regions in the *TH* proximal promoter may be to modulate activation of *TH* transcription by recruiting bZIP proteins through the clusters of CRE2-like sites. The original CRE2 site is not conserved, but this study showed that a degenerate consensus sequence incorporating this site recruits CREB and is present in all mammalian species examined. The CREB binding affinity for individual sites varies, but appears considerably lower than the canonical CRE motif. The affinity of individual CRE2-like sites is likely dependent on the specific site sequence as well as flanking nucleotides. Preliminary studies indicate that these sites can also recruit Fos-containing bZIP proteins, but affinity for these sites relative to CREB has not been established (data not shown). We predict that clusters of lower affinity binding sites enable graded induction levels of *TH* transcription in response to variable stimuli intensities. At basal or low stimulus intensities, minimal levels of activated bZIP factors (such as phosphorylated Ser-133 CREB) would be expected to occupy the highest affinity sites (such as the CRE) on the *TH* promoter. Increased levels of activated bZIP factors generated by larger stimulus intensities could generate proportional increases in *TH* promoter activity by targeting the lower affinity CRE2-like sites. The number and position of CRE2-like sites may be largely dependent on species-specific requirements. The strong conservation of sequence and location for the CRE2-like site overlapping the proximal GGTGG box, however, suggests that some sites have stricter functional requirements. Most mammalian species also have CRE2-like sites adjacent to the AP-1 motif, suggesting these pair of sites function as tandem bZIP regulatory regions. Such an arrangement suggests that interactions between the AP-1 and CRE2-like sites underlie the functionality of the reported dyad motif encompassing the AP-1 site (Yoon and Chikaraishi, 1992). The degeneracy of the CRE2-like motif accounts for the lack of dyad sequence conservation through mammals. Further studies are required to establish the physiologic significance of specific CRE2-like motifs and the mechanistic role they serve in regulating *TH* transcription.

The G:C-rich regions also resemble portions of several oncogene promoters, including *c-myc* and Bcl-2, that adopt single-stranded DNA secondary structures (Qin and Hurley, 2008). Melting of the genomic duplex DNA during the initiation of gene transcription induces torsional strain that can promote strand separation in the upstream promoter (Brooks and Hurley, 2009). Recent studies indicate that the G: C-rich regions can also adopt single-stranded DNA secondary structures that influence *TH* promoter activity, and bZIP proteins facilitate recruitment of factors that stabilize the separated single *TH* promoter DNA strands (manuscript in preparation). Thus, the G:C-regions may mediate multiple transcription regulatory mechanisms that include recruiting bZIP and EGR-1 transcription factors for stimulus-induced *TH* expression and formation of DNA secondary structure.

All transcription mechanisms regulating *TH* expression converge on the core promoter. As discussed above, the CRE and TATA box within the core promoter are highly conserved throughout vertebrate evolution. Given the essential function of recruiting TBP, conservation of the TATA box was expected. The considerably weaker conservation of the uBRE and dBRE, despite their critical role in recruiting TFIIB, was unexpected. The recruitment and/or stability of TFIIB on the *TH* promoter may be more dependent on interactions with tissue-specific transcription factors or co-activators. The strong conservation of the CRE is consistent with reports showing mutation of this site blocks reporter gene expression in all adult brain catecholaminergic regions (Trocmé et al., 1998). CREB is the canonical bZIP protein that recognizes the CRE, but the specific bZIP proteins bound to the *TH* promoter CRE *in vivo* are likely cell-type dependent since Fos, Jun and ATF family proteins can also bind this site (Sabban, 1997; Liu et al., 1999; Yukimasa et al., 1999; Suzuki et al., 2002). Partially overlapping the CRE is a putative and conserved PITX3 site. *In vitro* EMSAs have suggested that PITX3 binds this site (Cazorla et al., 2000), but the established PITX3 recognition motif is considerably different and ChIP-seq experiment indicates that PITX3 binds *TH* at a position much further upstream from the core promoter (Jacobs et al., 2009). By contrast, the putative PITX3 site and its 5′ flanking nucleotides form a site that is very close to a CRE2-like consensus site established in this study. The CRE and this potential CRE2-like site form a tandem arrangement that is similar to the AP-1 and CRE2-like site upstream. Clearly, further studies are required to establish the sequence degeneracy of CRE2-like site in order to better define the *cis*-regulatory architecture within the *TH* promoter.

Conclusions

This study has identified several regions and motifs that are conserved within the mammalian *TH* promoter. These conserved regions and motifs likely form the basis for *TH* transcription regulatory mechanisms shared across mammals. The *TH* promoter, however, is necessary, but not sufficient to drive gene expression in all catecholaminergic regions (Min et al., 1994; Liu et al., 1997). The complete spatial and temporal regulation of *TH* expression likely requires interactions between the promoter and distal *cis*-regulatory regions. The genomic locations and sequences of these distal enhancer or repressor domains remain to be established. In addition to transcriptional regulation, a complete understanding of functional TH protein production also requires integration with post-transcriptional and post-translational mechanisms (recently reviewed in Lenartowski and Goc, 2011; Dickson and Briggs, 2013).

This study has focused on nucleotide conservation, but species-specific differences are also important to establish. Identification of human-specific transcription regulatory mechanisms are likely to have important consequences for advancing development of therapies to treat catecholaminer-gic-related disorders. The present focus on elucidating regions of conservation, however, may facilitate the design and improvement of non-human mammalian models of catecholaminergic systems.

Acknowledgements

References

Akiba N, Jo S, Akiba Y, Baker H, Cave JW. Expression of EGR-1 in a subset of olfactory bulb dopaminergic cells. J Mol Histol. 2009; 40(2):151–155. [PubMed: 19387849]

Baker H, Kawano T, Margolis FL, Joh TH. Transneuronal regulation of tyrosine hydroxylase expression in olfactory bulb of mouse and rat. J Neurosci. 1983; 3(1):69–78. [PubMed: 6130133]

Baker H, Liu N, Chun HS, Saino S, Berlin R, Volpe B, Son JH. Phenotypic differentiation during migration of dopaminergic progenitor cells to the olfactory bulb. J Neurosci. 2001; 21(21):8505–8513. [PubMed: 11606639]

Baker H, Morel K, Stone DM, Maruniak JA. Adult naris closure profoundly reduces tyrosine hydroxylase expression in mouse olfactory bulb. Brain Res. 1993; 614(1–2):109–116. [PubMed: 8102310]

Barrett TJ, Vedeckis WV. Occupancy and composition of proteins bound to the AP-1 sites in the glucocorticoid receptor and c-jun promoters after glucocorticoid treatment and in different cell types. Recept Signal Transduct. 1996; 6(3–4):179–193. [PubMed: 9259052]

Best JA, Chen Y, Piech KM, Tank AW. The response of the tyrosine hydroxylase gene to cyclic AMP is mediated by two cyclic AMP-response elements. J Neurochem. 1995; 65(5):1934–1943. [PubMed: 7595476]

Brooks TA, Hurley LH. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. Nat Rev Cancer. 2009; 9(12):849–861. [PubMed: 19907434]

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. and the NISC Comparative Sequencing Program. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 2003; 13(4):721–731. [PubMed: 12654723]

Burbach JP, Smits S, Smidt MP. Transcription factors in the development of midbrain dopamine neurons. Ann N Y Acad Sci. 2003; 991(1):61–68. [PubMed: 12846974]

Byrd CA. Deafferentation-induced changes in the olfactory bulb of adult zebrafish. Brain Res. 2000; 866(1–2):92–100. [PubMed: 10825484]

Cambi F, Fung B, Chikaraishi D. 5′ flanking DNA sequences direct cell-specific expression of rat tyrosine hydroxylase. J Neurochem. 1989; 53(5):1656–1659. [PubMed: 2571679]

Cave JW, Akiba Y, Banerjee K, Bhosle S, Berlin R, Baker H. Differential regulation of dopaminergic gene expression by Er81. J Neurosci. 2010; 30(13):4717–4724. [PubMed: 20357122]

Cave JW, Baker H. Dopamine systems in the forebrain. Adv Exp Med Biol. 2009; 651:15–35. [PubMed: 19731547]

Cazorla P, Smidt MP, O'Malley KL, Burbach JP. A response element for the homeodomain transcription factor Ptx3 in the tyrosine hydroxylase gene promoter. J Neurochem. 2000; 74(5):1829–1837. [PubMed: 10800925]

Chen YC, Priyadarshini M, Panula P. Complementary developmental expression of the two tyrosine hydroxylase transcripts in zebrafish. Histochem Cell Biol. 2009; 132(4):375–381. [PubMed: 19603179]

Dawson SJ, Yoon SO, Chikaraishi DM, Lillycrop KA, Latchman DS. The Oct-2 transcription factor represses tyrosine hydroxylase expression via a heptamer TAATGARAT-like motif in the gene promoter. Nucleic Acids Res. 1994; 22(6):1023–1028. [PubMed: 7908738]

Dickson PW, Briggs GD. Tyrosine hydroxylase: regulation by feedback inhibition and phosphorylation. Adv Pharmacol. 2013; 68:13–21. [PubMed: 24054138]

Filippi A, Mahler J, Schweitzer J, Driever W. Expression of the paralogous tyrosine hydroxylase encoding genes th1 and th2 reveals the full complement of dopaminergic and noradrenergic

neurons in zebrafish larval and juvenile brain. J Comp Neurol. 2010; 518(4):423–438. [PubMed: 20017209]

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004; 32(Web Server issue):W273–W279. [PubMed: 15215394]

Goridis C, Rohrer H. Specification of catecholaminergic and serotonergic neurons. Nat Rev Neurosci. 2002; 3(7):531–541. [PubMed: 12094209]

Hack MA, Saghatelyan A, de Chevigny A, Pfeifer A, Ashery-Padan R, Lledo PM, Götz M. Neuronal fate determinants of adult olfactory bulb neurogenesis. Nat Neurosci. 2005; 8(7):865–872. [PubMed: 15951811]

Hebert MA, Serova LI, Sabban EL. Single and repeated immobilization stress differentially trigger induction and phosphorylation of several transcription factors and mitogen-activated protein kinases in the rat locus coeruleus. J Neurochem. 2005; 95(2):484–498. [PubMed: 16190871]

Hwang DY, Ardayfio P, Kang UJ, Semina EV, Kim KS. Selective loss of dopaminergic neurons in the substantia nigra of Pitx3-deficient aphakia mice. Brain Res Mol Brain Res. 2003; 114(2):123–131. [PubMed: 12829322]

Ichikawa H, Mo Z, Xiang M, Sugimoto T. Brn-3a deficiency increases tyrosine hydroxylase-immunoreactive neurons in the dorsal root ganglion. Brain Res. 2005; 1036(1–2):192–195. [PubMed: 15725417]

Iwawaki T, Kohno K, Kobayashi K. Identification of a potential nurr1 response element that activates the tyrosine hydroxylase gene promoter in cultured cells. Biochem Biophys Res Commun. 2000; 274(3):590–595. [PubMed: 10924322]

Jacobs FM, van Erp S, van der Linden AJ, von Oerthel L, Burbach JP, Smidt MP. Pitx3 potentiates Nurr1 in dopamine neuron terminal differentiation through release of SMRT-mediated repression. Development. 2009; 136(4):531–540. [PubMed: 19144721]

Kessler MA, Yang M, Gollomp KL, Jin H, Iacovitti L. The human tyrosine hydroxylase gene promoter. Brain Res Mol Brain Res. 2003; 112(1–2):8–23. [PubMed: 12670698]

Kim KS, Kim CH, Hwang DY, Seo H, Chung S, Hong SJ, Lim JK, Anderson T, Isacson O. Orphan nuclear receptor Nurr1 directly transactivates the promoter activity of the tyrosine hydroxylase gene in a cell-specific manner. J Neurochem. 2003a; 85(3):622–634. [PubMed: 12694388]

Kim SM, Yang JW, Park MJ, Lee JK, Kim SU, Lee YS, Lee MA. Regulation of human tyrosine hydroxylase gene by neuron-restrictive silencer factor. Biochem Biophys Res Commun. 2006; 346(2):426–435. [PubMed: 16764822]

Kim TE, Park MJ, Choi EJ, Lee HS, Lee SH, Yoon SH, Oh CK, Lee BJ, Kim SU, Lee YS, Lee MA. Cloning and cell type-specific regulation of the human tyrosine hydroxylase gene promoter. Biochem Biophys Res Commun. 2003b; 312(4):1123–1131. [PubMed: 14651989]

Kobayashi K, Kaneda N, Ichinose H, Kishi F, Nakazawa A, Kurosawa Y, Fujita K, Nagatsu T. Structure of the human tyrosine hydroxylase gene: alternative splicing from a single gene accounts for generation of four mRNA types. J Biochem. 1988; 103(6):907–912. [PubMed: 2902075]

Kobayashi K, Morita S, Sawada H, Mizuguchi T, Yamada K, Nagatsu I, Hata T, Watanabe Y, Fujita K, Nagatsu T. Targeted disruption of the tyrosine hydroxylase locus results in severe catecholamine depletion and perinatal lethality in mice. J Biol Chem. 1995; 270(45):27235–27243. [PubMed: 7592982]

Kobayashi K, Noda Y, Matsushita N, Nishii K, Sawada H, Nagatsu T, Nakahara D, Fukabori R, Yasoshima Y, Yamamoto T, Miura M, Kano M, Mamiya T, Miyamoto Y, Nabeshima T. Modest neuropsychological deficits caused by reduced noradrenaline metabolism in mice heterozygous for a mutated tyrosine hydroxylase gene. J Neurosci. 2000; 20(6):2418–2426. [PubMed: 10704516]

Kohwi M, Osumi N, Rubenstein JL, Alvarez-Buylla A. Pax6 is required for making specific subpopulations of granule and periglomerular neurons in the olfactory bulb. J Neurosci. 2005; 25(30):6997–7003. [PubMed: 16049175]

Lenartowski R, Goc A. Epigenetic, transcriptional and post-transcriptional regulation of the tyrosine hydroxylase gene. Int J Dev Neurosci. 2011; 29(8):873–883. [PubMed: 21803145]

Lewis-Tuffin LJ, Quinn PG, Chikaraishi DM. Tyrosine hydroxylase transcription depends primarily on cAMP response element activity, regardless of the type of inducing stimulus. Mol Cell Neurosci. 2004; 25(3):536–547. [PubMed: 15033181]

Liu J, Merlie JP, Todd RD, O'Malley KL. Identification of cell type-specific promoter elements associated with the rat tyrosine hydroxylase gene using transgenic founder analysis. Brain Res Mol Brain Res. 1997; 50(1–2):33–42. [PubMed: 9406915]

Liu N, Cigola E, Tinti C, Jin BK, Conti B, Volpe BT, Baker H. Unique regulation of immediate early gene and tyrosine hydroxylase expression in the odor-deprived mouse olfactory bulb. J Biol Chem. 1999; 274(5):3042–3047. [PubMed: 9915843]

Lo L, Morin X, Brunet JF, Anderson DJ. Specification of neurotransmitter identity by Phox2 proteins in neural crest stem cells. Neuron. 1999; 22(4):693–705. [PubMed: 10230790]

Min N, Joh TH, Kim KS, Peng C, Son JH. 5′ upstream DNA sequence of the rat tyrosine hydroxylase gene directs high-level and tissue-specific expression to catecholaminergic neurons in the central nervous system of transgenic mice. Brain Res Mol Brain Res. 1994; 27(2):281–289. [PubMed: 7898312]

Moll JR, Acharya A, Gal J, Mir AA, Vinson C. Magnesium is required for specific DNA binding of the CREB B-ZIP domain. Nucleic Acids Res. 2002; 30(5):1240–1246. [PubMed: 11861917]

Nagamoto-Combs K, Piech KM, Best JA, Sun B, Tank AW. Tyrosine hydroxylase gene promoter activity is regulated by both cyclic AMP-responsive element and AP1 sites following calcium influx. Evidence for cyclic amp-responsive element binding protein-independent regulation. J Biol Chem. 1997; 272(9):6051–6058. [PubMed: 9038229]

Nakashima A, Ota A, Sabban EL. Interactions between Egr1 and AP1 factors in regulation of tyrosine hydroxylase transcription. Brain Res Mol Brain Res. 2003; 112(1–2):61–69. [PubMed: 12670703]

Nunes I, Tovmasian LT, Silva RM, Burke RE, Goff SP. Pitx3 is required for development of substantia nigra dopaminergic neurons. Proc Natl Acad Sci USA. 2003; 100(7):4245–4250. [PubMed: 12655058]

Papanikolaou NA, Sabban EL. Sp1/Egr1 motif: a new candidate in the regulation of rat tyrosine hydroxylase gene transcription by immobilization stress. J Neurochem. 1999; 73(1):433–436. [PubMed: 10386997]

Papanikolaou NA, Sabban EL. Ability of Egr1 to activate tyrosine hydroxylase transcription in PC12 cells. Cross-talk with AP-1 factors. J Biol Chem. 2000; 275(35):26683–26689. [PubMed: 10842166]

Paskin TR, Iqbal TR, Byrd-Jacobs CA. Olfactory bulb recovery following reversible deafferentation with repeated detergent application in the adult zebrafish. Neuroscience. 2011; 196:276–284. [PubMed: 21933699]

Patankar S, Lazaroff M, Yoon SO, Chikaraishi DM. A novel basal promoter element is required for expression of the rat tyrosine hydroxylase gene. J Neurosci. 1997; 17(11):4076–4086. [PubMed: 9151724]

Patel P, Nankova BB, LaGamma EF. Butyrate, a gut-derived environmental signal, regulates tyrosine hydroxylase gene expression via a novel promoter element. Brain Res Dev Brain Res. 2005; 160(1):53–62.

Qin Y, Hurley LH. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie. 2008; 90(8):1149–1171. [PubMed: 18355457]

Rao F, Zhang L, Wessel J, Zhang K, Wen G, Kennedy BP, Rana BK, Das M, Rodriguez-Flores JL, Smith DW, Cadman PE, Salem RM, Mahata SK, Schork NJ, Taupenot L, Ziegler MG, O'Connor DT. Tyrosine hydroxylase, the rate-limiting enzyme in catecholamine biosynthesis: discovery of common human genetic variants governing transcription, autonomic activity, and blood pressure in vivo. Circulation. 2007; 116(9):993–1006. [PubMed: 17698732]

Sabban EL. Control of tyrosine hydroxylase gene expression in chromaffin and PC12 cells. Semin Cell Dev Biol. 1997; 8(2):101–111. [PubMed: 15001085]

Saucedo-Cardenas O, Quintana-Hau JD, Le WD, Smidt MP, Cox JJ, De Mayo F, Burbach JP, Conneely OM. Nurr1 is essential for the induction of the dopaminergic phenotype and the survival of ventral mesencephalic late dopaminergic precursor neurons. Proc Natl Acad Sci USA. 1998; 95(7):4013–4018. [PubMed: 9520484]

Schimmel JJ, Crews L, Roffler-Tarlov S, Chikaraishi DM. 4.5 kb of the rat tyrosine hydroxylase 5′ flanking sequence directs tissue specific expression during development and contains consensus sites for multiple transcription factors. Brain Res Mol Brain Res. 1999; 74(1–2):1–14. [PubMed: 10640671]

Smeets WJ, González A. Catecholamine systems in the brain of vertebrates: new perspectives through a comparative approach. Brain Res Brain Res Rev. 2000; 33(2–3):308–379. [PubMed: 11011071]

Smidt MP, Smits SM, Bouwmeester H, Hamers FP, van der Linden AJ, Hellemons AJ, Graw J, Burbach JP. Early developmental failure of substantia nigra dopamine neurons in mice lacking the homeodomain gene Pitx3. Development. 2004; 131(5):1145–1155. [PubMed: 14973278]

Son JH, Chun HS, Joh TH, Cho S, Conti B, Lee JW. Neuroprotection and neuronal differentiation studies using substantia nigra dopaminergic cells derived from transgenic mouse embryos. J Neurosci. 1999; 19(1):10–20. [PubMed: 9870933]

Stefano L, Al Sarraj J, Rössler OG, Vinson C, Thiel G. Up-regulation of tyrosine hydroxylase gene transcription by tetradecanoylphorbol acetate is mediated by the transcription factors Ets-like protein-1 (Elk-1) and Egr-1. J Neurochem. 2006; 97(1):92–104. [PubMed: 16515541]

Suzuki T, Yamakuni T, Hagiwara M, Ichinose H. Identification of ATF-2 as a transcriptional regulator for the tyrosine hydroxylase gene. J Biol Chem. 2002; 277(43):40768–40774. [PubMed: 12196528]

Tan X, Zhang L, Zhu H, Qin J, Tian M, Dong C, Li H, Jin G. Brn4 and TH synergistically promote the differentiation of neural stem cells into dopaminergic neurons. Neurosci Lett. 2014; 571:23–28. [PubMed: 24769320]

Trocmé C, Sarkis C, Hermel JM, Duchateau R, Harrison S, Simonneau M, Al-Shawi R, Mallet J. CRE and TRE sequences of the rat tyrosine hydroxylase promoter are required for TH basal expression in adult mice but not in the embryo. Eur J Neurosci. 1998; 10(2):508–521. [PubMed: 9749713]

Venters BJ, Pugh BF. Genomic organization of human transcription initiation complexes. Nature. 2013; 502(7469):53–58. [PubMed: 24048476]

Verbeek MM, Steenbergen-Spanjers GC, Willemsen MA, Hol FA, Smeitink J, Seeger J, Grattan-Smith P, Ryan MM, Hoffmann GF, Donati MA, Blau N, Wevers RA. Mutations in the cyclic adenosine monophosphate response element of the tyrosine hydro-xylase gene. Ann Neurol. 2007; 62(4): 422–426. [PubMed: 17696123]

Willemsen MA, Verbeek MM, Kamsteeg EJ, de Rijk-van Andel JF, Aeby A, Blau N, Burlina A, Donati MA, Geurtz B, Grattan-Smith PJ, Haeussler M, Hoffmann GF, Jung H, de Klerk JB, van der Knaap MS, Kok F, Leuzzi V, de Lonlay P, Megarbane A, Monaghan H, Renier WO, Rondot P, Ryan MM, Seeger J, Smeitink JA, Steenbergen-Spanjers GC, Wassmer E, Weschke B, Wijburg FA, Wilcken B, Zafeiriou DI, Wevers RA. Tyrosine hydroxylase deficiency: a treatable disorder of brain catecholamine biosynthesis. Brain. 2010; 133(Pt 6):1810–1822. [PubMed: 20430833]

Yamamoto K, Ruuskanen JO, Wullimann MF, Vernier P. Two tyrosine hydroxylase genes in vertebrates new dopaminergic territories revealed in the zebrafish brain. Mol Cell Neurosci. 2010; 43(4):394–402. [PubMed: 20123022]

Yamamoto K, Vernier P. The evolution of dopamine systems in chordates. Front Neuroanat. 2011; 5:21. [PubMed: 21483723]

Yamanoue Y, Miya M, Inoue JG, Matsuura K, Nishida M. The mitochondrial genome of spotted green pufferfish Tetraodon nigroviridis (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. Genes Genet Syst. 2006; 81(1):29–39. [PubMed: 16607039]

Yang C, Kim HS, Seo H, Kim KS. Identification and characterization of potential cis-regulatory elements governing transcriptional activation of the rat tyrosine hydroxylase gene. J Neurochem. 1998; 71(4):1358–1368. [PubMed: 9751166]

Yoon SO, Chikaraishi DM. Tissue-specific transcription of the rat tyrosine hydroxylase gene requires synergy between an AP-1 motif and an overlapping E box-containing dyad. Neuron. 1992; 9(1): 55–67. [PubMed: 1352985]

Yukimasa N, Isobe K, Nagai H, Takuwa Y, Nakai T. Successive occupancy by immediate early transcriptional factors of the tyrosine hydroxylase gene TRE and CRE sites in PACAP-stimulated PC12 pheochromocytoma cells. Neuropeptides. 1999; 33(6):475–482. [PubMed: 10657527]

Zetterström RH, Solomin L, Jansson L, Hoffer BJ, Olson L, Perlmann T. Dopamine neuron agenesis in Nurr1-deficient mice. Science. 1997; 276(5310):248–250. [PubMed: 9092472]

Zhou QY, Quaife CJ, Palmiter RD. Targeted disruption of the tyrosine hydroxylase gene reveals that catecholamines are required for mouse fetal development. Nature. 1995; 374(6523):640–643. [PubMed: 7715703]
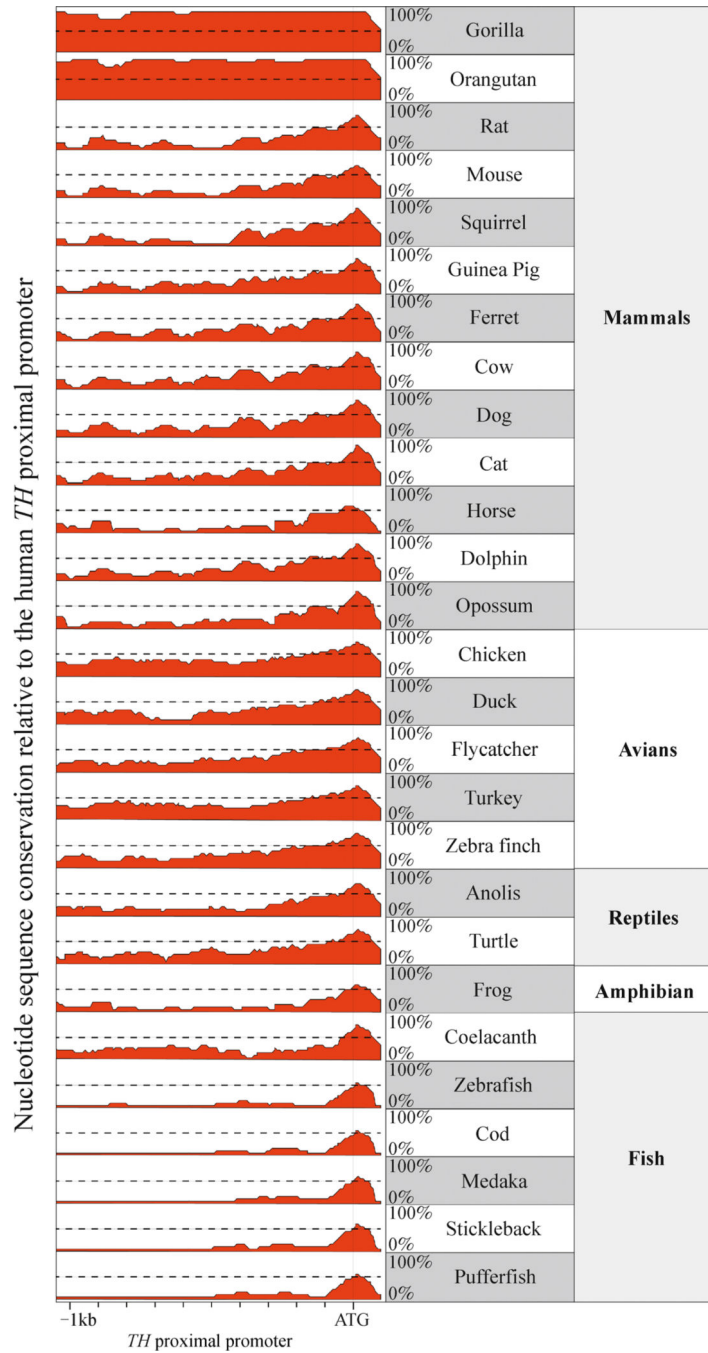
Figure 1.
Conservation of the vertebrate *TH* promoter. The comparative homology of the 1kb *TH* promoter from each species relative to the human sequence reveals that only about 100 bp upstream from the translation start site (indicated by ATG) is highly conserved across vertebrates.

Figure 2.
Conserved nucleotide motifs in the vertebrate *TH* promoter. In all species, a TATA box (blue) and CRE (green) were conserved. Nearly all species also contained a GGTGG box (red). Nucleotide numbering is relative to the translation start site.
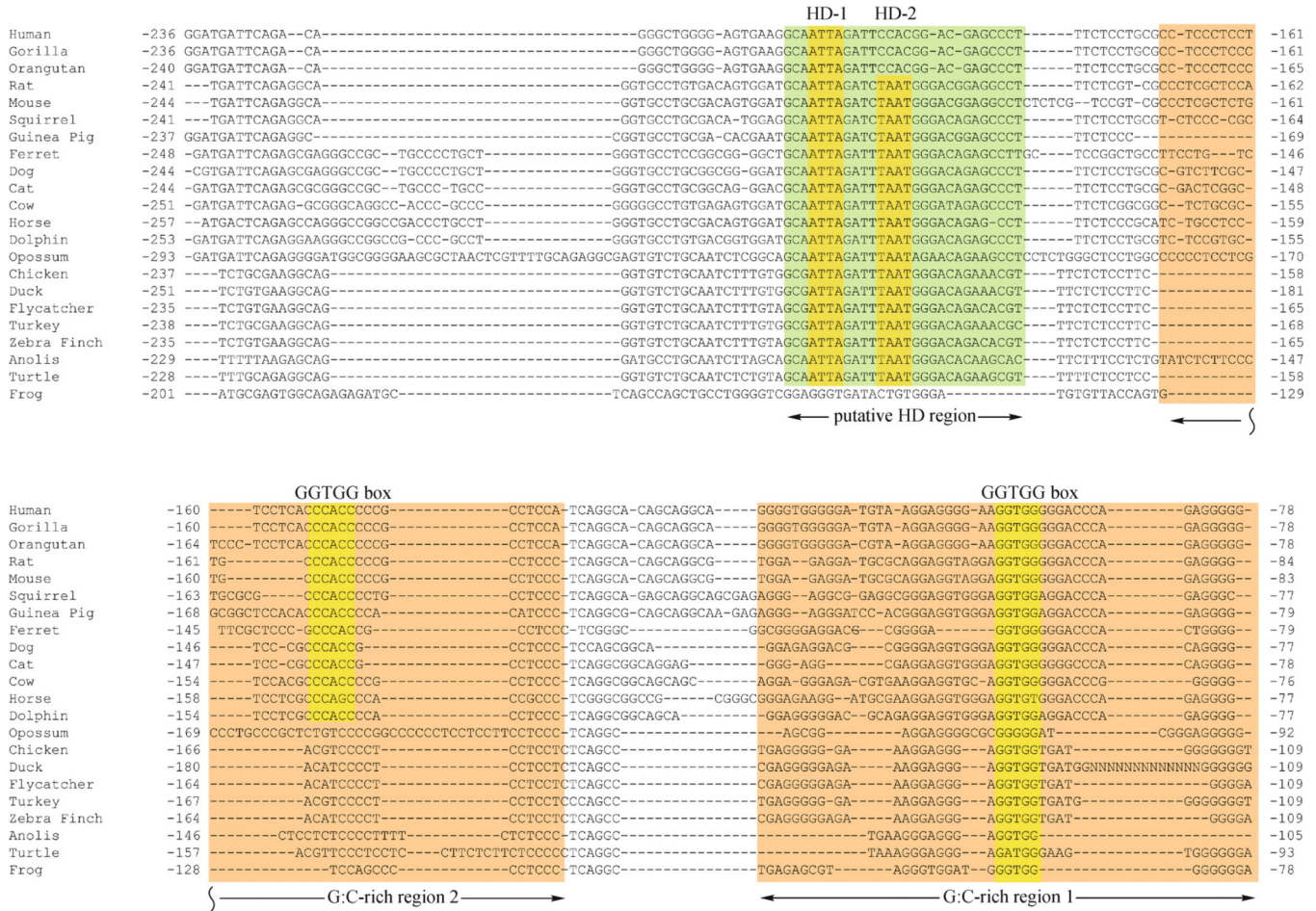
Figure 3.

Nucleotide sequence conservation in the tetrapod *TH* promoter. In addition to the TATA box and CRE (see Fig. 2), tetrapod *TH* promoters share additional regions of homology that include two G:C-rich regions (boxed in orange) and a region containing two canonical core homeodomain (HD) recognition motifs (the region is in green and core motifs are highlighted in yellow). All tetrapods contain a GGTGG box (highlighted in yellow) within the G:C-rich region 1, but only mammals had a GGTGG box in G:C-rich region 2 (highlighted in yellow). Nucleotide numbering is relative to the translation start site.
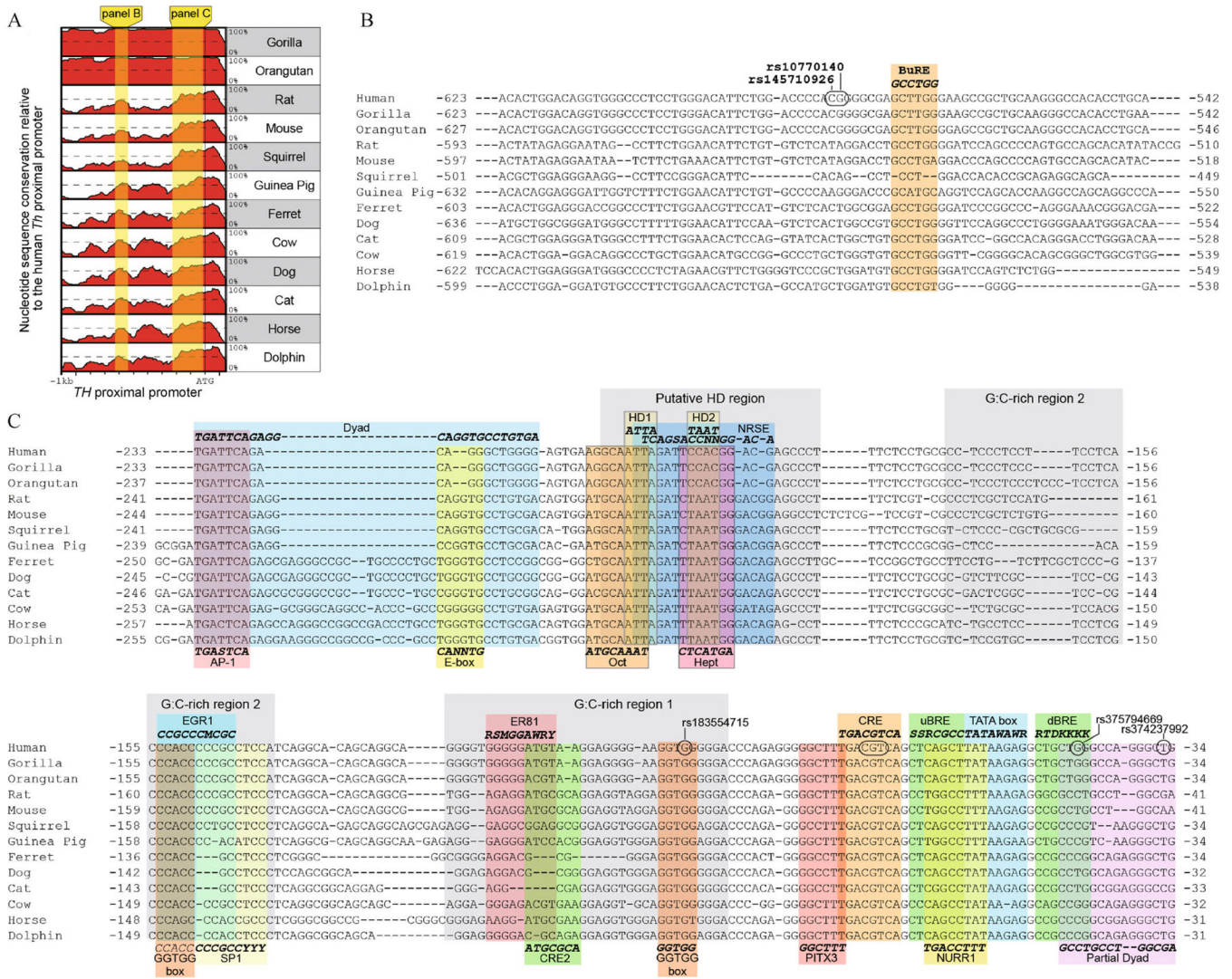
Figure 4.
Nucleotide sequence conservation within the mammalian *TH* promoter. (A) Graphical comparison of the percent homology for mammalian *TH* promoters relative to the human gene. Two regions of high conservation are shaded in yellow. (B and C) Nucleotide conservation within the distally and proximally conserved regions of *TH* promoter, respectively. Putative *cis*-regulatory elements previously reported in either the human or rodent *TH* promoter and their consensus sequence are positioned either above or below the alignments. Known single nucleotide polymorphisms in the human *TH* promoter are circled. Nucleotide numbering is relative to the translation start site.
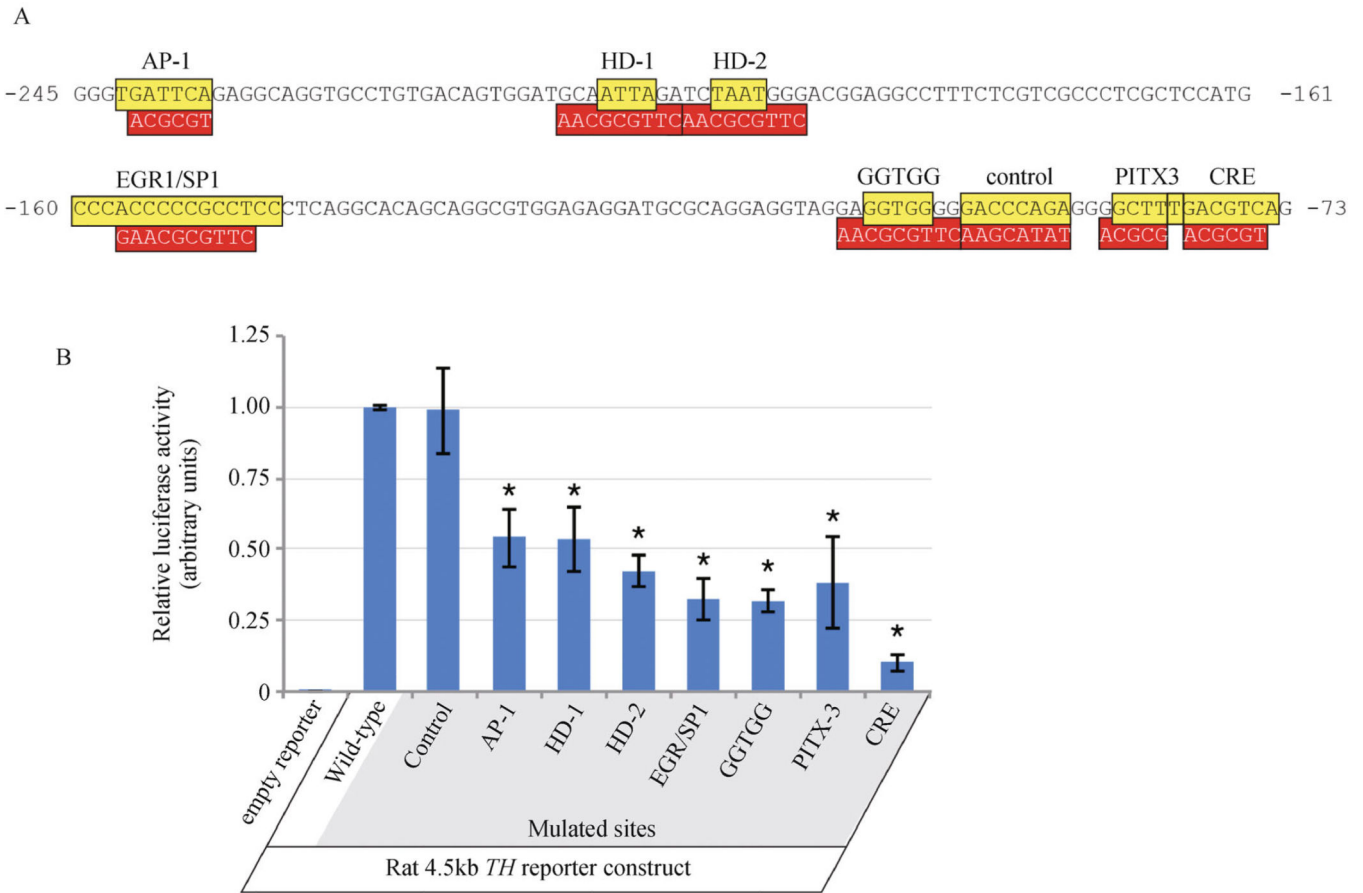
Figure 5.
Transcription reporter assays to test the functionality of highly conserved motifs in the mammalian *TH* promoter. (A) sequences and relative positions of highly conserved motifs (colored yellow) in the rat *TH* proximal promoter. Mutated sequences of the conserved motifs are boxed in red below the wild-type sequence. Nucleotide numbering is relative to the translation start site. (B) relative luciferase activities in transcription reporter assays in a *TH* -expressing olfactory bulb cell line using the rat 4.5kb *TH* promoter containing either the wild-type sequence or conserved motifs individually mutated. Mutation of all highly conserved motifs significantly reduced ($p < 0.05$) promoter activity relative to the wild-type reporter. By contrast, mutation of a region with low conservation ("control") did not significantly affect promoter activity.

A

AP-1   HD-1   HD-2
-248 TGAGGGTGATTCAGAGGCAGGTGCCTGTGACAGTGGATGCAATTAGATCTAATGGGACGGAGGCCTTTCTCGTCGCCCTCGCTCCATG -164

        EGR1/SP1   GGTGG   PITX3 CRE
-163 ATGCCCACCCCCGCCTCCCTCAGGCACAGCAGGCGTGGAGAGGATGCGCAGGAGGTAGGAGGTGGGGGACCCAGAGGGGCTTTGACGTCAGCCTG -69
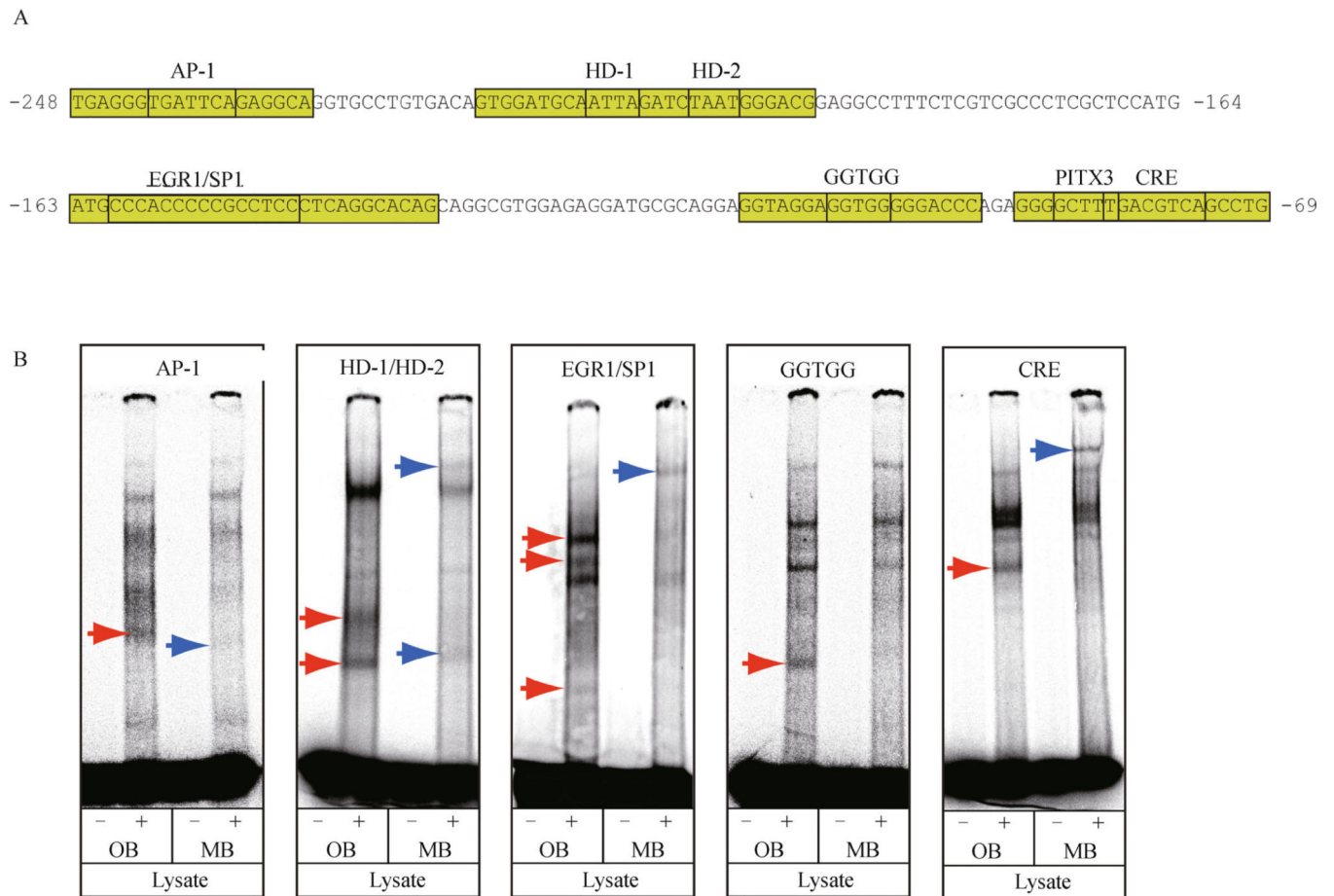
Figure 6.
Formation of brain region-specific and independent complexes on conserved motifs from the *TH* promoter. (A) portions of the rat *TH* proximal promoter used for oligonucleotide target sequences with highly conserved motifs in EMSAs are in green boxes. (B) EMSAs with oligonucleotides derived from the rat *TH* promoter and nuclear lysate from either the adult mouse olfactory bulb (OB) or midbrain (MB). OB and MB-specific complexes are indicated by red and blue arrows, respectively.

A

GGTGG box CRE

CRE-WT 5'-CAGGAGGTAGGA`GGTGG`GGGACCCAGAGGGGCTT`TGACGTCA`GCCTGGCC-3

```
a
  b
    c
```

CRE-mut 5'-CAGGAGGTAGGA`GGTGG`GGGACCCAGAGGGGCTTTGA`AAAA`AGCCTGGCC-3

```
a
  b
    c
```

a   5'-AGGTAGGA-3'
b   5'-AGGAGGTG-3'
c   5'-AGGTGGGG-3'
CRE2 5'-AGGCGTGG-3'

CREB − + − +
CRE-WT CRE-mut

B

AP-1 HD-1 HD-2

-257 GCCTCGGGCTGAGGG`TGATTCAG``AGGCAGGT`GCCTGTGACAGTGGATGCA`ATTA`GATC`TAAT` -196
CRE2-g

EGR-1/SP-1

-195 GGGACGGAGGCCTTTCTCGTCGCCCTCGCTCCATG`CCCAC``CCCCGCCTCCCT`CAGGCACAGC -134
CRE2-f CRE2-e

GGTGG CRE

-133 `AGGCGTGG`AG`AGGATGCG`CAGG`AGGTAGGA``GGTGGGG`GACCCAGAGGGGCTTT`TGACGTCA`GC -72
CRE2 CRE2-d CRE2-a PITX-3
CRE2-b
CRE2-c

C

```
CRE       5'-GGGCTTTGACGTCAGCCTGG-3'
CRE2      5'-CACAGCAGGCGTGGAGAGGA-3'
CRE2-a/b  5'-GCGCAGGAGGTAGGAGGTGG-3'
CRE2-b/c  5'-GGTAGGAGGTGGGGGACCCA-3'
CRE2-d    5'-GTGGAGAGGATGCGCAGGAG-3'
CRE2-e/f* 5'-TGAGGGAGGCGGGGGTGGGC-3'
CRE2-g    5'-AGAGGCAGGTGCCTGTGACA-3'
AP-1      5'-CTGAGGGTGATTCAGAGGCA-3'
HD-1      5'-GTGGATGCAATTAGATCTAA-3
```


CRE   CRE2   CRE2-a/b

CREB 0   0   0

CRE2-b/c   CRE2-d   CRE2-e/f   CRE2-g   AP-1   HD-1

CREB 0   0   0   0   0   0

Figure 7.
Identification of CRE2-like sites within the mammalian *TH* proximal promoter. (A) oligonucleotides containing the GGTGG box (green boxes) and either the wild-type and mutant CRE (yellow and red boxes, respectively). Both wild-type and mutant oligonucleotides (CRE-WT and CRE-mut, respectively) were bound by purified CREB in EMSAs. Within both the wild-type and mutant oligonucleotides, three partially overlapping motifs (labeled a, b, and c) that resemble the previously reported CRE2 site are underlined. The three individual overlapping motifs are compared to the CRE2 sequence below. (B)

inspection of the rat *TH* proximal promoter revealed a total of seven CRE2-like motifs (boxed in yellow and labeled CRE2-a through CRE2-g). Nucleotide numbering is relative to the translation start site. (C) EMSAs showed that purified CREB specifically bound oligonucleotides containing the CRE2-like sites with variable affinities, but always with a lower affinity than the canonical CRE target site. CREB was also able to bind an oligonucleotide with the AP-1 site, but not with the putative HD-1 homeodomain site.

Figure 8.
Conservation of CRE2-like sites in the mammalian *TH* promoter. CRE2-like sites matching the consensus of 5′-AGGNNGGG-3′ and allowing for only one of the three 3′ G positions to vary are boxed in yellow. The CRE2-like sites cluster in the G:C-rich regions, but are also found adjacent to the highly conserved AP-1 site. Nucleotide numbering is relative to the translation start site.

Table 1

Conservation of putative *cis*-regulatory motifs in the *TH* proximal promoter region

Motif	Rodent to primate conservation	Pan-mammalian conservation
Butyrate response element (BuRE)	+/−	+/−
Dyad	−	−
AP-1	+	+
E-box	−	−
Octamer (Oct)	+	+
Heptamer (Hept)	−	+/−
Homeodomain-1 (HD1)	+	+
Homeodomain-2 (HD2)	−	+/−
Neuron restrictive		
Silencing element (NRSE)	−	−
EGR-1 site	+	+/−
SP-1 site	+	+/−
ER81 site	−	−
cAMP response element-2-like (CRE2-like)	+/−	+/−
PITX3 site	+	+
cAMP response element (CRE)	+	+
NURR1	−	−
upstream TFIIB recognition element (uBRE)	+	+/−
TATA box	+	+
downstream TFIIB recognition element (dBRE)	−	−
Partial dyad	+/−	+/−

The degree of motif conservation is indicated by: (+), complete or near complete conservation; (+/−), modest conservation with several substitutions; (−), little or no conservation.