The regulated expression of chimeric tyrosine hydroxylase–insulin transcripts during early development

Catalina Hernández-Sánchez*, Óscar Bártulos, Ana I. Valenciano[1], Alicia Mansilla and Flora de Pablo

Group of Growth Factors in Vertebrate Development, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas (CSIC), Ramiro de Maeztu 9, E-28040 Madrid, Spain and [1]Department of Animal Physiology II, Facultad de C. Biológicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain

ABSTRACT

Biological complexity does not appear to be simply correlated with gene number but rather other mechanisms contribute to the morphological and functional diversity across phyla. Such mechanisms regulate different transcriptional, translational and post-translational processes and include the recently identified transcription induced chimerism (TIC). We have found two novel chimeric transcripts in the chick and quail that result from the fusion of tyrosine hydroxylase (TH) and insulin into a single mature transcript. The *th* and insulin genes are located in tandem and they are generally transcribed independently. However, it appears that two chimeric transcripts containing exons from both the genes can also be produced in a regulated manner. The TH–INS1 and TH–INS2 chimeras differ in their insulin gene content, and they encode two novel isoforms of the TH protein with markedly reduced functionality when compared with the canonical TH. In addition, the TH–INS1 chimeric mRNA generates a small amount of insulin. We propose that TIC is an additional mechanism that can be employed to further regulate TH and insulin expression according to the specific needs of developing vertebrates.

INTRODUCTION

The basis of an organism's functional and behavioral complexity is a question that currently remains unresolved. After completing the sequencing of multiple genomes, it is accepted that this complexity is not merely determined by the number of genes. Irrespective of the size of the proteome, the sophisticated regulation of gene expression is an aspect that clearly contributes to increased functional complexity across the phyla. Thus, the 20 000–25 000 genes estimated currently in humans can give rise to ∼1 million proteins, which in combination with other molecules generate >250 cell types in a human being (1) (http://www.hupo.org). Alternative splicing is a widespread mechanism used to generate protein diversity in metazoans, and it has been estimated that ≈70% of the genes undergo alternative splicing in humans (2). This mechanism has the potential to introduce tremendous variation as witnessed by the ≈38 000 isoforms that can be generated from the *Drosophila* gene Down Syndrome Adhesion Molecule gene (Dscam) (3). Indeed, comparative analyses have shown similar rates of alternative splicing in vertebrates and invertebrates (4). Nevertheless, other processes may also be employed to generate complexity in vertebrates, and multiple transcription start sites (5), 3′ end processing (6), pre-mRNA editing (7,8) and post-translational protein modifications (9) are all important sources of protein diversity.

Recently, novel mechanisms of co- and post-transcriptional regulation have been identified. As such, microRNAs have been shown to influence mRNA stability or translation, thereby regulating protein production and playing important roles in invertebrate and vertebrate development (10–12). Another quite unexpected phenomenon that may add complexity to a genome is the generation of chimeric transcripts from two adjacent, apparently independent genes that share the same orientation. Recently, this phenomenon has been termed transcription induced chimerism (TIC) (13,14), and it is estimated that between 2 and 5% of tandem human genes may be transcribed into chimeric mRNAs (13–15).

The tyrosine hydroxylase (*th*) and insulin genes are well-characterized independent genes that are situated in tandem, a syntenic organization that is conserved across phyla. While studying the presence of alternative insulin mRNA variants during development, we discovered two chimeric TH–insulin transcripts in avian species. We have characterized previously two embryonic mRNA isoforms of insulin

*To whom correspondence should be addressed. Tel/Fax: +34 91 534 9201; Email: chernandez@cib.csic.es

that differ from the pancreatic transcript (16,17) and that have a very marked functional impact at the level of translational regulation. Here we have characterized these two chimeric mRNAs, which are developmentally regulated and tissue-specific. These chimeras encoded two TH protein isoforms whose functionality was somewhat diminished with respect to the canonical TH. Our results underscore an additional aspect in the control of *th*/insulin gene expression involving TIC. In our previous reports, we referred to the insulin gene and mRNA as 'proinsulin' since the primary protein product is proinsulin, which is proteolytically processed to insulin in the pancreas but that remains as proinsulin in extrapancreatic tissues. In the current study, we prefer to use the term 'insulin' since this is how it is annotated in the databases (AY377922; NM_205222).

MATERIALS AND METHODS

Embryos

Fertilized White Leghorn eggs (Granja Rodríguez-Serrano, Salamanca, Spain) and quail eggs (kindly supplied by Dr V. García-Martínez, Universidad de Extremadura, Badajoz) were incubated at 38.4°C and 60–90% relative humidity for the time periods indicated, and the embryos were staged according to Hamburger and Hamilton (18). Mouse embryos (CD1) were removed from the uterus of pregnant females after 8.5 days of development (E) and subsequently dissected from the deciduum. All animals were handled according to European Union Guidelines for animal research.

Cell lines and transfection

HEK293T and NIH3T3 cells were cultured in DMEM supplemented with 10% fetal bovine serum, 2 mM glutamine and antibiotics (all from Invitrogen, Carlsbad, CA), at 37°C and 5% CO_2. Cells were transfected using Lipofectamine Plus (Invitrogen) and analyzed 24 h after transfection.

RNA isolation, RT–PCR and genomic PCR studies

Total RNA from whole embryos, embryonic tissues and HEK293T cells was isolated using Trizol reagent (Invitrogen). The RT reaction was typically performed with 5 μg RNA, the Superscript III Kit and oligo(dT) primer (all from Invitrogen), followed by amplification with the Expand High fidelity (Roche Diagnostics, Mannheim, Germany). The primers used for PCR amplification are listed in Table 1. Quantitative PCR was performed with LUX™ fluorogenic primers and Platinum Quantitative PCR SuperMix-UDG (all from Invitrogen). PCR was carried out in a real-time PCR apparatus ABI Prism 7700 (Applied Biosystems, Foster City, CA). Genomic PCR was performed using the Elongase kit (Invitrogen).

Plasmids

Chicken TH cDNAs were generated by RT of stage 10 total embryonic RNA using the Invitrogen Superscript III kit and an oligo(dT) primer, followed by PCR with the specific primers listed in Table 1. For TH amplification we used primers P1 and P2, and for the TH–INS1 chimera primers P1 and P3. The PCR products were cloned into the pCRII TOPO shuttle

Table 1. Primers used for PCR amplification

Primers	Location
P1: 5′-CACGAGTGAAGATGCCAAC-3′	Exon 1 of TH
P2: 5′-GTTACAGACTAGGAACAGGG-3′	Exon 13 of TH
P3: 5′-GCTAGTTGCAGTAGTTCTCCAGTT-3′	Exon 3 of insulin
P4: 5′-GACATTGGGCTTGCATCTCT-3′	Exon 10 of TH
P5: 5′-ACAGCATCGAGCTCCTGGACA-3′	Exon 13 of TH
P6: 5′-TGCCTGCAAGGAGTACCTAG-3′	Exon 7 of TH
P7: 5′-GTTGCAGTAGTTCTCCAGTT-3′	Exon 3 of insulin
P8 5′-GCAGGGCTCCTGTCTTCCTA-3′	Exon 11 of TH
P9: 5′-TCAGTGATGCCAAGGACAAG-3′	Exons 12 and 13 of TH
P10: 5′-ATCAAAGGGTCCAGCCAC-3′	Exon 13 of TH
P11: 5′-GGCTTCTTCTACACACCCA-3′	Exon 2 of insulin
P12: 5′-CAGTAGTTCTCCAGCTGGTA-3′	Exon 3 of insulin
P13: 5′-CCCTGTTCCTAGTCTGTAAC-3′	Exon 13 of TH
P14: 5′-CTGCACTGGCTCGAAGTTGA-3′	5′-Flanking region of insulin
P15: 5′-CCTTCTAGGAGTTTCTTCCC-3′	5′-Flanking region of insulin

vector. The TH–INS2 chimera cDNA was generated in two steps: (i) PCR with the P6 and P3 primers and cloning of the 843 bp amplified product into the pCRII TOPO shuttle vector; and (ii) excision of the HindIII fragment of the TH cDNA cloned into pCRII TOPO shuttle vector, which was subsequently cloned into the similarly digested TH-INS2 pCRII TOPO construct. The cDNA for each TH isoform was excised from the respective pCRII TOPO construct with EcoRI and cloned into a similarly digested pCI-neo mammalian expression vector (Promega, Madison, WI).

For the V5-fused construct, the TH–INS1 cDNA was amplified with the P1 and P7 primers and the PCR product was then cloned, in-frame, 5′ to the V5 epitope into pcDNA3.1/ V5-His TOPO vector (Invitrogen).

Western blotting

Cells were lyzed in a buffer containing 50 mM Tris–HCl (pH 7.5), 300 mM NaCl, 10 mM EDTA, 1% Triton X-100 and protease inhibitor cocktail (Roche). The homogenates were clarified by centrifugation, and equal amounts of protein were separated on a 10% NuPAGE gel and transferred on to the immobilon-P membranes. For the detection of the TH isoforms, we used a mouse anti-TH epitope antibody (1/1000; Chemicon Temecula, CA) followed by anti-mouse Ig–HRP antibody (1/10 000; Sigma, St Louis, MO). Antibody binding was visualized by chemiluminescence (Pierce, Rockford, IL). For insulin, the mouse anti-V5 epitope antibody (1/5000; Invitrogen) followed by anti-mouse Ig–HRP antibody was used. After stripping, the membranes were analyzed using either mouse anti-β-tubulin antibody (1/10 000; Sigma) or goat anti-actin antibody (1/10 000; Santa Cruz Biotechnology, Palo Alto, CA) followed by anti-mouse or anti-goat (1/10 000; Santa Cruz Biotechnology), as protein loading controls.

Northern blot

Equal amounts of RNA were separated by electrophoresis on formaldehyde–agarose gels and transferred onto nylon membranes. For TH, the membranes were hybridized with a ^{32}P-labeled chicken TH cDNA probe (1071 bp) that includes exons 4–12. After stripping, the membranes were

hybridized with either a ^{32}P-labeled mouse GAPDH (16) or 18S rRNA probes. TH mRNA, GAPDH mRNA and 18S rRNA levels were quantified using a Fuji image analyzer.

Southern blot

E4 chicken genomic DNA was digested with the restriction enzymes indicated and separated by electrophoresis on an agarose gel. The DNA was transferred onto nylon membranes that were hybridized with a ^{32}P-labeled genomic TH probe. After stripping the membrane, it was then hybridized with a ^{32}P-labeled cDNA TH probe (1071 bp) that includes exons 4–12. The genomic TH probe was generated by PCR of E4 chicken genomic DNA using the primers P8 and P2 located in exons 11 and 13 of TH, respectively. The 2891 bp amplification product was cloned into the pCRII TOPO vector.

Pulse labeling and immunoprecipitation

Transfected HEK293T cells were pulse-labeled with 50 μCi ^{35}S-Met/Cys ProMix (Amersham Pharmacia Biotech, Essex, UK) 20 h after transfection in 300 μl of medium per well in an M12 plate. After incubation for 1 h at 37°C, the culture medium was replaced by non-radioactive medium and the cells were lyzed in 50 mM Tris–HCl (pH 7.5), 120 mM NaCl, 0.5% NP-40 and protease inhibitor cocktail at the time periods indicated. The homogenates were clarified by centrifugation and equal amounts of proteins were immunoprecipitated with anti-TH antibody (at 4°C for 16 h), recovered with protein A–Sepharose (Amersham) and resolved in a 10% polyacrylamide NuPAGE gels (Invitrogen). TH levels were quantified using a Fuji image analyzer.

Determination of L-DOPA

Transfected cells were harvested in 0.3 N HClO$_4$ containing 0.4 mM sodium bisulphite and 0.4 mM EDTA. DHBA (100 pmol/ml) was added as an internal standard. Samples were sonicated and centrifuged at 15 000 g for 5 min at 4°C. The L-DOPA content in 20 μl of the supernatant fraction was quantified by high-performance liquid chromatography (HPLC) with colorimetric detection as described by de Pedro *et al.* (19), except for the following modifications: the mobile phase consisted of 10 mM phosphoric acid, 0.1 mM EDTA, 0.4 mM sodium octanesulfonate and 3% acetonitrile (pH 3.1); the potential of the analytical electrode was 200 mV. The HClO$_4$-insoluble proteins were quantified following resuspension in 0.5 M NaOH by using the Lowry method.

RESULTS AND DISCUSSION

### Identification of two novel chimeric transcripts

Through 5′ RACE and RT–PCR, we have isolated two transcripts from the chick embryo that are chimeras of the TH and insulin mRNAs. Both chimeras contain the first 12 of the 13 exons that make up the *th* gene, as well as the 5′ portion of the last exon excluding the stop codon. This fragment of the *th* gene is fused to exons 2 and 3 of the insulin gene in the first chimera but to only exon 3 of the insulin gene in the second chimera (Figure 1A and B). The last exon of the *th* gene contains a consensus sequence for an internal 5′ donor splice site that appears to be used by the TIC pre-mRNA to splice to the 3′ acceptor site in either exon 2 or 3 of the
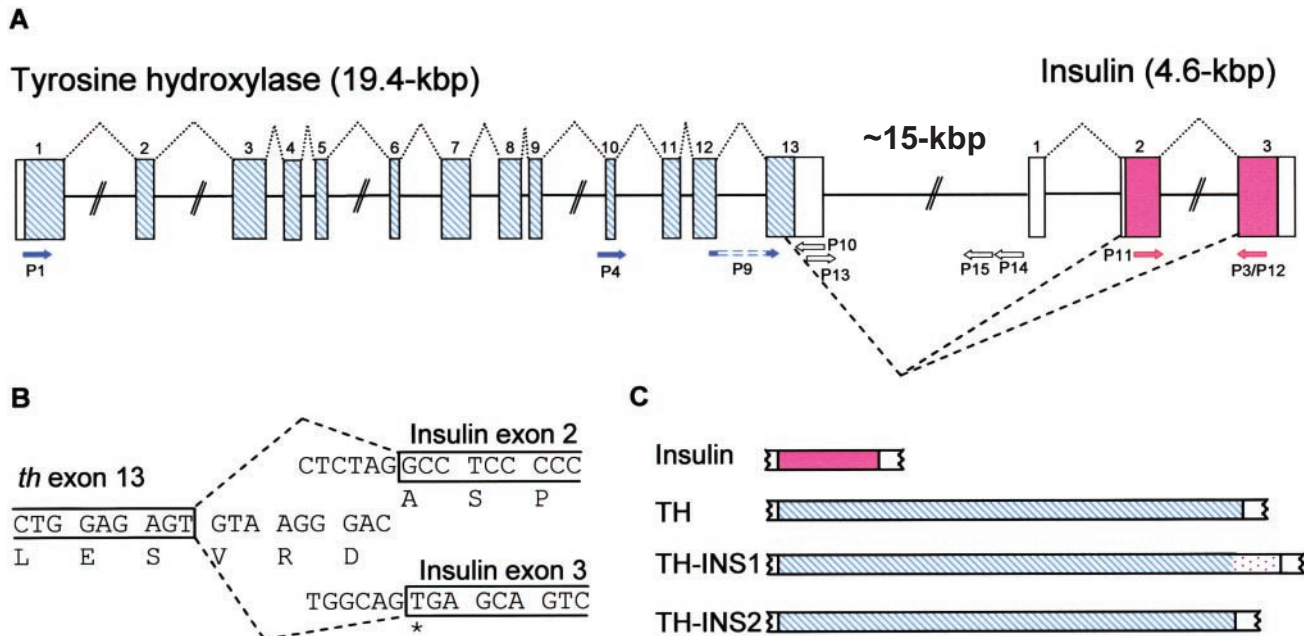
Figure 1. Genomic organization of the chicken *th*/insulin locus on chromosome 5. (**A**) Schematic representation of the *th* and insulin genes. Each box represents one exon and the exons are numbered. Open boxes indicate non-coding regions, the blue hatched boxes represent the *th* coding exons and the pink solid boxes correspond to the insulin coding exons. Dashed lines represent RNA processing. Primers (P) used in PCR are indicated. (**B**) Schematic representation of the out-of-frame splicing between *th* exon 13 and either exon 2 or 3 of the insulin gene. Nucleotide sequences at the end of *th* exon 13 and the beginning of insulin exon 2 or 3 are shown together with the partial amino acid sequence of the TH isoforms; Asterisk indicates a stop codon. (**C**) Diagram of the insulin, TH and TH–INS transcripts. The stretch of 67 new amino acids of the TH–INS1 chimera is represented by the dotted box.

insulin pre-mRNA, thereby generating the TH–INS1 and TH–INS2 chimeras, respectively (Figure 1A and B). This internal splicing of *th* exon 13 avoids the stop codon, extending the open reading frame (ORF) of *th* into the insulin gene. The TH–INS1 fusion gives rise to a putative TH protein with an altered C-terminus. Thus, the TH–INS1 isoform shares regulatory and catalytic domains with the protein encoded by the full-length TH mRNA, but it lacks the last 16 amino acids of the tetramerization domain (20,21). These are replaced by an extended stretch of 67 new amino acids since the insulin ORF overlaps with that of TH in this chimera. The extra 67 amino acids are not coding a part of the insulin protein since the insulin ORF is not kept. As for TH–INS2, a premature stop codon is introduced by the chimeric fragment that generates a truncated TH lacking the last 16 amino acids (see Supplementary Figure 1).

We performed a Southern blot analysis to determine whether the chimeric mRNAs are induced by the transcription of two independent genes or whether they are rather the product of retroposition of a chimeric gene. Chicken genomic DNA hybridized with either a cDNA or genomic TH probe rendered identical restriction enzyme pattern to that predicted in the recently released chicken genome (22) (Figure 2A). In addition, PCR of E4 chicken genomic DNA with the forward primer targeted to the 3′-untranslated region (3′-UTR) of the *th* gene and the reverse primer in the 5′-flanking region of the insulin gene generated a single amplified fragment of
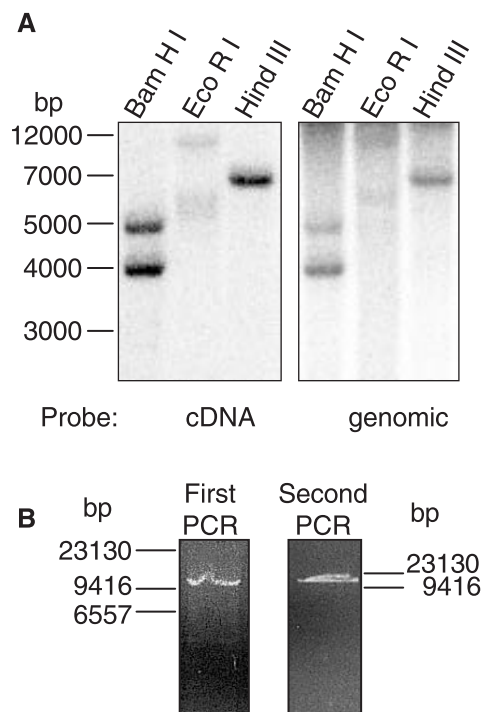

Figure 2. The *th* and insulin genes are unique in the chicken genome. (**A**) Southern blot analysis of genomic DNA digested with the restriction enzymes indicated and hybridized with either a TH cDNA probe or a TH genomic probe. BamHI generates bands of 5000 and 3880 bp; EcoRI generates bands of 12 020 and 5539 bp; HindIII generates a band of 7411 bp. (**B**) PCR of the *th*-insulin intergenic region. The first PCR was performed with the P13 (located in the *th* 3′-UTR) and P14 (located in the insulin 5′-flanking region); for the second PCR, the P13 and P15 primers (located in the insulin 5′-flanking region) were used.

∼14 kb. A semi-nested PCR of the first amplification reaction rendered also a single amplimer (Figure 2B). Sequencing of the PCR product confirmed that this corresponds to the intergenic region. These results confirm the existence of unique TH and insulin gene in the chick genome.

We further analyzed the nature of these chimeric mRNAs by performing RT–PCR on oligo(dT) primed cDNAs. When RT–PCR amplification was carried out with a forward primer located either in the first exon of *th* (P1; see primers in Materials and Methods) and a reverse primer in the last insulin exon (P3), two chimeric mRNAs were amplified (Figure 3A). These transcripts were not unique to the chick since the two chimeras were also found in quail cDNA. Indeed, RT–PCR analysis of RNA from stage 10 chick and quail embryos with the forward primer located in the 10th exon of *th* (P4) and the reverse primer in the last insulin exon (P3) demonstrated that the TH–INS1 and TH–INS2 chimeras were expressed in both avian species (Figure 3B). Sequencing of the amplification products confirmed their identity and the species of origin. Note that this combination of primers (P4 + P3) permitted better separation of the two chimeras than primers P1 and P3. Moreover, we could also detect the TICs in insulin-expressing quail embryonic retina cells (RTC5; data not shown) (16). However, we failed to detect the putative TH–INS chimeras by RT–PCR in RNA from 8.5 day mouse embryos (equivalent to chicken stage 10) using several primer combinations, despite the independent expression of both TH and insulin in these embryos (Figure 3C). Although these results indicate that the TH/insulin chimerism may not occur in mice, we cannot exclude the possibility that the expression of these chimeras in this species or at this stage is below the limits of detection of this technique. Furthermore, genome-wide analysis of the UCSC database (http://genome.ucsc.edu/) failed to identify any expressed sequences, EST or mRNA, that correspond to these TICs in other vertebrate species (Genis Parra and Roderic Gigó, personal communication). This would suggest that either chimera formation is confined to avian species or that the expression of these TICs is highly restricted to a specific tissue and/or developmental stage in other species. Accordingly, TICs identified in the human genome were observed at a single or low-EST copy number (14). Moreover, it is well accepted that EST libraries represent only a fraction of the diverse transcripts and splice variants (23,24).

Phylogenetic analysis of the *th*/insulin synteny

The insulin-like and the aromatic amino acid hydroxylase (AAAH) genes can be traced to the early metazoans (25–27). Moreover, paralogs of each family are located in the genome forming clusters. In humans, the tryptophan hydroxylase and tyrosine hydroxylase genes are situated on chromosome 11 followed by the insulin and insulin-like growth factor (*igf*)-2 genes. Similarly, chromosome 12 contains the phenylalanyl hydroxylase and *igf-1* tandem. The existence of a paralogous region containing an AAAH gene followed by an insulin-like gene is conserved throughout the vertebrate lineage, even in *Amphioxus* that has been proposed as the archetypal vertebrate (26). Hence, we analyzed the TH–insulin synteny in early metazoa. In *Drosophila melanogaster*, the *th* ortholog *ple* is located in tandem with several
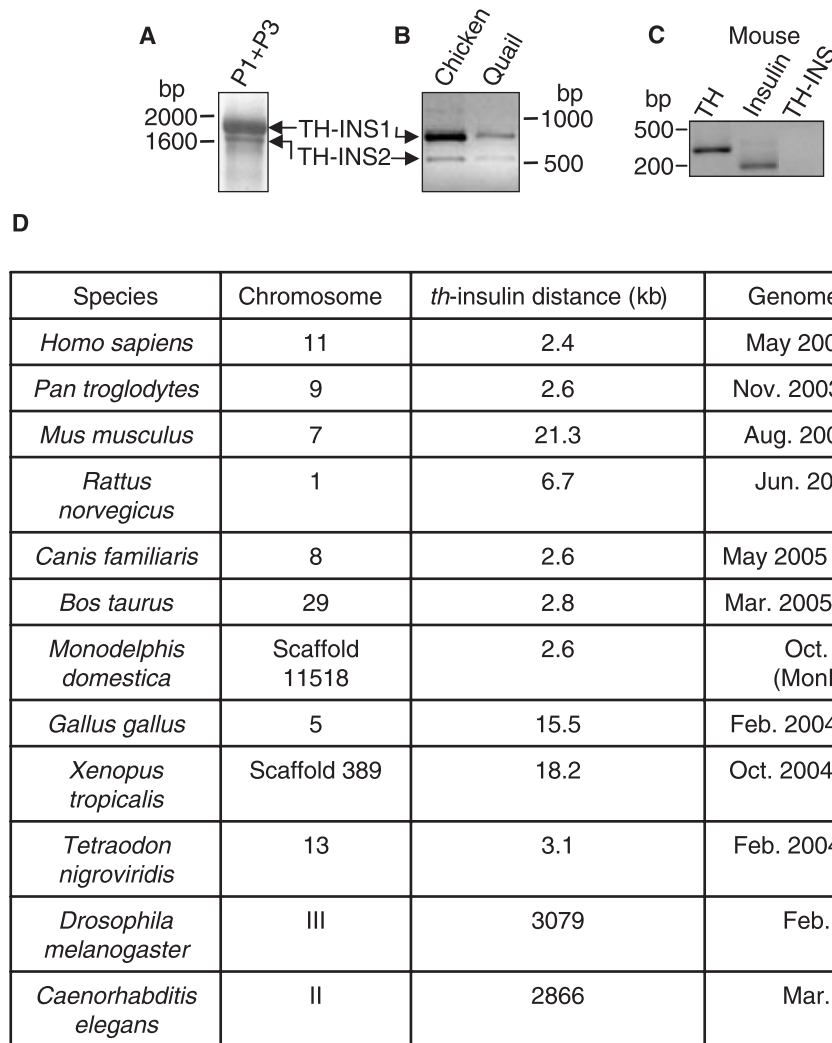
Figure 3. Expression of the TH–INS chimeric transcripts. (**A**) RT–PCR with primers P1 and P3 on RNA from stage 10 chicken embryo. (**B**) RT–PCR with primers P4 and P3 on RNA from stage 10 chicken and quail embryos. (**C**) RT–PCR on RNA from mouse E8.5 embryo. PCR was performed with the following primers: P9 and P10 for TH; P11 and P12 for insulin; and P9 and P12 for TH–INS. (**D**) *th* and insulin synteny across phyla. The distance is represented in kb. The distance for *Drosophila* is between *ple* and *dilp-1* and for *C.elegans* it is between *cat-2* and *ins-19*. The distances between *th* and insulin were extracted from the genome annotations provided by the UCSC genome server (http://genome.ucsc.edu/).

insulin family orthologs (*dilp1-5*) on chromosome 3. Similarly, in *Caenorhabditis elegans* the *th* ortholog *cat-2* is located in tandem with 10 insulin family orthologs (*ins 2–6, 11–15, 19, 20, 31, 32*) on chromosome II (Figure 3D). Thus, it is tempting to speculate that the *th*/insulin synteny is subject to selective pressure owing to the formation of chimeric transcripts.

Two plausible mechanisms might explain how these chimeras are generated. On the one hand, it has been suggested that TIC generation is due to regulated transcriptional read-through of the upstream gene, as frequently occurs in viruses (28–33). Alternatively, *trans*-splicing between the individual pre-mRNAs may result in chimeric transcripts, as it commonly occurs in nematodes and as has been described in a few mammalian situations (34–36). We favor the read-through mechanism, since the *th* and insulin genes are adjacent and in the same orientation. In addition, the intergenic distance between *th* and insulin is relatively short and well-conserved among vertebrates, a fact that could favor the

run-off transcription of TH into the insulin gene. In this respect, tandem genes that produce chimeric transcripts have a tendency to reside closer in the genome than the entire gene pair population (13). However, at present the *trans*-splicing hypothesis cannot be ruled out and more detailed molecular studies will be required to confirm either hypothesis.

The regulation TH–INS1 expression

Using quantitative RT–PCR, we studied the expression of the most abundant chimera, TH–INS1, throughout development. The levels of TH–INS1 mRNA increased between gastrulation and the beginning of neurulation (Figure 4A; stage 4 versus stage 8), maintaining a similar level of expression throughout neurulation (stages 8 and 10; Figure 4B). A similar pattern of expression was also observed for total insulin and TH mRNA. We then compared the relative abundance of the TH–INS1 transcript in stage 8 embryo to that in the
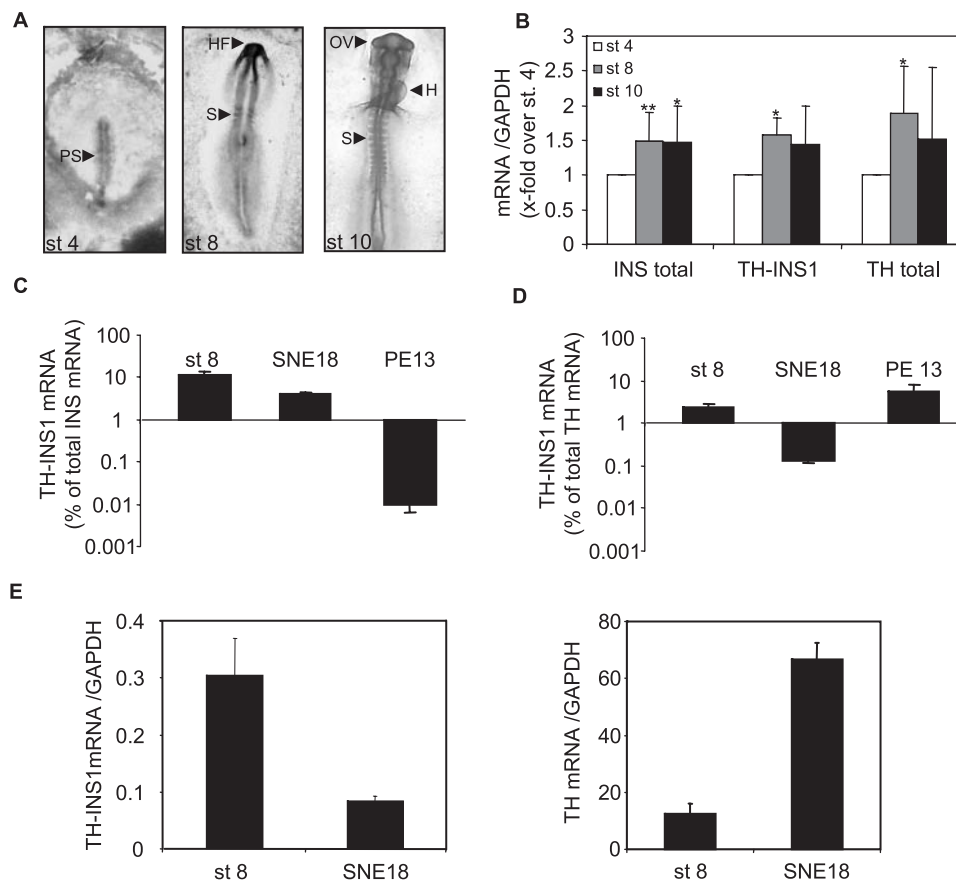
Figure 4. Regulation of chimeric TH–INS1 mRNA expression. (**A**) Photos of chick embryos at different stages (st). Stage 4 corresponds to gastrulation (≈18 h of incubation), stage 8 to early neurulation (≈24 h of incubation) and stage 10 to the end of neurulation (≈36 h of incubation). (**B**) Quantitative RT–PCR of RNA from chick embryos at the developmental stages indicated. The levels of each transcript were normalized to the levels of GAPDH mRNA. The results represent the mean ± SD. *$P < 0.05$; **$P < 0.01$; when compared to stage 4 calculated with the C Dunnett's test comparing all experimental groups. (**C–E**) Quantitative RT–PCR on RNA from stage 8 embryos, from an E13 (PE13) pancreas and from the *substantia nigra* of an E18 embryonic brain (SNE18). The results represent the mean ± SD. H, heart; HF, head fold; OV, optic vesicle; PS, primitive streak; and S, somites.

pancreas where insulin gene transcription is high, and to that in the *substantia nigra* where the *th* gene is preferentially expressed. The TH–INS1 mRNA was expressed in different ratios in the three tissues studied. With respect to total insulin mRNA, the highest relative abundance of the TH–INS1 transcript arose in stage 8 embryos (11.2%) whereas it was lowest in embryonic day (E) 13 pancreas (0.01%; Figure 4C). Conversely, the ratio of the TH–INS1 transcript with respect to total TH mRNA was greatest in E13 pancreas (5.4%) whereas it was the lowest in E18 *substantia nigra* (0.12%; Figure 4D).

These expression profiles suggest that the TH–INS chimeras are not the result of transcriptional leakage, where the transcriptional machinery accidentally ignores the termination of the *th* gene and transcribes through the insulin gene. Such leakage would be an unregulated stochastic event. In contrast, we found that the levels of the TH–INS1 chimera are regulated throughout development and that they are also tissue-specific. In addition, the TH–INS1 mRNA produced by the neurulating embryo (stage 8) was 3.5-fold higher than that generated in the *substantia nigra*, despite the total TH mRNA being 5.3-fold lower in stage 8 embryo than in *substantia nigra* (Figure 4E). These results argue against transcriptional leakage, which would

be reflected by higher TIC formation in the tissue with more active upstream gene transcription.

## Analysis of the protein products from chimeric transcripts

As mentioned, the TH–INS1 chimera appears to encode a C-terminal isoform of the TH enzyme, and TH–INS2 a truncated isoform that lacks the last 16 amino acids of TH (Supplementary Data). To test these predictions, we analyzed HEK293T cells expressing the chimeric mRNAs by western blotting. Immunoblotting with a monoclonal antibody generated against the whole TH protein detected three TH isoforms, which presented slight different electrophoretic mobilities in accordance with their predicted molecular masses (Figure 5A). No TH immunoreactivity was observed in untransfected cells or cells transfected with an unrelated construct (data not shown). Interestingly, the amount of protein recognized following transfection of the TH–INS1 construct was much lower than that observed in cells transfected with either TH or TH–INS2 construct. However in northern blots, the mRNAs for the three TH variants were expressed at comparable levels relative to endogenous

A

B

C
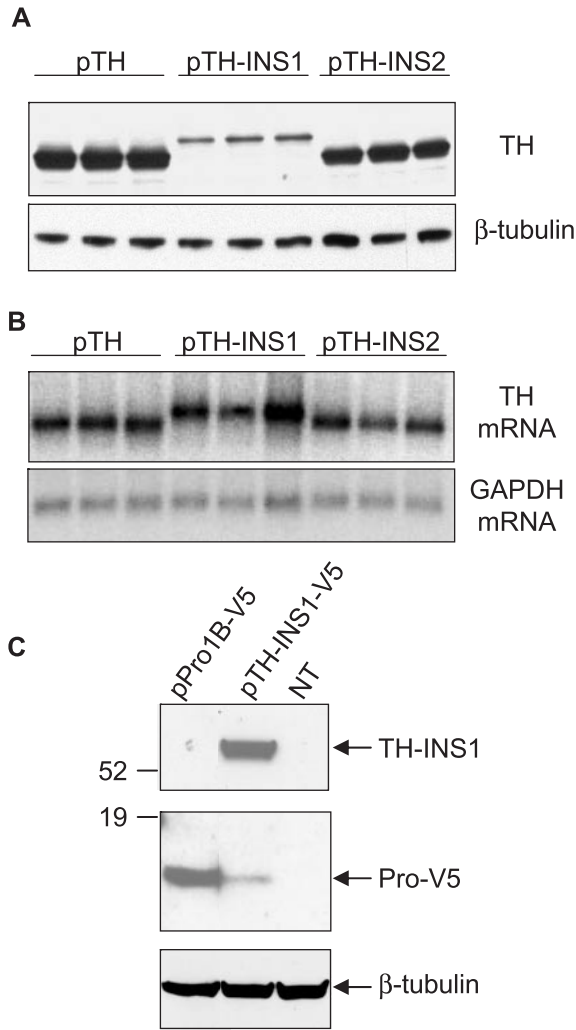
Figure 5. Translational products of the TH transcripts. (**A**) Western blot analysis with an anti-TH antibody of HEK293T cells transfected with the TH, TH–INS1 or TH–INS2 constructs. β-Tubulin was used as a loading control. (**B**) Northern blot of HEK293T transfected cells. Membranes were probed for TH and GAPDH mRNA. TH variant mRNA levels normalized against GAPDH mRNA were 1.95 ± 0.21 for TH, 2.34 ± 0.45 for TH–INS1 and 1.77 ± 0.32 for TH–INS2. (**C**) Co-expression of the TH–INS1 isoform and insulin from the TH–INS1 chimera. NIH3T3 cells transfected with the TH–INS1–V5 construct were analyzed by western blotting for insulin expression using the anti-V5 epitope antibody and for TH–INS1 isoform with anti-TH. Pro1B-V5 is the embryonic insulin Pro1B variant fused to the V5 epitope. β-Tubulin was used as a loading control.

GAPDH mRNA (Figure 5B). Hence, this difference in expression is produced at the post-transcriptional level (see below). Similar results were obtained in NIH3T3 cells (data not shown).

The TH–INS1 chimera also includes the insulin ORF. To assess whether this chimera could behave as a bicistronic transcript, the TH–INS1 cDNA was fused to the V5 epitope tag in phase with the insulin ORF to facilitate the detection of insulin. NIH3T3 cells transfected with the TH–INS1 chimera produced the TH–INS1 protein isoform as well as detectable but low levels of insulin (Figure 5C). The generation of multiple gene products from overlapping reading frames, although common in bacteria and viruses, is rather rare in higher eukaryotes. One of the few examples described

A

B

C

D
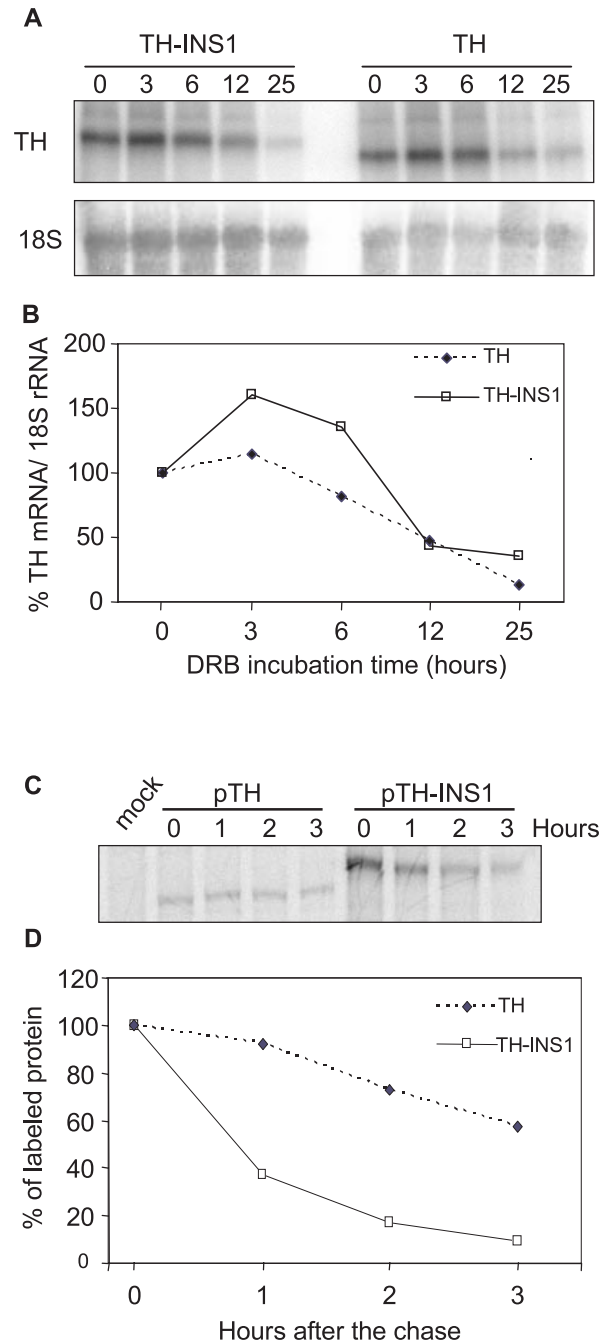
Figure 6. Stability of TH–INS1 mRNA and protein. (**A**) Northern blot of HEK293T cells transfected with the TH or TH–INS1 constructs followed by treatment with the RNA polymerase II inhibitor DRB. Membranes were probed for TH mRNA (upper panel) and 18S rRNA (lower panel). (**B**) Decay curves. TH mRNA and 18S rRNA levels were quantified using a Fuji image analyzer and TH mRNA levels corrected to that of endogenous 18S rRNA levels. (**C**) ^{35}S-Met pulse–chase labeling of HEK293T cells transfected with the constructs indicated. Control cells were transfected with the empty vector. Cell lysates were immunoprecipitated with anti-TH antibody and separated in a polyacrylamide gel. (**D**) Decay curves. TH and TH–INS1 isoforms were quantified using a Fuji image analyzer.

in vertebrates is the INK4a/ARF locus which generates two alternative transcripts encoding proteins with distinct reading frames (37). Here we showed a novel transcript in vertebrates, the TH–INS1 chimera, that contains two overlapping

reading frames encoding different proteins. However, translation of the first cistron imposed strong translation constriction to the second cistron, similarly to the effect of the upAUGs in the pro1A embryonic insulin transcript (16).

To study whether the relative decrease in TH–INS1 protein was mediated by a difference in mRNA stability, we studied mRNA degradation. The half-life of the TH and TH–INS1 chimera transcripts were identical (12 h; Figure 6A and B), indicating that the extension of the TH mRNA produced by the fusion of the insulin mRNA did not influence transcript stability. These results prompted us to check the translational activity and protein stability of the chimera. Pulse–chase experiments showed that although the translational activity of TH–INS1 chimera was higher (compare the time 0 h for both TH and TH–INS1 in Figure 6C), the half-life of the TH–INS1 protein was one-third that of the TH protein (45 min versus 150 min, Figure 6C and D). These differences in protein stability could well account for the marked decrease in steady-state protein levels of TH–INS1 with respect to TH and TH–INS2. Considering that the only difference between the TH–INS1 and TH–INS2 proteins is the extension of 67 amino acids in its C-terminal extension, it is plausible that these 67 additional amino acids confer instability to the TH–INS1 protein and therefore they act as a 'degron'.

Activity of the novel TH isoforms

TH is the rate-limiting enzyme in the synthesis of cathecholamines, catalyzing the conversion of tyrosine into L-DOPA. The holoenzyme is a homotetramer and this assembly is required to generate a fully functional enzyme (20,38). The TH–INS1 and TH–INS2 proteins only differ from the canonical TH in the tetramerization domain, both of them lacking the C-terminal leucine-zipper motive involved in the assembly of TH tetramers. Therefore, we analyzed how these modifications might affect the enzymatic activity of the protein. We measured L-DOPA accumulation in HEK293T cells expressing the TH–INS chimeras and compared it to that found in cells expressing the canonical TH. In order to generate equivalent levels of TH and TH–INS1 protein, the cells were transfected with different amounts of plasmid. The levels of L-DOPA in cells expressing TH–INS1 are ∼5-fold lower than those found in cells expressing similar or lower amounts of TH protein (Figure 7A). Similarly, cells transefected with the TH–INS2 construct produced ∼6-fold less L-DOPA than cells expressing the canonical TH isoform (Figure 7B). Taken together, these results indicate that the TH–INS1 chimera encodes an unstable TH isoform that presents considerably lower enzymatic activity than the canonical TH. Although the TH–INS2 generates similar levels of a truncated TH isoform, the enzymatic activity is also greatly diminished. It could be hypothesized that expression of the two chimeric isoforms might represent a mechanism to downregulate TH function. Hence, expression of TH–INS1 would result in a severe downregulation, decreasing both the level of protein and its activity, whereas the TH–INS2 would generate an intermediate situation affecting only enzymatic activity. It is also noteworthy that the TH–INS1 chimera also generates insulin, although at a much lower level than the embryonic Pro1B insulin transcript
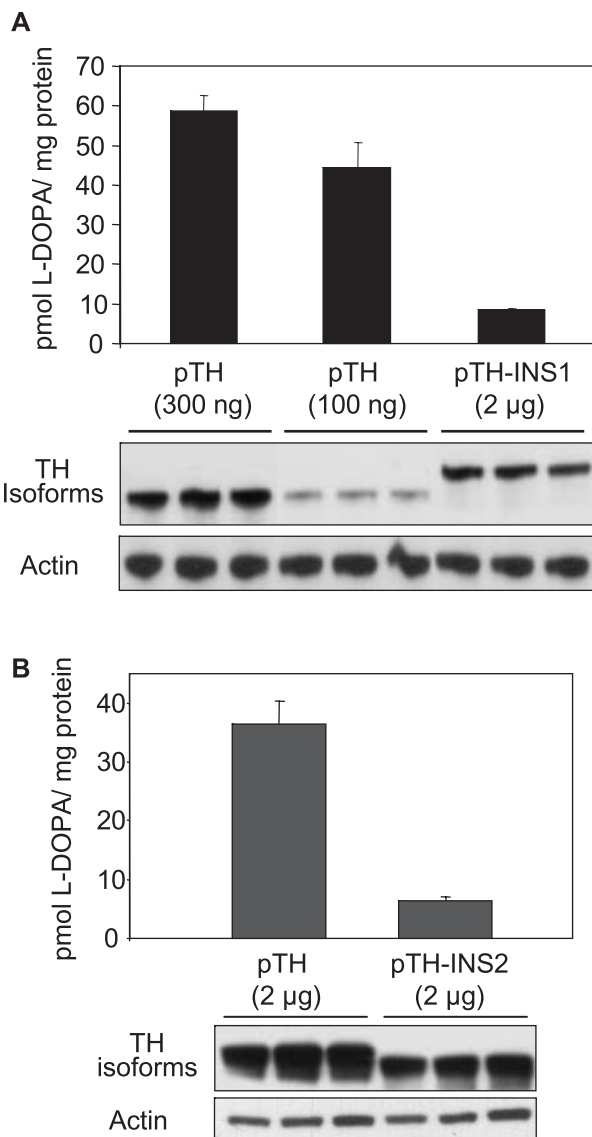
Figure 7. Enzymatic activity of TH isoforms. Extracts of transfected HEK293T cells were analyzed by HPLC for L-DOPA production. (**A**) HEK293T cells were transfected with different amounts of TH and TH–INS1 expressing constructs to obtain similar levels of the TH protein isoforms. The results represent the mean ± SD of three replicate experiments. A representative western blot is shown probed with both an anti-TH antibody and an antibody against actin as a loading control (lower panel). (**B**) HEK293T cells were transfected with the same amount of TH and TH–INS2 constructs. The results represent the mean ± SD of three replicate experiments. A representative western blot is shown that was probed with an anti-TH antibody and for actin as a loading control (lower panel).

(16). In this context, it represents the third alternative insulin transcript found in early embryos (16,17).

CONCLUSIONS

In this study, we demonstrate the formation of chimeras between the syntenic *th* and insulin genes. These genes have well-characterized independent expression patterns and physiological functions in postnatal organisms. TH expressed in neurons and chromaffin cells of the adrenal medulla is critical for the synthesis of essential neurotransmitters, the

catecholamines. Conversely, insulin produced and secreted into the blood by the pancreas is a vital anabolic hormone responsible for maintaining glucose homeostasis. However, in recent decades new biological functions together with novel regulatory mechanisms have been identified for proinsulin/insulin [reviewed in (39)]. During embryonic development and before pancreas formation, insulin acts as a survival factor and it is involved in early morphogenesis. Similarly, it has been suggested that cathecholamines might be involved in morphogenesis during development, preceding neuronal differentiation (40). These novel physiological roles appear to be paralleled by specific ways to regulate gene expression (16,17). Here, we have identified two TH–INS chimeras which, to our knowledge, are the first examples of TIC in avian species. Our results suggest that these TICs may represent an additional mechanism of regulating the post-transcriptional gene expression of two adjacent genes. Although, further studies are required to determine the physiological implications of the formation of TH–INS chimeras, we envisage that they fulfill a relevant function based on the changes in expression levels of both genes. Concomitant with the reduction of full-length TH and insulin transcripts, TIC formation by read-through would restrict the initiation of transcription from the insulin promoter owing to transcriptional interference (41). Finally, studies have begun to reveal that variation in the regulatory regions of particular genes might be associated with species differences (42). In this context, we speculate that changes in the regulatory elements that permit the formation of TH–INS chimeras in avian species may have contributed to their evolution.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

REFERENCES

1. Humphery-Smith,I. (2004) A human proteome project with a beginning and an end. *Proteomics*, **4**, 2519–2521.
2. Modrek,B. and Lee,C.J. (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature Genet.*, **34**, 177–180.
3. Schmucker,D., Clemens,J.C., Shu,H., Worby,C.A., Xiao,J., Muda,M., Dixon,J.E. and Zipursky,S.L. (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, **101**, 671–684.
4. Brett,D., Pospisil,H., Valcarcel,J., Reich,J. and Bork,P. (2002) Alternative splicing and genome complexity. *Nature Genet.*, **30**, 29–30.
5. Hashimoto,S., Suzuki,Y., Kasai,Y., Morohoshi,K., Yamada,T., Sese,J., Morishita,S., Sugano,S. and Matsushima,K. (2004) 5′-End SAGE for the analysis of transcriptional start sites. *Nat. Biotechnol.*, **22**, 1146–1149.
6. Beaudoing,E. and Gautheret,D. (2001) Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data. *Genome Res.*, **11**, 1520–1526.
7. Athanasiadis,A., Rich,A. and Maas,S. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.*, **2**, e391.
8. Blow,M., Futreal,P.A., Wooster,R. and Stratton,M.R. (2004) A survey of RNA editing in human brain. *Genome Res.*, **14**, 2379–2387.
9. Banks,R.E., Dunn,M.J., Hochstrasser,D.F., Sanchez,J.C., Blackstock,W., Pappin,D.J. and Selby,P.J. (2000) Proteomics: new perspectives, new biomedical opportunities. *Lancet*, **356**, 1749–1756.
10. Meister,G. and Tuschl,T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature*, **431**, 343–349.
11. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
12. Giraldez,A.J., Cinalli,R.M., Glasner,M.E., Enright,A.J., Thomson,J.M., Baskerville,S., Hammond,S.M., Bartel,D.P. and Schier,A.F. (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science*, **308**, 833–838.
13. Akiva,P., Toporik,A., Edelheit,S., Peretz,Y., Diber,A., Shemesh,R., Novik,A. and Sorek,R. (2006) Transcription-mediated gene fusion in the human genome. *Genome Res.*, **16**, 30–36.
14. Parra,G., Reymond,A., Dabbouseh,N., Dermitzakis,E.T., Castelo,R., Thomson,T.M., Antonarakis,S.E. and Guigo,R. (2006) Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res.*, **16**, 37–44.
15. Kim,N., Kim,P., Nam,S., Shin,S. and Lee,S. (2006) ChimerDB—a knowledgebase for fusion sequences. *Nucleic Acid Res.*, **34**, D21–D24.
16. Hernandez-Sanchez,C., Mansilla,A., de la Rosa,E.J., Pollerberg,G.E., Martinez-Salas,E. and de Pablo,F. (2003) Upstream AUGs in embryonic proinsulin mRNA control its low translation level. *EMBO J.*, **22**, 5582–5592.
17. Mansilla,A., Lopez-Sanchez,C., de la Rosa,E.J., Garcia-Martinez,V., Martinez-Salas,E., de Pablo,F. and Hernandez-Sanchez,C. (2005) Developmental regulation of a proinsulin messenger RNA generated by intron retention. *EMBO Rep.*, **6**, 1182–1187.
18. Hamburger,V. and Hamilton,H.L. (1951) A series of normal stages in the development of the chick embryo. *J. Morphol.*, **88**, 49–92.
19. de Pedro,N., Alonso-Gomez,A.L., Gancedo,B., Valenciano,A.I., Delgado,M.J. and Alonso-Bedate,M. (1997) Effect of alpha-helical-CRF[9-41] on feeding in goldfish: involvement of cortisol and catecholamines. *Behav. Neurosci.*, **111**, 398–403.
20. Vrana,K.E., Walker,S.J., Rucker,P. and Liu,X. (1994) A carboxyl terminal leucine zipper is required for tyrosine hydroxylase tetramer formation. *J. Neurochem.*, **63**, 2014–2020.
21. Goodwill,K.E., Sabatier,C., Marks,C., Raag,R., Fitzpatrick,P.F. and Stevens,R.C. (1997) Crystal structure of tyrosine hydroxylase at 2.3 A and its implications for inherited neurodegenerative diseases. *Nature Struct. Biol.*, **4**, 578–585.
22. Hillier,L.W., Miller,W., Birney,E., Warren,W., Hardison,R.C., Ponting,C.P., Bork,P., Burt,D.W., Groenen,M.A., Delany,M.E. *et al.* (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**, 695–716.
23. Kapranov,P., Cawley,S.E., Drenkow,J., Bekiranov,S., Strausberg,R.L., Fodor,S.P. and Gingeras,T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, **296**, 916–919.
24. Johnson,J.E., Stromvik,M.V., Silverstein,K.A., Crow,J.A., Shoop,E. and Retzel,E.F. (2003) TableView: portable genomic data visualization. *Bioinformatics*, **19**, 1292–1293.
25. Steele,R.E., Lieu,P., Mai,N.H., Shenk,M.A. and Sarras,M.P. (1996) Response to insulin and the expression pattern of a gene encoding an insulin receptor homologue suggest a role for an insulin-like molecule in regulatin growth and patterning in Hydra. *Dev. Gene Evol.*, **206**, 247–259.

26. Patton,S.J., Luke,G.N. and Holland,P.W. (1998) Complex history of a chromosomal paralogy region: insights from amphioxus aromatic amino acid hydroxylase genes and insulin-related genes. *Mol. Biol. Evol.*, **15**, 1373–1380.

27. Chan,S.J. and Steiner,D.F. (2000) Insulin through the ages: phylogeny of a growth promoting and metabolic regulatory hormone. *Am. Zool.*, **40**, 213–222.

28. Magrangeas,F., Pitiot,G., Dubois,S., Bragado-Nilsson,E., Cherel,M., Jobert,S., Lebeau,B., Boisteau,O., Lethe,B., Mallet,J. *et al.* (1998) Cotranscription and intergenic splicing of human galactose-1-phosphate uridylyltransferase and interleukin-11 receptor alpha-chain genes generate a fusion mRNA in normal cells. Implication for the production of multidomain proteins during evolution. *J. Biol. Chem.*, **273**, 16005–16010.

29. Thomson,T.M., Lozano,J.J., Loukili,N., Carrio,R., Serras,F., Cormand,B., Valeri,M., Diaz,V.M., Abril,J., Burset,M. *et al.* (2000) Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene. *Genome Res.*, **10**, 1743–1756.

30. Pradet-Balade,B., Medema,J.P., Lopez-Fraga,M., Lozano,J.C., Kolfschoten,G.M., Picard,A., Martinez,A., Garcia-Sanz,J.A. and Hahne,M. (2002) An endogenous hybrid mRNA encodes TWE-PRIL, a functional cell surface TWEAK-APRIL fusion protein. *EMBO J.*, **21**, 5711–5720.

31. Poulin,F., Brueschke,A. and Sonenberg,N. (2003) Gene fusion and overlapping reading frames in the mammalian genes for 4E-BP3 and MASK. *J. Biol. Chem.*, **278**, 52290–52297.

32. Roginski,R.S., Mohan Raj,B.K., Birditt,B. and Rowen,L. (2004) The human GRINL1A gene defines a complex transcription unit, an unusual form of gene organization in eukaryotes. *Genomics*, **84**, 265–276.

33. Hardy,R.W. and Wertz,G.W. (1998) The product of the respiratory syncytial virus M2 gene ORF1 enhances readthrough of intergenic junctions during viral transcription. *J. Virol.*, **72**, 520–526.

34. Blumenthal,T. (1998) Gene clusters and polycistronic transcription in eukaryotes. *Bioessays*, **20**, 480–487.

35. Caudevilla,C., Serra,D., Miliar,A., Codony,C., Asins,G., Bach,M. and Hegardt,F.G. (1998) Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc. Natl Acad. Sci. USA*, **95**, 12185–12190.

36. Takahara,T., Tasic,B., Maniatis,T., Akanuma,H. and Yanagisawa,S. (2005) Delay in synthesis of the 3′ splice site promotes trans-splicing of the preceding 5′ splice site. *Mol. Cell*, **18**, 245–251.

37. Quelle,D.E., Zindy,F., Ashmun,R.A. and Sherr,C.J. (1995) Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell*, **83**, 993–1000.

38. Lohse,D.L. and Fitzpatrick,P.F. (1993) Identification of the intersubunit binding region in rat tyrosine hydroxylase. *Biochem. Biophys. Res. Commun.*, **197**, 1543–1548.

39. Hernández-Sánchez,C., Mansilla,A., de la Rosa,E.J. and de Pablo,F. (2006) Proinsulin in development: new roles for an ancient prohormone. *Diabetologia*, **49**, 1142–1150.

40. Pendleton,R.G., Rasheed,A., Roychowdhury,R. and Hillman,R. (1998) A new role for catecholamines: ontogenesis. *Trends Pharmacol. Sci.*, **19**, 248–251.

41. Proudfoot,NJ. (1986) Transcriptional interference and termination between duplicated alpha-globin gene constructs suggests a novel mechanism for gene regulation. *Nature*, **322**, 562–565.

42. Wittkopp,P.J., Vaccaro,K. and Carroll,S.B. (2002) Evolution of yellow gene regulation and pigmentation in Drosophila. *Curr. Biol.*, **12**, 1547–1556.