SYMPOSIUM

Alternative Splicing in Development and Function of Chordate Endocrine Systems: A Focus on Pax Genes

Linda Z. Holland* and Stephen Short†

*Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093-0202, USA; †Institute of Marine Sciences, School of Biological Sciences, Portsmouth University, Portsmouth, PO4 9LY, UK


Synopsis

Genome sequencing has facilitated an understanding of gene networks but has also shown that they are only a small part of the answer to the question of how genes translate into a functional organism. Much of the answer lies in epigenetics—heritable traits not directly encoded by the genome. One such phenomenon is alternative splicing, which affects over 75% of protein coding genes and greatly amplifies the number of proteins. Although it was postulated that alternative splicing and gene duplication are inversely proportional and, therefore, have similar effects on the size of the proteome, for ancient duplications such as occurred in the Pax family of transcription factors, that is not necessarily so. The importance of alternative splicing in development and physiology is only just coming to light. However, several techniques for studying isoform functions both in vitro and in vivo have been recently developed. As examples of what is known and what is yet to be discovered, this review focuses on the evolution and roles of the Pax family of transcription factors in development and on alternative splicing of endocrine genes and the factors that regulate them.

Introduction

In the past 10 years, as the complete genome sequences of a wide range of organisms have become available, a major focus of comparative genomics has been on evolutionary changes in cis-regulation as a driving force in animal evolution (Carroll et al. 2004; Davidson and Erwin 2010). The best-studied gene network is that mediating patterning of the sea urchin endoderm (Peter and Davidson 2010). The best-studied gene network is that mediating patterning of the sea urchin endoderm (Peter and Davidson 2010). Comparisons with the network patterning the endoderm in starfish have identified both conserved kernels of gene networks together with many diverged connections (McCauley et al. 2010). Recently, comparative genomics has received a major boost from “next generation” sequencing, which provides much shorter reads (18–400 or more bases) than the conventional Sanger sequencing, but provides high throughput at a comparatively low cost. These shorter reads can be assembled into a complete genome sequence by matching to the genome of a related organism (reference genome) or used to look for conserved noncoding DNA or particular motifs such as miRNAs or transposable elements. Data generated by these methods alone and in combination with various types of gene chips/microarrays together with high-throughput sequencing (Solexa/Illumina, Solid) are giving insights into direct targets of individual transcription factors (ChIP on chip, ChIP-seq) and alternative splice forms expressed at particular times in development or under different physiological conditions (exon-microarrays, RNA-seq).

These new technologies are helping to show that changes in cis-regulation, such as have occurred repeatedly in the Pitx gene of stickleback fish, causing reduction of the pelvic girdle (Chan et al. 2010) are just one of many driving forces in evolution. Protein-coding sequences can also change, and these changes can contribute to evolution of development, reproductive isolation and speciation [reviewed by Lynch and Wagner (2008) and...
Palumbi (2008). For example, mutation of C-terminal serine/threonine phosphorylation sites in the homeodomain protein Ubx has been implicated in evolution of a domain that represses formation of abdominal appendages in six-legged arthropods (Ronshaugen et al. 2002), while evolution of placental development in mammals correlates with positive selection on the HoxA-11 protein, which is a direct regulator of prolactin (Lynch et al. 2008). Moreover, positive selection acting on gamete-interaction proteins has played a role in speciation of sea urchins (Landry et al. 2003; Moy et al. 2008). However, in addition to these genetic changes, a wide range of epigenetic changes has been implicated as driving evolutionary change. There are several definitions of what constitutes epigenetics. The most common one is heritable traits not directly encoded by the genome. However, this definition is imperfect. Some epigenetic traits such as DNA methylation are often, but not always, inherited, while the proteins that mediate some epigenetic phenomena such as alternative-splicing are encoded by the genome. In addition, genomic elements such as transposons can be subject to epigenetic modifications such as methylation, with the degree of methylation varying with the location of the transposons in the genome and among individuals (Reiss et al. 2010). In addition to methylation, epigenetic mechanisms include a variety of DNA modifications and posttranscriptional and posttranslational modifications. The present review concerns the evolution and the roles in development and physiology of just one of these mechanisms, alternative-splicing, with particular attention to the Pax family of transcription factors and the role of alternative splicing in endocrine systems.

**Alternative splicing of pre-mRNAs is the rule for eukaryotes**

The organization of genes into exons and introns, which are spliced out of pre-mRNAs to create the mature mRNA, is ubiquitous in eukaryotes (Koonin 2009). Although some pre-mRNAs of multi-exon genes are constitutively-spliced—that is, each intron is spliced out and all the exons are incorporated into the mature mRNA—it has been estimated that >90% of human pre-mRNAs are alternatively spliced (Blencowe et al. 2009). The more exons, the greater is the number of potential splice forms. Thus, while the number of protein-coding genes in the Bilateria ranges from about 14,000 in *Drosophila melanogaster* to 21,900 in the cephalochordate *Branchiostoma floridae* to 25,000–30,000 in humans (Claverie 2001; Putnam et al. 2008), the proteome is typically much larger. Splicing is mediated by a suite of five small nuclear ribonucleoproteins (snRNPs) plus several proteins that recognize the sequences at splice junctions [reviewed by Ritchie et al. (2009)]. The canonical splice junction consists of AG/GTRAGT and YYTTYYYYYNCA/G at the donor (5′) and acceptor (3′) splice sites, respectively, where R is A or G, Y is C or T, and N is any base. The canonical GT–AG donor and acceptor sequences occur over 98% of the time (Burset et al. 2001). However, there are exceptions such as GC–AG and AT–AC. Alternative splicing can occur at both the canonical and noncanonical sites (Rovescalli et al. 2000). Some exons may have internal donor or acceptor splice-sites that can mediate alternative splicing. A database of alternative splice events and the isoform splice patterns of genes from human and other species is available at http://www.ebi.ac.uk/asd/. The determinants of which alternative splice-sites are used in a given cell types are very complex. They include splice-site recognition by splicing factors, splicing silencers, inhibitors, and activators. The specific sequences that bind these factors and their distance from the splice junction are important parameters for regulation of alternative splicing. Up-to-date reviews on the splicing mechanisms and techniques for identifying alternatively spliced exons are from Chen and Manley (2009), Ritchie et al. (2009), and Licatalosi and Darnell (2010).

**The evolution of chordate Pax genes**

Pax genes code for transcription factors with two DNA-binding domains (the paired domain and the homeodomain), a co-factor-interacting domain (the octapeptide) and C-terminal activation and repression domains. Due to gene duplications in pre-bilaterians, there were five Pax genes (*Pax1/9, Pax2/5/8, Pax3/7, Pax4/6, and Pax neuro*) at the base of the chordates. Pax1/9 genes have lost the homeodomain, and Pax2/5/8 genes have only a partial homeodomain. *Pax neuro* is present in amphioxus, but has been lost from tunicates and vertebrates. Early in the vertebrate lineage, the genome underwent two rounds of whole-genome duplication with subsequent loss of many gene duplicates (Putnam et al. 2008; van de Peer et al. 2009). Thus, the basal chordate amphioxus has about 21,999 genes, while the human only has 25,000–30,000. However, duplicates of transcription factors and genes in signaling pathways were preferentially retained (Putnam et al. 2008). Consequently, there are nine Pax genes in mammals and, due to an additional genome duplication at the base of the
teleosts (Meyer and Van de Peer 2005; Van de Peer et al. 2009) there are 12 Pax genes in teleost fish.

Pax genes are interesting from the point of view of chordate embryogenesis because they have key roles in the development of such organs as the eye, muscles, central nervous system (CNS), kidney, pituitary, the pharynx and its derivatives, and the gut and related organs. Moreover, not only are Pax genes involved in development of several endocrine organs, they also help regulate endocrine genes. Developmental expression of Pax genes is highly conserved in amphioxus and vertebrates. In both chordates, Pax1/9 genes are expressed in the pharyngeal endoderm and developing muscles, and Pax2/5/8 in the pharyngeal arches/forming gill slits in aquatic vertebrates and amphioxus, the endostyle and/or its homolog, the thyroid, the developing kidney, the hindbrain and spinal cord, and pigment cells of the eye in vertebrates and anterior photoreceptor in amphioxus (Heller and Brandli 1997, 1999; Kozmik et al. 1999) (Figs. 1 and 2). Pax3/7 is expressed in the edges of the neural plate and in developing muscle in both amphioxus and vertebrates (Goulding et al. 1991; Holland et al. 1999), while Pax4/6 is expressed in the olfactory epithelium, anterior photoreceptor (amphioxus) or eye (vertebrates), the lamellar body (amphioxus) and its homolog the pineal in vertebrates, and in the pituitary (vertebrates) or its homolog, Hatscheck’s diverticulum (amphioxus) (Glardon et al. 1998; Rath et al. 2009). Pax3/7 is also expressed in the pharyngeal endoderm in amphioxus, but apparently not in vertebrates (Holland et al. 1999).

Additional expression domains for Pax genes arose in vertebrates subsequent to gene duplication and the evolution of new structures. In mammals, Pax1 and

![Image of Pax gene expression](https://academic.oup.com/icb/article-abstract/50/1/22/734457/24)

**Fig. 1** Expression of Pax2/5/8 genes shows both evolutionary conservation and subfunctionalization and neofunctionalization subsequent to gene duplication in vertebrates. (A) In amphioxus, the single Pax2/5/8 gene is broadly expressed in the hindbrain and spinal cord (hb + sc), pigment cells of the frontal eye (fe), the pronephric kidney (pk), endostyle (e), and around the forming gill slits [branchial arches (ba)] as well as in Hatscheck’s anterior left diverticulum, which will fuse with the ciliated pit to become Hatscheck’s pit, homologous to the vertebrate pituitary. (B–D) All but the last domain is conserved in one or more of the three Xenopus Pax2/5/8 genes. Pax2/5/8 genes have also acquired new domains in the vertebrate midbrain–hindbrain boundary (mhb) and in the otic placode and vesicle (ov). Pax2 (B) has retained all of the presumed ancestral vertebrate domains, while Pax5 (C) has lost all the domains except for that at the midbrain–hindbrain boundary and otic vesicle, and Pax8 (D) has lost all the domains except for those in the otic vesicle, hindbrain and spinal cord and pronephric kidney. Thy, thyroid (homolog of the endostyle); os, optic stalk; va, visceral arches; homologous to the branchial arches of amphioxus. Not shown is conserved expression of amphioxus Pax2/5/8 and Xenopus Pax2 in the forming anus. Amphioxus data from Kozmik et al. (2001). Xenopus data from Heller and Brandli (1997, 1999).
In addition to expressing Ttf1, lamprey embryo expresses Pax9 (Wallin et al. 2002). The lamprey endostyle ultimately metamorphoses into a thyroid gland (Marine, 1913). The larval lamprey endostyle as shown by immunostaining. T4 and Pax2.1 are also expressed in the early zebrafish thyroid (Wendl et al. 2005). Tff1 is also expressed in the forebrain. (E and F) By stage Tahara 29, thyroxine (T4 thyroid hormone) is secreted by the larval lamprey endostyle as shown by immunostaining. T4 and Pax2.1 are also expressed in the early zebrafish thyroid (Wendl et al. 2002). The lamprey endostyle ultimately metamorphoses into a thyroid gland (Marine, 1913).

Pax9 are expressed in the primordia of the thymus and parathyroid, which are derivatives, in part, of the pharyngeal endoderm (Liu et al. 2007). Pax1 is essential for development of the thymus (Wallin et al. 1996). In addition, Pax9 is expressed in mammalian thyroid cell lines, although it does not mediate transcription from a 2-kb calcitonin (CT) promoter (Suzuki et al. 2007) (see Table 1 for a list of endocrine organs and their secretions). In teleost fish, Pax1 and Pax9 are also expressed in the ultimobranchial gland, a pharyngeal derivative (Suzuki et al. 2005, 2007). Pax2 is also expressed in pancreatic islets, where it transcriptionally activates glucagon (Ritz-Laser et al. 2000). Both Pax2 and Pax5 are expressed at the midbrain–hindbrain boundary (Schwarz et al. 1997), where they are involved in regulation of Engrailed 2 (Li Song and Joyner 2000), and Pax2, Pax5, Pax8, and Pax3 are expressed in the otic and other neurogenic placodes (Heller and Brandli 1999; Schlosser 2006; Nechiporuk et al. 2007; McCabe and Bronner-Fraser 2008). Both Pax4 and Pax6 are expressed in the pancreas and intestine in vertebrates (Ritz-Laser et al. 2002). Pax-6 is required for development of α cells, which secrete glucagon, whereas Pax-4 is required for insulin secretion and development of β and δ cells, which produce somatostatin (Ritz-Laser et al. 2002). Pax4 protein, which acts as a transcriptional repressor, binds to a regulatory element in the insulin promoter, and suppresses transcription (Campbell et al. 1999). Pax4 and Pax6, which is a transcriptional activator, also bind to the glucagon promoter (Ritz-Laser et al. 2002), and it may be that they fine-tune transcription by competing with one another.

**Alternative splicing of chordate Pax genes**

Pax genes are ideal for studies of alternative splicing because they exhibit varying degrees of splicing, roughly proportional to the number of exons, and because they have several functional domains, each of which can be altered by alternative-splicing. Using exon-to-exon PCR with first-strand cDNAs from several embryonic stages and adults of amphioxus, we identified two splicing events in Pax1/9, which has 5 exons, 12 splicing events for Pax2/5/8, which has 11 exons, five events for Pax3/7, which has 6 exons, and 18 isoforms of Pax4/6, which has 13 exons (Glardon et al. 1998; Short and Holland 2008) (Fig. 3). These alternative splice forms include exon-skipping, use of alternative 3′- and 5′-splice sites and intron-inclusion.

Based on bioinformatic analyses of mammalian EST sequences, it was proposed that alternative splicing and gene duplication are inversely proportional with singleton genes having more splice forms than duplicate ones (Su et al. 2006). To test this idea, we compared alternative splicing of the Pax genes in amphioxus with published data on vertebrates. We found there was at least as much alternative splicing in each of the vertebrate Pax genes as in their unduplicated amphioxus homologs (Short and Holland 2008). Moreover, although we found several evolutionarily conserved splice forms, including one of Pax2/5/8 conserved with vertebrate Pax5 that would be predicted to result in a truncated isoform lacking the activation and repression domains, many isoforms were not conserved in amphioxus and vertebrates.

The disparity between our results and results from EST sequences may be due in part to the inability of EST analyses to detect alternative splicing involving the C-terminal half of the coding sequence, since 3′-ESTs typically only cover the 3′-UTR, and in
Comparisons of EST data from mouse and human, which diverged ~96 million years ago (mya) (Nei et al. 2001), showed that the amount of alternative splicing is larger in small gene families and larger in duplicated genes than in single ones (Jin et al. 2008). However, while Su et al. (2006) agreed that for human genes, smaller gene families have more duplicate genes, they argued that shortly after gene duplication, splice forms tended to be lost while Jin et al. (2008) concluded that newly duplicated genes have fewer constraints and, thus, are free to undergo more alternative-splicing. Evidently, the relationship between alternative splicing and gene duplication depends on many variables such as the age of duplicates, the size of the gene families and the organisms being compared. However, our study comparing amphioxus and vertebrates, whose lineages split over 500 mya, concerned the whole-genome duplications giving rise to multiple Pax genes in the vertebrate lineage that probably occurred ~400–500 mya (Schubert et al. 2006). It may be that such old gene duplicates have had time to evolve myriad splice forms.

### Alternatively spliced isoforms are differentially expressed in development

Alternatively spliced isoforms of developmental genes can be differentially expressed at various stages in embryogenesis and in different tissues. However, relatively little is known about the functions of specific isoforms in vivo. Perhaps the best examples of roles of isoforms in development are in the central and peripheral nervous systems. For example, in the vertebrate central nervous system, two isoforms of Fgf8...
(Fgf8a and Fgf8b) are expressed at the midbrain/hindbrain boundary, together with Fgf17 and Fgf18, which derived from duplication of an ancestral chordate Fgf8/17/18 gene. Of these four proteins, which have different affinities for Fgf receptors, only Fgf8b can induce Gbx2 expression in the hindbrain (Liu et al. 2003; Olsen et al. 2006).

An extreme example of the roles of alternative splice forms in neuronal development is the Down syndrome cell adhesion molecule1 (Dscam1) in the immunoglobulin superfamily. Experiments have shown that in Drosophila, thousands of isoforms of Dscam1 must be expressed during development for neurons to have normal branching patterns (Hattori et al. 2009). In vertebrates, a nervous-system-specific splicing factor (nSR100) was found that regulates splicing of brain-specific alternative exons of many genes (Calarco et al. 2009). Mutation of nSR100 inhibits differentiation of nerve cells in both cell culture and in the zebrafish brain. The human mid-fetal brain expresses ~76% of the total genes, of which 28% showed region-specific expression of specific splice forms (Johnson et al. 2009). Interestingly, these alternatively spliced genes did not include DSCAM or the related DSCAML1.

Differential expression of isoforms of a number of genes in the CNS and peripheral nervous systems of other mammals has also been described. These include transcription factors (e.g., Prrx1, which regulates processing of nociceptive information), cell-adhesion structural proteins (protocaderin-α genes involved in projections of serotonergic neurons), and signaling molecules (e.g., GSK3β involved in Wnt/β-catenin signaling) (Katori et al. 2009; Rebello et al. 2009; Wood-Kaczmar et al. 2009).

For Pax genes, the two isoforms of amphioxus Pax1/9, which use alternative 5’-splice sites in the C-terminal exon, one of which causes a frame shift and premature stop codon, are differentially expressed in development. The isoform with the premature stop codon is dominant at all developmental stages, while the other is expressed at a low level at

Fig. 3 Alternative splicing of the single amphioxus Pax2/5/8 pre-mRNA is extensive, but no more so than that of the separate vertebrate Pax2, Pax5, and Pax8 mRNAs. (A) While several splice forms are evolutionarily conserved, others are not. Double vertical lines indicate alternative stop codons. In the Xenopus genes, exon 4 is split into two exons, while Xenopus Pax5 has evolved exon 6.1, and Xenopus Pax8 has evolved a new exon 2. Thus, exon 6 in amphioxus Pax2/5/8 is comparable to exon 7 in Xenopus Pax2 and Pax5 and to exon 8 in Xenopus Pax8. Alternative splicing of exons 7 and 8 in amphioxus Pax2/5/8 and exons 8 and 9 in Xenopus Pax2 results in intron inclusion and a premature stop codon in the intron. Vertical stripes: paired box; open box: octapeptide; and diagonal stripes: partial homeodomain. Data from Kozmik et al. (1993), Heller and Brandli (1997), and Short and Holland (2008). (B) Exon 8 in amphioxus [Branchiostoma florideae (B. f)] is longer than the comparable exon in Xenopus laevis (X. l) and Danio rerio (D. r), but the core sequence is highly conserved, suggesting that the function of this exon is also conserved.
the early larval stage only, when expression in the muscle is at its highest (Short and Holland 2008). Similarly, Pax2 isoforms are temporally expressed differentially during development in *Xenopus* (Heller and Brandli 1997). In addition, two isoforms of Pax2 are expressed in the vertebrate pancreas. One of these, Pax2B, activates transcription at a higher level than the other, Pax2A. Both of these bind to the G3 regulatory element in the glucagon gene and, with lower affinity to the G1 element, which has a high affinity for Pax6 (Ritz-Laser et al. 2000). A third Pax2 isoform (Pax2D2) is also expressed in cell-lines that produce insulin; however, Pax2 does not regulate transcription from the insulin gene. In chordates, Pax4/6 genes are expressed in the developing pineal gland in vertebrates (Estivill-Torrués et al. 2001) and its homolog, the lamellar body in amphioxus (Glardon et al. 1998). Pax6 is required for development of the pituitary (Estivill-Torrués et al. 2001). Amphioxus Pax4/6 has at least six isoforms, while for vertebrate Pax4 and Pax6 genes 5 and 13 splicing events, respectively, have been identified (Bandah et al. 2007; Short and Holland 2008). Only two of four isoforms of Pax4 identified in the mouse are expressed in the developing pituitary (Rath et al. 2009), indicative of the roles that differential expression of isoforms play in development.

**Unique constellations of isoforms of Pax genes can be expressed in cancers and genetic disorders**

Considerable information on alternative splice forms has come from cancers and genetic defects. Mutants of all vertebrate Pax genes, except for Pax4, have been associated both with cancers and birth defects (reviewed by Wang et al. 2008). While many of these involve point mutations resulting in truncated proteins, several involve splice forms. For example, in pancreatic cancer cell lines, an alternatively-spliced isoform of Pax6, which has a 14-amino acid insertion in the paired domain, is expressed at higher levels than is the isoform lacking the insertion (Mascarenhas et al. 2009), while in a multiple-myeloma cell line, a Pax5 isoform with a variant transactivation domain is expressed at lower levels than in normal B-lymphocytes (Borson et al. 2006). For Pax5, many isoforms have been found expressed in B-lymphocytes, which are produced in the bone marrow and migrate to the thymus (Arseneau et al. 2009). Most of these isoforms are also expressed in lymphomas. Two of them code for isoforms lacking the DNA-binding domains, octapeptide and activation and repression domains. Two have the paired domain and octapeptide, but not the homeodomain and the others are C-terminal variants.

In vertebrates, congenital defects caused by mis-splicing are known for Pax6 and Pax3. Pax6 produces at many isoforms (41 in the pigeon), several of which are expressed in specific regions of the retina (Bandah et al. 2007). Three of these isoforms, one of which lacks a paired domain, are produced by use of an alternative promoter and ATG start codons. A second contains a 14 amino-acid insertion at the N-terminal end of the paired domain. Overexpression of the second isoform in the mammalian eye results in extreme microphthalmia (Kim and Lauderdale, 2008). Moreover, in humans, a mutation at a splice junction that removes exon 4 resulting in an altered paired-domain with 13 extra amino acids congenital nystagmus (involuntary eye movements) and various defects of the iris (Vincent et al. 2004). Similarly, in the mouse Sey (small eye) mutant, a point mutation results in intron retention and a premature stop codon C-terminal of the homeodomain of Pax4 (Stuart et al. 1993). The original Splotch mouse, which has defective neural tube closure, has an A to T transversion at the 3′-AG splice acceptor of intron 3 of Pax3, which prevents normal splicing of that intron, thereby creating four mis-spliced mRNAs. Due to use of cryptic 3′-splice sites in the downstream exon, two of these have deletions that cause an altered reading frame. The other two result in retention of intron 3 and deletion of exon 4, respectively (Epstein et al. 1993). Mutants of other Pax genes with defective splicing have not yet been described.

**Alternative splicing and endocrine systems**

Alternative splicing affects pre-mRNAs of hormones, hormone receptors and proteins that regulate their expression. Although many endocrine hormones are not subject to alternative-splicing, several are. An alternative splice form of thyroid stimulating hormone β-subunit occurs in the human pituitary and in leukocytes in the peripheral blood and thyroid, whereas expression of the major isoform is limited to the pituitary (Schaefer and Klein 2009). Moreover, several peptide hormones can be derived from a single pre-mRNA by alternative splicing. An example is the *Calca* gene, which encodes two peptides, CT and α-CT gene-related peptide (α-CGRP) by alternative splicing of pre-mRNA. This gene has six exons. Inclusion of exon 4, which results in a premature stop codon and a polyA signal at the 3′-end of the exon, creates CT, while skipping of exon 4 and
inclusion of exons 5 and 6 generates α-CGRP (Chew 1997). Splicing is tissue-specific, with CT, which regulates bone resorption, being produced mainly by the thyroid (although there is some expression in motor neurons), while α-CGRP, which is a neuropeptide produced by the cells in both the central and peripheral nervous systems, regulates vascular properties (Terrado et al. 1998; Huebner et al. 2008). Not only does Calca mRNA undergo alternative splicing, but so does the CT receptor, which has at least seven different isoforms with three alternate promoters, which show tissue-specific expression (Anusaksathien et al. 2001). All seven are expressed in osteoclasts, but not at equal levels, while only some of the isoforms are expressed in kidney and brain.

Another example of alternative-splicing of primary transcripts for encoding peptide hormones is the proglucagon gene. The proglucagon mRNA is cleaved into several peptide hormones. The 29-amino-acid glucagon is secreted from pancreatic islet A-cells and helps regulate metabolism. Mammalian glucagon-like peptide 1 (GLP-1) is secreted from L-cells in the intestine and potentiates insulin secretion from pancreatic islet B-cells. Glucagon-like peptide 2 (GLP-2) is known only from mammals, where it can promote growth of intestinal cells (reviewed by Irwin 2001). There is no evidence for alternative splicing of proglucagon mRNA in lampreys, although in teleost fish, there is apparently alternative splicing as the processed mRNA in the pancreas codes only for GLP-1, not GLP-2, whereas the reverse is true of mRNAs in the intestine (Irwin and Wong 1995). The situation in the chicken and frog is similar (Yeung and Chow 2001).

The enzymes that cleave endocrine peptides from their precursor proteins are also alternatively spliced. The major gene family involved in cleaving these peptides includes the subtilisin-like proprotein convertases of which there are seven members in mammals—SPC1-7 (reviewed by Steiner 1998). The primary neurohormone processors are SPC3 (PC1/PC3) and SPC2 (PC2). Moreover, SPC6 (PC5/PC6) is expressed in the brain, the adrenal gland and the intestine and is alternatively spliced, as is SP1 (PACE4), which is expressed in the testis (Rouillé et al. 1995; Steiner 1998). These enzymes typically cleave the hormone precursors at lysine/arginine or arginine/arginine boundaries. The family radiated early in metazoan evolution. Blast searches indicate that these proprotein convertases are present in chordates other than vertebrates, but only a few have been studied. The basal chordate amphioxus (Branchiostoma) has homologs of PC2 and PC1/PC3 as well as PC6 (Oliva et al. 1995, 2000). PC6 is alternatively spliced to yield three cDNAs differing in the C-terminal. However, there is no evidence for alternative splicing of SPC2 and SPC3.

Receptors for neuroendocrine peptides are also typically alternatively spliced with different splice forms expressed in different cells. For example, isoforms of the lutenizing hormone receptor (LH-R) are differentially expressed in cumulus cells and theca cells in the cow ovary whereas all isoforms are expressed in granulose cells, although the levels of expression vary depending on cell size (Robert et al. 2003). Similarly, splice forms of the estrogen related receptor β (ERRβ), only one of which can increase transcription of estrogen receptor α (ERα), are differentially expressed in the human endometrium (Bombail et al. 2008, 2010).

For deuterostomes other than jawed vertebrates, there is comparative little information concerning alternative splicing of hormone receptors and the pre-mRNAs of hormones. However, in the sea lamprey, the GnRH-1 pre-mRNA produces three mature mRNAs by alternative splicing of the 3'-end (Suzuki et al. 2000). All three of the translated proteins are cleaved to yield the same 10-amino-acid GnRH-1 peptide. However, the carboxy terminal half of each protein is cleaved to yield a different GnRH-associated peptide (GAP). The function of these three GAP peptides from the lamprey GnRH-1 gene is unknown, but the single GAP peptide cleaved from the human GnRH precursor protein can inhibit secretion of lutenizing hormone (LH), follicle stimulating hormone (FSH), and prolactin from cultures of rat pituitary (Nicolics et al. 1985).

Future directions
Thanks to the sequencing of numerous eukaryotic genomes and the development of new technologies such as exon microarrays and high throughput sequencing of cDNAs (RNA-seq), the pervasiveness of epigenetic phenomena has come to light. However, the roles of epigenetic phenomena in development and homeostasis have only begun to be appreciated. Much of the recent focus on epigenetics has been on the roles of miRNAs in posttranscriptional gene silencing. With the realization that over 75% of all genes are alternatively spliced, comes an appreciation that the emerging picture of the most studied gene networks, such as that for the sea urchin endoderm, as complicated as it is, is only a rough framework. Nearly every gene in the network is really represented...
by several proteins, expression of which is tightly regulated in time and space within an organism.

Techniques for investigating the function of particular isoforms include in vitro analyses. Transactivation and DNA-binding assays can be used for transcription factors. In transactivation assays, the activation or repression domains of transcription factors are linked to a DNA binding domain (typically Gal4) and transfected into tissue-culture cells together with a reporter plasmid with Gal4 binding sites and the luciferase coding sequence. The amount of luminescence produced when the cell lysate is mixed with ATP, and luciferin is proportional to the ability of the protein to activate transcription (Hsieh and Hayward 2001). These assays depend on the particular cell line having the requisite co-factors for transactivation. Not all cell lines are adequate for all transcription factors. In addition, binding assays can be used to determine the relative ability of isoforms with alternatively spliced DNA-binding domains to interact with target sequences, while chromatin immunoprecipitation (ChIP) can reveal the proteins that bind to a particular DNA-binding domain (Collas 2010). There are various modification of ChIP such as ChIP on microarrays and ChIP-seq that allow binding sites of particular DNA-binding proteins to be determined. Arrays for mammals are commercially available, but could be constructed for any organism for which the genome has been sequenced.

There is a wide range of techniques for studying protein–protein interactions (Phizicky and Fields 1995; Li and Wu 2009), which theoretically could be applied to individual isoforms. Some methods such as protein-affinity chromatography may require a relatively large amount of protein, while others such as affinity blotting, similar to western blotting except that a labeled protein rather than an antibody is used, require smaller amounts. Immunoprecipitation is used to isolate a protein complex. Recently, peptide arrays have been developed for identification of proteins that bind particular modular domains (Li and Wu 2009).

It is more difficult to study protein functions in vivo. However, in vivo function of some isoforms can be studied by overexpression or knock-down of particular isoforms. For organisms with genetics such as the mouse, it can be possible to engineer mutants defective in splicing of particular sequences. For example, in the mouse, Fgf8 produces two major isoforms, Fgf8a and Fgf8b, which differ in that due to use of an alternative 5'-splice site in exon 4, Fgf8a is 11 amino acids shorter than Fgf8b, resulting in different affinities of the two for the FGF receptors (MacArthur et al. 1995). A mutation that eliminates all isoforms containing the longer exon 4 has the same effect as null mutation of Fgf8 in that the midbrain and cerebellum are defective. However, a mutation that eliminates only isoforms with the short exon 4 has no apparent effect on development of the mouse (Guo et al. 2010). For organisms without genetics such as amphioxus, splice-blocking antisense morpholino oligonucleotides (morpholinos) promise to allow knockdown of some isoforms. UV-light-activated morpholinos offer the opportunity to inject morpholinos into early embryos but delay the effects until later in development (Tomasini et al. 2009). This technique can be used with splice-blocking morpholinos. Techniques have also been developed for morpholino-knock-downs in adult animals (Moulton and Jiang 2009); these could facilitate an investigations of the roles of splice forms of endocrine peptides and their receptors.

Finally, oligonucleotide probes synthesized with locked nucleic acids, which have a methylene bridge between the 2’O and the 4’C on the ribose ring, have the potential for demonstrating tissue-specific expression of particular isoforms. To date, locked nucleic acid probes have been extensively used to identify tissue-specific expression of miRNAs and for fluorescent in situ hybridization to determine the location of particular genes on chromosomes. However, the technique has recently been extended to identify tissue-specific expression of alternatively splice exons (Darnell et al. 2010).

In summary, there are many techniques for investigating protein functions both in vitro and in vivo that are at least theoretically applicable to specific isoforms. Several of these techniques have been developed only recently. With the realization that multiple isoforms are the rule rather than the exception, it can be anticipated that our understanding of the roles of isoforms in modulating development and physiology will increase exponentially in the next few years.

Funding
This work was supported by the National Science Foundation (grants IOS07-43485 and MCB06-20019 to L.Z.H.).

References


family is required for serotonergic projections to appropriately innervate target brain areas. J Neurosci 29:9137–47.


