Short linear motif acquisition, exon formation and alternative splicing determine a pathway to diversity for NCoR-family co-repressors

Stephen Short1,†,‡, Tessa Peterkin2,†, Matthew Guille3,4, Roger Patient2 and Colin Sharpe3,4

1Institute of Marine Sciences, School of Biological Science, University of Portsmouth, Portsmouth PO1 2DY, UK
2The Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK
3Institute of Biomolecular and Biomedical Science, School of Biological Sciences, University of Portsmouth, Portsmouth PO1 2DY, UK
4European Xenopus Resource Centre, University of Portsmouth, St Michael’s Building, Portsmouth PO1 2DT, UK

Vertebrate NCoR-family co-repressors play central roles in the timing of embryo and stem cell differentiation by repressing the activity of a range of transcription factors. They interact with nuclear receptors using short linear motifs (SLiMs) termed co-repressor for nuclear receptor (CoRNR) boxes. Here, we identify the pathway leading to increasing co-repressor diversity across the deuterostomes. The final complement of CoRNR boxes arose in an ancestral cephalochordate, and was encoded in one large exon; the urochordates and vertebrates then split this region between 10 and 12 exons. In Xenopus, alternative splicing is prevalent in NCoR2, but absent in NCoR1. We show for one NCoR1 exon that alternative splicing can be recovered by a single point mutation, suggesting NCoR1 lost the capacity for alternative splicing. Analyses in Xenopus and zebrafish identify that cellular context, rather than gene sequence, predominantly determines species differences in alternative splicing. We identify a pathway to diversity for the NCoR family beginning with the addition of a SLiM, followed by gene duplication, the generation of alternatively spliced isoforms and their differential deployment.

1. Introduction

Vertebrates may be intuitively described as more complex than invertebrates, but the molecular basis for this distinction, and the pathways by which it is achieved, are less apparent. Because total gene counts are often comparable, it has been suggested that increases in the number and type of regulatory DNA elements, combined with an increased diversity in the composition of the transcription factor complexes with which they interact, may begin to account for the increasingly complex patterns of gene expression seen over evolutionary time (reviewed in [1,2]). In contrast, even small changes to the sequence and structure of transcription factors themselves are likely to disrupt their activity and have deleterious effects. The recent identification, however, of short linear motifs (SLiMs), defined as functional peptide modules 3–10 amino acids in length [3,4] which act as modular components within a larger protein, points to these as independent targets for evolutionary change, because the gain, loss or alteration of one motif is less likely to compromise the activity of others [5]. In addition, many genes use multiple promoters and alternative splicing to make several transcripts from one gene that can then be translated into isoforms with distinct functions [6,7]. Using alternative splicing to generate
isoforms that differ in their complement of SLiMs will generate related proteins with diverse functions that may contribute to organismal complexity [3].

Vertebrate nuclear co-repressors NCoR1 and NCoR2, also known as silencing mediator for retinoid or thyroid-hormone receptors (SMRT), are large proteins whose genes are derived from a common ancestor. Co-repressor activity is reflected in their structure, in which the 50 amino acid amino-terminal SANT domains (named after Swi3, Ada2, NCoR and TFIIB) [8] are core to regions that interact with histone deacetylases to put chromatin into a compact, transcriptionally inactive state [9–13]. Sequences that mediate the interaction with the nuclear receptor transcription factors, however, are found as SLiMs, termed co-repressor for nuclear receptor (CoRNR) boxes, embedded within a carboxy-terminal region that lacks significant structural organization [14–18]. Type II nuclear receptors, such as the retinoid receptors, bind DNA as heterodimers with a common, RXR partner [19–22], and each co-repressor is thought to interact with a nuclear receptor dimer [18,22,23]. To achieve this, NCoR1 uses any two of its three CoRNR boxes to bind directly to the receptors, but only in the absence of the receptor’s ligand, such as retinoic acid [14–17,24,25]. The human, mouse and Xenopus NCoR2 genes also encode three CoRNR boxes, equivalent to those in NCoR1 but, through alternative splicing, produce protein isoforms with variable numbers of these motifs [24–29]. The co-repressors bind to a wide range of nuclear receptors and the different in vitro affinities of the CoRNR boxes for nuclear receptors and their distribution between the NCoR2 isoforms demonstrate that alternative splicing generates diverse isoforms that preferentially interact with specific subsets of nuclear receptors [25,28–31].

Each co-repressor acts as a platform for the assembly of multi-protein complexes [32,33] that actively repress a remarkably wide range of transcription factors including, mos, if not all, of the type II nuclear receptors and, among others, the transcription factors Pit1, PLZF, Bel-6, Nfkb, SRF, CBF-1 and ETO (reviewed in [28]). Not surprisingly, NCoR1 and NCoR2 have been implicated in diverse biological processes. NCoR1 knockouts in mice have lethal defects in erythropoiesis [34], while NCoR2 knockouts die from defects in cardiac development [35,36]. NCoR1 and NCoR2 also affect embryonic development [24,37], neural stem cell differentiation [35,38], homeostasis [39], oxidative metabolism and ageing [40], adipocyte differentiation [31] and embryonic blood formation [41]. Altered interactions between the co-repressors and mutated retinoid receptors underlie acute promyelocytic leukaemia [42–44] and primary myelofibrosis [45], while NCoR2 has been implicated in the progression of glioblastoma in animal models [46].

In vertebrates, the 3′ part of the gene encoding the carboxy-terminal region of each co-repressor is divided between 10 exons. In NCoR2, this structure underpins alternative splicing to generate isoforms with different numbers of CoRNR boxes. For example, exon 37 encodes a CoRNR box, but the use of an internal splice donor generates an isoform lacking this motif. The capacity for exon 37 alternative splicing in NCoR2 is conserved between Xenopus, mice and humans [27]. While both exon 37 isoforms are found at roughly equivalent levels in Xenopus tissues, in mice the outcome of exon 37 alternative splicing is tissue specific, with the CoRNR box-containing isoform (37b+) predominant in the brain and the CoRNR box excluded isoform (37b−) found in most tissues [27,29]. Unlike NCoR2, there is no detectable alternative splicing of this exon in Xenopus NCoR1, but a distinct isoform has been reported in mammals [27,28].

Significant differences in function between NCoR2 isoforms have been demonstrated in vitro [25,29,30]. The exclusion of NCoR2 exon 37b in vivo, during Xenopus development, results in embryos with deformed heads, disturbed axon guidance and the repression of some early thyroid hormone-responsive genes, indicating this alternative splicing event is significant for embryogenesis [24]. In addition, mice engineered to express NCoR2 with defective CoRNR boxes show a range of mutant phenotypes [40,45,47]. These results indicate that the CoRNR boxes are not redundant, because a full complement is required for normal function.

The gain and loss of SLiMs in proteins involved in transcriptional control is a significant mechanism in vertebrate evolution [5]. In addition, the direct correlation between intrinsically disordered regions (IDRs) and alternatively spliced exons [48], combined with the frequent presence of SLiMs in IDRs, indicates a mechanism by which the assortment of SLiMs between tissue-specific isoforms can contribute to functional complexity at the level of the cell (reviewed in [3]). Using the nuclear co-repressors as a test case, we extend this concept from cells to organisms by demonstrating a transformative increase in the diversity of these proteins from sea urchin to frog. The pathway to diversity, involving progressive SLiM acquisition, augmented by a striking exon fragmentation and the deployment of alternatively spliced isoforms, defines a direct mechanism by which the complexity of interactions of a family of transcription-associated proteins is enhanced over evolutionary time.

2. Results

2.1. The acquisition of short linear motifs

The vertebrate paralogues NCoR1 and NCoR2 are defined by two SANT domains [9–13], three CoRNR box motifs that mediate interactions with nuclear receptors [14–17,24,26] and a carboxy-terminal domain that interacts with SHARP, a transcriptional repressor [49] (figure 1a). Alignment of vertebrate NCoR1 and NCoR2 C-terminal sequences identified four additional conserved short motifs (figure 1a,b and electronic supplementary material, figure S1) that will be targets for future functional analysis.

SLiMs, such as CoRNR boxes, in the carboxy-terminal region mediate many of the interactions of the NCoR-family co-repressors with transcription factors [25,28–31]. Because additional isoform diversity, particularly in NCoR2, is generated by alternative splicing of the primary transcript in this region, we next looked at the organization of exons encoding the C-terminal interaction domains and mapped the SLiMs to their encoding exons (figure 1c). The organization of the paralogues is highly conserved in vertebrates with most having 10 exons, from that encoding the first CoRNR box (exon 37) to the stop codon (exon 46). An exception is zebrafish NCoR2, which lacks exon 38, although many other actinopterygians have the standard vertebrate organization (data not shown).

The presence of the domains and motifs was used to confirm the annotation of NCoR-family proteins encoded in invertebrate deuterostome genomes and identified that a
Figure 1. The NCoR-family conserved motifs and exon structure. (a) The NCoR-family proteins in the vertebrates typically contain two SANT domains (green bar), followed by three CoRNR boxes, nuclear receptor interaction motifs (yellow bars) and a carboxy-terminal SHARP-binding motif (red bar). Alignment of vertebrate NCoR1 and NCoR2 sequences identifies a further four conserved motifs (blue bars, lower diagram). Full sequence alignments are in the electronic supplementary material, figure S1. (b) Identity of conserved vertebrate sequences using the motif notation. Yellow bars overlie the consensus CoRNR box motif $L/I.X.X.X/I.H.L.X.X.X/I.L$ ([50,51]) that is embedded in each of motifs 1, 2 and 5. The C-terminal SHARP-binding sequence is overlined in red as part of motif 8. (c) Exon organization of the 3′ end of representative NCoR-family genes. The regions encoding the motifs have been mapped onto the relevant exons maintaining the colour scheme in (a). In parentheses is the number of exons in this region of the gene. The C-terminal motifs are encoded by one large exon encoding 843 amino acids in the sea urchin, but 12 exons encoding 365 amino acids in the sea squirt. (d) Summary of C-terminal motif acquisition across the representative deuterostome panel. All contain motifs 1, 2 (CoRNR boxes 1 and 2) and 8, but motif 5 (the third CoRNR box) is not present in the echinoderm and incomplete in the hemichordate and urochordate. Two of the four vertebrate specific motifs (3 and 4) are represented by partial motifs in the urochordate and cephalochordate (motif 4 only). Sequence alignments of the motifs are in the electronic supplementary material, figure S2.
biases alternative splicing to produce predominantly the short 37b—isoform, without altering the overall level of NCoR2 transcripts either in the whole embryo or in the range of tissues examined. This experimental bias generates a distinct mutant phenotype, indicating the functional significance of exon 37 alternative splicing in embryonic development [24].

Alignment of the 3' part of *Xenopus* exon 37 in NCoR1 and NCoR2 shows extensive sequence conservation, apart from the internal splice donor, which in NCoR1 is a GA rather than the active GT dinucleotide seen in NCoR2 (figure 2a). Including the equivalent region of the *Ciona* NCoR-family gene in the comparison (figure 2a) shows a GT at the corresponding position suggesting that the common ancestor of *Ciona* and the vertebrates had a potential splice donor dinucleotide.

Because an effective splice donor requires sequences in addition to the conserved GT [54], we next tested whether the presence of a GT, rather than the GA, at the internal site in NCoR1 can reconstitute an effective splice donor. First, we used the program MAXENTSCAN [55] to quantify the effectiveness, as splice donors, of the sequences surrounding all dinucleotides in NCoR1 exon 37 that could be converted to a GT by a single base change, and then compared these with the range of scores found for all other validated *Xenopus* NCoR1 exon splice donors. The average score for the confirmed splice donors is just over 8, and this approach identified three sites within exon 37 with a greater score, and these sites are predicted to form strong splice donors when the core dinucleotide is mutated to a GT. Of these three sites, one was at the GA corresponding to the internal splice donor, which in NCoR2 and the other two were approximately 50 and 90 bp further upstream (figure 2b).

To determine, experimentally, whether the sequence context of the equivalent site in NCoR1 reconstitutes an effective splice donor, we used site-directed mutagenesis to convert the GA to a GT and then cloned the exon, and flanking intron sequences, into the splicing minigene, pTBNde1 [56]. Because it is possible that any GT that has a surrounding sequence calculated to be a strong splice donor might permit alternative splicing, we addressed specificity using an NCoR1 exon 37 minigene in which the first predicted strong site upstream of the equivalent site was also converted to a GT (figure 2c).
Consensus splice donor. (a) Representative of each vertebrate group and when present is part of a high-scoring consensus. A site 1 GT is present in the sea squirt, but is part of a poor consensus splice donor context in the precursor. Following gene duplication, it was subsequently lost from NCoR1, by point mutation, but was subsequently gained in NCoR2. The simpler explanation, given the presence of the equivalent GT in both the precursor, followed by the gain of the obligatory GT which an effective splice donor context arose in the NCoR-family precursor, is that exon 37 alternative splicing arose in the precursor but was subsequently lost from NCoR1, by point mutation, following gene duplication.

2.3. The conservation of NCoR2 exon 37 alternative splicing

Because the alternative splicing of NCoR2 exon 37 has been characterized in *Xenopus*, mouse and humans, and generates isoforms that differ in a functional CoRNR box motif [24], we next investigated the conservation of the internal splice donor across nine species of fish, two lampreys and *Ciona*. The position and splice donor strength of each GT dinucleotide across 115 bases of the 3' UTR of NCoR2 exon 37, centred on the internal splice donor, was calculated by MAXEntScan (figure 3a).

While the NCoR-family gene in both *C. intestinalis* and *C. savignyi* (sea squirts) has a GT at site 1, the equivalent position to the internal splice donor in *Xenopus* NCoR2, (figure 3a), it does not score well as a predicted splice donor and there is no published transcriptomic evidence for its use. The agnathostomes *Petromyzon marinus* (sea lamprey) and *Leithenteron japonicum* (Japanese lamprey) each have two paralogues and one, like NCoR1, lacks the equivalent internal splice donor while it is present in the second, where it is predicted to be a strong splice donor in the correct frame for productive splicing. EST data and limited RT-PCR analysis (figure 3b) in *Petromyzon marinus* (sea lamprey), though, suggest that the internal site is not commonly used.

Of three cartilaginous fish examined, only *Squalus acanthias* (dogfish) has site 1. Although the *Leucoraja erinacea* (little skate) has a site further upstream that is predicted to be an effective splice donor, the corresponding transcripts are not present in the reported transcriptome. Of the ray-finned fish, NCoR2 site 1 is present in four out of six genomes examined, being absent in two related catfish species. It is likely that the internal splice donor site is active in *Oryzias latipes* (medaka) because it

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Species & Exon 37 1 & Exon 37 2 & Score & Evidence \\
\hline
Mammal & *Mus musculus* & 10.5 & G/T & \\
Amphibian & *Xenopus laevis* & 10.5 & G/T & \\
Lobe finned fish & *Latimeria chalumnae* & 10.5 & G/T & \\
Ray finned fish & *Lepisosteus oculatus* & 10.5 & G & \\
Teleosts & *Danio rerio* & 10.5 & G & \\
& *Oryzias latipes* & 10.5 & G & \\
& *Ictalurus punctatus* & 10.0 & G & \\
& *Ictalurus furcatus* & — & G & \\
& *Callorhinchus milli* & — & G/T & \\
& *Squalus acanthias* & 8.91 & G/T & \\
& *Leucoraja erinacea* & — & G/T & \\
Jawless fish & *Petromyzon marinus* & 9.5 & G/T & \\
& *Lampetra japonica* & 9.5 & G/T & \\
Sea squirt & *Ciona intestinalis* & —1.7 & G/T & \\
\hline
\end{tabular}
\caption{Site score and evidence for alternative splicing across a range of species. The size of the PCR product generated from lamprey cDNA indicates that it includes exon 37b while that from 48 h zebrafish embryos corresponds predominantly to 37b transcripts. *Xenopus*, in contrast, generates both isoforms as shown by the two bands in the RT-PCR.}
\end{table}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3}
\caption{The conservation of NCoR2 exon 37 alternative splicing. NCoR2 exon 37 in *Xenopus* and mouse has two splice donors (red bars), site 1 is internal and site 2 is at the end of the extended exon. Exon 37b, between the two sites, encodes the CoRNR box 1 motif. (a) The potential site 1 GT splice donor is present in at least one representative of each vertebrate group and when present is part of a high-scoring consensus. A site 1 GT is present in the sea squirt, but is part of a poor consensus splice donor. (b) RT-PCR analysis of exon 37 alternative splicing across a range of species. The size of the PCR product generated from lamprey cDNA indicates that it includes exon 37b while that from 48 h zebrafish embryos corresponds predominantly to 37b transcripts. *Xenopus*, in contrast, generates both isoforms as shown by the two bands in the RT-PCR.}
\end{figure}
is closely followed by an in-frame stop codon that would otherwise produce a truncated protein with compromised function (electronic supplementary material, figure S3). An assessment of 40-h post-fertilization (hpf) Danio rerio (zebrafish) embryos indicates that site 1 is predominantly used (figure 3b) and this is supported by EST data, but there is also a low level of the longer transcripts that use site 2. We next examined, in more detail, why the observed use of site 1 differs between species such as zebrafish and Xenopus, when the consensus splice-donor sequences are identical.

### 2.4. The acquisition of distinct patterns of alternative splicing in NCoR2 exon 37

During early development, zebrafish uses site 1 to produce solely the short (37b–) isoform, however a low level of the longer exon 37b+ transcripts can be detected by embryonic day 5 (figure 4a). This is likely to represent the production of NCoR2 exon 37b+ transcripts in neural tissue, as they are also found in the dissected brain and eyes of adult fish, but not in other tissues examined (figure 4b). This is similar to the tissue-specific pattern seen in mice [27]. Consequently, while both Xenopus and zebrafish use alternative splicing to generate NCoR2 exon 37 isoforms, strategies for isoform deployment differ in that the expression of both isoforms is widespread in Xenopus, but temporally, and spatially, regulated in zebrafish.

To determine whether the intrinsic sequence of the internal splice donor or its cellular context plays the greater role in determining the splicing pattern of NCoR2 exon 37, we generated splicing minigenes containing either zebrafish or Xenopus exon 37, together with flanking intron sequences, in the pTBNde1 minigene [56]. The minigenes were each injected into Xenopus laevis embryos either have a suppressor to inhibit the terminal splice donor or a splice-promoting protein to enhance the use of the internal donor that operates less efficiently than the intrinsic sequence of the splice donor, determines the outcome of exon 37 alternative splicing.

Because placing either minigene in a Xenopus context mimics the endogenous Xenopus pattern of alternative splicing, we next repeated the analysis, injecting the minigenes into zebrafish embryos. The Xenopus minigene again produced two bands, but this time with a significant bias towards the short 37b– isoform. This indicates that the pattern of alternative splicing is strongly influenced by the cellular context. S, size markers; U, uninjected; X, injected Xenopus construct; Z, injected zebrafish construct.

in approximately equal amounts, in contrast to the total exclusion of the longer form seen for the endogenous gene in fish at an equivalent developmental stage. This suggests that the cellular context provided by the Xenopus embryos, rather than the intrinsic sequence of the splice donor, determines the outcome of exon 37 alternative splicing.
3. Discussion

The vertebrate nuclear receptor co-repressors, NCoR1 and NCoR2, play important roles in physiological [22,24,31,34–37, 39–41] and pathological conditions [42–45] by interacting with a wide variety of transcription factors and other DNA binding proteins [28]. NCoR1 and NCoR2 interact with nuclear receptors via short sequence motifs called CoRNR boxes, located in the intrinsically disordered carboxy-terminal part of the co-repressor [14–18]. Alternative splicing, particularly in NCoR2, determines the complement of motifs in the protein and so generates diverse isoforms, each with specific binding capabilities [25,28–31]. As a result, NCoR1 and NCoR2 conform to a model where the selection of SLiMs by alternative splicing, from within an IDR of a protein, plays a significant role in the generation of functional diversity [3]. Here, we combine comparative and experimental approaches to analyse the origins of co-repressor diversity across the deuterostomes.

3.1. Diversity through motif acquisition

Strongylocentrotus purpuratus (sea urchin) has a single NCoR-family gene with only limited sequence homology to NCoR1 and NCoR2, but encoding two indicative SANT domains and three of the eight vertebrate NCoR-family motifs. These include two CoRNR boxes [14,50,51] that are typical SLiMs and a SHARP interacting motif at the carboxy-terminal of the protein. The remaining motifs may indicate regions that interact with other transcription factors or act as sites for the protein. The remaining functions of the co-repressor protein in vivo [3,57–59]. In comparison, Branchiostoma floridae (amphioxus) produces a co-repressor with three complete CoRNR boxes. Increasing the number of motifs will increase the functional diversity of the co-repressor, because in vitro experiments using mouse or Xenopus proteins have shown that different CoRNR boxes have different affinities for specific nuclear receptors [14–17,25]. It is likely, however, that lifting repression by the ligand-dependent displacement of the co-repressor will be more significant than imposing repression by binding, because this mechanism would set ligand concentration thresholds for nuclear receptor activation that are dependent on the CoRNR box complement of the co-repressor. This concept is illustrated, in exaggerated fashion, in acute promyeloctic leukaemia, in which specific NCoR2 isoforms are displaced from the pathological RAR fusion protein at distinct concentrations of retinoic acid [44].

Changes to the cis-regulatory elements in the promoter of a transcription factor have been directly associated with evolutionary events [60]. Because most promoters are a collection of independent elements that each control a limited aspect of gene expression, a mutation in one element is likely to affect expression of the gene in only one component of its pattern. In contrast, mutations that affect the protein coding sequence of a transcription factor itself will tend to affect, often calamously, the expression of all downstream targets [60]. The protein sequence changes seen in the NCoR family, however, illustrate how the consequences of changes to the protein coding sequence can be mitigated. By encoding functional SLiMs within IDRs, the gain or loss of a SLiM has an incremental effect, because the remaining functions of the protein are essentially maintained [5]. The insect Fitz protein, and its ability to interact with Ftz-F1, typically illustrates this interaction and involves a SLiM closely related to the core CoRNR box sequence [52].

3.2. Fragmentation of the invertebrate NCoR-family terminal exon

The entire C-terminal region, encoded by exons 37–46 in Xenopus, is encoded by a single exon in sea urchins and is predicted to have the same organization in acorn worms and amphioxus. In C. intestinalis (sea squirt), however, this part of the gene is divided into 12 exons and is consistent with chordate phylogeny, which predicts the tunicates, rather than amphioxus, are most closely related to the vertebrates [61]. A similar degree of discrepancy in exon number and exon boundary location between C. intestinalis and humans is seen in the huntingtin gene [62]. The trigger and mechanism for this remarkable and extensive fragmentation of the NCoR-family gene terminal exon is unknown.

3.3. Diversity through gene duplication

Across the deuterostomes analysed, a complement of two NCoR-family genes is first seen in the genome of the lampreys. Gene duplication opens the possibility for a form of sub-functionalization and neofunctionalization in which altered cis-regulatory events, alternative splicing and protein sequence changes happen within one parologue on the background of an initially redundant second sequence [63,64]. Following gene duplication, the amino acid sequences of the paralogues have (apart from the identified motifs) diverged extensively in the C-terminal region such that NCoR1 and NCoR2 have less than 40% identity in humans (data not shown). Importantly, gene knockout studies in mice show that the two paralogues are no longer equivalent [22,34–36].

3.4. Diversity through alternative splicing: the case of NCoR2 exon 37

Comparisons between Xenopus NCoR1 and NCoR2 show a high degree of nucleotide sequence conservation across the latter half of exon 37. One difference, however, is the GT that forms the conserved core dinucleotide of the NCoR2 internal splice donor that is a GA in NCoR1. A GT at the equivalent position in the single gene in both C. intestinalis and C. savignyi suggests that the GT may be the ancestral form that changed to GA in the NCoR1 gene after duplication. A point change that restores the GT to the internal NCoR1 splice donor recovers the splicing activity of this site. There is more to the activity of this site, however, than just the dinucleotide and the immediate surrounding sequence, because the introduction of a GT upstream in the same exon, that generates a site predicted to be an efficient splice donor, is inactive in Xenopus embryos. Su et al. [53] have suggested that the loss of pre-existing alternative splicing in one parologue, and the generation of more diversity in the other, may not be uncommon, and this seems a plausible scenario for NCoR1 and NCoR2.

Alternative splicing at exon 37b varies the number of CoRNR boxes in NCoR2 and this has functional significance in Xenopus laevis embryonic development [24]. Unlike Xenopus, the equivalent exon in NCoR1 is alternatively
spliced in mammals to generate isoforms with different numbers of CoRNR boxes, though from a different splice donor [28] (and see NM_001190440). This is consistent with the idea that alternative splicing of exons that contain SLiMs within an IDR is an efficient mechanism for the generation of isoforms with different activities that can progressively contribute to the complexity of the cellular functions during evolution [3,65].

3.5. Diversity through the deployment of alternative splicing

With two splice donors in exon 37, zebrafish has the capacity for alternative splicing, but in the early embryo uses only the internal splice donor, and so the resulting isoform excludes one of the CoRNR boxes. It is only later in development, and in the adult, that alternative splicing is deployed, but restricted to neural tissues (figure 4). In contrast, Xenopus NCoR2 37b+ and 37b− isoforms are readily found in all embryonic and adult tissues analysed [27]. The activity of trans-acting factors [66] in zebrafish, but not Xenopus, embryos may prevent splicing from the external site either directly, or indirectly by promoting the use of the internal site. This is supported by the observation that a zebrafish exon 37 minigene introduced into Xenopus embryos gave approximately equal amounts of 37b+ and 37b− transcripts. The final outcome of alternative splicing, however, is likely to depend on a combination of the intrinsic strength of the splice sites, determined by nucleotide sequence, and the activity of a number of trans-acting factors.

Analyses of differences in alternative splicing patterns between humans and mice have largely come to a different conclusion. Using transgenic mice that contain part of human chromosome 21, and looking at genes whose splicing patterns differ between mice and humans, Barbosa-Morais et al. [67] found that the human genes maintain the human pattern, even in the mouse context, concluding that species-specific patterns of alternative splicing are driven by differences within the genes rather than by changes in the trans-acting factors [67]. The results presented here indicate that differences in the activity of trans-acting factors between species can also play a significant role.

A difference between vertebrates and other deuterostomes may lie in the increased complexity of their gene regulatory networks [68]. The vertebrate co-repressors NCoR1 and NCoR2 exemplify this because they interact with an impressively broad range of transcription factors by generating isoforms in which the interaction domains contain different complements of the CoRNR box motifs. In contrast, the sea urchin co-repressor is much simpler with one fewer CoRNR boxes and a lack of carboxy-terminal isoforms. In this paper, we detail the pathway leading to the increased diversity of vertebrate co-repressor isoforms (figure 6), highlighting the role of SLiMs located within IDRs, and their deployment by alternative splicing. We therefore identify a mechanism that generates functional diversity in a transcription-associated protein, a critical contributory factor in determining organismal complexity.
4. Material and methods

4.1. Sequence alignment

The accession numbers of genes used in the comparisons are listed in the electronic supplementary material, table S1. Where the invertebrate NCoR-family orthologue was not annotated, candidates were identified by BLAST comparisons using vertebrate NCoR-family motifs. Candidates with at least two CoRNR box motifs and a C-terminal SHARP interaction motif in the correct order were further validated by the presence of two upstream SANT domains. The sequence of the protein was then inferred from a combination of manual annotation and reference to online annotation. Multiple sequences were aligned using CLUSTALW2 and CLUSTAL OMEGA (EBI-EMBL) using standard criteria.

4.2. RT-PCR, cloning and sequencing

Zebrafish (AB mixed with Tubingen) total RNA was isolated from five to ten embryos at the developmental stages described in the text or at 26–28 hpf using TRI Reagent® (Sigma) and purified using RNeasy Micro Kit (Qiagen). *Xenopus* total RNA was isolated from three to five neurula stage embryos by phenol extraction and precipitation. Alternative splicing was assessed by conversion of RNA into cDNA using Superscript III reverse transcriptase and random nonameric primers, followed by PCR. PCR used species and exon-specific oligonucleotides primers (electronic supplementary material, table S2) and Platinum Taq polymerase (ThermoBioscience) or ReadyMix™ Taq (Sigma). Where described, PCR products were resolved on 1–2% agarose, 1× TBE or 0.5× TAE gels, cloned directly into the vector pCR2.1 (TA cloning, Invitrogen) and sequenced (Source Bioscience).

4.3. Cloning of *Xenopus laevis* and zebrafish exon 37 genomic regions

Total nucleic acid was prepared from 50 *X. laevis* tailbud embryos [70] and treated with RNase. Sets of primers were designed from the *X. laevis* genome assembly v6 on Xenbase [71] to amplify exon 37 of NCoR2 and NCoR1 with approximately 250 base pairs of upstream and downstream intron sequence. The PCR products were cloned into pCR2.1 (Invitrogen, TA cloning) and sequenced (Source Bioscience). The genomic fragments were then blunt-end cloned into the Nde1 site (blunted) of the splicing vector pTBNde1 [56,72]. This vector is based on pBluescript and contains the CMV enhancer driving the expression of human globin and fibronectin exons separated by an intron. The Nde1 site is located centrally within the intron. The orientation of the cloned insert was determined by sequence.

In the zebrafish genome, NCoR2 exon 36 is separated from exon 37 by a short intron of 97 basepairs. We therefore used a primer 249 basepairs upstream of exon 36, spanning a naturally occurring Nde1 site and a reverse primer 474 basepairs downstream of exon 37 that incorporated an Nde1 site. Fragments were cloned into pCR2.1 (Invitrogen, TA cloning) excised with Nde1, cloned into the Nde1 site of splicing vector pTBNde1 and the orientation checked by sequencing.

4.4. Site-directed mutagenesis of NCoR1 exon 37

A single base change was introduced into the NCoR1 exon 37 sequence by site-directed mutagenesis of the clone in pCR2.1 using overlapping oligonucleotides carrying the required mutation. Amplification of the mutated sequence used Vent polymerase (New England Biolabs) to limit further mutation and the final construct checked by sequencing. The fragments from pCR2.1 were then blunt-end cloned into pTBNde1 [72] as described above.

4.5. Splicing assays

The NCoR1 and NCoR2 exon 37 constructs in pTBNde1 were grown in media and isolated (plasmid midi-prep kit, Machery-Nagel). Approximately 200 pg of plasmid at 20 pg nl⁻¹ was injected into each *X. laevis* embryo at the two-cell stage, and the embryos grown to the mid neurula stage (Nieuwkoop and Faber, stage 16) [73]. Zebrafish embryos were injected at the one-cell stage with 200 pg of plasmid at 400 pg nl⁻¹, grown overnight at 32°C and collected at stage 27/28 hpf. Total nucleic acid was then extracted [70] and DNA removed by RNase-free DNase digestion. The remaining nucleic acid was precipitated, resuspended and converted to cDNA using reverse transcriptase. The splicing status of the transcripts from the clones was assayed by PCR using forward and reverse primers against the human exons [72] or one human and one *Xenopus-* or zebrafish-specific sequence and the products resolved on 1.5–2% agarose gels.

Ethics. Research in this paper was covered by local ethical review and is covered by Home Office project licences (C.S., R.P.).

Data accessibility. Additional datasets supporting this article have been uploaded as part of the electronic supplementary material.

Authors’ contributions. T.P. and C.S. carried out the molecular laboratory work, S.S. and C.S. carried out sequence alignments, participated in the design of the study and drafted the manuscript; M.G., T.P. and R.K.P. participated in data analysis and helped draft the manuscript. All authors gave final approval for publication.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by the BBSRC (M.G., C.S.), MRC (R.K.P., T.P.), BHF CRE (R.K.P.) and CRM (R.K.P.) and by the University of Portsmouth, Institute of Biomolecular and Biomedical Science (C.S., S.S.).

Acknowledgements. We thank Maggie Walmsley for critical reading of the manuscript, Prof. F. E. Baralle for the splicing plasmid pTBNde1, Clare Baker and Dorit Hockman for sea lamprey RNA.

References


Weston AD, Blumberg B, Underhill TM. 2003 Active repression by unliganded retinoid receptors in development: less is sometimes more. *J. Cell Biol.* **161**, 223 – 228. (doi:10.1083/jcb.200211117)


