Origins of Cdx1 regulatory elements suggest roles in vertebrate evolution

STEPHEN J. GAUNT* and YU-LEE PAUL

Department of Development and Genetics, The Babraham Institute, Babraham, Cambridge, U.K.

ABSTRACT  Cdx1, an upstream regulator of Hox genes, is best characterized for its homeotic effects upon the developing axial skeleton, particularly in the neck. It responds to retinoic acid (RA) in both mouse embryos and embryonal carcinoma (EC) cells. By use of β-galactosidase chemiluminescence, we show that a mouse Cdx1/lacZ reporter expressed in P19 EC cells responds to RA by the combined activities of an intron retinoic acid response element (RARE) and an upstream RARE. In contrast, a chicken Cdx1/lacZ reporter responds only by activity of the intron RARE. Database analyses upon Cdx1 from twenty three vertebrate species reveal that the intron RARE is structurally conserved in amniotes (eutherian mammals, marsupials, birds and Anole lizard), but not in Xenopus or fish. The upstream RARE is structurally conserved only in eutherian mammals. We conclude that the intron RARE originated at around the amphibian/amniote division, and the upstream RARE appeared around the marsupial/eutherian mammal division. In view of the site of action of Cdx1, we propose that acquisition of the intron RARE may have facilitated the substantial changes that occurred in the neck and anterior thorax at the advent of the amniotes. We present evidence that Cdx1 is also a developmental regulator of the female urogenital system, and we suggest that acquisition of the upstream RARE may have contributed to morphological divergence of marsupial and eutherian mammals.

KEY WORDS: Cdx1, retinoic acid response element, evolution

Cdx proteins (Cdx1, Cdx2 and Cdx4) are upstream regulators of vertebrate Hox genes (Deschamps and van Nes, 2005; Houle et al., 2003a). Cdx1 knockout (Subramanian et al., 1995; van den Akker et al., 2002) or over-expresser (Gaunt et al., 2008) mice display homeotic mutations within the axial skeleton, particularly in cervical and anterior thoracic regions. The axial defects in Cdx1 mutants extend forward to the skull/atlas level. The other reported anatomical defects in Cdx1 mutants are in forelimb bud size (Gaunt et al., 2008) and, when combined with Cdx2 mutation, in spinal ganglia (van den Akker et al., 2002).

Mouse Cdx1 is activated by retinoic acid (RA) in both embryos (Houle et al., 2000) and P19 or F9 embryonal carcinoma cells (Beland and Lohnes, 2005; Houle et al., 2000). RA response reporter studies show that RA is present in the embryonic regions known to express Cdx1: the primitive streak of 7.5 day mouse embryos and, by 8.5 days, the developing somites (Rossant et al., 1991). RA administered to pregnant mothers induces homeotic mutations in the axial skeleton (Kessel and Gruss, 1991), and the range of these resemblances those induced by Cdx1 protein over-expression (Gaunt et al., 2008).

RA activates Cdx1 by a retinoic acid response element (RARE) located upstream of the coding region (Houle et al., 2000). Mutation of this RARE results in homeotic shifts in cervical vertebrae, indicating that this element plays a functional role (Houle et al., 2003b). In classical RA transactivation, all-trans RA first interacts with a protein dimer comprising a RA receptor and a retinoid X receptor. This complex regulates gene transcription by binding to a RARE, which typically consists of a pair of [T][G][A][C/A][C][T/C] hexamer repeats separated, most commonly, by either two (DR2-type RARE) or five (DR5) nucleotides, and these may lie with equal frequency on the coding or the non-coding strand (Balmer and Blomhoff, 2005). The Cdx1 upstream RARE (Houle et al., 2000) is atypical in containing only one clear consensus element.

A DR-2 type RARE motif was subsequently found in the first

Abbreviations used in this paper: EC, embryonal carcinoma; RA, retinoic acid; RARE, retinoic acid response element.
intron of both mouse and chicken Cdx1 (Gaunt et al., 2003). This
was putative (Houle et al., 2003b) since it had not been shown to
respond to RA. The role of the intron RARE motif was examined
using a chicken Cdx1/lacZ reporter construct expressed in trans-
genic mouse embryos (Gaunt et al., 2003). Mutation in the motif
resulted in a more posterior expression boundary in neural tube
and greatly reduced expression in somites. Both chicken and
mouse intron RARE motifs are adjacent to a Lef/Tcf motif, which is
a response element usually used by the canonical Wnt pathway.
Mutation of this in the Cdx1/lacZ reporter resulted in substantial
loss of lacZ expression in transgenic embryos (Gaunt et al., 2003).

We now provide direct evidence that the intron RARE motif in
Cdx1 is responsive to RA. Database analyses indicate that this
element has a much earlier origin in vertebrate evolution (around
the amphibian/amniote split) than the upstream RARE (around the
marsupial/eutherian mammal split). We consider how origin of
each RARE may have contributed to a different step in the
morphological evolution of vertebrates.

Results

β-galactosidase chemiluminescence to detect Cdx1 reporter
activity in P19 cells

Figs 1A, 2A show construct #1, the mouse Cdx1/lacZ reporter
that provides a normal Cdx1-like pattern of expression in trans-
genic mouse embryos (Gaunt et al., 2003). The construct contains
the upstream RARE (Houle et al., 2000) and also the putative intron
RARE earlier identified as conserved in chicken Cdx1 (CdxA)
(Gaunt et al., 2003). We first examined luciferase (luc) reporters
since these are most commonly used in culture assays. Construct
#2 contains Cdx1 sequence upstream of the first exon while
construct #3, designed to express a Cdx1/luc fusion protein, also
includes parts of intron 1 and exon 2 (Fig. 1A). Construct #2 is
stimulated by RA (Fig. 1B). However, construct #3 is inactive (Fig.
1B), showing that the Cdx1/luc fusion protein is non-functional.
This raised the problem of how to test for activity in the intron RARE
motif whilst maintaining its normal position in the gene.

We overcame this by using the Cdx1/lacZ reporter itself (con-
struct #1), measuring lacZ (β-galactosidase) activity by a sensitive
chemiluminescence technique. Fig. 1C shows that Cdx1/lacZ
reporter is stimulated by RA, whereas Cdx2/lacZ and Cdx4/lacZ,
both expressed in transgenic mice (Gaunt et al., 2005), show little
evidence of stimulation.

Functional assay of intron and upstream RAREs in mouse
and chicken Cdx1/lacZ reporters

Following transfection of construct #1 into P19 cells, β-galac-
tosidase activity is stimulated 20- to 30-fold by RA (Figs 1C, 2B).
Mouse constructs #4, #5 and #6 are similar to construct #1 but
contain, respectively, mutations within the intron RARE, upstream
RARE, and both RAREs. Constructs #4 and #5 show only about 5-
and 8-fold stimulations by RA, whereas construct #6 shows only 2-
fold stimulation (Fig. 2B). This shows that the putative intron RARE motif is indeed functional, and that the upstream and intron RAREs together account for almost all of the RA stimulation of mouse Cdx1 reporter in P19 cells.

Chicken construct #8 (Fig. 2A), encoding a chicken Cdx1/lacZ fusion protein, is expressed in transgenic mouse embryos with a similar pattern to construct #1 although the level of expression in somites is reduced (Gaunt et al., 2003). Construct #8 in P19 cells is stimulated about 20-fold by RA (Fig. 2C). Construct #9, which differs from construct #8 only by mutation in the intron RARE, is stimulated only 1.5-fold. Addition of a further 6.8kb of upstream sequence (construct #10) does not increase response to RA (Fig. 2C). We conclude that the chicken Cdx1 reporter responds to RA mainly through the intron RARE.

**Fig. 3. Cdx1 retinoic acid response elements (RAREs) in vertebrates.** (A) Alignment of Cdx1 intron sequences. Hexamer repeats of the RARE, beginning position 7, are underlined. Conserved Leu/Tcf motifs are also underlined (dotted). (B) Alignment of Cdx1 upstream RAREs. This RARE (underlined) is atypical, and its overall limits uncertain. Numbers in brackets in A and B indicate positions of nucleotide 1 relative to start of the coding sequence. Yellow highlight, fully conserved bases; red, highly conserved. Turkey and dog sequences are unavailable in regions of intron and upstream RAREs, respectively. (C) Origins of Cdx1 RAREs mapped on an evolutionary tree.

amphibian (*Xenopus tropicalis*) and two fish (Fugu and Tetraodon). Homeodomain amino acid sequences were aligned to facilitate identification of genes as Cdx1 (supplementary Fig. 1). Amino acids consistently conserved in Cdx1 are Asp at position 2, Tyr at 22, Ser at 23, and Asn at 60.

Fig. 3 A-B shows alignments of Cdx1 nucleotide sequences around the two RAREs. The Cdx1 intron RARE (repeats underlined in Fig. 3A) is conserved in all mammals and birds but we were unable to detect it in *Xenopus* or fish. The Anole lizard has a RARE-like motif but this differs in one base that is usually, although not invariably, conserved (Balmer and Blomhoff, 2005). Moreover, the lizard RARE is located much further from exon1 and lies within an intron that is much greater in length than in other species examined. We consider, however, that this is the homologous RARE sequence in the lizard since (i) surrounding regions show sequence similarities with birds and mammals, including the adjacent Leu/Tcf binding motif (dotted underline in Fig. 3A) (Gaunt et al., 2003), (ii) the mouse sequence retains RARE activity even after its mutation to the Anole sequence (construct #7), although activity is reduced by about 50% (Fig 2D), (iii) it is

**Structural survey of Cdx1 RAREs in vertebrate genome databases**

We examined sequences from fourteen eutherian (placental) mammals, two marsupials (wallaby and opossum), three birds (chicken, zebra finch and turkey), one lizard (Green Anole), one
known that non-coding DNA in Anole may be extended by insertion of retrotransposable elements (Di-Poi et al., 2009), (iv) lizards are commonly considered as a sister group to birds (Fig. 3C) (Eernisse and Kluge, 1993) and would therefore be expected to possess the intron RARE, at least in their ancestry.

The upstream RARE (underlined in Fig. 3B) is atypical, containing only one clear consensus hexamer (Houle et al., 2000). It is conserved in all eutherian mammals, although it appears to be diverging in the Afrotherian sub-group (elephant, lesser hedgehog tenrec and hyrax) (Kriegs et al., 2006; Murphy et al., 2001). We were unable to detect the upstream RARE in marsupials, birds, Anole lizard, Xenopus or fish. Fig. 3C shows the inferred origins of the intron and upstream RAREs marked on a commonly adopted vertebrate evolutionary tree (Eernisse and Kluge, 1993).

Defects in the female urogenital systems of Cdx1 over-expressor mice

Fig. 3C raises the question of what are the anatomical changes occurring at these points in evolutionary history that might be relevant to a change in Cdx1 regulation? The main defects so far recorded in Cdx1 mutant mice lie within the cervical and anterior thoracic vertebrae and the ribs. We now present evidence that Cdx1 is also involved in development of the female urogenital system.

From examination of the thoracic and abdominal organs of heterozygous female OE1, OE2 and OE4 mice (which over-express, respectively, Cdx1, Cdx2 and Cdx4 under control of their own promoter elements) (Gaunt et al., 2008) only the urogenital systems of OE1 and OE2 mice show obvious defects. About 16% of both OE1 (10/62) and OE2 (4/25) heterozygous mice show defects in their reproductive tracts. The range of defects is apparently identical for two independently-derived OE1 lines and one OE2 line. Four different phenotypes are seen. First, there is absence of one uterine horn, with persistence of an undeveloped Müllerian duct-like structure which can be traced from the uterine bifurcation to the ovary (Fig. 4A). Second, there is absence of one uterine horn with no detected remnants of the Müllerian duct (Fig. 4B). Third, there is absence of the posterior region of one uterine horn with the anterior part swollen due to retained uterine fluid; the extent of the retained anterior part is variable (Fig. 4 C-E,H). Fourth, there is posterior shift in the level of the uterotubal junction, resulting in elongation of the oviduct (Fig. 4F). In addition, about 8% of OE1 and OE2 mice show defects in their urinary system. This usually appears as hydronephrosis due to an apparent absence/blockage of the ureter (not shown). The hydronephrotic kidney is vesicular and appears swollen (Fig. 4G), shrunken (Fig. 4D), or haemorrhagic (Fig 4H). In other cases there is no kidney (Fig. 4 B,E) or a small kidney (Fig. 4A). A blocked ureter may be swollen (hydroureter) and haemorrhagic (Fig. 4H). Where reproductive and urinary defects occur together they are usually located on the same side (Fig. 4 B,D,E,H). Because gain of Cdx1 and Cdx2 function may lead to these anomalies, we propose a role for these genes in the network that normally regulates development of the female reproductive and urinary systems.

Discussion

**lacZ fusion constructs assayed in P19 cells**

In our study, we found Cdx1/lacZ fusion protein to be active while Cdx1/luc fusion protein was not. This allowed, in conjunction with use of a sensitive β-galactosidase chemiluminescence assay, lacZ reporters to be readily examined for presence of regulatory elements within Cdx1 introns. lacZ fusion reporters have been a commonly used tool in transgenic mouse studies and so we suggest that the methods now described might have more general applicability for in vitro analysis of constructs already proven in vivo. We show that mouse Cdx1 has both intron and upstream functional RAREs, while chicken Cdx1 has only the intron RARE activity.

**Origins and possible early roles of the Cdx1 RAREs**

Our database searches on the Cdx1 genes of twenty three different vertebrates provide evidence that the intron RARE originated at around the division between amphibiaens and amniotes, whereas the upstream RARE originated later, at around the split between marsupial and eutherian mammals. We now consider anatomical changes occurring at these points in evolutionary history that may have been directed by changes in Cdx1 regulation.
Living amphibians (frogs/toads, salamanders and caecilians) have three features in their neck and anterior thorax that distinguish them from amniotes (Wake, 1979). First, they have a single neck vertebra (atlas). Second, there is absence of an odontoid peg/axis, which in amniotes permits rotational movement at the atlas/axis joint. Third, their sternum is joined to the pectoral girdle but not to the ribs. The extensively reported involvement of both Cdx1 (Gaunt et al., 2008; Subramanian et al., 1995; van den Akker et al., 2002) and RA signalling (Kessel and Gruss, 1991; Lohnes et al., 1994) in development of the vertebræ and ribs in mice leads us to propose that origin of the Cdx1 intron RARE at around the amphibia/amniote split may have contributed to evolutionary changes in the neck and thorax, including an enhanced ability to lift and/or rotate the head. Xenopus Cdx1 (Xcad2) does not respond to retinoid signalling (Shiotsugu et al., 2004). It is not possible to predict whether the Cdx1 intron RARE arose within the early amniotes or in their amphibian ancestors. Multiple neck vertebrae were already present in some Carboniferous amphibians, while the odontoid peg first appeared at around the transition from amphibian to amniote, also in the Carboniferous (Carroll, 2001; Evans, 1939).

Instead, or in addition, it is possible that some other role of Cdx1 may have become facilitated by acquisition of the intron RARE at the amphibia/amniote split. The adult metanephric kidney, for example, first appeared with the amniotes and, as we have shown, this structure may be developmentally regulated by Cdx1. However, the amphibian opisthnonephric kidney already includes a metanephric-like region (Nelson, 1953). There is no evidence that Cdx1 regulates development of extra-embryonic membranes, a defining feature of the amniotes.

At the marsupial/eutherian split there is no consistent accompanying change in the arrangement and numbers of vertebrae (Narita and Kuratani, 2005). The most obvious difference between these mammalian groups lies in the structure and function of their reproductive tracts. In particular, the paired uteri of a marsupial remain unfused and lead separately into a single vaginal sinus. This structure connects laterally with paired vagi-

Many of the urogenital defects that we find in Cdx1 over-expressers (partial or complete agenesis of the uterus and ureter, renal agenesis, hypoplasia and hydronephrosis, and hydroureter) are also seen in mice with double mutations in RA receptor genes (Ghyselinck et al., 1997; Mendelsohn et al., 1994). It might be expected that Cdx1 expression is reduced in these mice, although this is not established. Together, the mouse studies might then indicate that correct dosage of Cdx1 protein is crucial for normal urogenital development, with over- or under-expression both generating similar phenotypes, a finding also made for Pax6 protein (Manuel et al., 2008). In this scenario it is not clear why urogenital defects are not reported in Cdx1 knockout mice. One possibility is overlapping function of Cdx2 (Gaunt et al., 2008; van den Akker et al., 2002), a gene which our over-expressor mice also imply in urogenital development. However, it would then not be clear why Cdx2 cannot compensate in RA receptor mutants. These points might be elucidated by re-examination of Cdx1 knockout mice for urogenital defects, and by measuring Cdx1 protein levels in RA receptor mutants. Over-expression of Cdx4 was not associated with urogenital defects. This might be due to a different pattern of Cdx4 expression in urogenital pro-
genitors, or to a different property of the protein.

A major challenge is to identify the genetic changes responsible for appearance of new animal groups during the course of evolution. To our knowledge, we now provide the first evidence that origins of regulatory elements in a homoeotic gene, Cdx1, correspond with evolutionary transitions between vertebrate groups. The Cdx genes, as upstream regulators of Hox genes, seem ideally suited for effecting large scale evolutionary changes in body plan. It is unlikely that modifications in expression of Cdx1 alone could have resulted in the morphological changes that we have discussed. Cdx1 modification may have been one of several changes, each acquired step-wise and providing some selective advantage. Sexual reproduction will have accelerated the accumulation of changes necessary for large scale morphological evolution.

Materials and Methods

Plasmid

Cdx1: luciferase reporters are shown in Fig. 1A compared with construct #1, the Cdx1/lacZ reporter previously named construct 7 (Gaunt et al., 2003). Construct #2 contains the upstream sequence of Cdx1, extending to 21 bases 5' of ATG, subcloned into pGL3-Basic (promoterless luciferase vector; Promega). Construct #3 has the Cdx1 component of construct #1 spliced in-frame to luciferase. Cdx2/lacZ and Cdx4/lacZ reporters are constructs 5 and 1 described earlier (Gaunt et al., 2005).

Cdx1/lacZ reporters (Fig. 2A) have the second exon of Cdx1 spliced in-frame to lacZ/SV40 polyA and express Cdx1/lacZ fusion proteins. Construct #4 resembles construct #1 but with mutation of the intron RARE from TGAACCTTGGACC to CATATGGCAGACT. Construct #5 resembles construct #1 but with mutation of the upstream RARE from GAAGGTTGTGACACCCCTAA to GAAGGTTGTGACACCCCTA. Construct #6 is similar but with the mutations in both RAREs. Construct #7 resembles construct #5 but the intron RARE is mutated to TGAACTCTGGACT, like that in Anole lizard. Constructs #8 and #9 are chicken Cdx1/lacZ reporters previously named constructs 1 and 5 (Gaunt et al., 2003); construct #9 has the intron RARE mutated from TGAACCTCTGGACC to CTCGAGCCGACT. Construct #10 is similar but with the 6.8kb Xhol/Hindlll fragment (Gaunt et al., 2003) extending the upstream sequence. Mutations were introduced by PCR or by QuikChange in vitro mutagenesis (Stratagene).

Cdx1 reporter expression

P19 cells, from ATCC, were cultured on gelatinized surfaces in Dulbecco’s MEM with glutamine and 10% foetal bovine serum. For reporter assays, cells were grown in 24-well plates. Plasmid DNA was transfected uncut into monolayers using Lipofectamine 2000 (Invitrogen) and following manufacturer’s instructions. Medium was replaced after 5 hours, with or without 2.5μM all-trans retinoic acid (RA) (Sigma). After a further 20 hours, luciferase or β-galactosidase activities were assayed by chemiluminescence. The Luciferase Assay System (Promega) or the Galacto-Light Plus System (Applied Biosystems) were used, respectively. Methods were as instructed by manufacturers. Light levels were
assayed using a LB 96V MicrolumatPlus luminometer.

**Sequence analysis**

RARE and other conserved motifs were identified in database sequences (Accession numbers in supplementary Table 1) visually or by use of the EMBOSS fuzznuc tool. We searched, in both orientations, up to 10kb of upstream sequence and the entire first intron. Sequence alignments were made using Vector NTI (Invitrogen).

**Acknowledgements**

We thank Stephen Brown for help in database analysis, and the BBSRC for financial support.

**References**


Further Related Reading, published previously in the *Int. J. Dev. Biol.*


**cdx4/lacZ and cdx2/lacZ protein gradients formed by decay during gastrulation in the mouse**
Stephen J. Gaunt, Deborah Drage and Richard C. Trubshaw
*Int. J. Dev. Biol.* (2005) 49: 901-908

**Additional enhancer copies, with intact cdx binding sites, anteriorize Hoxa-7/lacZ expression in mouse embryos: evidence in keeping with an instructional cdx gradient**
Stephen J. Gaunt, Adam Cockley and Deborah Drage
*Int. J. Dev. Biol.* (2004) 48: 613-622

**Initiation, establishment and maintenance of Hox gene expression patterns in the mouse**
J Deschamps, E van den Akker, S Forlani, W De Graaff, T Oosterveen, B Roelen and J Roelfsema
*Int. J. Dev. Biol.* (1999) 43: 635-650

**Role of the retinoic acid receptor beta (RARbeta) during mouse development**
N B Ghyselinck, V Dupé, A Dierich, N Messaddeq, J M Garnier, C Rochette-Egly, P Chambon and M Mark
*Int. J. Dev. Biol.* (1997) 41: 425-447

**Evolution of cis-regulatory regions versus codifying regions**
Francisco Rodríguez-Trelles, Rosa Tarrío and Francisco J Ayala
*Int. J. Dev. Biol.* (2003) 47: 665-673

5 yr ISI Impact Factor (2009) = 3.253