Development 125, 3111-3121 (1998)
Printed in Great Britain © The Company of Biologists Limited 1998
DEV2302

Conserved and divergent roles for members of the *Snail* family of transcription factors in the chick and mouse embryo

Mark Sefton, Silvia Sánchez and M. Angela Nieto*

Instituto Cajal, CSIC, Avenida Doctor Arce, 37, 28002 Madrid, Spain
*Author for correspondence (e-mail: anieto@cajal.csic.es)

Accepted 2 June; published on WWW 21 July 1998

SUMMARY

The members of the Snail family of zinc-finger transcription factors have been implicated in the formation of distinct tissues within the developing vertebrate and invertebrate embryo. Two members of this family have been described in higher vertebrates, *Snail* (*Sna*) and *Slug* (*Slu*), where they have been implicated in the formation of tissues such as the mesoderm and the neural crest. We have isolated the mouse homologue of the *Slu* gene enabling us to analyse and compare the amino acid sequences and the patterns of expression of both *Sna* and *Slu* in the chick and mouse. We have detected features in the sequences that allow the unequivocal ascription of any family member to the *Sna* or *Slu* subfamilies and we have observed that, during early stages of development, many of the sites of *Slu* and *Sna* expression in the mouse and chick embryo are swapped. Later in development, the sites of expression of *Slu* and *Sna* are conserved between these two species. These data, together with the data available in other species, lead us to propose that *Slu* and *Sna* arose as a duplication of an ancestor gene and that an extra duplication in the fish lineage has given rise to two *Sna* genes. Furthermore, several early sites of *Slu* and *Sna* expression have been swapped in the avian lineage. Our analysis of the Snail family may also shed new light on the origin of the neural crest.

Key words: *Slug*, *Snail*, Mouse, Chick, Zinc-finger transcription factor, Neural crest, Mesoderm, Gene duplication, Chordate, Vertebrate, Evolution

INTRODUCTION

Much of the progress made in our understanding of developmental processes over recent years has been the result of the identification of genes responsible for mutant phenotypes in the *Drosophila* embryo. The subsequent cloning of vertebrate homologues of such genes and the analysis of their functions has in many cases established the importance of these genes in fundamental developmental events. However, the increasing use of transgenic technology has brought to light the existence of compensatory mechanisms that exist within gene families whereby the absence of a specific family member may be compensated for by another. Such redundancy within gene families, along with the overlapping domains of expression displayed in some cases by distinct members of the same gene family, has given rise to the need to define the roles of the individual members of gene families and to compare their function in distinct species.

The zinc-finger transcription factor, *snail* (*sna*), was initially identified in the *Drosophila* embryo (Grau et al., 1984; Boulay et al., 1987), embryos carrying mutations in *sna* showing defects in mesoderm formation (Alberga et al., 1991). The cloning of vertebrate homologues of *sna* confirmed the possible role of this gene in mesoderm specification in distinct species (Sargent and Bennet, 1990; Nieto et al., 1992; Smith et al., 1992). Further vertebrate members of the Snail family have now been isolated and their expression patterns analysed (Hammerschmidt and Nüsslein-Volhard, 1993; Nieto et al., 1994; Thisse et al., 1993, 1995; Mayor et al., 1995). To date, it appears that, based on sequence analysis, only two members of this family exist in vertebrates, *Snail* (*Sna*) and *Slug* (*Slu*). However, in zebrafish, the two genes isolated have been called *Sna1* (Thisse et al., 1993; Hammerschmidt and Nüsslein-Volhard, 1993) and *Sna2* (Thisse et al., 1995), although the authors mention a higher degree of similarity between *Sna2* and *Slu* (Thisse et al., 1995).

Whilst interest in *Sna* was originally due to its proposed role in mesoderm formation (Sargent and Bennet, 1990; Nieto et al., 1992; Smith et al., 1992), more recently, a *Sna*-related chick gene has been implicated in the control of vertebrate left-right asymmetry (Isaac et al., 1997). *Slu* has also generated a great deal of interest since it was first identified in the chick embryo (Nieto et al., 1994). *Slu* is a useful marker for premigratory neural crest (Nieto et al., 1994; Mayor et al., 1995) as well as being critical for the emigration of the neural crest from the neural tube and of the early mesoderm from the primitive streak. Indeed, this property led to the suggestion that c*Slu* may be required to release cells from epithelial structures permitting them to migrate, a process known as the epithelial-to-mesenchyme transition (EMT, Nieto et al., 1994). Despite the

importance of these two genes during chick development, the data available regarding their patterns of expression during development are limited to discrete tissues.

Recently, a mouse homologue of *Slu* has been identified from a murine cell line and employed to further demonstrate the role of *Slu* in driving EMT in tissue culture cells (Savagner et al., 1997). However, no expression studies of this gene have been carried out during development. In addition, we have isolated m*Slu* from embryonic mouse tissue and analysed its pattern of expression in detail. We have completed the expression studies of the two chick genes, enabling us to compare the expression pattern of the two vertebrate members of the Snail family in both chick and mouse.

As a result of this analysis, we have observed that, at various sites during early development, the expression of *Slu* and *Sna* is inverted between chick and mouse, notably in the neural crest and the mesoderm. Other sites of expression for each of the genes are conserved between the two species. When taken together, the combined expression sites for the chick and mouse Snail family members is the same in each species. Together with these expression studies, sequence comparison analysis has allowed us to identify residues that are characteristic of either Sna or Slu proteins and thus, diagnostic for the identification of subfamily members. Finally, the comparison of both sequence and expression patterns of this gene family from *Drosophila* to mammals has allowed us to propose how Snail proteins may have evolved. Indeed, our data may shed new light on the origin of tissues of evolutionary significance such as the neural crest.

MATERIALS AND METHODS

Embryos

The embryos used throughout this study were obtained from natural matings of Balb-C (Harlan) and SJL mouse strains (Jackson), and White Leghorn chickens (Granja Rodriguez-Serrano, Salamanca, Spain). Chick eggs were incubated and opened, and the embryos staged according to Hamburger and Hamilton (1951). For mouse embryos, ages were determined as days post-coitum (d.p.c.), the day on which the vaginal plug was detected being designated 0.5 d.p.c.

Isolation of m*Slu* by PCR

Degenerate primers were designed, based upon the first and last 6 amino acid sequences of the chick and *Xenopus* Slu proteins (Nieto et al., 1994; Mayor et al., 1995), to amplify cDNA isolated from 9.5 d.p.c. mouse embryos. A fragment that approximated in size to the full-length chick and *Xenopus* cDNAs was cloned into the pGEM T-vector (Promega). When sequenced, this cDNA showed a high degree of similarity to the chick sequence and a predicted 92% identity at the amino acid level.

In situ hybridisation

Whole-mount in situ hybridisation was carried out in chick and mouse embryos at various stages of development as previously described (Nieto et al., 1996). Digoxigenin-labelled probes were synthesized from the complete coding sequence of m*Slu* or fragments of the cDNAs corresponding to c*Slu*, m*Sna* and c*Sna* as follows: m*Slu*, nucleotides 1-807 (complete coding sequence); c*Slu*, nucleotides 1-360; m*Sna*, nucleotides 433-824; c*Sna*, nucleotides 258-767. Following hybridisation, the embryos were embedded in fibrowax and sectioned at 15 μm.

RESULTS

Sequence analysis of members of the Snail family

We have cloned the coding region of the mouse *Slu* gene enabling us to compare the sequences and patterns of expression of the two members of the Snail family of zinc-finger transcription factors, *Sna* and *Slu*, during early stages of chick and mouse development. The m*Slu* sequence is identical to that cloned by Savagner et al. (1997) and encodes a protein that shows 92% identity to the predicted cSlu protein and 65% identity to cSna protein (Table 1). Comparing the amino acid sequence of the vertebrate Slu and Sna proteins, several points become immediately obvious (Fig. 1). Of the eight vertebrate Snail family members isolated to date, three of them showed very little divergence between species and corresponded to mouse, chick and *Xenopus* Slu proteins (mSlu and xSlu show 92% and 91% identity to cSlu, respectively). The conservation is particularly notable in the zinc-finger domain where mSlu and xSlu show 99% and 98% identity to cSlu respectively, across the last four fingers (Table 1). The remaining five proteins correspond to the Sna subfamily, which similarly show a high degree of conservation in the zinc-finger domain, although less than that shown by Slu proteins (between 81% and 95% identity when compared to cSna). However, they are much more variable in the amino terminal portion of the protein (between 40% and 55% identity when compared to cSna).

When we analysed the amino terminal portion of the protein in greater detail, we encountered information that may prove critical for the identification of distinct family members and that may provide clues as to the evolution of this protein family. We identified a stretch of 29 amino acids immediately

Table 1. A comparison of the identity between vertebrate members of the Snail family

Gene	No. of fingers	Total		Four fingers		Non-fingers	
		cSlu	cSna	cSlu	cSna	cSlu	cSna
c*Slu*	5	100	68	100	90	100	56
m*Slu*	5	92	65	99	91	84	41
x*Slu*	5	91	65	98	92	84	40
c*Sna*	5	68	100	90	100	56	100
m*Sna*	4	50	58	81	85	29	40
x*Sna*	5	69	75	93	95	48	55
z*Sna1*	4	47	54	81	81	38	41
z*Sna2*	5	53	57	83	83	32	41

The percentage identity of each vertebrate member of the Snail family is shown with respect to the chick Slu (left) and Sna (right) proteins. The number of fingers contained within each gene is shown in the first column. We have calculated the identity that exists across the whole protein (Total) and in the last four fingers only (Four fingers). We restricted our analysis to the last four fingers in order to compensate for the proteins in which the first finger is absent. For the same reason we have compared each vertebrate protein with the chick proteins. We also calculated the identity in the 5′ region, that is from the methionine at the amino terminal of the protein to the amino acid immediately preceding the first zinc-finger domain or the corresponding amino acid in proteins containing only four fingers. The data presented show that each of the vertebrate proteins shows a higher degree of identity to its chick homologue than to the other chick members of the family. This is true in both the 5′ and finger region, as well as in the protein as a whole. Moreover, the Slu homologues show a greater degree of identity between themselves than is seen for Sna homologues. The accession numbers are indicated in the legend to Fig. 1.

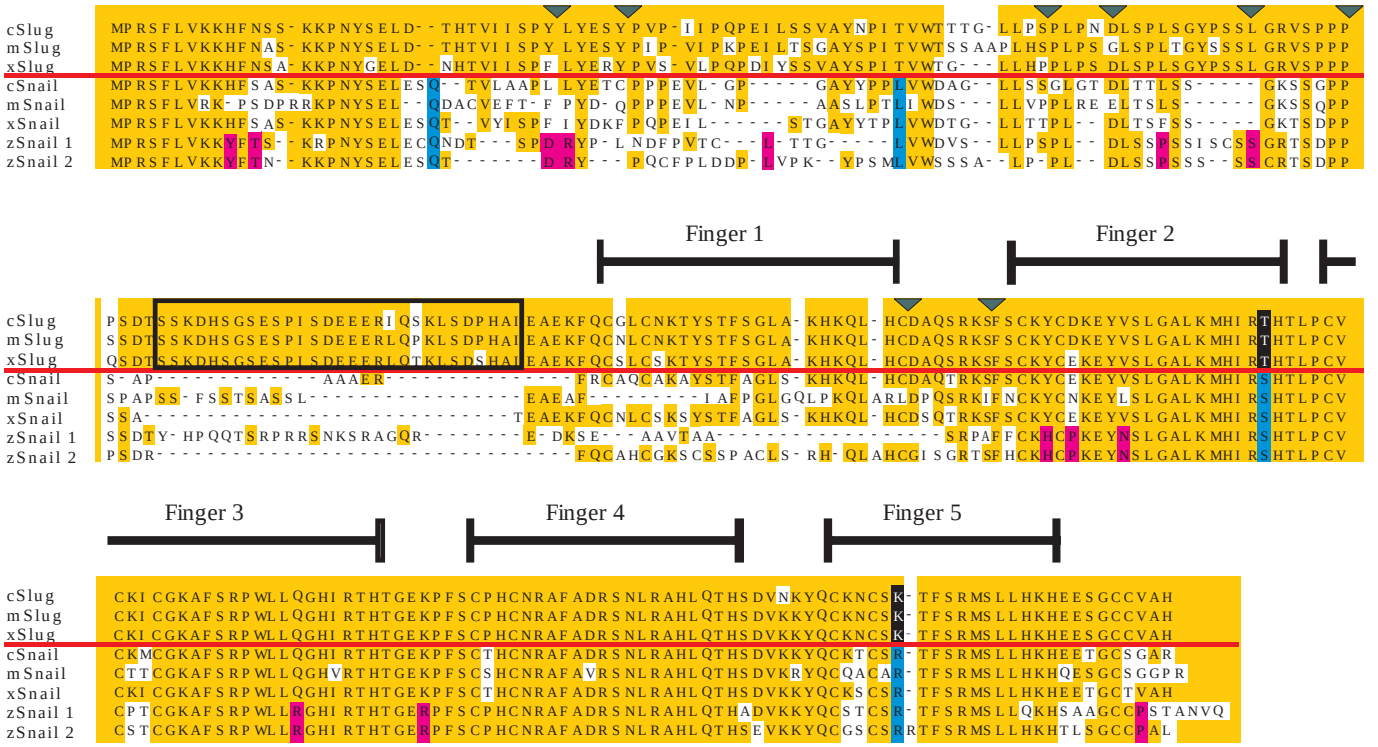

Fig. 1. Sequence comparison of the vertebrate homologues of the Snail family of zinc-finger transcription factors. The amino acid sequences of the three vertebrate Slu proteins in mouse, chick and *Xenopus* are compared with the vertebrate Sna proteins from mouse, chick, *Xenopus* and zebrafish. The grey arrowheads above the sequences indicate the sites where stretches of amino acids in Sna proteins that do not correspond to sequences within the Slu proteins have been excluded. Amino acids that correspond to the consensus of the Slu proteins are shaded in yellow and those that do not in white. Where no consensus exists between Slu homologues, all amino acids are shaded in white. The zinc-finger domains are indicated in the figure from the first cystidine of the finger domain to the last histidine. The box identifies a 29 amino acid stretch that is highly conserved, and that we consider diagnostic for Slu homologues. The amino acids shaded in black are those in the zinc-finger region that are conserved between ascidian and vertebrate Slu proteins and are different from those in Sna proteins. The amino acids that are distinct in Sna from Slu but that are conserved in all vertebrate Sna proteins are shaded in blue and those specific to the two zebrafish homologues are shaded in red. The accession numbers for the sequences shown are: c*Slu*, X77572; m*Slu*, U97059; x*Slu*, X80269; c*Sna*, Y09905; m*Sna*, X67253; x*Sna*, X53450; z*Sna* 1, X74790; z*Sna* 2, U24225.

preceding the zinc-finger domain, which is exclusive to and highly conserved in vertebrate Slu proteins (boxed in Fig. 1). This sequence may be diagnostic for vertebrate Slu proteins. Furthermore, the vertebrate Sna proteins appear to contain several amino acid stretches of variable size at conserved positions that are absent in the Slu protein sequences (grey arrowheads in Fig. 1). The identity between each of these short sequences is not sufficiently conserved as to enable us to consider these stretches of amino acids diagnostic for Sna proteins. However, we have identified a few amino acids that are distinct in Sna from Slu but that are conserved in all Sna proteins (shaded in blue in Fig. 1). Taking all this data together, it became evident that the two genes isolated from zebrafish embryos belong to the Sna subgroup.

When compared to the three members of the Snail family in *Drosophila*, *sna* (Boulay et al., 1987), *escargot* (*esg*, Whiteley et al., 1992) and *scratch* (*scr*, Roark et al., 1995), it appears that all vertebrate members of the family show a slightly greater degree of identity to the product of *esg* (between 50 and 59% identity) than to that of *sna* (between 44 and 55% identity), *scr* being the most distant relative of the *Drosophila* genes (Table 2). The region 5′ to the zinc-finger domains in each of the *Drosophila* proteins is notably larger than that of the vertebrate homologues as is the case for the only Snail family homologues described in sea urchin and ascidians (Illingworth et al., 1992; Corbo et al., 1997). The vertebrate proteins of the Sna and Slu subfamilies contain sequences in this 5′ region that are specific to both esg or sna. These data indicate that the vertebrate and *Drosophila* genes arise from a common progenitor.

Analysis of the expression of Snail family members in the neural crest

We have determined the distribution of m*Slu* transcripts at various stages of development by in situ hybridisation and compared the distribution of both *Slu* and *Sna* transcripts in equivalent developmental stages of chick and mouse development. The whole-mount in situs comparing the sites of expression between these genes are shown in Figs 2 and 3. A more detailed analysis with respect to their expression in specific tissues is presented in the remaining figures.

The *Slu* gene was first described in the chick embryo where it is expressed at high levels in both the premigratory and migratory neural crest (Nieto et al., 1994). Indeed, loss-of-function experiments indicated that this gene is critical for the emigration of the neural crest from the neural tube. For this

Table 2. A comparison of the identity between vertebrate members of the Snail family and two *Drosophila* members of the family, sna and esg

Gene	No. of fingers	Total		Four fingers		Non-fingers	
		cSlu	cSna	cSlu	cSna	cSlu	cSna
		sna	esg	sna	esg	sna	esg
c*Slu*	5	55	57	75	84	30	30
m*Slu*	5	50	57	74	84	30	34
x*Slu*	5	47	57	73	84	26	33
c*Sna*	5	53	55	69	83	37	27
m*Sna*	4	44	52	67	77	32	41
x*Sna*	5	48	59	68	83	37	37
z*Sna1*	4	45	50	70	74	32	37
z*Sna2*	5	46	52	67	77	39	34

When we compared the similarity between the vertebrate members of the Snail family and the two family members from *Drosophila*, sna and esg, a slightly greater identity to esg was observed. This was true for all vertebrate proteins and in all regions of the protein except for the non-finger domain of zSna2 and cSna. The accession numbers are indicated in the legend to Fig. 1 and those corresponding to the *Drosophila* proteins are sna, Y00288; esg, M83207.

reason, we first analysed the expression of m*Slu* in the cephalic neural crest, and compared it with the expression of *Sna* in both chick and mouse embryos. In the chick, strong expression of c*Slu* in the premigratory and migratory crest was observed at all stages examined (stage 8-18: Figs 2C,E, 3A,C, 4A,C,K). However, when we analysed the expression of m*Slu* in embryos at 8.5 and 9.5 d.p.c., transcripts were observed in migrating neural crest cells but not in the premigratory neural crest (Figs 2I,K, 3E, 4E,G,I). In contrast, the expression of *Sna* in the mouse neural crest appeared to be very similar to that observed for c*Slu*, both in premigratory (compare Fig. 2J,L with C,E and Fig. 4H with A) and migratory cephalic neural crests cells (compare Fig. 4J with C). Transcripts of m*Sna* were expressed in a greater number of migratory crest cells than m*Slu* (Fig. 4I,J).

Conversely, c*Sna* transcripts were not observed in the chick premigratory neural crest (Figs 2D,F, 4B,D) or in the migratory crest cells at early stages (Fig. 2F). Up to stage 13, c*Sna* transcripts were completely absent in migratory hindbrain crest cells (Fig. 3B) but c*Sna* expression was detected in a subpopulation of migratory hindbrain neural crest from stage 14 (Fig. 4D). At stage 18, the majority of both *Slu*- and *Sna*-expressing crest had arrived at their destination although expression of both genes could still be detected in some migrating cells (Fig. 3C,D). This could be clearly seen in the hindbrain-derived crest that migrates to the branchial arches. It is interesting to note that the number of crest cells that express c*Slu* is significantly greater than those expressing c*Sna* (Fig. 4C,D,K,L). Thus, it appears that, in the neural crest, m*Sna* expression is very similar to that of c*Slu*, and the expression of c*Sna* and m*Slu* is also very similar to each other.

The expression of Snail family members in the primitive streak and early mesoderm

In the chick embryo, both *Slu* and *Sna* are expressed from early on in development. At stage 5, c*Slu* transcripts were restricted to the primitive streak and the ingressing mesodermal cells, whilst c*Sna* transcripts were absent from these cells (Nieto et al., 1994; Isaac et al., 1997; Fig. 2A,B). At an equivalent stage of mouse postimplantation development, no m*Slu* transcripts were detected in 7.5 d.p.c. embryonic tissues (Fig. 2G). In contrast, m*Sna* expression is particularly strong in the mesoderm as it migrates from the primitive streak (Nieto et al., 1992; Smith et al., 1992; Fig. 2H). The expression of m*Sna* and the absence of m*Slu* transcripts both in the primitive streak and early mesoderm were also observed in mouse embryos at 8.5 d.p.c. (Fig. 5E,F). Similarly, in the posterior region of stage 8

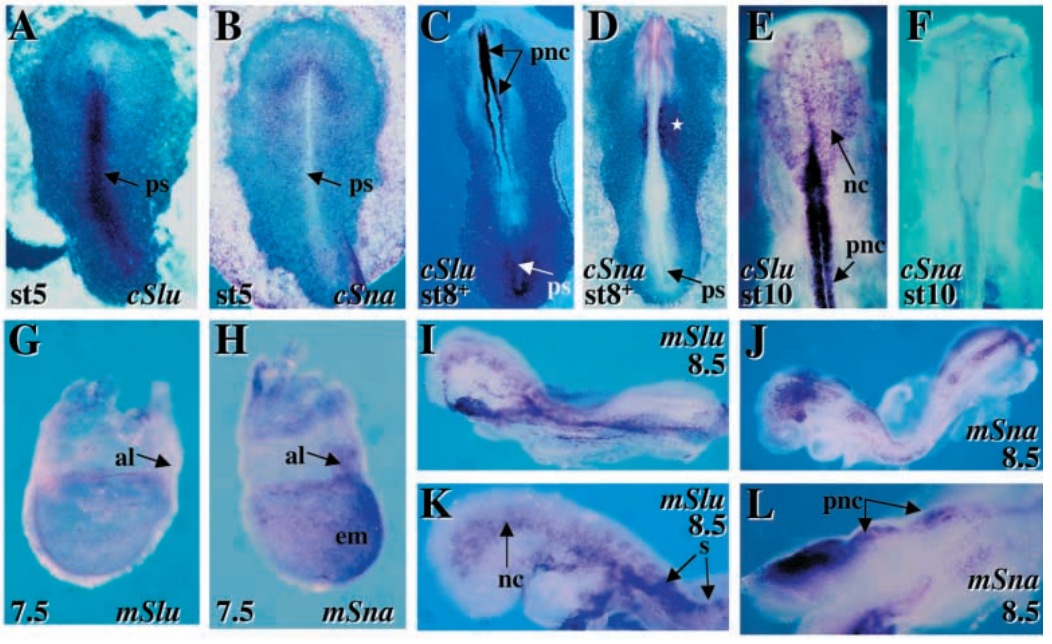

Fig. 2. Whole-mount in situ hybridisations showing the early expression of members of the Snail family in chick and mouse embryos. In all the figures, whole-mount hybridisations in chick embryos are labelled with the gene name preceded by a 'c' (c*Slu* A,C,E and c*Sna* B,D,F) and in mouse embryos by a 'm' (m*Slu* G,I,K and m*Sna* H,J,L). The developmental stages of each embryo is shown, according to Hamburger and Hamilton (1951) in the case of chick embryos and in days post coitum (d.p.c.) for mouse embryos. In the figure, hybridisations of equivalent developmental stages of chick (stages 5-10; A-F) and mouse embryos (7.5 and 8.5 d.p.c.; G-L) are presented. Note the differences of expression between embryos in the primitive streak, early mesoderm, and neural crest. The star in the whole mount of a stage 8+ chick embryo indicates the asymmetric left-right expression of *Snail* in the lateral plate mesoderm (star in D). al, allantois; em, early mesoderm; nc, neural crest; pnc, premigratory neural crest; ps, primitive streak; s, somites.

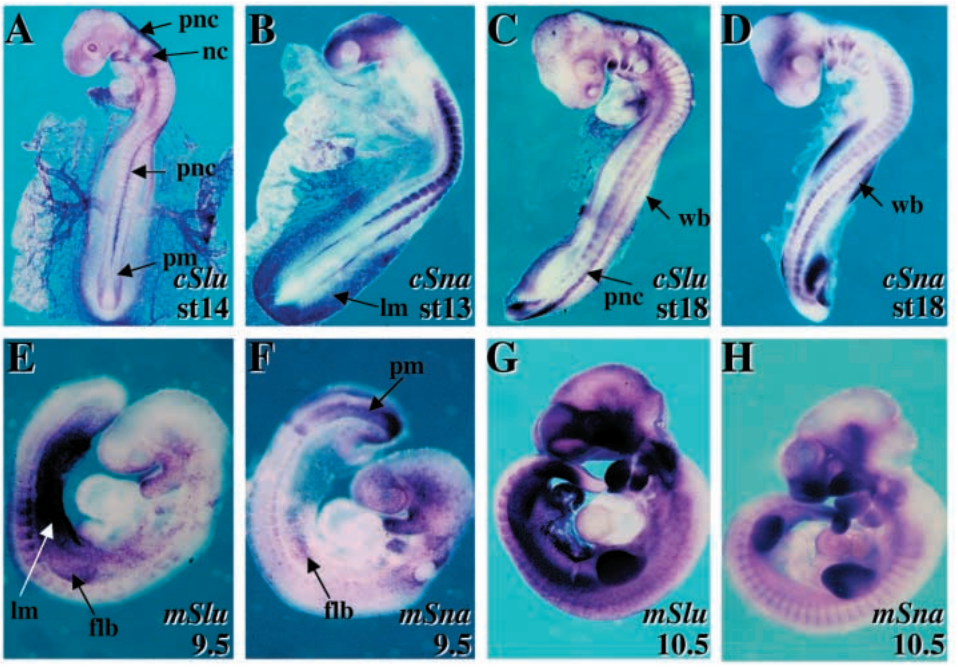

Fig. 3. Whole-mount in situ hybridisations showing the expression of the Snail family members in chick and mouse at later stages of development. The whole-mount hybridisations are labelled as outlined in Fig. 2. Note the complementary expression of *Slu* and *Sna* in the paraxial and lateral mesoderm between chick embryos and mouse embryos. As well as the complementary expression between the same species (A,B and E,F), an inversion in the sites of expression exists between chick and mouse. flb, forelimb bud; lm, lateral mesoderm; nc, neural crest; pnc, premigratory neural crest; pm, paraxial mesoderm; wb, wing bud.

chick embryos, the expression of c*Slu* and the absence of c*Sna* transcripts in these tissues was readily appreciated both in whole mounts and in sections (Fig. 5A-D). As described above for the neural crest, *Sna* expression in the primitive streak and early mesoderm in the mouse is similar to that of *Slu* in the chick.

Mesodermal expression of *Slu* and *Sna*

As development proceeds and the mesoderm segregates into distinct populations, *Snail* family members are expressed in a more complex fashion reflecting these processes. m*Slu* expression was detected exclusively in the lateral mesoderm (Figs 3E, 6E,G) whereas, in contrast, m*Sna* transcripts were observed in the paraxial mesoderm (Figs 3F, 6F,H).

The pattern of mesoderm expression of these genes in the chick was somewhat more complex. In whole-mount chick embryos at stage 13, the inverted pattern of expression of both c*Slu* and c*Sna* with respect to their mouse counterparts can be readily appreciated (compare Fig. 6A,B with E,F). c*Slu* transcripts were restricted to the paraxial mesoderm in caudal regions, whereas c*Sna* transcripts were detected in lateral plate mesoderm (Fig. 6A,B). However, in more rostral regions but

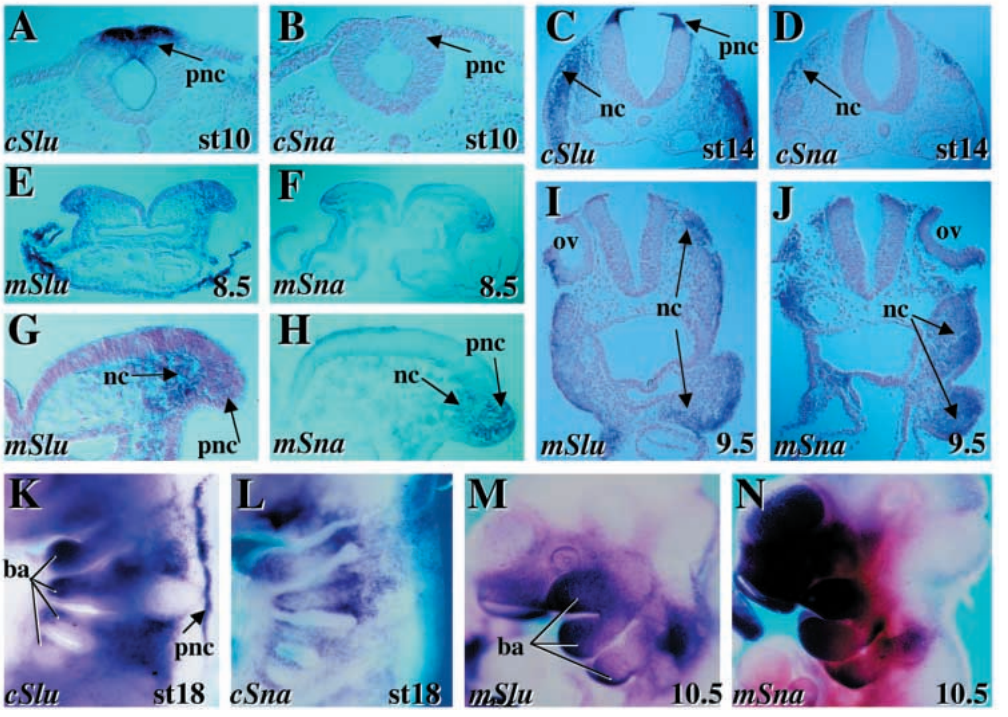

Fig. 4. Expression of chick and mouse members of the *Snail* family in the cephalic neural crest and its derivatives. Transverse paraffin sections at the level of the hindbrain of whole-mount in situs showing the expression of *Slu* and *Sna* in the premigratory and migratory neural crest (A-J). Note the expression of c*Slu* and m*Sna* in the premigratory neural crest (A,F,H) and at later stages, the expression of both *Slu* and *Sna* in the migratory crest in both species (C,D,I,J). Also note the difference in the number of migratory neural crest cells expressing *Sna* or *Slu*. G and H are high-power images of the sections in E and F to show more clearly the presence of *Sna* but not *Slu* transcripts in mouse premigratory neural crest. The lower panels show high-power images of the branchial arches in chick (K,L) and mouse embryos (M,N). Note the expression of both genes in derivatives of the neural crest within the branchial arches. ba, branchial arches; nc, neural crest; ov, otic vesicle; pnc, premigratory neural crest.

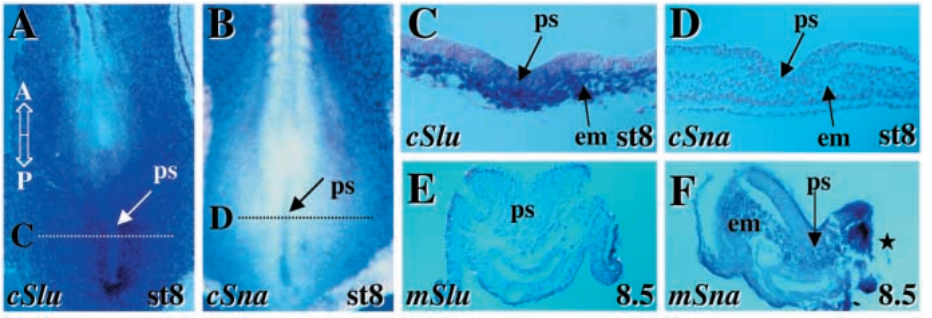

Fig. 5. Expression of chick and mouse members of the Snail family in the early mesoderm. High-power images of the posterior region of whole-mount in situs highlight the expression c*Slu* (A) and the absence of c*Sna* (B) in the primitive streak and early mesoderm of stage 8 chick embryos. This is more clearly seen in transverse paraffin sections of these whole mounts, taken at the level indicated by the dotted line, as shown in C and D. The equivalent paraffin sections below demonstrate the inverted expression of *Slu* (E) and *Sna* (F) in the early mesoderm and primitive streak of mouse embryos. The star in F emphasises the asymmetric left-right expression of *Sna* in the mesoderm. The open arrows in A indicate the orientation of the anteroposterior axis in A and B. em, early mesoderm; ps, primitive streak.

caudal to the last-formed somite, more complicated patterns of expression were observed. In the paraxial mesoderm, c*Slu* continued to be expressed in the absence of c*Sna* transcripts (Fig. 6C,D). However, in the lateral mesoderm, both genes appeared to be expressed, although in complementary domains (Fig. 6C,D). The more medial lateral mesoderm cells expressed c*Slu* but not c*Sna* transcripts, whereas, in the more lateral regions of the lateral plate mesoderm, cells do not express c*Slu* but c*Sna* transcripts were detected. These domains of expression appear to be complementary. This applied both to the somatic and splanchnic mesoderm although the boundary between *Slu*- and *Sna*-expressing cells is positioned at different points along the mediolateral axis in each tissue.

A distinct domain of expression in the paraxial mesoderm was observed immediately prior to the last-formed somite, which is separated from the rest of the paraxial mesoderm by a non-expressing domain. This transitory expression domain could be identified in whole mounts of mouse embryos at 8.5 and 9.5 d.p.c. labelled for *Sna* (Figs 2J, 6F) or conversely in stage 13 chick embryos labelled for *Slu* (Fig. 6A).

Somitic expression of *Slu* and *Sna*

The somites form as a result of the segmentation and epithelial transformation of the paraxial mesoderm. Each somite subsequently subdivides into distinct domains that differentiate and give rise to different tissues. Both *Slu* and *Sna* are expressed within the somites, their patterns of expression changing depending on the differentiated state of the somite. In the early somites, m*Slu* transcripts were detected from as early as 8.5 d.p.c. in the whole of the somite (Figs 2I,K, 7B,G). This expression pattern was similar to that observed for c*Sna*, which was also expressed ubiquitously across the somite between stage 8 and 14 (Isaac et al., 1997; Figs 2D, 3B, 7C,F). The expression of both c*Slu* and m*Sna* was confined to cells situated ventrally in the somites at these early stages of somite development, most probably cells undergoing EMT (Fig. 7A,E,D,H). By 9.5 d.p.c., m*Slu* was detected in sites corresponding to the rostral halves of the somites (Fig. 3E,G). This corresponded to the neural crest cells migrating across the sclerotome as was confirmed in sections of the embryos (Fig. 7K). At these stages, m*Sna* transcripts were detected in the neural crest and in the cells of the somite proper, being excluded from the dermatome whilst being expressed in the myotome and sclerotome (Nieto et al., 1992; Smith et al., 1992; Fig. 7L). The distribution of both *Sna* and *Slu* transcripts in somites at these stages were conserved in chick embryos of an equivalent age, c*Sna* being expressed in the myotome and sclerotome (Fig. 7J), whilst

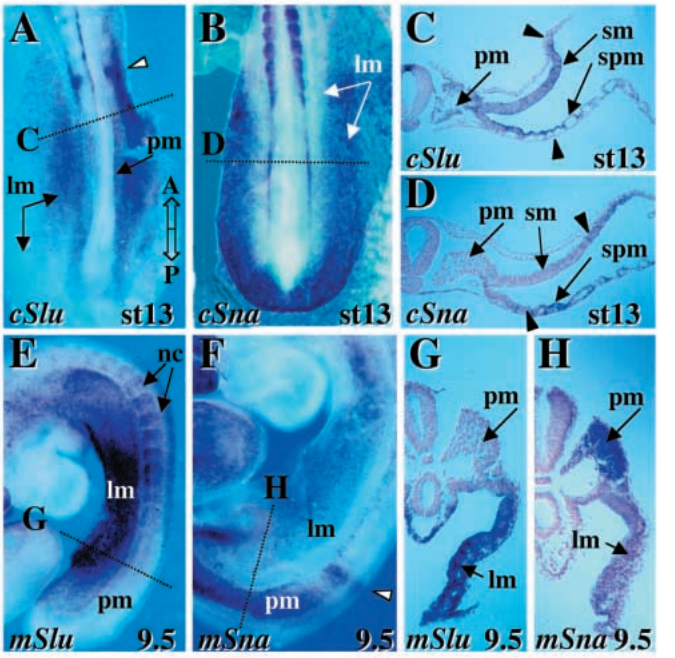

Fig. 6. Expression of chick and mouse members of the Snail family during the development of the mesoderm. High-power images are shown of the posterior halves of whole-mount in situs of chick (A,B) and mouse embryos (E,F) to demonstrate the complementary expression of *Slu* and *Sna* in the lateral and paraxial mesoderm in both species. Also note the inversion of the patterns of expression in the mesoderm between chick and mouse for both *Slu* and *Sna*. A separate domain of expression in the paraxial mesoderm of c*Slu* and m*Sna* is indicated by the white arrowhead in A and F, respectively. Transverse paraffin sections of the whole mounts in A,B and E,F at the level indicated by the dotted lines are shown in C,D and G,H. The sections highlight the complementary patterns of expression of the two genes within each species and the inversion in the expression of each gene between species. The arrowheads in the somatic and splanchnic mesoderm in C and D indicate the boundary along the mediolateral axis of the lateral mesoderm for c*Slu* and c*Sna* expression. The open arrows in A indicate the orientation of the anteroposterior axis in A and B. lm, lateral mesoderm; pm, paraxial mesoderm; sm, somatic mesoderm; spm, splanchnic mesoderm.

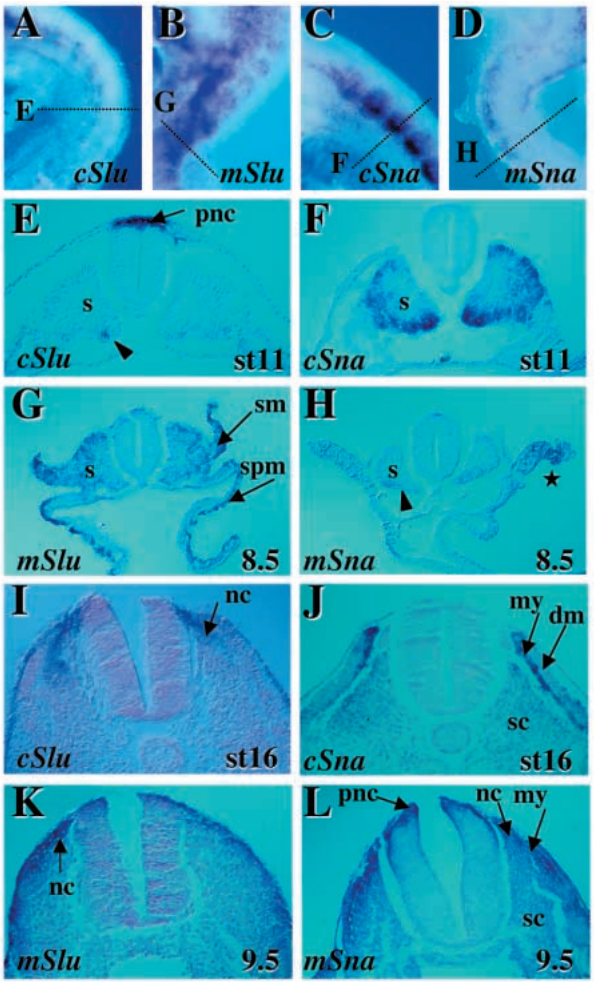

Fig. 7. Expression of chick and mouse members of the Snail family during somite development. High-power images at the level of the somites of whole-mount in situs of *Slu* (A,B) and *Sna* (C,D) in chick (A,C) and mouse embryos (B,D). In transverse paraffin sections of these whole mounts at the level indicated by the dotted lines, the expression of c*Slu* (E) and m*Sna* (H) restricted to cells in the ventral domain of the early somite can be clearly appreciated. The ubiquitous expression of m*Slu* (G) and c*Sna* (F) in the early somite is also shown. The arrowheads in E and H indicate the ventrally located cells within the somites undergoing a transition from mesenchymal to epithelial cells. In H the asymmetric left-right expression of c*Sna* in the lateral plate mesoderm is again indicated by a star. The transverse paraffin sections below show the expression of *Slu* in the trunk neural crest emigrating from the neural tube in both chick (I) and mouse (K). The expression of *Sna* in the myotome and sclerotome in both chick and mouse embryos can be observed in J and L, respectively. dm, dermatome; my, myotome; nc, neural crest; pnc, premigratory neural crest; s, somite; sc, sclerotome; sm, somatic mesoderm; spm, splanchnic mesoderm .

c*Slu* transcripts were only observed in migratory trunk neural crest cells (Fig. 7I).

Slu and *Sna* expression during limb bud development

We have previously described a highly dynamic pattern of expression for c*Slu* in the developing limb bud (Ros et al., 1997) and we therefore investigated whether this pattern of expression might be conserved in the mouse. The expression of m*Slu* and m*Sna* in the limb bud primordium was observed from 9.5 d.p.c. (Fig. 3E,F). As the limb bud developed, m*Slu* transcripts were observed in the ventral and dorsal ectoderm of the limb bud, as well as in the forelimb mesenchyme (Fig. 8G,H). This pattern of m*Slu* expression in the limb bud is very similar to that seen in the chick (compare Fig. 8C,D with G,H). However, the principal difference was the presence of m*Slu* transcripts in the ventral ectoderm of the limb bud where c*Slu* transcripts were not detected at these stages. Transcripts of m*Sna* were observed in a more extensive domain of the limb bud mesenchyme whilst they appeared to be excluded from the limb bud ectoderm (Fig. 8E,F). Transcripts of c*Sna* were essentially expressed in a similar extensive mesenchymal domain to that observed in the mouse and were excluded from the limb bud ectoderm (Fig. 8A,B). *Sna* transcripts were detected before the appearance of *Slu* expression (see Fig. 3). A further observation was the presence of m*Slu* but not m*Sna* transcripts in the apical ectodermal ridge (AER; Fig. 8H), whilst neither gene was expressed in the AER of the chick limb

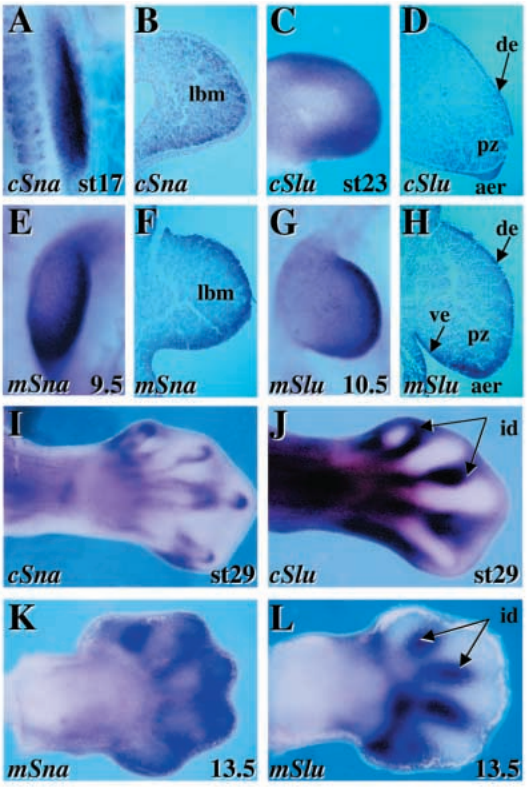

Fig. 8. Expression of chick and mouse members of the Snail family during limb development. The high-power images of whole-mount in situs of limb buds (A,C,E,G) and the corresponding longitudinal sections (B,D,F,H) show the expression of *Sna* and *Slu* in the early limb bud. The expression of *Sna* in much of the limb bud mesenchyme is conserved between chick (B) and mouse (F). Similarly, the more restricted expression of *Slu* in the progress zone was also observed in both species (D,H). In the lower half of the figure (I-L), the conservation of expression between chick and mouse at later stages of limb development is also shown. aer, apical ectodermal ridge; de, dorsal ectoderm; id, interdigital area; lbm, limb bud mesenchyme; pz, progress zone; ve, ventral ectoderm.

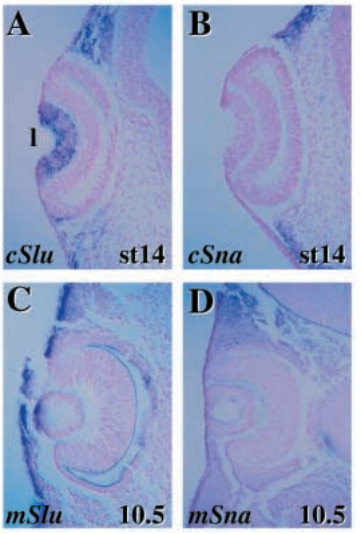

Fig. 9. Expression of chick and mouse members of the Snail family in the developing lenses. Parasagittal paraffin sections highlight the expression of *Slu* in the developing chick (A) and mouse (C) lens. The adjacent chick (B) and mouse (D) sections demonstrate the absence of *Sna* transcripts in the lens or other parts of the eye at equivalent stages of development. l, lens.

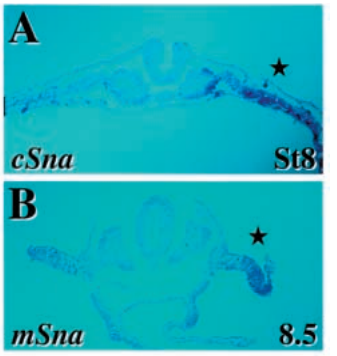

Fig. 10. The transient left-right asymmetry of *Sna* expression is conserved in the chick and mouse. Transverse paraffin sections show the asymmetric distribution of expression of *Sna* transcripts in the lateral mesoderm of a stage 8 chick (A) and 8.5 d.p.c. mouse (B) embryo. The star emphasises the higher levels of expression on the right hand side.

bud. Moreover, neither of these genes were observed to be restricted to the zone of polarising activity in the mouse as has been reported during the early stages of c*Slu* expression in the limb bud (Ros et al., 1997; Buxton et al., 1997).

At later stages of forelimb development, a similar distribution of m*Slu* transcripts was observed at 13.5 d.p.c. to that observed at HH stage 32 in the chick limb (Fig. 8J,L), *Slu* transcripts being restricted to the interdigital regions of the limb. A conservation in the pattern of expression was also observed for *Sna* in both mouse and chick limbs (Fig. 8I,K).

Further observations

In our examination of the expression pattern of mouse *Slu*, we observed transcripts in the lens of the developing eye at 10.5 d.p.c. (Fig. 9C). This expression in the lens was conserved in chick embryos between stage 14 and stage 18 (Fig. 9A). We were unable to detect any mouse or chick *Sna*-expressing cells in the eye (Fig. 9B,D).

Another facet of *Sna* and *Slu* expression that seems to have been conserved between these two species is the asymmetric left-right expression of *Sna* described in the chick, and thought to be related to the establishment of left-right asymmetry (Isaac et al., 1997). The transient asymmetric expression of *Sna* in the lateral mesoderm was observed in both mouse and chick embryos, expression being notably higher in the right side of the embryo (Figs 2D, 10). We were unable to detect an asymmetric distribution of *Slu* transcripts in any tissue at any of the stages examined.

DISCUSSION

The Snail family of zinc-finger transcription factors has been implicated in the formation of distinct tissues during the early development of the vertebrate embryo. We have cloned the mouse homologue of *Slu*, and compared its sequence and expression with that of the other mouse and chick Snail family members.

Sequence comparisons of the vertebrate Snail family members

Comparing the amino acid sequences of the existing vertebrate members of the Snail family, it is apparent that they are best grouped into two subfamilies corresponding to the *Sna* and *Slu* genes already described (Fig. 1). The predicted Slu proteins show a degree of identity across their whole sequence considerably higher than that between Sna proteins, suggesting a greater evolutionary divergence among the latter. This divergence is particularly evident in the 5′ region of the Sna proteins, the zinc-finger domains of all family members being highly conserved. Members of the Slu subfamily contain a highly conserved stretch of 29 amino acids 5′ to the zinc-finger domains that we consider diagnostic for Slu proteins. Indeed, this sequence is absent from the vertebrate Sna proteins and, interestingly, is also absent from the two Snail family members isolated from zebrafish (Hammerschmidt and Nüsslein-Volhard, 1993; Thisse et al., 1993, 1995). The absence of this sequence, together with the presence of several amino acid stretches at conserved positions in Sna but not in Slu homologues, may be diagnostic for Sna proteins. In summary, we have detected diagnostic hints that permit the immediate identification of members of the Sna or Slu subfamily. Thus, with respect to the existing Snail family members, we conclude that mouse, chick and *Xenopus* contain one *Sna* and one *Slu* gene, whereas zebrafish has two *Sna* genes.

A peculiarity of the Snail family of proteins identified in distinct vertebrates is the variability in the number of zinc-finger domains that they contain. Whilst all Slu proteins identified contain 5 fingers, two of the five vertebrate Sna proteins, mSna and zSna1, contain only 4 fingers. It has been suggested that there might be a certain relationship between the number of fingers and the sites of expression that could reflect a subdivision for the family members (Thisse et al., 1995). The information that we have compiled in conjunction with the expression patterns of these genes indicates that this is not the case. The absence of the first finger in mSna and

zSna1 brings into question its functional significance. With respect to this, the zinc-finger protein GL1 contains 5-finger domains of a similar structure to that of the Snail family, and it has been shown that the first of these domains does not interact with DNA (Pavletich and Pabo, 1993).

The data that we present here also enables us to clarify the confusion over the identity of certain genes within the family, for example that of the chick *Snail*-related gene (c*SnR*). Despite the high degree of identity at the amino acid level to the existing vertebrate *Sna* genes, it remained unclear, owing to differences observed by the authors in the patterns of expression between c*SnR* and m*Sna* at early stages, as to whether this gene represented the true c*Sna* homologue. We show that, structurally speaking, c*SnR* belongs to the *Sna* subfamily and that the combined sites of expression of c*SnR* and c*Slu* are the same as those identified in other vertebrate species for the Snail family members. On this basis, we consider it correct to assume that the c*SnR* does indeed represent the true chick *Sna* homologue, *cSna*.

The evolution of the Snail family

In *Drosophila*, three members of the Snail family have been identified: s*na*, e*sg* and s*cr* (Boulay et al., 1987; Whiteley et al., 1992; Roark et al., 1995). Whilst s*na* appears to be functionally more closely related to the vertebrate genes, e*sg* shows a slightly greater sequence similarity to both vertebrate S*na* and *Slu*. Short stretches of amino acids that are found independently in sna or esg correspond to amino acids in the 5′ region of the vertebrate genes. Our interpretation of these sequence comparisons is that the *Drosophila* and vertebrate genes have descended from a common progenitor. It is possible that the duplication giving rise to the two vertebrate genes might have occurred at the described major phase of gene duplication at the origin of vertebrates (Holland et al., 1994). In support of this idea, only a single *Sna*/*Slu* homologue has been identified in the sea urchin (Illingworth et al., 1992) and ascidians (Corbo et al., 1997). The predicted protein sequence of the sea urchin gene is clearly identifiable as a member of this family owing to the absolute conservation of the first 9 amino acids and the high degree of conservation within the five zinc-finger domains. The conserved diagnostic sequence in Slu is not present in the sea urchin or in the ascidian protein, suggesting that this sequence may have arisen later in the Slu subfamily. The conservation of certain amino acids specific to the Slu protein in the zinc-finger domains is evidence that both these proteins also show a similarity to Slu.

Inversions of sites of expression for Slu and Sna between the chick and mouse

In our analysis of embryos at early developmental stages, we observed that m*Slu* expression was absent in tissues that express *Slu* in the chick. Surprisingly, in many of the sites where chick *Slu* was expressed, we observed *Sna* expression in the mouse. Moreover, at sites of c*Sna* expression, *Slu* transcripts are expressed in the mouse. Thus, an inversion in the expression of Snail family members appears to have occurred between the chick and mouse in the premigratory neural crest, early mesoderm and during early somite formation. The expression of *Xenopus Sna* and *Slu* is similar to that of both genes in the mouse (Essex et al., 1993; Mayor et al., 1995), indicating that expression at early stages of development is inverted in the chick. Thus, we propose that this inversion must have occurred in the avian lineage after the divergence between birds and mammals. This is corroborated by the fact that the sum of the expression sites of both genes is conserved in vertebrates.

The inversion in expression sites between *Slu* and *Sna* seems likely to be the result of recombination events between the regulatory sequences of both genes. Thus, our data indicate the existence of modulatory elements that can independently regulate the temporal and spatial expression of these genes. In support of this hypothesis, distinct elements have been identified in the promoter of x*Sna* that are required for its mesodermal and ectodermal expression (Mayor et al., 1993). One explanation for our observations is that a reshuffling of these elements has occurred in the avian lineage. This would be consistent with the swapping of only some sites of expression and the conservation of others. Thus, it would be of interest to analyse the promoter regions of the family members in several species, the prediction being that some of the vertebrate regulatory sequences of m*Sna* would be regulating c*Slu* and vice versa.

The expression of different family members in a particular tissue in distinct species is not unusual and probably evolves from a situation where both family members are co-expressed in such a tissue following gene duplication (see below). However, to our knowledge, the swapping of expression sites between different species as observed here, has not been described for other gene families. One implication of these findings relates to the analysis of knockouts and to the results of loss- or gain-of-function experiments referring to this family. For example, according to the data regarding *Slu* function in the chick (Nieto et al., 1994), one might expect that the mouse knockout for *Slu* should show defects in the emigration of the neural crest cells from the neural tube or in the delamination of the early mesoderm. However, in the light of our results one would more expect this phenotype for a mouse deficient in *Sna* function, or possibly, for the double mutant mice.

Phylogenetic and ontogenetic early and late expression sites of the *Snail* gene family

As discussed by Cooke et al. (1997), gene duplication offers an opportunity for the acquisition of new roles for different members of gene families. The expression of duplicated genes can be gradually modified such that they might be inactivated in some tissues and recruited to others where new morphological features unique to the vertebrate lineage would form (Holland and García-Fernández, 1996). This is likely to have been the case for the Snail family.

Recent duplicates would originally have overlapping expression patterns at ancestral sites that could gradually be modified. Thus, the expression of one of the duplicates in a specific tissue might be lost and the acquisition of new expression sites in an independent manner for each duplicate might also occur (see below). Once the expression of one family member were lost in a tissue, it is not difficult to imagine that evolutionary pressure would maintain the expression of the other. Otherwise, the loss of function of the two Snail family members in any tissue might give rise to serious problems during development and possible lethality. According to this, the expression sites of the ancient gene might have been distributed between the two duplicates, as

seems to be the case for the mesodermal expression of Snail genes. This expression in the mesoderm has been conserved from *Drosophila* (Alberga et al., 1991) to vertebrates (Nieto et al., 1992; Smith et al., 1992; Essex et al., 1993; Hammerschmidt and Nüsslein-Volhard, 1993; Thisse et al., 1993; Mayor et al., 1995), where different mesodermal populations express *Sna* or *Slu*. Indeed, we observe a complementarity in the mesodermal expression of the two genes within one species in the newly formed mesoderm and its derivatives (paraxial and lateral mesoderm).

The two *Snail* genes isolated from zebrafish (*Sna1* and *Sna2*) are very likely the result of an additional duplication in the fish genome, which is believed to have undergone all the evolutionary changes of the vertebrates as well as some additional ones (see Cooke et al., 1997). Indeed, there are several conserved residues specific to these two zebrafish Sna proteins. Being more recent duplicates, they maintain redundant expression sites (e.g. in the early mesoderm, paraxial mesoderm, migratory neural crest and the somites), as well as differential expression sites such as in the premigratory neural crest (Hammerschmidt and Nüsslein-Volhard, 1993; Thisse et al., 1993, 1995). Indeed, the sum of the expression sites of the two zebrafish *Sna* genes is equivalent to the sum of the expression sites of *Slu* and *Sna* in the other vertebrates that we have analysed. Thus, it appears that the two *Sna* genes identified in zebrafish are capable of performing the functions that *Slu* and *Sna* perform in other vertebrates.

Regarding the recruitment of genes to new functions following duplication, each gene may acquire new expression sites independently. This seems to be the case for the Snail family, *Sna* being recruited to play a role in chondrogenesis (Nieto et al., 1992; Smith et al., 1992) and *Slu* to play a role in lens morphogenesis. As discussed by Duboule and Wilkins (1998), the increase in developmental complexity during evolution originates not only from the appearance of new genes, but also from the recruitment of old genes to accomplish additional functions, thus generating new expression sites during development rather than providing them with emergent biochemical activities. *Slu* has been implicated in the process of EMT during the formation of the neural crest and the early mesoderm (Nieto et al., 1994). Whilst these processes are early patterning events, it seems likely that *Slu* might also have been co-opted to participate in EMT in later differentiation processes such as in the formation of the endocardial cushions (our unpublished observations) or in the growth-factor-induced EMT in a rat bladder carcinoma cell line (Savagner et al., 1997).

Considering that the history of gene families could reflect the phylogeny of structures (Duboule and Wilkins, 1998), it is interesting to note that the interchanged sites of expression between chick and mouse are phylogenetically ancient. This is compatible with their being regulated by ancestral promoter sequences, whereas many of the sites where the family member expressed is conserved, are new expression sites. These sites might have been independently acquired for each gene (*Sna* or *Slu*), possibly by the acquisition of additional *cis*-regulatory elements, as suggested for the *Hox* genes (Holland, 1992).

Is *Snail* a marker of the precursors of the neural crest in the chordate lineage?

One of the main sites of expression of the vertebrate members of the Snail family is the neural crest. The neural crest is a tissue that, along with the placodes, is believed to have been crucial in the formation of complex sensory organs giving rise to the 'new head' of vertebrates as proposed by Gans and Northcutt (1983) and, thus, it has classically been considered as a vertebrate character. However, it is interesting to note that the neural crest first appears at the edges of the neural folds, precisely the region where the ascidian *snail* homologue is expressed (Corbo et al., 1997). This raises the possibility that these cells are the evolutionary precursors of the neural crest that also appear to be present in the cephalochordata (for a review see Baker and Bronner-Fraser, 1997; Corbo et al., 1997). Thus, the study of this gene family may be of great help in the understanding of the mechanisms that generated the neural crest during evolution.

We are grateful to Michael Sargent and Pierre Savagner for sharing sequences of partial clones of the mouse *Slug* gene before publication and to Alberto Ferrús, Jordi García-Fernández and the anonymous referees, for helpful advice and discussions. This work was supported by grants from the Spanish Ministry of Education and Culture (DGICYT-PM95-0024), the Comunidad Autónoma de Madrid (08.1/0020/97) and the European Union (FMRX-CT96-0065) to M. A. N.; M. S. was also the recipient of a Wellcome Trust Travel fellowship and S. S. is the recipient of a predoctoral fellowship from the Spanish Ministry of Education and Culture.

REFERENCES

Alberga, A., Boulay, J. L., Kempe, E., Dennefield, C. and Haenlin, M. (1991). The *snail* gene required for mesoderm formation in *Drosophila* is expressed dynamically in derivatives of all three germ layers. *Development* **111**, 983-992.

Baker, C. V. H. and Bronner-Fraser, M. (1997). The origins of the neural crest. Part II: an evolutionary perspective. *Mech. Dev.* **69**, 13-29.

Boulay, J. L., Dennefield, C. and Alberga, A. (1987). The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding domains. *Nature* **330**, 395-398.

Buxton, P. G., Kostakopoulou, K., Brickell, P., Thorogood, P., and Ferretti, P., (1997). Expression of the transcription factor *Slug* correlates with the growth of the limb bud and is regulated by FGF-4 and retinoic acid. *Int. J. Dev. Biol.* **41**, 559-568.

Cooke, J., Nowak, M. A., Boerjist, M. and Maynard-Smith, J. (1997). Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends Genet.* **13**, 360-364.

Corbo, J. C., Erives, A., Di Gregorio, A., Chang, A. and Levine, M. (1997). Dorsoventral patterning of the vertebrate neural tube is conserved in a protochordate. *Development* **124**, 2335-2344.

Duboule, D. and Wilkins, A. S. (1998). The evolution of 'bricolage'. *Trends Genet.* **14**, 54-58.

Essex, L. J., Mayor, R. and Sargent, M. G. (1993). Expression of *Xenopus Snail* in mesoderm and prospective neural fold ectoderm. *Dev. Dyn.* **198**, 108-122.

Gans, C. and Northcutt, R. G. (1983). Neural crest and the evolution of vertebrates: a new head. *Science* **220**, 268-274.

Grau, Y., Carteret, C. and Simpson, P. (1984). Mutations and chromosomal rearrangements affecting the expression of *snail*, a gene involved in embryonic patterning in *Drosophila*. *Genetics* **108**, 347-360.

Hamburger, V. and Hamilton, H. (1951). A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49-92.

Hammerschmidt, M. and Nüsslein-Volhard, C. (1993). The expression of a zebrafish gene homologous to *Drosophila snail* suggests a conserved function in invertebrate and vertebrate gastrulation. *Development* **119**, 1107-1118.

Holland, P. W. H. (1992). Homeobox genes in vertebrate evolution. *BioEssays* **14**, 267-273.

Holland, P. W. H., García-Fernández, J., Williams, N. A. and Sidow, A.

(1994). Gene duplications at the origin of vertebrate development. *Development* **1994 Supplement** 125-133.

Holland, P. W. H. and García-Fernández, J. (1996). *Hox* genes and chordate evolution. *Dev. Biol.* **173**, 382-395.

Illingworth, C., Cheetham, J. and Hardin, J. (1992). Cloning and characterization of homologues of twist and snail in the sea urchin embryo. *Mol. Biol. Cell (suppl)* **3,** 110a.

Isaac, A., Sargent, M. G. and Cooke, J. (1997). Control of vertebrate left-right asymmetry by a *Snail*-related zinc finger gene. *Science* **275,** 1301-1304.

Mayor, R., Essex, L. J., Bennett, M. F. and Sargent M. G. (1993). Distinct elements of the x*Sna* promoter are required for mesodermal and ectodermal expression. *Development* **119**, 661-671.

Mayor, R., Morgan, R. and Sargent M. G. (1995). Induction of the prospective neural crest of *Xenopus*. *Development* **121**, 767-777.

Nieto, M. A., Bennet, M. F., Sargent, M. G. and Wilkinson, D. G. (1992). Cloning and developmental expression of *Sna*, a murine homologue of the *Drosophila snail* gene. *Development* **116**, 227-237.

Nieto, M. A., Sargent, M. G. Wilkinson, D. G. and Cooke, J. (1994). Control of cell behaviour during vertebrate development by *Slug*, a zinc finger gene. *Science* **264**, 835-839.

Nieto, M. A., Patel, K. and Wilkinson, D. G. (1996). In situ hybridisation analysis of chick embryos in whole mount and tissue sections. *Meth. Cell Biol.* **51**, 220-235.

Pavletich, N. P. and Pabo, C. O. (1993). Crystal structure of a five-finger GL1-DNA complex: new perspectives on zinc fingers. *Science* **261,** 1701-1707.

Roark, M., Sturtevant, M. A., Emery, J., Vaessin, H. and Grell, E. (1995). *scratch*, a pan-neural gene encoding a zinc finger protein related to *snail*, promotes neural development. *Genes Dev.* **9**, 2384-2398.

Ros, M. A., Sefton, M. and Nieto, M. A. (1997). *Slug*, a zinc finger gene previously implicated in the early patterning of the mesoderm and the neural crest, is also involved in chick limb development. *Development.* **124**, 1821-1829.

Sargent, M. G. and Bennet, M. F. (1990). Identification in *Xenopus* of a structural homologue of the *Drosophila* gene *snail*. *Development* **109**, 963-973.

Savagner, P., Yamada, K. M. and Thiery, J. P. (1997). The zinc finger protein *Slug* causes desmosome dissociation, an initial and necessary step growth factor-induced epithelial-mesenchymal transition. *J. Cell. Biol.* **137**, 1403-1419.

Smith D. E, Del Amo, F. F. and Gridley, T. (1992). Isolation of *Sna*, a mouse gene homologous to the *Drosophila* gene *snail* and *escargot*: its expression pattern suggests multiple roles during postimplantation development. *Development* **116**, 1033-1039.

Thisse, C., Thisse, B., Schilling, T. F. and Postlethwait, J. H. (1993). Structure of the zebrafish *Snail1* gene and its expression in wild-type, *notail* and *spadetail* mutant embryos. *Development* **119**, 1203-1215.

Thisse, C., Thisse, B. and Postlethwait, J. H. (1995). Expression of *Snail2*, a second member of the zebrafish Snail family, in cephalic mesendoderm and presumptive neural crest of wild-type and *spadetail* mutant embryos. *Dev. Biol.* **172**, 86-99.

Whiteley, M., Noguchi, P. D., Sensaburgh, S. M., Odenwald, W. F. and Kassis, J. A. (1992). The *Drosophila* gene *escargot* encodes a zinc-finger motif found in *snail*-related genes. *Mech. Dev.* **36**, 117-127.