Nucleotide sequence, chromosomal localization and developmental expression of the mouse int-1-related gene

JILL A. McMAHON and ANDREW P. McMAHON

Department of Cell and Developmental Biology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110, USA

Summary

cDNA clones encoding the murine int-1-related protein (m-irp) were isolated from an 8.5-day mouse embryo library. m-irp and its human counterpart, h-irp, share extensive nucleotide homology in coding (92%) and 3' untranslated (69%) regions. At the amino acid level, m-irp and h-irp share 97% of amino acids including all 24 cysteine residues, which are highly conserved among members of the int-1 family. However, in contrast to h-irp and int-1, the predicted m-irp protein sequence did not contain a signal peptide sequence. Analysis of polymerase chain reaction, amplified cDNA, and genomic sequences strongly suggests that a single-base substitution has created a new 5' splice site 17 bp 5' of a highly conserved splice site. Splicing at this new site generates a mRNA-encoding an amino-terminal truncated protein. Splicing at the conserved splice site generates a mRNA species encoding a protein with a signal peptide sequence similar to h-irp. Close linkage between m-irp and the met oncogene maps m-irp sequences to proximal mouse chromosome 6. Adult and fetal expression of m-irp was examined by RNA blot analysis. Adult expression of m-irp is restricted to lungs and heart, and fetal expression, to placental tissue and to all stages of fetal development examined. In situ hybridization localized early fetal m-irp expression to the pericardium of the heart, to the umbilicus and associated allantoic mesoderm, and to the ventral lateral mesenchyme tissue surrounding the umbilical and allantoic veins in the fetus. These results suggest a role for m-irp in the development of fetal allantoic communication.

Key words: mouse int-1-related gene, cDNA cloning, differential splicing, fetal expression, ventral body wall, allantois.

Introduction

The int-1 family currently comprises only two members, int-1 and the int-1-related protein (irp). int-1 was first identified as a gene frequently activated in mouse mammary tumor virus-induced mammary tumors due to insertion of viral sequences (Nusse and Varmus, 1982; Nusse et al. 1984). That int-1 itself is involved in tumor formation has been demonstrated in cell culture (Brown et al. 1986; Rijsewijk et al. 1987a) and in transgenic mice (Tsukamoto et al. 1988). Studies on the normal expression of int-1 suggest that this gene may be an important regulator of developmental processes in several organisms (for review, see McMahon and Moon, 1989a). Drosophila int-1 is encoded by the segment polarity gene wingless (Rijsewijk et al. 1987b; Cabrera et al. 1987) and is essential for normal segmental pattern formation. The murine int-1 gene is expressed during later stages of sperm development (Shackleford and Varmus, 1987) and in the developing mouse neural tube (Wilkinson et al. 1987a); however, the role of int-1 in mouse development is unknown.

Frog int-1 is also expressed at early neural stages (Noordemeer et al. 1989), and recent studies indicate that inappropriate expression of int-1 in developing frog embryos severely alters the normal body plan (McMahon and Moon, 1989b).

Our interest in int-1 initiated a search for int-1-related genes that may also function in a developmental context. However, attempts to date to identify homologous sequences using reduced stringency screening with nucleic acid probes have been unsuccessful (McMahon, unpublished data). Recently, Wainwright et al. (1988), while examining genes expressed in human lung and closely linked to the cystic fibrosis locus on human chromosome 7q, identified a gene with significant homology to int-1. Although overall amino acid homology is quite low (38%), cysteine residues are highly conserved between the int-1-related protein (h-irp) and int-1 from several different species. Thus, it is clear that h-irp and int-1 are distantly related, although nothing is known of the normal role of irp.

We report here the isolation of cDNA and genomic clones encoding mouse irp (m-irp). These results indicate that m-irp and its human counterpart h-irp share considerable nucleotide and amino acid homology. However, unlike h-irp, differential splicing of m-irp transcripts may generate altered forms of the protein.
m-irp expression was restricted to the adult lung and heart and to the developing fetus. Examination of fetal expression by in situ hybridization suggests a role for m-irp in the development of the allantoic component of the fetal–maternal connection.

Materials and methods

Isolation of cDNA clones

Approximately $5 \times 10^6$ pfu of an 8.5-day mouse embryo cDNA library, constructed in ρgt10 (the gift of Brigid Hogan), were screened with a 469-bp PsI EcoRI fragment (641–1110 in Wainwright et al. 1988) and a 595-bp EcoRI-XbaI fragment (1110–1705 in Wainwright et al. 1988) that recognize 5' and 3' coding sequences. Hybridization was in 50% formamide at 37°C, as described by Maniatis et al. (1982). Post-hybridization filter washes were at room temperature in 2×SSC, 0.5% SDS and at 60°C in 1×SSC, 0.5% SDS. Only plaques that were strong positives with either or both probes were picked and plaque purified. EcoRI fragments were isolated and subeloned into pGEM4 (Promega) or into M13 mp18 (Yanisch-Perron et al. 1985).

DNA sequencing

All clones were sequenced on both strands using a combination of subeloning of the original fragments into M13 mp18 or 19 (Yanisch-Perron et al. 1985) or by strand-specific primers. In all cases, sequence was determined by a modified dideoxy chain termination procedure (Sanger et al. 1980) using a Sequenase Kit (USB) following manufacturer's recommended protocol.

RNA isolation and hybridization

Embryo and adult RNAs were isolated using the lithium chloride urea procedure (Ruffray and Rougeon, 1980). Total RNA (10μg) was size-fractionated on a 1% formaldehyde agarose gel (Maniatis et al. 1982), transferred to Gene Screen (DuPont), cross-linked with ultraviolet light (Church and Gilbert, 1984), and hybridized with 3×10^6 cts min^-1 μg^-1 of high specific activity (10^9 cts min^-1 μg^-1), randomly primed (Feinberg and Vogelstein, 1984), 32P-labelled probe in 50% formamide at 45°C (Maniatis et al. 1982). Filters were washed at a final stringency of 0.2×SSC at 65°C. A 600-bp 3' fragment of clone 2 was used in all Northern.

In situ hybridization

Fetal samples at 9.5 and 10.5 days post-coitum were processed for in situ hybridization, as previously described (Wilkinson et al. 1987b). Single-stranded, 32P-labelled RNA probes in sense and antisense orientations were derived from a 3' terminal fragment of clone 2 using T7 or SP6 RNA polymerase (Melton et al. 1984).

PCR amplification of m-irp genomic sequences

Polymerase chain reaction (PCR) (Scharf et al. 1986) was used to amplify mouse genomic sequences. The oligonucleotide primers used in the PCR reaction were as follows: CCGGGCCACCGATGGCCCGG (17–36 in cDNA sequence, Fig. 2) and AAAGCTTCCCTGGAGGAGC-CACCTGTAGC (174–153 in cDNA sequence, Fig. 2). Reactions were performed using a Perkin-Elmer Cetus Kit according to manufacturer's recommendations, using 1.2μg of each oligonucleotide, 0.5μg of genomic DNA, 2.5 units of Taq DNA polymerase in 100 μl of reaction mix. Samples were heated to 94°C for 1.5 min followed by 30 cycles at 72°C for 1.5 min and 94°C for 1 min. Samples were extracted with phenol–chloroform and chloroform. DNA was then ethanol precipitated, 5' overhangs filled in with the Klenow fragment of DNA polymerase, kinased and the single 2.1 kb PCR product subeloned into EcoRV-cut, dephosphorylated pSP73 (Promega). Termini were sequenced directly in pSP73 or after subeloning into M13mp18.

PCR amplification of 9.5-day m-irp cDNA sequences

Oligo (dT)-primed, first-strand cDNA was synthesized according to the manufacturer's instructions from 10μg of 9.5-day total fetal RNA using an Invitrogen kit in 40μl of reaction mix. Two microliters of this reaction was PCR-amplified exactly as described above. PCR products were phosphorylated by T4 polynucleotide kinase and blunt-end ligated into dephosphorylated Smal-cleaved pGEM4 (Promega). Inserts were sequenced from both ends by the dideoxy chain termination procedure (Sanger et al. 1980) using T7 and SP6 oligonucleotide primers and T7 DNA polymerase (Pharmacia), according to manufacturer's instructions.

Results

Isolation of mouse embryo cDNA clones

DNA fragments that included 5' and 3' coding sequences for h-irp were used to screen approximately 5×10^5 plaque-forming units of an 8.5-day mouse embryo cDNA library. Three clones were isolated. Clone 7 hybridized specifically to 5' h-irp sequences; clone 4, to 3' sequences only; and clone 2, to 5' and 3' sequences. The relationship between these clones and the original h-irp lung clones is shown in Fig. 1.

DNA sequence analysis

The composite sequence of clones 2, 4, and 7 is shown in Fig. 2. All of clone 4 and all but the 5' terminal 41 nucleotides of clone 7 were present in the 1458-bp clone 2. Overlapping sequences were identical. The m-irp composite sequence encompasses nucleotides 219–1742 of the human sequence (Wainwright et al. 1988) and includes one long, open reading frame starting at the ATG at position 147, and ending at 1142 (Fig. 2). Nucleotide homology is extensive throughout the sequence. Homology in the 5' untranslated and coding regions is approximately 92% and, in the 3' untranslated region, approximately 69%. However, one important difference between the human and mouse sequence is a short 17-bp deletion at the 5' end of m-irp. (Fig. 2). Although two independently isolated cDNA clones (7 and 2) showed this 5' deletion, it is possible...
that the deletion resulted from a cDNA cloning artefact. To eliminate this possibility, we have examined the
m-irp genomic sequence in this region.

**Differential splicing generates 5' heterogeneity**

A 2.1-kb genomic fragment was isolated following polymerase chain reaction (PCR) amplification of mouse genomic DNA with oligonucleotides, 105 nucleotides 5' and 52 nucleotides 3' of the deletion. The large size of the amplified region suggested that an intron occurs within the sequence. This prediction was confirmed by DNA sequencing of the genomic fragment (Fig. 3). Moreover, analysis of the DNA sequence strongly suggests that the 17-bp deletion in the m-irp cDNA clones resulted from the creation of a new 5' splice donor site.

At the genomic level, m-irp contains the 17-base sequence (italics, Fig. 3) present in h-irp, but absent in m-irp 8.5-day cDNA clones. 3' of this sequence is a 2.1-kb intron surrounded by 5' and 3' splice sites (solid arrows, Fig. 3). The 5' splice donor sequence (shaded, Fig. 3) has a 7/9 nucleotide match to the consensus 5' donor sequence (Mount, 1982). Moreover, the se-
quence at this splice site is identical to the first 5' splice donor site in mouse int-1 (van Ooyen and Nusse, 1984) and occurs at the same position as the first intron in mouse (van Ooyen and Nusse, 1984), human (van Ooyen et al. 1985), and Drosophila (Rijsewijk et al. 1987b) int-1. Although the genomic sequence of h-irp is not available, h-irp shows a characteristic conservation of sequences encoding two tryptophan residues (amino acids 27 and 28, Wainwright et al. 1988). These tryptophan residues, which are conserved in int-1 genes of different species, are encoded by the nucleotide sequence that spans this 5' splice site. Therefore, conservation of this sequence probably reflects a conservation of a similar 5' splice donor sequence in the h-irp gene. From the above sequence analysis, it is clear that splicing of m-irp transcripts at this 5' splice site (shaded, Fig. 3) would generate processed RNA containing the 17-bp sequence present in h-irp cDNA. However, a single-base substitution (boxed, Fig. 3) that changes an A to G (position 365 in h-irp cDNA, Fig. 2) within this sequence has created a new 5' splice donor site. This sequence (stippled, Fig. 3) now shares a 6/9 nucleotide match to the 5' splice donor consensus sequence (Mount, 1982). Splicing at this site (open arrow) would delete the 17-bp sequence, producing a mature transcript identical to the m-irp cDNA clones isolated. Thus, splicing at either the new 5' splice donor (stippled, Fig. 3) or the int-1-conserved splice donor (shaded, Fig. 3) would generate mature m-irp RNAs that differ by the absence or presence of the 17-bp sequence. Interestingly, this deletion occurs within a sequence homologous to the proposed amino terminal sequence of the human clone. Thus, if m-irp translation initiated from the same ATG as proposed for the human sequence (Wainwright et al. 1988, shaded, Fig. 3), the 17-bp deletion would cause a frame shift producing a severely truncated protein (only 35 amino acids). Initiation at the ATG at position 147 in m-irp (stippled, Fig. 3) would generate a 37-kilodalton polypeptide, 97% homologous to the predicted h-irp protein (Fig. 4), but lacking a signal peptide sequence. Thus, it seems unlikely that either of these protein species would function in a way similar to the human counterpart. In contrast, splicing from the conserved splice donor (shaded, Fig. 3) would generate a m-irp protein with a signal peptide that initiates and terminates at the same positions as predicted for h-irp (Fig. 4).

To directly examine RNAs from fetal tissues, cDNA was synthesized from 9.5-day fetal RNA. m-irp sequences were specifically amplified by PCR with the same oligonucleotide pair used in the genomic amplification described earlier. A fragment identical in size to the PCR product produced on amplification of m-irp clone 2 (157 bp) and a second, more abundant, slightly larger fragment were amplified (data not shown). Amplified products were 'shot-gun' cloned, and three independent recombinants were sequenced. All three encoded identical sequences that contained the 17-base
pair sequence. Thus, at 9.5 days, splicing occurred predominantly at the conserved 5' splice donor (shaded, Fig. 3) producing processed RNAs encoding full-length m-irp protein with a signal peptide sequence (Fig. 4). The full-length m-irp protein shares 96% amino acid homology (346 of 360 amino acids) with h-irp. All 24 cysteine residues are absolutely conserved. m-irp and h-irp share one putative asn-linked glycosylation site (boxed in Fig. 4). A second site towards the amino terminus is not conserved in m-irp (boxed in Fig. 4). Three of the 12 amino acid changes are clustered in the carboxyl-eight amino acids, a part of the protein not shared with mouse int-1 or the Drosophila int-1 gene wingless, suggesting that extreme conservation of this carboxy-terminal region is not essential.

Chromosomal localization

h-irp was originally isolated as a gene on human chromosome 7q, closely linked with the met oncogene and the cystic fibrosis locus (Wainwright et al. 1988). In order to ascertain whether m-irp was linked to met in the mouse, segregation of a BglII restriction fragment length polymorphism for m-irp was compared to that of met among 43 progeny of an interspecific Mus spretus back-cross (Roberts et al. 1985). No recombinants between met and m-irp occurred among these animals (data not shown). Thus, m-irp is tightly linked to met in the mouse genome and is part of a syntenic group between proximal mouse chromosome 6 and human 7q (Dean et al. 1987; Bucan et al. 1986).

RNA blot analysis of fetal adult expression

m-irp clones were originally isolated from an 8.5-day embryo library, indicating that m-irp is expressed during mouse development. RNA blot analysis of 9.5- and 10.5-day fetal RNA detected low levels of two m-irp transcripts of approximately 2.0 and 2.2 kb (Fig. 5). Low levels of m-irp RNA were also detected in placent al samples at 14.5 days, but not in either yolk sac or decidual tissue. A more extensive analysis of fetal and adult tissues detected low levels of m-irp transcripts at all stages of fetal development as well as in adult lung and heart, but not in testes, brain, spleen, liver, or kidney (data not shown). It should be noted that, whereas we do not know the significance of the two transcripts, Wainwright et al. (1988) have demonstrated two RNA transcripts for the human gene that result from the use of alternate poly A addition sites.

In situ analysis of m-irp expression

To further characterize fetal expression of m-irp, in situ hybridization was performed to 9.5- and 10.5-day sections. At these stages, the relative amounts of m-irp RNA are highest in total fetal RNA. Specific hybridization to m-irp antisense RNA probes was detected at both stages, and the pattern of hybridization was essentially the same. m-irp transcripts were present in the pericardium, the outer mesodermal tissue surrounding the heart (Fig. 6C,D). However, inner myocardial and endocardial tissues do not express m-irp (Fig. 6A–D). From the heart, m-irp expression extended caudally along the ventral lateral aspects of the body wall (Fig. 6A,B,E,F). Expression was localized to stellate, mesenchyme tissue surrounding the paired umbilical veins and ductus venosus. Expression continues into the allantois and allantoic stalk from its base of attachment to the body wall (Fig. 6A,B,G,H) and into the umbilicus, which is formed from the allantois and maintains fetal-placental communication. No expression of this gene was observed. No expression was seen using sense RNA probes (data not shown).

Discussion

Comparison of nucleotide and amino acid homologies indicates that h-irp and m-irp, like the human and mouse int-1 genes (Nusse et al. 1984; van Ooyen et al. 1985), are highly conserved. In particular, all cysteine residues, a key feature of these cysteine-rich proteins, are conserved between h-irp and m-irp, and 22 of the 24, between m-irp and mouse int-1. However, a distinguishing feature of the m-irp gene, not previously reported for h-irp or int-1, is the potential for amino terminal heterogeneity due to alternative splicing of fetal RNAs. We have described two splice donor sequences separated by 17 bp, both of which are used. The more 5' site has been created by a single-base substitution relative to h-irp. The other site is highly conserved in all int-1 genes. Processed messages differ in the presence or absence of the 17-bp sequence separating these sites. In two 8.5-day cDNA clones encompassing this region, the more 5' splice site is used, and this sequence is not present. In contrast, PCR analysis at 9.5 days indicated that the more 3' splice site is preferentially used and the 17-bp sequence is present.
in the processed message. Differential splicing has profound consequences for the production of m-irp proteins. RNA in which the 17-bp sequence is removed cannot encode a full-length m-irp protein. Depending on where translation initiation occurs, either a short protein that consists of little more than a signal peptide and a near full-length protein with no signal peptide sequence would be produced. Unlike int-1 (Papkoff et al. 1987; Brown et al. 1987; McMahon and Moon, 1989a) and h-irp (McMahon and McMahon, unpublished observations), this form of the protein does not enter the secretory pathway (McMahon and McMahon, unpublished observations). Thus, the two cDNA clones isolated at 8.5 days post-coitum encode presumably non-functional proteins while at 9.5 days post-coitum, PCR analysis indicates that the predominant transcript encodes the full-length and presumably functional m-irp protein. Whether there is an actual switch in splice preference at this time remains to be unambiguously demonstrated.

Although little is known of the properties of m-irp protein, the finding that m-irp is expressed in specific cell populations suggests that m-irp, like int-1, may have an important role in development. Interestingly, while int-1 is present in distantly related species such as mammals and Drosophila, the distribution of irp is more restricted. No homologous sequences can be detected in Drosophila (Wainwright et al. 1988; Struhl, personal communication); however, homologous sequences are readily detected in several mammalian and non-mammalian vertebrates, including chicken and Xenopus (Estivill et al. 1987). In humans, irp expression has been reported in adult and fetal lung and in term placenta (Wainwright et al. 1988). Our studies indicate that as in humans, m-irp is expressed in adult lung and placenta. In addition, we detected m-irp RNA in adult heart and in total RNA from fetuses between 9.5 and 17.5 days of gestation. In situ hybridization at 9.5 and 10.5 days indicates that at this time heart-associated m-irp expression is limited to the pericardial membrane that surrounds the heart and not to myocardial or endocardial tissue. The predominate site of expression, however, is clearly associated with the developing allantois and its fetal–maternal connections.

The allantois arises from a mesodermal outgrowth at the base of the primitive streak at approximately 8 days post-coitum. The allantois grows upward through the amnion, contacts the chorion, and gives rise to the chorio-allantoic placenta typical of the mouse and other mammalian species. Fetal placental circulation is established in allantoic blood vessels that run through the umbilicus. In this way, the allantois is responsible for establishing the exchange of materials between the fetus and the maternal circulation.

m-irp expression in the lateral ventral body wall is limited to mesenchymal tissue surrounding the umbilical vein. Whereas we cannot rule out a role for m-irp in establishment of the umbilical vein, m-irp expression is not generally associated with blood vessel formation. No m-irp transcription is seen around other major blood vessels, including the dorsal aorta. Rather, m-irp expression seems to be localized in a morphologically similar group of cells that extend along the ventral body wall of the fetus and are continuous with the allantoic membrane and allantoic-derived umbilical cord. It seems likely that these cells in the fetus and allantois have a similar role, the establishment of fetal–allantoic–placental communication. It will be interesting to examine whether irp in non-mammalian vertebrates is also involved in the formation of equivalent tissues in these species, tissues that are not formed during embryonic development of organisms such as Drosophila, which lack irp sequences.

The exact role of m-irp in mouse development remains unknown. The sequence homology between irp and int-1 suggests that functional homology may exist between these proteins. Moreover, it is clear that both m-irp and int-1 protein products contain a signal peptide sequence, and, therefore, the proteins presumably operate through a similar pathway. It will be of interest to determine whether ectopic expression of m-irp leads to a similar phenotype as misexpression of mouse int-1 (McMahon and Moon, 1989b). However, a better understanding of the role of m-irp awaits a thorough characterization of all protein species and their biological properties.

We thank the St. Mary's Cystic Fibrosis group for the gift of the h-irp clone, Brigid Hogan for the 8.5-day-cDNA library,
Paul Kelly for initial screening of the 8.5-day library, Jeffrey Friedman for chromosomal mapping of \textit{m-irp}, and Sharon Perry and Alice O’Connor for expert typing of the manuscript. We are particularly grateful to Don Nuss for carefully reviewing the manuscript.

References


(Accepted 24 August 1989)