We report the genomic organization of the mouse orphan receptor related to tyrosine kinases (Ryk), a structurally unclassified member of the growth factor receptor family. The mouse Ryk protein is encoded by 15 exons distributed over a minimum of 81 kilobases. Genomic DNA sequences encoding a variant protein tyrosine kinase ATP-binding motif characteristic of Ryk are unexpectedly found in two separate exons. A feature of the gene is an unmethylated CpG island spanning exon 1 and flanking sequences, including a TATA box-containing putative promoter and single transcription start site. Immunohistochemical examination of Ryk protein distribution revealed widespread but developmentally regulated expression, which was spatially restricted within particular adult organs. Quantitative reduction of Southern blotting stringency for the detection of Ryk-related sequences provided evidence for a retroprocessed mouse pseudogene and a more distantly related gene paralogue. Extensive cross-species reactivity of a mouse Ryk kinase subdomain probe and the cloning of a Ryk orthologue from Caenorhabditis elegans demonstrate that Ryk and its relatives encode widely conserved members of a novel receptor tyrosine kinase subfamily.

Members of the receptor tyrosine kinase (RTK)1 family of transmembrane signal transduction proteins participate in the regulation of diverse cellular activities such as mitogenesis, motility, differentiation, survival, metabolism, adhesion, fasciculation, morphogenesis, and oncogenesis (1). Structurally, and usually functionally, related RTKs are grouped into distinct subfamilies largely on the basis of characteristic combinations of protein motifs in their ligand-binding extracellular domains. Unique in this classification system is the Ryk (for related to tyrosine kinase) receptor subfamily which, aside from two leucine-rich motifs in the mammalian orthologues and a possible protease cleavage site common to all Ryks, is comprised of members devoid of recognizable protein modules in their extracellular domains characteristic of other RTKs (2–4).

The cytoplasmic protein tyrosine kinase (PTK) activity of RTKs is classically activated by homodimerization of receptor protomers mediated by the binding of an extracellular growth factor ligand (5). More recently, RTK activation via heterodimerization (6, 7), extracellular matrix (8), and ligand-independent routes (9) have been recognized as physiologically significant modes of signal generation. Twelve conserved peptide sequence motifs, or subdomains, are the signature of the PTK catalytic domain, including some 13 invariant residues, which participate in the phosphotransfer reaction (10, 11).

A subclass of RTKs that lack demonstrable phosphotransferase activity has emerged recently whose members display substitutions in one or more of the conserved and catalytically important PTK motifs. These include the human CCK4 (12) and chicken KLG (13) orthologues, ERBB3 (14), the human and mouse Eph-like proteins HEP (15) and MEP (16), and the human Ryk protein (3). The likely functional importance of Ryk subfamily-specific amino acid substitutions to otherwise universally conserved PTK motifs is underscored by their occurrence in such phylogenetically diverse species as Drosophila (Doughnut2 and Derailed, Ref. 4), human (3), mouse (2) and zebrafish Ryk.3

Although largely functionally uncharacterized, transmembrane proteins such as Ryk which bear an apparently catalytically inactive PTK domain may well modulate or relay growth-regulatory cues present in the extracellular environment to cytoplasmic signaling and/or effector molecules. For instance, inactivation of the PTK activity of the Caenorhabditis elegans Eph receptor VAB-1 results in weak mutant phenotypes (relative to extracellular domain mutations), suggesting that a subset of RTKs may execute kinase-independent functions (17). These may include recruitment of heterologous protein kinase-competent receptor subunits, as in the case of a kinase-defective EGFR mutant still capable of activating mitogen-activated protein kinase (18). That mice homozygous for a deletion of exons encoding the PTK domain of VEGFR1 (Flt-1), an RTK essential for embryonic angiogenesis, exhibit normal vessel development and are fully viable suggests that the primary

---

1 The abbreviations used are: RTK, receptor-type tyrosine kinase; Ryk, receptor related to tyrosine kinases; Ryk-EX-FLAG, FLAG-tagged human Ryk extracellular domain; PTK protein tyrosine kinase; EGFR, epidermal growth factor receptor; VEGFR, vascular endothelial growth factor receptor; FGFR, fibroblast growth factor receptor; INSR, insulin receptor; LRME, leucine-rich motif; BAC, bacterial artificial chromosome; RT-PCR, reverse transcriptase-polymerase chain reaction; nt, nucleotide; kb, kilobase pair(s); bp, base pair(s).

2 Oates, A. C., Bonkovsky, J. L., Irvine, D. V., Thomas, J. B., and Wilks, A. F. (1998) Mech. Dev. 78, 165–169.

3 C. P. Do, A. C. Oates, and A. F. Wilks, manuscript in preparation.
biological role of this receptor is in ligand binding rather than signal transduction (19).

Identification of derailed in a P element enhancer trap screen for fasciculating neurons demonstrated a requirement for growth cone pathfinding cues transduced by this Ryk subfamily member during Drosophila nervous system development (4). Furthermore, defective somatic muscle insertion site selection into the epidermis of derailed mutants suggests that the mechanism for muscle insertion target recognition is biochemically similar to that used in axon pathfinding (20). A pleiotropic role for ERBB3 in cardiac and central nervous system development has been demonstrated in ErbB3-null mice (21, 22). However, a role for other PTK-inactive RTKs in growth regulation remains to be demonstrated.

Here we report the structure of the mouse Ryk transcription unit, including the likely location of its promoter, as identified by an unmethylated CpG island. The unusual organization of domains in the mouse Ryk protein extends to the genomic structure that displays features unique within the RTK/growth factor receptor gene family. Southern blotting results suggestive of the existence of a retroprocessed mouse Ryk pseudogene are also presented. Immunohistochemical analysis has been undertaken to define the spatial distribution of Ryk expression within individual adult mouse organs. Conserved structural elements that define the Ryk subfamily are identified in five metazoan proteins, including one from C. elegans, indicating an early evolutionary origin of these RTK-like molecules. Consistent with subsequent duplications of the ancestral vertebrate genome, the mouse, human, and zebrafish genomes also contain putative Ryk subfamily paralogues that we find to be phylogenetically conserved molecules with likely roles in cellular growth regulation.

EXPERIMENTAL PROCEDURES

Isolation and Characterization of Genomic Clones—Lambda phage from a mouse 129/Sv genomic DNA library (Stratagene, La Jolla, CA) were screened with probes generated from the mouse Ryk cDNA (2) by PCR. The extracellular domain probe was amplified using the primers 5’-TGGGCTTCTCGGGCCACG-3’ (human cDNA nt 1021–1042) and 5’-ACACAGCTGTTGGAAGCTGTTGGC-3’ (mouse cDNA nt 737–715), and the kinase domain probe was amplified using the primers 5’-TTGGGCTTCTTTTTCATGGA-3’ (mouse cDNA nt 1083–1103) and 5’-TGACCCGAGCTGTTGAACTA-3’ (mouse cDNA nt 1842–1823). Hybridizing phage were plaque-purified, and phage DNA was prepared using the algorithms of Gardiner-Garden and Frommer (31). The CpG island was analyzed with the EGCG program CpGplot (http://www.enson.de/PSORT). and cleavage sites in the N termini of the Ryk proteins was performed using the Doughnut extracellular domain revealed homology with an open reading frame on genomic cosmid C16B8 designated C16B8.1 (27), here referred to as Ceryk. To verify the exon-intron structure predicted by the Genefinder program (28), oligonucleotides to putative Ceryk exons were used to amplify the entire coding sequence, and flanking regions, in overlapping RT-PCR products. Reverse transcription of mixed stage C. elegans total RNA (~5 mg) was performed using a first strand synthesis for RT-PCR kit (Amerham); cDNA amplification by touchdown PCR was performed as described (29). Detection of sorting signals and cleavage sites in the N termini of the Ryk proteins was performed using SignalP, and transmembrane domains were identified using PSORT.

RESULTS

Structure of the Mouse Ryk Gene—The mouse Ryk protein is encoded by 15 exons covering a minimum of 81 kb (Fig. 1a). Twenty four independent phage clones isolated from a 129/Sv
genomic lambda library and two bacterial artificial chromosome (BAC) clones identified by amplification with the primer pair BAC.01 and BAC.02 (Fig. 1) were analyzed to define the organization of mouse Ryk. A 4-kb XbaI fragment containing exon 1, a CpG island, the transcription start site, and putative promoter sequences (Fig. 1, a and b) was subcloned from BAC.Ryk.2. The exact size of intron 1 (.19 kb) has not been determined.

Two possible translation initiation codons (M1 and M2; see Fig. 2b) are located in exon 1 which has a G + C content of 83%. No in-frame ATG codons are present upstream of M1 before the next in-frame stop codon. Exons 1–5 encode the RYK extracellular domain (exons 2 and 4 each encode a putative leucine-rich motif, LRM), exon 6 the hydrophobic transmembrane domain, exons 7 and 8 the serine-threonine-rich juxtamembrane domain, and exons 8–15 the PTK-like domain. A 15-residue C-terminal tail is also encoded by exon 15.

Six independent groups have now cloned and sequenced the mouse Ryk cDNA (2, 37–41). When these sequences are aligned (not shown), nucleotide differences that predict amino acid substitutions are apparent. The major differences are found in the 5’ region of the cdNA and most likely reflect difficulties in sequencing this highly G + C-rich region. No nucleotide ambiguities were encountered in our genomic sequencing of this portion of the open reading frame, the sequence of which always matched a majority of published cDNAs, consistent with the interpretation that differences between the reported sequences are the result of sequencing errors. The long 5’-untranslated region published by Kelman et al. (39) appears to be part of a chimeric cDNA clone given that the sequence is unrelated to genomic DNA sequence in Fig. 2b beginning at position 1763 and extending upstream. Additionally, our BAC.Ryk clones do not hybridize with oligonucleotides complementary to the unique 5’-untranslated end reported by Kelman et al. (39). The kinase activation loop of mouse RYK, bounded by subdomains VII and VIII of the PTK catalytic core (10), is encoded by exon 13 and contains a conserved tyrosine residue at position 454 (Ref. 2; see Fig. 6) which is often involved in catalytic autoregulation of diverse PTK superfamily members (42). The two alternative polyadenylation signals reported by Kelman et al. (39), both of which perfectly match the consensus sequence AATAAA, reside in exon 15 (Fig. 1c).

The sequences of mouse Ryk exon-intron boundaries are presented in Table I. The mouse Ryk transcription start site (see below) defines the 5’ boundary of exon 1. All introns

---

**Fig. 1.** Organization of the mouse Ryk gene (drawn to scale). a, exon-intron distribution. Translated sequences are shown as black boxes and untranslated sequences as open boxes. b, partial restriction map of the Ryk locus. c, overlapping recombinant BAC and lambda genomic DNA clones. E, EcoRI; H, HindIII; X, XbaI.
FIG. 2. Structure and sequence of the 5' region of the Ryk gene. a, restriction map of the 4-kb XbaI fragment subcloned from BAC.Ryk.2. Sites for rare-cutting enzymes recognizing CpG-containing hexanucleotide sequences are shown in boldface. The 0.7-kb intron 1-derived probe used in Southern blotting is indicated. The BssHII site in brackets is either incompletely methylated or polymorphic in genomic DNA from W9.5 ES cells (see text and Fig. 3b). b, DNA sequence of the region surrounding exon 1. The boundaries of a CpG island are indicated by curved arrows and a TATA box plus other putative transcription factor binding motifs, with core sequences underlined, are boxed. Only the single best potential binding
conform to the 5'-gt...ag-3' motif, with the exception of exon 8 which has a modified 3' splice acceptor site (Table I).

Identification of a CpG Island Containing the Putative Mouse Ryk Promoter—The DNA sequence flanking exon 1 is shown in Fig. 2b. No consensus sequence for a splice acceptor site associated with this exon was found in the upstream region. This and the location of the transcription start site (see Discussion) suggest that the promoter resides at least partly within this sequence. The presence of a CpG island in this area, recognizable as a high density of sites for restriction enzymes with G+C-rich recognition sequences containing one or more CpG dinucleotides (Fig. 2a; Ref. 43), is a landmark for the Ryk promoter since a CpG island was always found to cover the whole or part of the promoter in a survey of 102-island-containing genes expressed in a wide variety of tissues (44).

CpG islands are defined as sequences >200 bp in size comprising a moving average G+C content >50% and an observed/expected CpG dinucleotide frequency >0.6 (31, 44). The 859-bp Ryk CpG island lies between nucleotides 1469 and 2347 and is characterized by elevated G+C content (75%), an oberved/expected CpG ratio of 1.004 (Fig. 3a), and is unmethylated at CpG residues (Fig. 3b). Mouse genomic DNA prepared from low passage embryonic stem cells (W9.5 line, derived from mouse 129/Sv C3—+/-+P, Jackson Laboratory stock number JR0090; Ref. 45), which express Ryk mRNA, was subjected to Southern blot analysis with methylation-sensitive restriction enzymes. Digestion with XbaI, or double-digestion with XbaI and either BssHII or EagI (both of which are blocked by methylation), protected from methylation.

The exon 1, 3'-end splice acceptor of exon 8 is underlined. The consensus sequence of the 3' splice acceptor is (11Py)ncaag and of the 5' splice donor is NN(C/A)AG (75). When a 0.7-kb probe hybridized with the expected 2.2-kb BssHII-XbaI fragment, an additional 1.4-kb BssHII-XbaI fragment was detected. This smaller fragment is generated by cleavage at a BssHII site 1.4 kb upstream of the most 3' XbaI site (Fig. 2a) and reflects either clonal variation in methylation of CpGs in the sequence 5'-GCGCGC-3', as previously reported for CpGs in DNA flanking the mouse liver aprt CpG island (46), or an allelic polymorphism. Allelic variation in simple sequence length polymorphisms has been reported in the 129 mouse substrains and ES cell lines derived from them (47), demonstrating that alleles are still segregating in these supposedly highly inbred mice. As judged by the status of four CpG dinucleotides within the Ryk CpG island, this region of DNA is unmethylated and is expected to constitutively maintain this property in expressing and nonexpressing cell types at all stages of development (48, 49). The unmethylated or variably methylated status of the BssHII site suggests that CpG dinucleotides immediately surrounding the island may also be protected from methylation.

Sequences immediately flanking the transcription start site (see “Discussion”) display potential binding motifs for a variety of transcriptional modulators (Fig. 2b). The MatInspector program (30) was used to search the sequence for high quality matches to a data base of position-weighted nucleotide distribution matrices. DNA elements potentially mediating the response to growth-regulatory stimuli through the sequence-specific binding of transcription factors are particularly abundant in the putative Ryk promoter (Fig. 2b). These include response elements for transcription factors activated by mitogenic (e.g. Ets family, AP-1, n-Myc, Egr family, E2F, RREB-1, NGFI-C, c-Myc), anti-mitogenic (IRF-1, ISRE), metabolotropic (T3 receptor, aryl hydrocarbon receptor, and nuclear translocator), and motif for each of the transcription factors shown is boxed (core similarity >0.70, matrix similarity >0.8, see “Experimental Procedures”). The transcription start site is marked by a square arrow, and the antisense 35-mer used for primer extension by Yee et al. (40) is represented by a half-arrow, and alternative translation initiation methionines (M1 and M2) are circled. The exon 1, 3'-end splice donor site (type I) is boldface. SRP, SRE-binding factor; NRF-2, nuclear respiratory factor 2; ISRE, interferon-stimulated response element; Tol-1, T-cell acute lymphocytic leukemia protein 1; ER, estrogen receptor; STAT, signal transducer and activator of transcription; AML-1, acute myeloid leukemia protein 1; T3R, thyroid hormone receptor; HNF-3b, hepatocyte nuclear factor 3b; IRF-1, interferon regulatory factor 1; EGR, early growth response protein; Ahr/ARNT, aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator; RREB-1, Ras-responsive element binding protein 1; NGFI-C, nerve growth factor-induced protein C; Gfi-1, growth factor independence protein 1.
differentiation (HNF-3β, GATA-1, Nkx-2.5, MyoD) signals. This is consistent with the likely function of RYK as a cell-surface growth factor receptor or co-receptor, the expression of which is expected to be under stringent spatial and temporal control (see "Immunohistochemical Analysis of RYK Expression in the Mouse," below).

Identification of Mouse Ryk-related Sequences by Southern Blotting—A screen of the mouse genome by Southern blotting using mouse Ryk cDNA-derived probes was initiated as a first step in resolving the issue of whether mammals have one or more Ryk gene paralogues (see Ref. 41). Restriction fragment length polymorphism mapping in the mouse (50) and human cytogenetic analysis (3) have previously identified loci on chromosomes 12 and 17p13.3, respectively, with high homology to the canonical mammalian Ryk cDNAs. Consistent with this, most RTKs belong to subfamilies of homologous receptors that engage in functional interactions (e.g. the ErbB subfamily, Refs. 6 and 51; the platelet-derived growth factor receptor, FGFR, and Trk subfamilies, Ref. 7; the INSR subfamily, Ref. 52).

By using a gradient of hybridization and washing stringencies originally employed to detect novel human RTKs by Southern blotting of genomic DNA (25, 53), we have detected two classes of mouse genomic DNA fragments related to the Ryk gene characterized herein. The first class is detectable at high stringency (Fig. 4c, arrows) using an Ryk kinase domain probe. However, only the hybridization signals expected from the structure of the mouse Ryk gene (Fig. 1a) are observed when genomic DNA digests are hybridized at high stringency with a probe to the Ryk extracellular domain (Fig. 4a). The nature of this first class of Ryk-related hybridization signals (high signal strength of the fragments, in particular the 3.8-kb EcoRI and 2.2-kb XbaI fragments, at high stringency) is consistent with a contiguous organization of kinase domain exons comprising a partially retroprocessed pseudogene (cRyk). At reduced stringency, hybridization signals not present at high stringency are visible using both extracellular and kinase domain probes from the Ryk cDNA (Fig. 4, b and d, arrowheads). Using a v-erbB PTK domain probe, this level of hybridization stringency has proven to be suitable for discriminating between RTKs highly related to the probe sequence, such as subfamily members, while more distantly related members of the PTK superfamily are not detected (25, 53). The detection of DNA fragments at reduced stringency with probes to the extracellular and kinase domains of mouse Ryk further suggests that the hybridization signals represent a genuine subfamily member. We predict that this second class of fragments is derived from a novel mouse member of the Ryk RTK subfamily. Screening of genomic DNA libraries enriched for cRyk, which we have found to be under-represented in supposedly complete genomic DNA libraries, and Ryk-related fragments is underway.

Immunohistochemical Analysis of RYK Protein Expression in the Mouse—Analysis of Ryk mRNA expression in a large sample of mouse and human tissues has been reported by us and others (2, 37–41, 54–57). These analyses show that Ryk mRNA is distributed ubiquitously. In order to assess the spatial distribution of the functional gene product, we generated an anti-RYK monoclonal antibody and performed immunohistochemi-
try on sections of embryonic and adult mouse tissues. The RYK-1 antibody was raised against the extracellular domain of human RYK (Fig. 5a) and is specific for the extracellular domain of this receptor (Fig. 5, b and c). The RYK-1 monoclonal antibody has been used in Western blotting analysis of cell lines overexpressing a variety of RTK extracellular domains (e.g. EGFR, VEGFR1, VEGFR2, VEGFR3, and Tie2) with no evidence of cross-immunoreactivity. The dual Western blotting signal from MCF-7 lysate at approximately 90 and 47 kDa (Fig. 5c) is also observed in reduced immunoprecipitates of metabolically labeled MCF-7 cells using an antiserum raised against the C terminus of human RYK (3) and is consistent with proteolytic processing of the receptor extracellular domain at the tetra-basic KRRK site. Disulfide bonding between Cys-156 and Cys-191 of the human receptor, which represent conserved residues that flank the KRRK motif in all RYK subfamily members (see Fig. 6), may be an important post-translational modification. The RYK-1 antibody cross-reacts with both mouse and zebrafish RYK but not with Drosophila Derailed.4

RYK protein expression was detected in a large proportion of adult mouse tissues and on cells from a variety of origins. In the kidney, strongest RYK staining is seen in tubules immediately under the cortex (Fig. 5d), which gradually declines in strength toward the pelvic region. Glomeruli within all regions of the kidney are uniformly unstained with the RYK-1 antibody. RYK expression is limited to epithelial cells lining the renal tubules. Strong and uniform RYK expression is evident in liver lobules, associated with the membranes of hepatocytes (Fig. 5f). The adrenal medulla shows no reactivity with the RYK-1 antibody, whereas high RYK expression is seen in the adrenal cortex (Fig. 5h). Strong and uniform RYK expression is present on embryonic day 14 myocardium (Fig. 5j); this represents the earliest expression of RYK that we have observed in the mouse (embryonic days 8–12 are unstained; not shown). Adult spleen shows staining with the RYK-1 antibody restricted to the red pulp (Fig. 5l). Within this compartment, megakaryocytes show membrane-localized immunoreactivity, whereas erythroblasts and lymphocytes are free of staining. In addition, reticular fibers and endothelial cells lining the venous sinuses are stained. The adult small intestine displays strong staining with the RYK-1 antibody on the epithelium and connective tissue/vascular core of villi, whereas the glandular crypts are unstained (Fig. 5n). The mucosa of the large intestine is stained on connective tissue surrounding intestinal glands, which are themselves free of RYK-1 immunoreactivity (Fig. 5p).

Phylogenetic Conservation of the Ryk Gene Subfamily—Complementary DNAs encoding RYK subfamily members have been isolated from mouse (2, 37–41), human (3, 54, 57), and zebrafish, and two orthologues, doughnut and derailed (4), have been cloned from Drosophila. We report here the cloning, by RT-PCR, of a C. elegans ryk orthologue (Ceryk) that bears the characteristic features of this RTK subfamily (Fig. 6). The cDNA predicts a 583-residue pre-protein containing an N-terminal hydrophobic signal sequence and a transmembrane domain.

To explore the range of evolutionary conservation of Ryk subfamily members, low stringency Southern blot hybridization of a mouse exon 13 probe to a “Zoo blot” was performed. Extensive and strong hybridization of the Ryk-specific kinase
FIG. 5. Specificity of the RYK-1 monoclonal antibody and immunohistochemical analysis of RYK expression in the mouse. a, Western blot of RYK-EX-FLAG probed with the RYK-1 antibody. A negative control lane containing RYK-EX-FLAG was probed with supernatant from a hybridoma which scored negative in the enzyme immunoassay screen. b, Western blot of alkaline phosphatase-tagged human RYK and VEGFR2 extracellular domains probed with the RYK-1 antibody. c, Western blot analysis of MCF-7 human breast carcinoma total cell lysate (10⁶ cells) with the RYK-1 antibody. The signal at 85,000–90,000 kDa (arrowhead) represents the RYK receptor. Kidney (d and e), liver (f and g), adrenal gland (h and i), embryonic day 14 heart (j and k), spleen (l and m), small intestine (n and o), and large intestine (p and q) were stained with RYK-1 (left column) or an isotype-matched (IgM) negative control antibody, respectively. Scale bars, 100 μm.
subdomain probe to genomic DNA from a wide range of metazoans was observed (Fig. 7a). The 3.6-kb mouse HindIII fragment represents the expected hybridization signal from mouse Ryk, and the strong ~0.7-kb signal, which is also seen at high stringency in Fig. 4c, is most likely derived from a mouse pseudogene. Weaker hybridization to other mouse genomic HindIII species is also visible (at ~1.3, ~1.5, and ~1.8 kb, see also Fig. 4d). To identify the origins of the two strongest hybridization signals from human genomic DNA (Fig. 7a), PCR was employed to map the organization of human RYK sequences recognized by the Zoo blot probe. Using primers to human RYK corresponding to those used to produce the mouse exon 13 Zoo blot probe, a 159-bp PCR product, which was uninterrupted by introns and not cleavable by HindIII (3), was amplified from genomic DNA (Fig. 7b). This result indicates that the mouse Zoo blot probe recognizes a single human RYK exon resistant to cleavage by HindIII within the limits of homology to the probe. Thus it is likely that the signals at 3.2 and 4.3 kb (Fig. 7a, human) reflect hybridization to human RYK and its pseudogene. Hybridization to RYK-related sequences in the human genome is also evident as weaker signals (e.g. at ~1.8 kb; Fig. 7a). Mapping of exon-intron boundaries of the zebrafish ryk gene predicts the mouse Zoo blot probe to detect only one fragment, such that the two strong zebrafish hybridization signals (Fig. 7a) can be confidently assigned to different loci.

Hybridization of the kinase domain probe to distinct genomic DNA fragments was observed in all species tested, with the exception of Escherichia coli (negative control). After prolonged exposure, three autoradiographic signals were visible from Drosophila genomic DNA (not shown). Ryk gene orthologues and paralogues, both characterized and uncharacterized, therefore seem to be widely conserved over the phylogenetic scale, from mammals (human, mouse, horse, dog, cow, sheep, and rabbit), birds (chicken), through lower vertebrates (Xenopus, zebrafish, and carp), to invertebrate metazoans (Drosophila and C. elegans).

**DISCUSSION**

We have determined in detail the genomic organization of the mouse member of the growing subfamily of growth factor receptors related to tyrosine kinases (Ryk). Although all subfamily members are currently orphan receptors, RYK molecules are expected to function in the transduction of growth-regulatory information across the plasma membrane by virtue of their prototypical RTK topology, as has been demonstrated for all other RTK subfamilies (5).

The mouse Ryk transcription unit resides on chromosome 9 (39, 50) and spans a minimum of 81 kb of genomic DNA. A large first intron is a feature of the mouse Ryk gene which is shared with zebrafish ryk and Drosophila derailed (4). A comparison of mouse Ryk with the structures of other mammalian RTK genes shows that a large first intron is not uncommon (e.g. human INSR, >25 kb, Ref. 34; human IGF1R, >20 kb, Ref. 58; human FMS, 26 kb, and human c-KIT, 37.4 kb, Ref. 59; human C-MET, 26 kb, Ref. 60). The GXGXXG motif vital for ATP binding in PTKs is modified to XXGXSG in the RYK subfamily. This motif is invariably encoded by a single exon; however, in the mouse and C. elegans Ryk genes a type I splice donor site interrupts the codon for the second consensus glycine residue, and a type O splice site separates the fourth and fifth residues of motif I in the zebrafish ryk gene (data not shown). Other mouse RYK PTK subdomains, with the exception of IX, are encoded by single exons, as is usually the case for other RTK genes. This supports the view that the RYK kinase-like domains from phylogenetically diverse species are evolutionarily related.

We speculate that the significance of exon splitting of RYK subdomains I and IX, together with unusual kinase subdomain sequences, may be that the ancestral Ryk gene arose very early in metazoan evolution and has since been subjected to selective pressure to maintain a modified PTK activity that relies heavily on atypical residues at normally conserved positions. Molecular modeling of the mouse RYK PTK domain indicates that the nucleotide-binding cleft is particularly large and may therefore indicate a preference for a phosphodonor substrate other than ATP. This prediction is currently being tested.

Primer extension analysis performed by Yee et al. (40) allowed us to identify the transcription start site at nucleotide 1627, within the CpG island (Fig. 2b). This result is consistent with the finding that almost every widely expressed gene transcribed by RNA polymerase II that has been examined, such as mouse Ryk, has a CpG island at the 5' end that includes the transcription start site (Refs. 31 and 44; CpG island data base 4.0). Assignment of the transcription start site leads us to predict the synthesis of two mouse Ryk mRNA species of 2.3 and 2.6 kb. Two polyadenylation sites in exon 15, which appear to be utilized at equal frequencies in most tissues (2, 37–41), define the predicted alternative 3'-mRNA ends. By using RT-

---

5 H. Treutlein, M. M. Halford, and S. A. Stacker, unpublished data.
6 Available on-line at the following address: ftp://ftp.no.embnet.org/cpgisle/relnotes4.0.doc.
Ryk Gene Structure

A transcriptionally active CpG island represents a domain of “open” chromatin structure characterized by core histone underacetylation, histone H1 depletion, and a nucleosome-free region (63). Constitutive binding of transcription factors is essential for the maintenance of a methylation-free CpG island (62). This region of the Ryk gene is also highly enriched for the CpG dinucleotide relative to the bulk mammalian genome, where CpG depletion to 20% of the frequency expected from base composition has resulted from methylation of CpG and the high mutability of 5-MeCpG to TpG or CpA (CpG suppression; Refs. 48 and 49). CpG dinucleotides in the Ryk promoter are protected from genomic DNA methylation and are present at the frequency predicted by base composition alone (i.e. no CpG suppression). These properties identify the sequence spanning exon 1 of the mouse Ryk gene as a CpG island 859 bp in size (Fig. 3).

Our immunohistochemical analysis of mouse Ryk further confirms localization of the functional mRNA to the epithelial and stromal compartments of human tissues such as ovary, brain, lung, colon, kidney, and breast by in situ hybridization. The immunohistochemical staining results presented here confirm localization of the functional Ryk gene product to primary transcript in 3T3 L1 cells.

7 M. M. Halford and S. A. Stacker, unpublished data.

8 Available on-line at the following address: ftp://ftp.ebi.ac.uk/pub/databases/cpgisle/.

FIG. 6. Deduced amino acid sequence of C. elegans Ryk and alignment with other members of the Ryk subfamily of proteins. Alternative ATG translation start codons are marked in the mouse sequence (*). N-terminal signal sequences and their predicted sites of cleavage are shown (*). The transmembrane domains (TM), mammalian leucine-rich motifs (LRMs), putative tetrameric cleavage sites (TBC), conserved cysteine residues (●), Hanks kinase motifs (I–VI), potential autoregulatory tyrosine residue homologous to that in pp60** (crossed circle), boundaries of the protein tyrosine kinase ligand consensus sequence (PDZL) are indicated. Residues that match the sequence of mouse Ryk are boxed. **M. M. Halford and S. A. Stacker, unpublished data. 

FIG. 7. The diagram shows the domains and motifs of mammalian Ryk protein. Domain boundaries are established by the alignment of Rhomboid-like2 (Rhomboid) and PDZL domains at the N- and C-termini, respectively.
the tubular epithelium in kidney and to the stromal compartment of the large intestine and indicate that mouse Ryk is localized to the stroma and epithelium of villi in the small intestine. However, organs including embryonic day 14 heart and adult liver show strong and homogenous expression of Ryk throughout. Furthermore, the spatially restricted localization of mouse Ryk to distinct parenchymal compartments of organs such as the adrenal gland and spleen suggests that specific differentiated cell types not related in embryonic origin may require the signal transduced by Ryk.

We have demonstrated the existence of multiple Ryk subfamily members in the mouse genome by Southern blotting: the canonical Ryk (2), a likely partial retroprocessed pseudogene derived from Ryk kinase domain exons, and a Ryk-related gene detectable at quantitatively reduced stringency using Ryk cDNA-derived probes to the extracellular and PTK domains. The human genome appears to share with the mouse the feature of a Ryk pseudogene, as well as the Ryk-related gene. We have detected Ryk subfamily-like sequences in the genomes of mammals, chicken, fish, and Drosophila by reduced stringency Southern blotting and in a nematode worm by cDNA amplification and sequencing. Although the stringency conditions used for Southern blot hybridization were relaxed, more stringent and extensive washing was performed, suggesting that the level of cross-species nucleotide sequence conservation in the kinase-like domain is high.

Southern blot analysis of the Danio rerio (zebrafish) genome showed hybridization signals representing the ryk locus plus a related sequence of unknown identity. From the simple nematode worm C. elegans, we have sequenced a Ryk cDNA orthologue, Ceryk, which predicts a transmembrane protein demonstrating structural conservation of Ryk subfamily-specific features. These include a compact extracellular domain containing a putative basic protease cleavage site flanked by universally conserved cysteine residues. Post-translational processing of the human Ryk exodomain into αβ disulfide-linked subunits analogous to the c-MET/HGF receptor (70), mouse STK/RON (71), the insulin receptor (72), and the insulin-like growth factor 1 receptor (73) is supported by our Western blotting and immunoprecipitation data (Fig. 5c and Ref. 3), although we have found human Ryk to be inconsistently processed in this manner. In the predicted Ceryk protein, a pro-
tein tyrosine kinase-like domain with strongest sequence homology to other RYKs, distinctive RYK-specific amino acid substitutions in PTK subdomains I and VII (11), and a C-terminus conforming to the consensus -Y(V/I)-COOH, representing a possible PDZ domain ligand (74), are identifiable in the intracellular domain (Fig. 6). No likely candidate C. elegans mutant phenotypes mapping to this CELC16B8 chromosomal locus are known.

The RYK proteins are structurally unique in at least two significant features. Few recognizable protein motifs are present in an unusually short extracellular domain, and phylogenetically conserved intracellular substitutions in at least two highly conserved PTK sequence elements responsible for cooperatively binding the Mg2+-ATP phosphate donor complex at the active site may abrogate, or more likely modify, catalytic activity. These changes perhaps reflect involvement of RYK in a unique cell-surface signal transduction complex, where it may interact with a novel family of extracellular ligands and/or be recruited into a heterodimeric PTK-competent receptor complex. Alternatively, RYK could conceivably function to attenuate signaling from such a complex by failing to transactivate the protein tyrosine kinase activity of the partner receptor in competition with homodimeric receptor formation. Third, atypical substrate contacts mediated by the nucleotide binding cleft of RYK subfamily proteins, comprising variant Hanks’ motifs I and VII, may indicate an altered specificity for the phosphatase, sugar, and/or divergent cation moieties of the phosphodonor complex, as proposed by Kelman et al. (39). We are further investigating the function of RYK in embryonic and early postnatal development, and its potential role in cancer, through the generation and analysis of RYK-deficient mice and a screen for ligands of the RYK receptor.

Acknowledgments—We thank Dr. Guo-Fen Tu for DNA sequencing, Dr. Andrew Runting for assistance with computing, Cuong Do for access to unpublished data, and Profs. Ashley Dunn and Antony Burgess for helpful comments on the manuscript.

References

1. van der Geer, P., Hunter, T., and Lindberg, R. A. (1994) Annu. Rev. Cell Biol. 10, 253–317.

2. Hovens, C. M., Stacke, S. A., Andres, A., Harper, A. G., Ziemiecki, A., and Wilks, A. F. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 11818–11822.

3. Stacke, S. A., Hovens, C. M., Vitali, A., Pritchard, M. A., Baker, E., Sutherland, G. R., and Wilks, A. F. (1993) Oncogene 8, 1347–1356.

4. Callahan, C. A., Moritzsch, M., Lindgren, S. E., Scully, A. L., and Thomas, J. B. (1995) Nature 376, 171–174.

5. Schlessinger, J., and Ullrich, A. (1992) J. Biol. Chem. 267, 487–493.

6. Gardiner-Garden, M., and Frommer, M. (1987) J. Mol. Biol. 196, 261–282.

7. Henikoff, S. (1987) Trends Genet. 3, 342–347.

8. Brundage, M. B., and Lehrach, H. (1996) Trends Biochem. Sci. 21, 216–226.

9. Sitnikova, M. D., Williams, K. L., Nakagawara, A., Enger, K. T., Gus, C., Brodeur, G. M., and Deuel, T. F. (1992) Biochem. Biophys. Res. Commun. 199, 1077–1083.

10. Chen, L. L., Williams, K. L., Nakagawara, A., Enger, K. T., Gus, C., Brodeur, G. M., and Deuel, T. F. (1992) Biochem. Biophys. Res. Commun. 200, 546–556.