Molecular and Cellular Probes (1999) **13**, 169–173
Article No. mcpr.1999.0228, available online at http://www.idealibrary.com on IDEAL®

Genomic structure of the *EPHA1* receptor tyrosine kinase gene

D. Owshalimpur and M. J. Kelley*

Lung Cancer Biology Section, National Cancer Institute, Bethesda, MD 20889-5105, USA

(Received 26 October 1998, Accepted 30 December 1998)

Some receptor tyrosine kinase genes are mutated in inherited and somatically acquired human cancers. To permit mutational analysis, the complete genomic structure of the human *EPHA1* gene on chromosome 7q34 was determined and oligonucleotide pairs were designed to amplify coding regions. The gene contains 18 exons, two more than the related tyrosine kinase, *EPHB2*. Presumed sequencing errors in the published cDNA sequence of *EPHA1* were identified in exons 10 and 11. Availability of this information will facilitate mutational analysis of *EPHA1*.

KEYWORDS: exon, intron, mutation, evolution, genomic structure.

INTRODUCTION

Receptor tyrosine kinases are a large family of transmembrane molecules that includes the platelet-derived growth factor (PDGF), epidermal growth factor (EGF), fibroblast growth factor/heparin binding growth factor (FGF/HBGF), and insulin receptor subfamilies. These molecules transduce extracellular signals involved in various cellular processes.[1] Mutation of tyrosine kinases have been implicated in pathogenesis of human cancer. Germline mutations of *MET* and *RET* proto-oncogenes have been found in hereditary papillary renal cancer[1] and multiple endocrine neoplasia type II,[2] respectively, and acquired mutations of MET have been found in sporadic papillary renal cell cancer.[1] Mutations of *MET* in hereditary papillary renal cell cancer are associated with trisomy of chromosome 7, on which *MET* is located, suggesting that duplication of the mutated copy of *MET* is a second alteration necessary for malignant transformation. Thus, trisomy may be associated with tyrosine kinase mutation.

About 30% of sporadic colon, lung and kidney tumours have trisomy of chromosome 7.[3] However, *MET* mutations were found in only 3 of 60 (5%) sporadic papillary renal tumours[1] and have not been reported in colon and lung tumours. Thus, mutations of other tyrosine kinase receptors on chromosome 7 may be found in these tumour types.

EPHA1 (also known as eph) is located at chromosome 7q34 and is the prototype for a 14 member subfamily of the tyrosine kinase receptor family.[4,5] When bound to the appropriate ligand (known as ephrins) these receptors have been shown to modulate cellular growth, differentiation and mitogenic activities.[6] Some members of the eph family, including *EPHA1*, have been implicated in carcinogenesis.[7] *EPHA1* is expressed in a subset of human cancer cell lines,[8] and transfection of the full length cDNA into NIH-3T3 cells results in increased colony formation in soft agar and development of tumours in nude mice.[8] The genomic structure of only the tyrosine kinase domain of the *EPHA1* has previously been determined.[8] Therefore, as a prerequisite to mutational analysis of

*Author to whom all correspondence should be addressed at: Hematology/Oncology (111G), Duke University/VA Hospital, 508 Fulton Street, Durham, NC 27705, USA. Tel: +1 919 286 6944; Fax: +1 919 286 6896; E-mail: kelleym@duke.edu

0890–8508/99/030169+05 $30.00/0

Table 1. Exon characteristics and amplification primers for *EPHA1*

Number	Exon size (bp)	Intron[a] size (bp)	Amplification primers	Annealing temperature	Amplified size[b] (bp)
1	82	ND	TCCCTTGCAACCTGGCGCTG GGTCGGTACCTTCCTTGGCG	70	154
2	68	ND	TTAGGAACAGGATGTTGCAGC ACTGGCCTTCCAGGAGTTTT	60	227
3	282	~1160	TCTACTGTCCTGGAGCTGAGG CATTCCCAAACTCAGTGCCA	65	1254
4	403	237	AAAATCACCTGGGCATGGTG GGAGAGAGAACGCAGCAGAG	65	1539
5	156	312	GAGATGTTTCCAGTAGAAGGG CAGAGAAGCTCAGGTTTCGG	65	605
6	345	152	ACATGGACACACCCCATTG GTCCAGAGGGATAAGGTTGG	65	932
7	125	250	GAAGCCCAAAATGGAGTGTC GGAAAACGAGAATCCCAAGC	65	1112
8	154	250	GAAGCCCAAAATGGAGTGTC GGAAAACGAGAATCCCAAGC	65	1112
9	97	198	AACTCAGCTCCTGTTCCACG GTAGCAAAACACAAGAGCCC	60	731
10	59	793	AACTCAGCTCCTGTTCCACG GTAGCAAAACACAAGAGCCC	60	731
11	126	~943	AAACAGGAGTCACCTTGGGAAAC AACTCTTGGACCAGAGAAGCCAG	65	375
12	186	136	GGGAGACAGTGGTGAATAGG CTGTGATGATCATGATCGGC	65	559
13	186	106	CAGTGGTGGAACTTCCTTCG GAAGCTTCTGGAAGTGTGGC	65	1669
14	207	464	CAGTGGTGGAACTTCCTTCG GAAGCTTCTGGAAGTGTGGC	65	1669
15	150	329	CAGTGGTGGAACTTCCTTCG GAAGCTTCTGGAAGTGTGGC	65	1669
16	194	ND	TTATCTTGTGTGACCGTCGG CGTCTTGCCTCATACACTGG	60	406
17	156	ND	TTCAGCTTCCTCTCAGACAGG AAGAATGCGCTTCTGGTGC	65	649
18	79		GACACCATGGAGTGTGTGCT CCCCACCTCCCTTTTAAAAC	65	386

[a]Size of intron following indicated exon; introns with exact sizes stated were completely sequenced, while others were partially sequenced.
[b]Exons 7–8, 9–10 and 13–15 are amplified with a single polymerase chain reaction each.
ND = no data

EPHA1 in cancer, we determined the complete intron/exon structure of this gene to allow amplification of the coding regions using genomic DNA as template.

MATERIALS AND METHODS

Polymerase chain reaction (PCR) was performed in 100 µl reaction volume containing 1X PCR buffer, 1·5 mM $MgCl_2$, 5 units *Taq* platinum (Life Technologies, Gaithersburg, MD, USA), 0·2 mM each dNTP, and 0·2 µM each oligonucleotide primer. Typical cycling conditions were initial denaturation of 94°C for 2 min followed by 30–40 cycles of 94°C for 30 s, annealing temperature (see Table 1) for 30 s, and 72°C for 2 min and final extension time of 10 min at 72°C. Polymerase chain reaction products were purified using QIAquick Spin PCR Purification Columns (Qiagen, Santa Clarita, CA, USA) per manufacturer's recommendations. An *EPHA1* BAC clone was obtained by screening a human BAC library (Research Genetics, Huntsville, AL, USA) with *EPHA1* oligonucleotides from both the 5′ and 3′ end of the cDNA. Approximately 400 ng of the PCR product or 2 µg of BAC plasmid DNA was sequenced using BigDye terminator kit (Perkin Elmer, Norwalk, CT, USA). DNA sequences have been submitted to GenBank (AF101165-AF101171).

Exon 1 gcccggagctATGGAGCGGCGCTGGCCCCTGGGGCTAGGGCTGGTGCTGCTGCTCT
GCGCCCCGCTGCCCCCGGGGGCGCGCGCCAAGGAAGgtaccgacc

Exon 2 ccactttcagTTACTCTGATGGACACAAGCAAGGCACAGGGAGAGCTGGGCTGGCT
GCTGGATCCCCCAAAAGATGGGgtaagtgtc

Exon 3 ctccttgcagTGGAGTGAACAGCAACAGATACTGAATGGGACACCCCTGTACATGT
ACCAGGACTGCCCAATGCAAGGACGCAGAGACACTGACCACTGGCTTCGCTCCAAT
TGGATCTACCGCGGGGAGGAGGCTTCCCGCGTCCACGTGGAGCTGCAGTTCACCGT
GCGGGACTGCAAGAGTTTCCCTGGGGGAGCCGGGCCTCTGGGCTGCAAGGAGACCT
TCAACCTTCTGTACATGGAGAGTGACCAGGATGTGGGCATTcagctccgac

Exon 4 ttctacccagGTAACCACGGTGGCTGCAGACCAGAGCTTCACCATTCGAGACCTTG
CGTCTGGCTCCGTGAAGCTGAATGTGGAGCGCTGCTCTCTGGGCCGCCTGACCCGC
CGTGGCCTCTACCTCGCTTTCCACAACCCGGGTGCCTGTGTGGCCCTGGTGTCTGT
CCGGGTCTTCTACCAGCGCTGTCCTGAGACCCTGAATGGCTTGGCCCAATTCCCAG
ACACTCTGCCTGGCCCCGCTGGGTTGGTGGAAGTGGCGGGGacctgcttgc

Exon 5 ctctctccagCCTGCCCTAGCGGCTCCTACCGGATGGACATGGACACACCCCATTG
TCTCACGTGCCCCCAGCAGAGCACTGCTGAGTCTGAGGGGGCCACCATCTGTACCT
GTGAGAGCGGCCATTACAGAGCTCCCGGGGAGGGCCCCCAGGTGGCATGCACAGgt
gagtccag

Exon 6 tctcccacagGTCCCCCCTCGGCCCCCCGAAACCTGAGCTTCTCTGCCTCAGGGAC
TCAGCTCTCCCTGCGTTGGGAACCCCCAGCAGATACGGGGGGACGCCAGGATGTCA
GATACAGTGTGAGGTGTTCCCAGTGTCAGGGCACAGCACAGGACGGGGGGCCCTGC
CAGCCCTGTGGGGTGGGCGTGCACTTCTCGCCGGGGGCCCGGGCGCTCACCACACC
TGCAGTGCATGTCAATGGCCTTGAACCTTATGCCAACTACAcctttaatgt

Exon 7 ggacccccagAGTCACTGTCAGGCCTGTCTCTGAGACTGGTGAAGAAAGAACCGAG
GCAACTAGAGCTGACCTGGGCGGGGTCCCGGCCCCGAAGCCCTGGGGCGAACCTGA
CCTATGAGCTGCACGTGCTGAACcaggtcagga

Exon 8 ctccaacccaCAGGATGAAGAACGGTACCAGATGGTTCTAGAACCCAGGGTCTTGC
TGACAGAGCTGCACCCTGACACCACATACATCGTCAGAGTCCGAATGCTGACCCCA
CTGGGTCCTGGCCCTTTCTCCCCTGATCATGAGTTTCGGACCAGCCCACCAGgtgg
ggtatc

Exon 9 gtgcctacagTGTCCAGGGGCCTGACTGGAGGAGAGATTGTAGCCGTCATCTTTGG
GCTGCTGCTTGGTGCAGCCTTGCTGCTTGGGATTCTCGTTTTCCGGTCCAGgtgcc
agctc

Exon 10 AccccacagGAGAGCCCAGCGGCAGAGGCAGCAGAGGCAGCGTGACCGCGCCACC
GATGTGGATCGAGgtgagtcggg

Exon 11 ctgcccatagAGGACAAGCTGTGGCTGAAGCCTTATGTGG_ACCTCCAGGCATACG
AGGACCCTGCACAGGGAGCCTTGGACTTTA_CCCGGGAGCTTGATCCAGCGTGGCT
GATGGTGGACACTGTCATAGGAGAAGgtgagtcctg

Exon 12 ttctacccagGAGAGTTTGGGGAAGTGTATCGAGGGACCCTCAGGCTCCCCAGCCA
GGACTGCAAGACTGTGGCCATTAAGACCTTAAAAGACACATCCCCAGGTGGCCAGT
GGTGGAACTTCCTTCGAGAGGCAACTATCATGGGCCAGTTTAGCCACCCGCATATT
CTGCATCTGGAAGGCGTCGTCACAAAGCgtacgatgag

Exon 13 tgtgtctcagGAAAGCCGATCATGATCATCACAGAATTTATGGAGAATGGAGCCCT
GGATGCCTTCCTGAGGgtgacgaggg

Exon 14 cctgctgcagGAGCGGGAGGACCAGCTGGTCCCTGGGCAGCTAGTGGCCATGCTGC
AGGGCATAGCATCTGGCATGAACTACCTCAGTAACTCACAATTATGTCCACCGGGAC
CTGGCTGCCAGAAACATCTTGGTGAATCAAAACCTGTGCTGCAAGGTGTCTGACTT
TGGCCTGACTCGCCTCCTGGATGACTTTGATGGCACATACGAAACCCAGgttagag
gcc

Exon 15 gccttcttagGGAGGAAAGATCCCTATCCGTTGGACAGCCCCTGAAGCCATTGCCC
ATCGGATCTTCACCACAGCCAGCGATGTGTGGAGCTTTGGGATTGTGATGTGGGAG
GTGCTGAGCTTTGGGGACAAGCCTTATGGGGAGATGAGCAATCAGGAGgtgagccc
ag

Exon 16 cgccgtgcagGTTATGAAGAGCATTGAGGATGGGTACCGGTTGCCCCCTCCTGTGG
ACTGCCCTGCCCCTCTGTATGAGCTCATGAAGAACTGCTGGGCATATGACCGTGCC
CGCCGGCCACACTTCCAGAAGCTTCAGGCACATCTGGAGCAACTGCTTGCCAACCC
CCACTCCCTGCGGACCATTGCCAACTTTGACCCCAGgtaaccatgc

Exon 17 tgtcccacagGGTGACTCTTCGCCTGCCCAGCCTGAGTGGCTCAGATGGGATCCCG
TATCGAACCGTCTCTGAGTGGCTCGAGTCCATACGCATGAAACGCTACATCCTGCA
CTTCCACTCGGCTGGGCTGGACACCATGGAGTGTGTGCTGGAGCTGACCGCTGAgt
aaggagct

Exon 18 ttgcccacagGGACCTGACGCAGATGGGAATCACACTGCCCGGGCACCAGAAGCGC
ATTCTTTGCAGTATTCAGGGATTCAAGGACTGATCCCTCCTCTCACCCCATGCCCA
GTCAGGGTGCAAGGAGCAAGGACGGGGCCAAGGTCGCTCATGGTCACTCCCTGCGC
CCCTTCCCACAACCTGCCAGACTAGGCTATCGGTGCTGCTTCTGCCCACTTTCAGG
AGAACCCTGCTCTGCACCCCAGAAAACCTCTTTGTTTTAAAagggaggtgg

Fig. 1. Partial Genomic Sequence of *EPHA1*. Intron sequence is in lower case, exon sequence in upper case. The bold nucleotides in exon 10 and 11 indicate the location of an inserted base pairs compared to the previously published cDNA sequence (GenBank accession number M18391). Underscores in exon 11 indicate the location of deletion of 1 bp and 25 bp, respectively. These sequences and additional intron sequence has been submitted to GenBank (accession numbers AF101165-AF101171).

RESULTS/DISCUSSION

To determine the intron/exon borders of *EPHA1*, we first predicted the location of introns within the *EPHA1* cDNA by alignment to *EPHB2* [cek5 (*C*hicken *e*mbryo *k*inase 5)], for which intron/exon borders have been determined.[9] Using oligonucleotides designed to amplify across the presumed introns, PCR products containing *EPHA1* cDNA sequence were amplified from genomic DNA. Inspection of DNA sequence identified intron/exon borders (Fig. 2).

No PCR products were obtained using this strategy for introns following exons 1, 2, 9, 10 and 17. Instead, a human genomic BAC clone, 494m18, which contains the entire *EPHA1* gene, was obtained and partially sequenced. Intron/exon borders were identified for all exons. Only 9 bp of sequence could be obtained from the 5′ end of intron 1, probably due to strong secondary structure in this area. Primers capable of

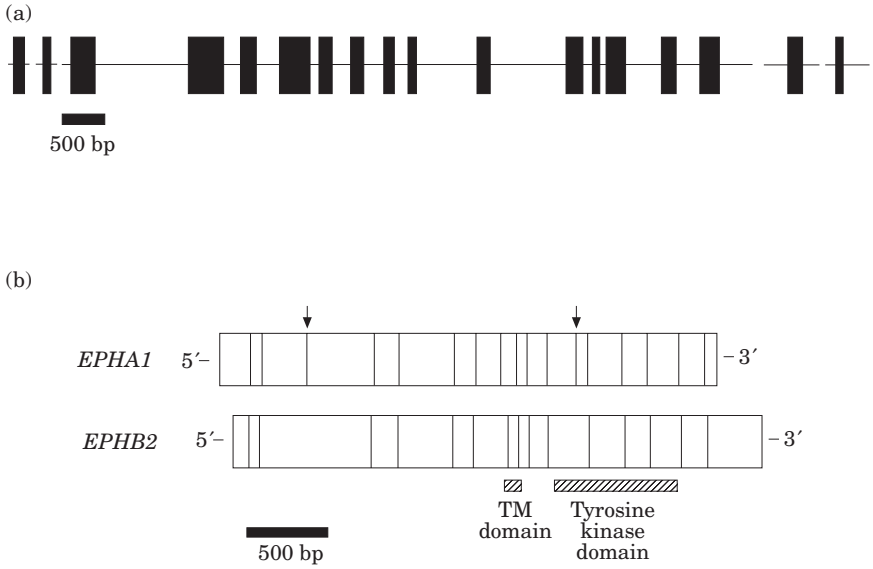

Fig. 2. Schematic representation of the relative size of *EPHA1* introns and exons (a) and relative size and position of *EPHA1* and *EPHB2* exons (b). Only coding regions of exons are shown. Arrows indicate position of *EPHA1* introns not found in *EPHB2*. *EPHB2* based on perviously published genomic structure from chicken.[9]

amplifying each exon and adjacent splice sites from genomic DNA were designed (Table 1), with the exception of exon 1, where the primers do not amplify the 11 bp at the 3′ end of the exon.

Eighteen coding exons were identified, two more than the related *EPHB2*. There may also be non-coding exons upstream of exon 1, as this region was not sequenced. Alignment of *EPHA1* and *EPHB2* cDNA's indicates that *EPHA1* exons 3–4 and 12–13 correspond to a single exon each in *EPHB2*. Thus, the two additional exons of *EPHA1* (one in the extracellular domain and one in the tyrosine kinase domain) arise by replacement of one *EPHB2* exon with two in *EPHA1*, or conversely, the joining of two exons of *EPHA1* into a single exon in *EPHB2*. Determination of the genomic structure of additional members of this subfamily might distinguish these alternative evolutionary events.

The coding regions contained multiple sequence variations from the published *EPHA1* cDNA sequence.[10] Four of these variations (insertion of G after nucleotide 1835, insertion of G after nucleotide 1876, deletion of nucleotide 1893, and deletion of nucleotides 1939–1963; Fig. 1) result in reading frame shifts and likely represent sequencing errors in the original cDNA sequence. Alternate splicing of the involved exons (10 and 11) was excluded by complete sequencing of the intervening introns. These four frameshifts are absent from an unreferenced partial *EPHA1* cDNA (GenBank accession number Z27409). Therefore, the correct *EPHA1* protein contains a unique stretch of 36 amino acids between the transmembrane and tyrosine kinase domains in place of 44 amino acids in the published sequence.

Additional variations noted were three missense changes (nucleotide 1286, GCG Ala to GGG Gly; nucleotide 2240, GCA ala to GGA Gly; and nucleotide 2340, GCA Ala to GGA Gly), three silent changes (nucleotide 282, C to G; nucleotide 780, C to G; and nucleotide 1929, T to C), and three changes in the 3′ untranslated region (nucleotide 3072, A to G; nucleotide 3175, G to A; and nucleotide 3180, A to C). Each of these variants may represent sequencing errors or polymorphisms. The determination of the correct cDNA sequence and complete genomic structure, as well as identification of primers capable of amplifying the coding regions, will now allow analysis of the role intragenic mutation of *EPHA1* plays in human cancer.

REFERENCES

1. Schmidt, L., Duh, F. M., Chen, F. *et al.* (1997). Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. *Nature Genetics* **16**, 68–73.
2. Mulligan, L. M., Kwok, J. B., Healey, C. S. *et al.* (1993). Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* **363**, 458–60.

3. Mertens, F., Johansson, B., Hoglund, M. & Mitelman, F. (1997). Chromosomal imbalance maps of malignant solid tumours: a cytogenetic survey of 3185 neoplasms. *Cancer Research* **57**, 2765–80.
4. Pasquale, E. B. (1997). The Eph family of receptors. *Current Opinion in Cell Biology* **9**, 608–15.
5. Zisch, A. H. & Pasquale, E. B. (1997). The Eph family: a multitude of receptors that mediate cell recognition signals. *Cell and Tissue Research* **290**, 217–26.
6. Sajjadi, F. G., Pasquale, E. B. & Subramani, S. (1991). Identification of a new eph-related receptor tyrosine kinase gene from mouse and chicken that is developmentally regulated and encodes at least two forms of the receptor. *New Biology* **3**, 769–78.
7. Tuzi, N. L. & Gullick, W. J. (1994). eph, the largest known family of putative growth factor receptors. *British Journal of Cancer* **69**, 417–21.
8. Maru, Y., Hirai, H., Yoshida, M. C. & Takaku, F. (1988). Evolution, expression, and chromosomal location of a novel receptor tyrosine kinase gene, eph. *Molecular and Cellular Biology* **8**, 3770–6.
9. Connor, R. J. & Pasquale, E. B. (1995). Genomic organization and alternatively processed forms of Cek5, a receptor protein-tyrosine kinase of the Eph subfamily. *Oncogene* **11**, 2429–38.
10. Hirai, H., Maru, Y., Hagiwara, K., Nishida, J. & Takaku, F. (1987). A novel putative tyrosine kinase receptor encoded by the eph gene. *Science* **238**, 1717–20.