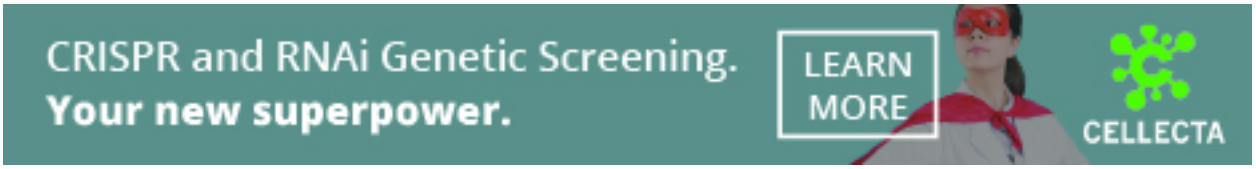
# The genomic structure of discoidin receptor tyrosine kinase.

M P Playford, R J Butler, X C Wang, et al.

*Genome Res.* 1996 6: 620-627
Access the most recent version at doi:10.1101/gr.6.7.620





LETTER

# The Genomic Structure of Discoidin Receptor Tyrosine Kinase

**Martin P. Playford, Robin J. Butler, Xiao Cun Wang, Roy M. Katso, Inez E. Cooke, and Trivadi S. Ganesan[1]**

Imperial Cancer Research Fund Molecular Oncology Laboratories, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

The discoidin domain receptor (DDR) is a new class of receptor tyrosine kinase that is distinguished by a unique extracellular domain homologous to the lectin Discoidin I found *Dictyostelium discoideum.* A cosmid was isolated from a human chromosome 6 cosmid library containing the DDR gene. A complete genomic contig of the DDR gene was constructed from seven subclones of the cosmid. The cosmid fragments were analyzed by PCR, sequencing, and comparison of genomic/cDNA sequence. The DDR gene is composed of 17 exons, ranging in size from 96 to 1014 bp, distributed along ~12 kb of genomic DNA. The extracellular domain is encoded by 8 exons of which three code for the discoidin domain. The transmembrane domain is encoded by 1 exon, the juxtamembrane domain by 3 exons, and the catalytic domain by 5 exons. The generation of the two splice variants of DDR, EDDR1 and EDDR2 are explained by the genomic structure. Exon 11 (111 bp in the juxtamembrane domain) is present in DDR and absent in the splice variant EDDR1. An inverted repeat of 20 bp was identified at the 3′ exon–intron junction of exon 11, which results in a lariat loop-like secondary structure. EDDR2 is generated because of a cryptic splice acceptor site that results in an extra 18 bp (6 amino acids) inserted 5′ of exon 14 in the catalytic domain. A polymorphic $(GT)_{17}$ repeat was identified in intron 5 with a hererozygosity of 0.71. The exon–intron structure of the DDR gene will be helpful in further understanding of its function and explains the possible structural basis for the two splice variants.

Receptor tyrosine kinases (RTKs) play an important role in cellular metabolism as key components of signal transduction pathways (Hunter 1987). They are composed of a unique extracellular domain, transmembrane domain, and a highly conserved catalytic domain (Hanks et al. 1988). In general subclasses of RTKs have been classified by the structure of the extracellular domains. These structural motifs vary between immunoglobulin-like domains, fibronectin domains, and cysteine-rich domains. RTKs are involved in cellular growth, differentiation, and development. Specific cognate ligands transmit signals through the receptor to the interior of the cell through defined pathways (Schlessinger and Ullrich 1992; Heldin 1995; Pawson 1995)

Recently, a new family of RTKs have been discovered, comprising of discoidin domain receptor (DDR), EDDR1 (also called cak, TrkE) (Dimarco et al. 1993; Johnson et al. 1993; Laval et al. 1994), and TKT (tyro 10) (Karn et al. 1993; Lai and Lemke 1994). They all share a unique extracellular region, which is homologous to a repeated domain found in factor VIII, factor V, milk fat globule membrane protein, and A5 antigen (Laval et al. 1994). This domain has similarity (49%) to a lectin called Discoidin I, found in *Dictyostelium discoideum* (Poole et al. 1981). The exact function of this protein is still unknown although it was initially suggested that it plays a role in cell adhesion because of its ability to agglutinate red blood cells in vitro (Rosen et al. 1973). This lectin was then shown to be an important protein involved in cell aggregation, cell–cell contact, and the promotion of cellular migration in *D. discoideum* on the basis of a minor homology to fibronectin (Springer et al. 1984). More recently this function has been questioned and instead it is thought that the protein is important in maintenance of morphology, cytoskeletal organization, and the ability to align with other cells during aggregation (Alexander et al. 1992). In analogy, the function of this discoidin family of RTKs is thought to be in cell–cell contact signaling pathways. This domain has also been shown to be important in the interaction of factor VIII with phospholipid and may have a

**[1]Corresponding author.**
**E-MAIL ganesan@icrf.icnet.uk; FAX 01865-222431.**



6:620–627 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

similar role in the milk fat globule membrane proteins (Arai et al. 1989; Stubbs et al. 1990). The receptor is expressed ubiquitously in all tissues, maximally in brain. Both DDR and EDDR1 receptors are expressed in epithelial cells, whereas TKT is expressed also in the stroma. DDR and EDDR1 are expressed as 116-kD proteins in intact cells, which are phosphorylated on tyrosine. Preliminary analysis of its expression at RNA level suggest that it is increased in malignant ovarian tumors and metastatic breast cancer (Laval et al. 1994; Alves et al. 1995; Barker et al. 1995).

DDR and EDDR1 are identical in all respects except for the presence of an additional exon (111 bp) in the juxtamembrane domain of DDR (Laval et al. 1994). An additional splice varaint (18 bp) of DDR was also reported (Alves et al. 1995). To understand the structural basis of these splice variants, the exon–intron structure of the entire gene was determined. Two mechanisms of splicing result in the production of the two variants EDDR1 and EDDR2. In addition, an intragenic polymorphic GT repeat was also indentified. The genomic structure of the DDR gene will be helpful in the further understanding of the normal function of this receptor family and its role in the pathogenesis of cancer.

## RESULTS

### Isolation of Cosmids

The entire cDNA of EDDR1 was used to probe a flow-sorted human chromosome 6 cosmid library, and three overlapping cosmids were identified. Cosmid 1 was shown to contain the entire cDNA by PCR using 5′- and 3′-specific primers, Southern blotting, and hybridization (Shelling et al. 1995). The entire cosmid was digested with restriction enzymes and subcloned into pBluescript for further analysis. Individual cosmid subclones containing the DDR gene were identified by restriction mapping, hybridization, and sequencing. Seven cosmid subclones encompassed the entire gene (Fig. 1).

### Intron–Exon Structure

The entire DDR gene is spread over ~12 kb and comprises 17 exons (Table 1). The 5′ and 3′ boundaries of each exon were identified by designing exon-specific primers (Table 2) and sequencing of cosmid subclones. The intron size was estimated by comparing size of PCR products using exon-specific primers on DDR cDNA and cosmid subclones (Table 1; Fig. 2). The entire extracellular domain was encoded by exons 1–8 of which 3 exons are coded for the discoidin domain. The transmembrane domain was encoded by a single exon (9). The juxtamembrane domain, which is unusually long, was encoded by 3 exons (10–12). The catalytic domain was encoded predominantly by 5 exons (13–17). The GT/AG rule was observed by all exons. The size of the exons varied from 96 to 1014 bp. The tran-

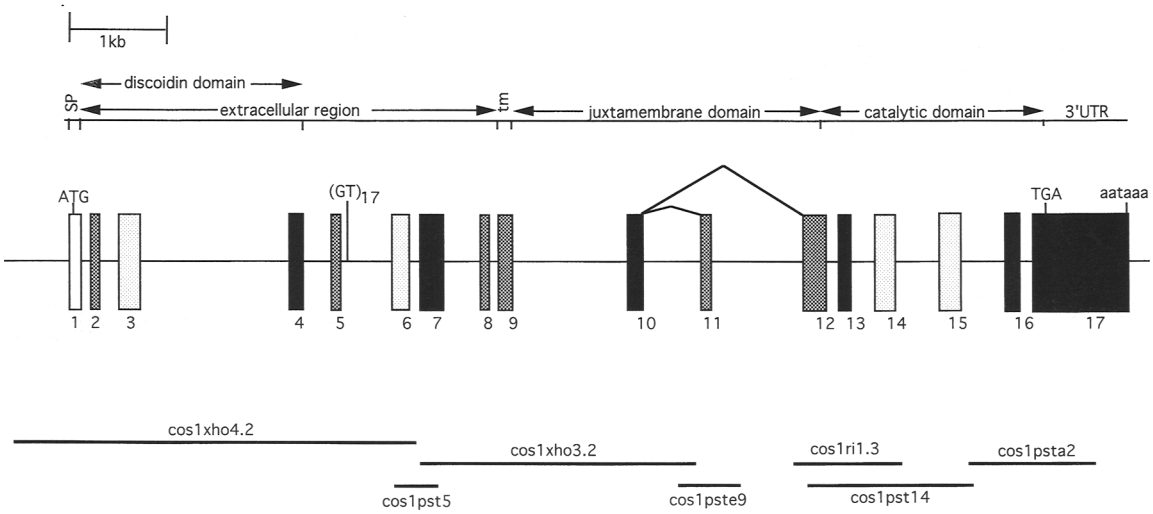



**Figure 1** Schematic diagram showing the domains of the receptor, the exon distribution, and the cosmid subclones. Exons are represented in boxes of three shades: Black, where the exon begins from the first base of the codon (phase 0); dark gray, from the second base (phase 1), and light gray, from the third base of the codon (phase 2). $(GT)_{17}$ indicates the position of the polymorphic repeat. (SP) Signal peptide. The alternative splicing of exon 11 leading to EDDR1 is indicated above exons 10–12.

**Table 1. Exon Structure of the DDR Gene**

| Exon No. | Exon size (bp) | Exon* location | Intron size (bp) | 5' intron/exon start sequence | 3' exon/intron end sequence |
|---|---|---|---|---|---|
| 1 | 127 | 295 | 93 | tcccctg**cag**AGATGC | ATCCTG**gt**gaggagac |
| 2 | 103 | 422 | 190 | ttttact**cag**CCAAGT | CAGCAG**gt**acttggca |
| 3 | 229 | 525 | 1560 | tcaccct**cag**GTTGGA | CAGGAG**gt**gagactgg |
| 4 | 148 | 754 | 284 | tggccct**cag**GTGATC | GGAGGG**gt**gagaggct |
| 5 | 100 | 902 | 537 | cctgccc**cag**ATGGAC | GGGCGG**gt**gtaagaaa |
| 6 | 187 | 1002 | 103 | tcttctc**cag**ACTGCA | ATGCAG**gt**gagtgagt |
| 7 | 247 | 1189 | 498 | ttccccc**cag**GTCCAC | TCTCTG**gt**aagccctg |
| 8 | 96 | 1436 | 90 | cttctcc**cag**ATGTGG | GCTTGG**gt**gagcaatct |
| 9 | 152 | 1532 | 1200 | cctccac**cag**AGCTGG | AGCAAG**gt**gggcacag |
| 10 | 166 | 1684 | 600 | ctatgac**aag**GCTGAA | GCTCTG**gt**aagacctt |
| 11 | 111 | 1850 | 950 | ctcccga**cag**CGTTGC | CCCAGG**gt**aagcccct |
| 12 | 245 | 1951 | 126 | accccg**cag**CCTACA | GGGGAG**gt**aaggagg |
| 13 | 128 | 2206 | 238 | ctcttct**cag**GTGCAC | TGCCAG**gt**gaggacca |
| 14 | 219 | 2334 | 450 | tgttctc**cag**GAATGA | CATCAG**gt**acctgctt |
| 15 | 235 | 2553 | 450 | cccttct**cag**CTACCC | CTCATG**gt**gagcagcc |
| 16 | 150 | 2788 | 110 | tgcatcc**cag**GGGAAG | CGGCAG**gt**cagagtgg |
| 17 | 1014 | 2938 | | cacaatg**cag**GTGTAC | ACGGTG/3'UTR |

(*) The exon location is from the EDDR1 sequence Z29093.

scriptional start site for the gene has not been determined, so there may be additional exons 5′ of exon 1. Introns ranged in size from 93 bp to ~1200 bp. There was no significant difference between the frequency of different phases in the splicing of exons.

### Splicing

The variant EDDR1 cDNA is produced as a result of alternative splicing of exon 10 to 12, resulting in deletion of exon 11. The preferred use of the splice acceptor site on the intron/exon boundary of exon 12 is possibly explained by the presence of an inverted repeat of 20 bp at the 3′ end of exon 11 (Fig. 3a). The predicted secondary structure of the pre-mRNA using the modeling program MFold (Jaeger et al. 1989; Zuker 1989) suggests a possible hairpin loop (lariat-like structure) because of the inverted repeat. On careful examination of the 5′ intron/exon boundary sequence of exon 14, it was apparent that there were two possible splice acceptor sites (Fig. 4). The use of the proximal 5′ acceptor site would result in the production of a cDNA with an extra 18 bp (coding for an extra 6 amino acids, FSLFSR) without disruption of the open reading frame (EDDR2). The use of the alternative cryptic splice site results in the substitution of the last amino acid serine by arginine in exon 13 of DDR. The predicted secondary structure of the pre-mRNA using the program MFold (Jaeger et al. 1989; Zuker 1989), corresponding to intron 13 and exon 14, demonstrated a long stable hairpin loop between the 5′ and 3′ sequence of intron 13 (Fig. 3b).

### Polymorphic CA Repeat

A polymorphic dinucleotide GT repeat was identified on sequencing of cosmid subclone cos1xho4.2 in intron 5. This has 17 GT repeats, and four alleles of 158 bp (25%), 160 bp (47%), 162 bp (3%), and 174 bp (25%) were identified in normal genomic DNA when amplified by PCR using flanking primers. The polymorphic repeat is heterozygous (71%) when examined on 18 unrelated individuals and is inherited in a Mendelian fashion when analyzed on three reference Centre d'Etudes du Polymorphisme Humain (CEPH) families (Gyapay et al. 1994).

## DISCUSSION

DDR was isolated from a number of sources, in-

**Table 2. Primer Sequences**

| Sense primer (5'-3') for 5' intron / exon boundary | Antisense primer (5'-3') for 3' exon / intron boundary |
|---|---|
| **312**ccttaggcccgagggatcag | gcagagatgtcactgtctgg**483** |
| **312**ccttaggcccgagggatcag | gagaactccttgcccagg**681** |
| **603**ggatctacaacgactgcacc | gtagaagcgaaccagtc **838** |
| **764**gcaatgaggaccctgag | cgcccacggtatgtcc**1001** |
| **958**gtgtacctcaacgactc | ctggagaagctgtggttgc**1230** |
| **1060**aggaagagtcaggagctgcg | agcaggagcaggatgatg**1632** |
| **1414**agcgaaatctccttcatctctg | agcaggagcaggatgatg**1632** |
| **1436**atgtggtgaacaattcctc | agcaggagcaggatgatg**1632** |
| **1436**atgtggtgaacaattcctc | gttgttgatgaggatagtgtc**1759** |
| **1740**cactatcctcatcaacaaccg | cctgggtgttggtgggtttg**1961** |
| **1850**cgttgctgctctccaatccagcctac | gggacgctgttctggggagg**2037** |
| **2133**gccccccagagtggatt | ctggttgaggtcgccgttctc**2470** |
| **2218**gaggtcgacagccctcaa | ctggttgaggtcgccgttctc**2470** |
| **2429**gcatgattactgactac | gtacaaagttgagtgtggc**2627** |
| **2582**agatcgcctccggcatgcg | tcgtcggtgagctgcccaaa**2889** |
| **2789**ggaagttcacgactgcga | ttgttttagtgtcactgg**3143** |

The position of the first nucleotide of each primer is shown in boldface type and numbered according to the EDDR1 sequence Z29093.

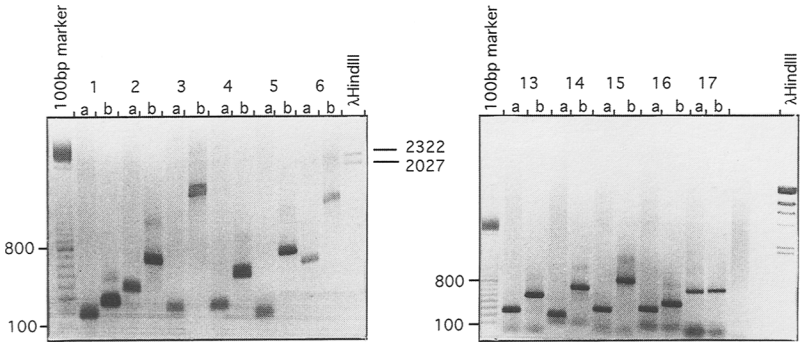



**Figure 2** Ethidium bromide-stained photograph of agarose gel (1%) showing separation of PCR-amplified products. The markers are a 100-bp ladder and λ *Hin*dIII digest; (a) PCR product from EDDR1/DDR cDNA; (b) PCR product from genomic cosmid clones using the same primer pair. (Lanes *1–6, 13–17*) The amplified products for each corresponding primer pair on cDNA and genomic cosmid subclones, delineating the intron size between each exon, which is the difference between the two products. Primer pairs were chosen from the DDR cDNA.

cluding placenta, metastatic breast cancer lymph nodes, breast cancer cell line MCF-7, and HeLa cell line (Johnson et al. 1993; Perez et al. 1994; Alves et al. 1995; Barker et al. 1995). The splice variant EDDR1 was isolated from at least two sources, an ovarian cancer cell line SKOV-3 and human keratinocytes (Dimarco et al. 1993; Laval et al. 1994). A third cDNA variant isolated from a human fetal brain library was reported (MCK-10c) recently (Alves et al. 1995). DDR and EDDR1 are identical except for an insertion of 37 amino acids in DDR in the juxtamembrane domain as reported previously (Laval et al. 1994). A possible mechanism for production of EDDR1 as a splice variant maybe explained by the identification of an inverted repeat at the 3′ end of the exon 11/intron boundary (Fig. 1). RNA secondary structure modeling using the program MFold (Jaeger et al. 1989; Zuker 1989) demonstrates a hairpin structure attributable to this repeat (Fig. 3a). Such a hairpin structure has been shown to be important in exon selection (Balvay et al. 1993). It is possible to speculate that although there are no such repeats at the 5′ end of exon 11, the presence of this repeat at the 3′ end explains the production of the alternatively spliced transcript EDDR1. This exon contains several motifs such as the NPXY sequence, the phosphotyrosine binding (PTB) domain, and SH3-domain protein binding motifs. The NPXY motif has been suggested as an important requirement for internalization in coated pits and is observed in the human lipoprotein receptor and in other receptor tyrosine kinases (Chen et al. 1990; Bansal and Gierasch 1991). However, in RTKs it may not be important for ligand-dependent internalization of receptor. Phosphorylation of the tyrosine in the NPXY motif is not important for the internalization of the insulin receptor, whereas it is necessary in the platelet-derived growth factor (PDGF) receptor (MacLeod et al. 1991; Kaburagi et al. 1993; Mori et

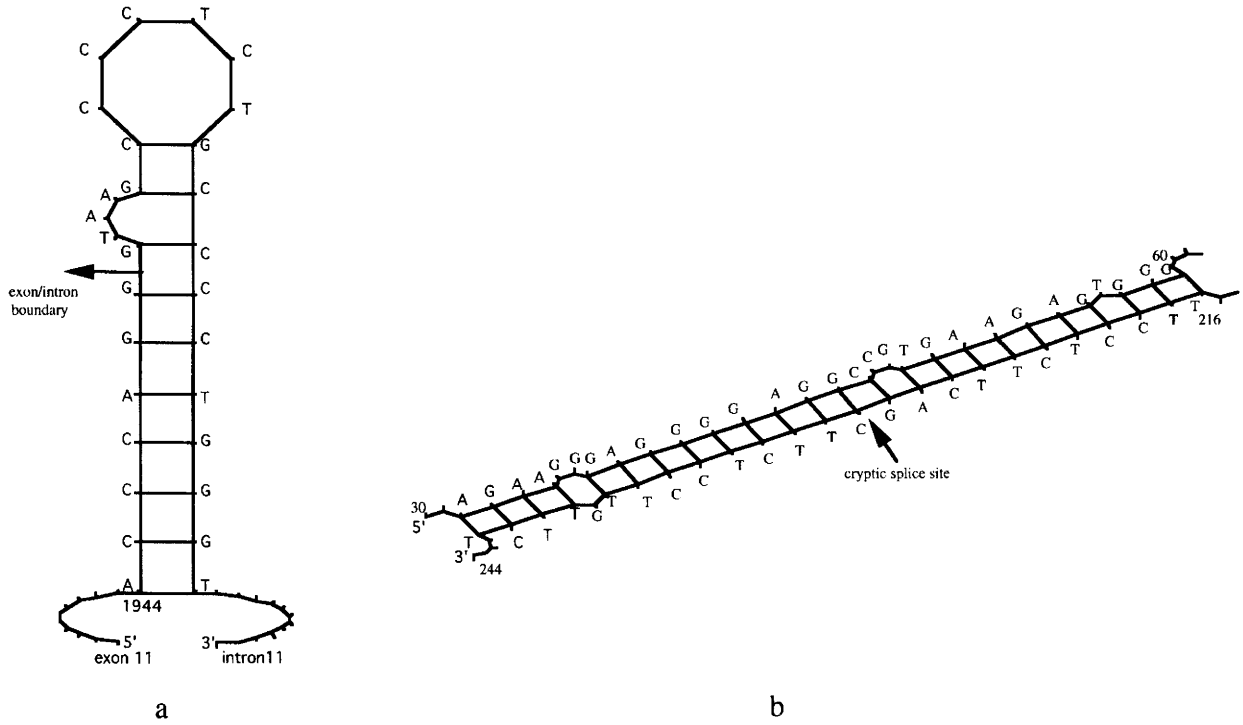



**Figure 3** Diagram showing squiggle plot of RNA secondary structure using the program MFold and PlotFold. (*a*) The displayed structure is from the nucleotide of exon 11 to the subsequent intron sequence (nucleotide 1944 of DDR on the diagram). (*b*) The displayed structure is of intron 13 from nucleotide 32 up to the intron/exon boundary of exon 14 showing the hairpin structure. The arrow points to the cryptic splice site.

al. 1993). Recently, it has been shown that the consensus PTB motif of the SH2 adaptor protein Shc is NXXY and the asparagine is an essential requirement (Kavanaugh et al. 1995; Laminet et al. 1996). This might be a universal mechanism by which proteins with a PTB domain bind to phosphorylated tyrosine residues. Motifs that recognize SH2 domain proteins are normally found outside the juxtamembrane region, whereas in DDR this might serve an analogous function as the carboxy-terminal tail in other RTKs. In addition, there is a putative binding motif for SH3 domain proteins within exon 11, PGPPTP, that has been described for non-RTKs. The presence of these important motifs in the alternatively spliced exon 11 suggests different functional roles for these two proteins, DDR and EDDR1. Previously, it was shown that at the RNA level the expression of these two isoforms is similar in a panel of cell lines and tissues (Laval et al. 1994), but the pattern of expression is different at a protein level (X.C. Wang, M.P. Playford, and T.S. Ganesan, unpubl.).

The second splice variant (EDDR2) uses a cryptic splice site 5′ of the preferred splice site in exon 14 (Fig. 4). This results in the addition of 18 bp (6 amino acids) to exon 14. Interestingly, modeling of intron 13 using the program MFold (Jaeger et al. 1989; Zuker 1989) shows that there is a long inverted repeat bringing together in apposition the 5′ end of intron 13 with the 3′ end with a long stem (48 nucleotides). The choice of 3′ splice sites is believed to be determined by the position of branch formation, the first encountered AG being selected by a 5′ → 3′ molecular scanning mechanism (Smith et al. 1989). Secondary structure around the first encountered splice site (CAG) can result in the second one chosen normally as was shown for the actin gene of *Kluveromyces lactis* (Deshler and Rossi 1991). Thus, in general transcription favors the natural splice site, and perhaps occasionally the cryptic splice site. This splice variant is normally associated with the DDR variant rather than EDDR1 (M.P. Playford and T.S. Ganesan, unpubl.). Thus, it is possible to speculate that production of both splice variants DDR may be explained by RNA folding and secondary structure. The accuracy of the predicted RNA secondary structure using MFold was shown to be 49% on average (Zuker

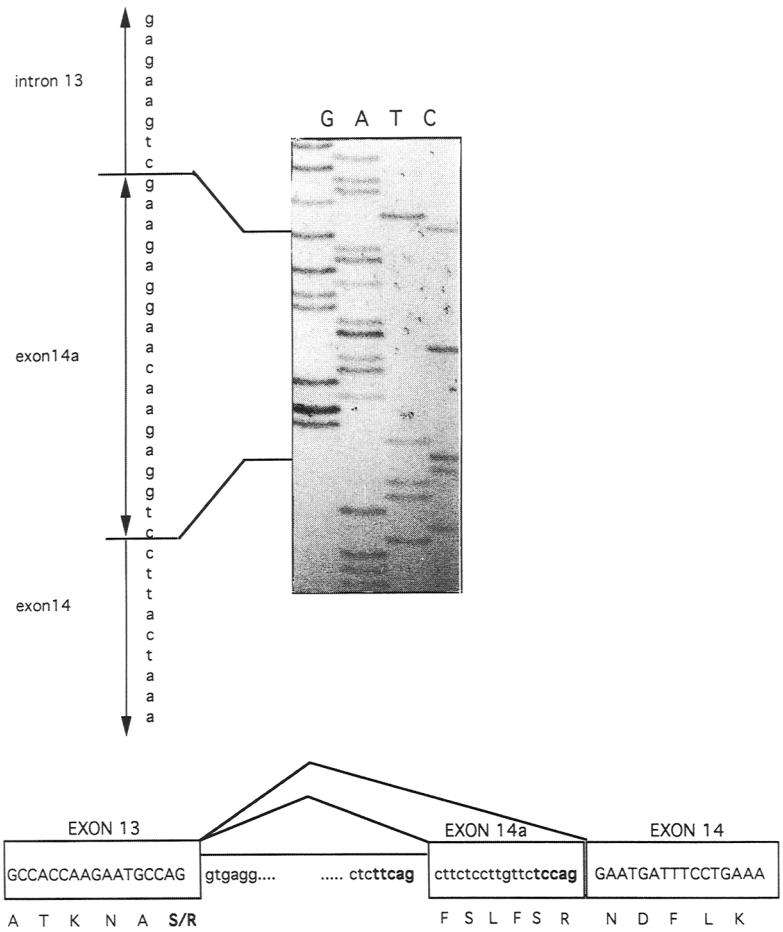



**Figure 4** Sequencing gel photograph (6%) demonstrating the sequence of the 5′ boundary for exon 14 determined using an exon 14-specific reverse primer and containing the two splice acceptor sites. The schematic diagram illustrates the extra exon 14a from the proximal splice acceptor site. Amino acids are shown in single-letter codes. Bases shown in lowercase and boldface letters represent the splice acceptor sites. The two possible amino acids are shown in boldface type.

and Jacobson 1995); therefore, site-directed mutagenesis of appropriate nucleotides that abolish the secondary structures are required to demonstrate formally that such a mechanism operates in vivo.

The identification of an intragenic polymorphic repeat that is heterozygous (71%) will be useful in genetic linkage analysis as the gene is mapped to 6q16.3, a region where several disease genes have been mapped. The exon–intron structure of only a few RTKs [epidermal growth factor (EGF) receptor, fibroblast growth factor (FGF) receptor, PDGF receptor, *fms,* c-*kit, FLT3,* and *Ret*] have been determined (Agnes et al. 1994; Kwok et al. 1993). The transmembrane domain is encoded by one exon as in others. There is no kinase insert in the catalytic domain of DDR, which is encoded by only 5 exons unlike other RTKs of class III. In general, the exon–intron structure of this novel RTK is similar to that of other tyrosine kinases. Splicing is a common mechanism for production of proteins with different functions. It remains to be ascertained whether such is the case for splice variants of DDR.

## METHODS

### Isolation and Characterization of Cosmids

The entire EDDR1 cDNA was used to probe a flow-sorted chromosome 6 library, and 3 cosmids were isolated. Cosmid 1 was shown to contain the gene by PCR and hybridization (Shelling et al. 1995). The cosmid was digested with several restriction enzymes (*Xho*I, *Pst*I, etc.) and cloned in pBluescript SKII.

### PCR

PCR was performed using a denaturation temperature of 94°C (90 sec), annealing temperature of 60°C (1 min), and extension at 72°C (2 min) using *Taq* polymerase, dNTPs, and magnesium under standard reaction conditions. Primers were used to amplify the cDNA and genomic clones, and products were separated on agarose gel electrophoresis (1%) to compare sizes and estimate introns (Fig. 2; Table 2). The sequences of the primers were chosen from the DDR and EDDR1 cDNA using the program Amplify 1.2 (Table 2). The conditions for the primers to amplify the GT repeat were identical as above except for a concentration of 3.5 mM of $MgCl_2$. The forward primer was 5′-TTTGGGGTGGGAGGAGGAC-3′; the reverse prim-er was 5′-TGGTCACCCTAGTGCTACCC-3′.

### Sequencing

Sequencing was performed using the Sequenase 2.0 (U.S. Biochemical) and Sequencing Pro (Cambridge Biosciences) kits, and gene specific/vector specific primers. Intron/exon boundaries were determined by comparison of cDNA and genomic sequences.

## ACKNOWLEDGMENTS

We thank Dr. S. Laval, Dr. A.N. Shelling, and other members of the laboratory for advice. This work is supported by

the Imperial Cancer Research Fund, Wellbeing, and a European Union Human Capital and Mobility Grant. The sequence data described in this paper have been submitted to the EMBL data library under accession no. X98208.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

Agnes, F., B. Shamoon, C. Dina, O. Rosnet, D. Birnbaum, and F. Galibert. 1994. Genomic structure of the downstream part of the human FLT3 gene: Exon/intron structure conservation among genes encoding receptor tyrosine kinases (RTK) of subclass III. *Gene* **145:** 283–287.

Alexander, S., L.M. Sydow, D. Wessels, and D.R. Soll. 1992. Discoidin proteins of Dictyostelium are necessary for normal cytoskeletal organization and cellular morphology during aggregation. *Differentiation* **51:** 149–161.

Alves, F., W. Vogel, K. Mossie, B. Millauer, H. Hofler, and A. Ullrich. 1995. Distinct structural characteristics of discoidin I subfamily receptor tyrosine kinases and complementary expression in human cancer. *Oncogene* **10:** 609–618.

Arai, M., D. Scandella, and L.W. Hoyer. 1989. Molecular basis of Factor VIII inhibition by human antibodies. Antibodies that bind to factor VIII light chain prevent the interaction of factor VIII with phospholipids. *J. Clin. Invest.* **83:** 1978–1984.

Balvay, L., D. Libri, and M.Y. Fiszman. 1993. Pre-mRNA secondary structure and the regulation of splicing. *BioEssays* **15:** 165–169.

Bansal, A. and L.M. Gierasch. 1991. The NPXY internalization signal of the LDL receptor adopts a reverse-turn conformation. *Cell* **67:** 1195–1201.

Barker, K.T., J.E. Martindale, P.J. Mitchell, T. Kamaalti, M.J. Page, D.J. Phippard, T.C. Dale, B.A. Gusterson, and M.R. Crompton. 1995. Expression patterns of the novel receptor-like tyrosine kinase, DDR, in human breast tumours. *Oncogene* **10:** 569–575.

Chen, W.-J., J. Goldstein, and M.S. Brown. 1990. NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. *J. Biol. Chem.* **265:** 3116–3123.

Deshler, J.O. and J.J. Rossi. 1991. Unexpected point mutations activate cryptic 3′ splice sites in perturbing a natural secondary structure within a yeast intron. *Genes & Dev.* **5:** 1252–1263.

Dimarco, E., N. Cutuli, L. Guerra, R. Cancedda, and M. Deluca. 1993. Molecular cloning of trkE, a novel trk-related putative tyrosine kinase receptor isolated from normal human keratinocytes and widely expressed by normal tissues. *J. Biol. Chem.* **268:** 24290–24295.

Gyapay, G., J. Morissette, A. Vignal, C. Dib, C. Fizames, P. Millasseau, S. Marc, B. Bernardi, M. Lathrop, and J. Weissenbach. 1994. The 1993-94 Genethon human genetic linkage map. *Nature Genet.* **7:** 246–339.

Hanks, S.K., A.M. Quinn, and T. Hunter. 1988. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* **241:** 42–52.

Heldin, C.-H. 1995. Dimerization of cell surface receptors in signal transduction. *Cell* **80:** 213–223.

Hunter, T. 1987. A thousand and one protein kinases. *Cell* **50:** 823–829.

Jaeger, J.A., D.H. Turner, and M. Zuker. 1989. Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci.* **86:** 7706–7710.

Johnson, J.D., J.C. Edman, and W.J. Rutter. 1993. A receptor tyrosine kinase found in breast carcinoma cells has an extracellular discoidin I-like domain. *Proc. Natl. Acad. Sci.* **90:** 5677–5681.

Kaburagi, Y., K. Momomura, R. Yamammotohonda, K. Tobe, Y. Tamori, H. Sakura, Y. Akamura, Y. Yazaki, and T. Kadowak. 1993. Site-directed mutagenesis of the juxta-membrame domain of the insulin receptor. *J. Biol. Chem.* **268:** 16610–16622.

Karn, T., U. Holtrich, U. Brauninger, B. Bohme, G. Wolf, H. Rubsamen-Waigmann, and K. Strebhardt. 1993. Structure, expression and chromosomal mapping of TKT from man and mouse: A new subclass of receptor tyrosine kinases with a factor VIII like domain. *Oncogene* **8:** 3443–3440.

Kavanaugh, W.M., C.W. Turck, and L.T. Williams. 1995 PTB domain binding to signaling proteins through a sequence motif containing phosphotyrosine. *Science* **268:** 1177–1179.

Kwok, J.B.J., E. Gardner, J.P. Warner, B.A.J. Ponder, and L.M. Mulligan. 1993. Structural analysis of the human Ret proto-oncogene using exon trapping. *Oncogene* **8:** 2575–2582.

Lai, C. and G. Lemke. 1994. Structure and expression of the Tyro 10 receptor tyrosine kinase. *Oncogene* **9:** 877–883.

Laminet, A.A., G. Apell, L. Conroy, and W.M. Kavanaugh. 1996. Affinity, specificity, and kinetics of the interaction of the SHC phosphotyrosine binding domain with Asparagine-X-X-phosphotyrosine motifs of growth factor receptors. *J. Biol. Chem.* **271:** 264–269.

Laval, S., R. Butler, A.N. Shelling, A.M. Hanby, R. Poulsom, and T.S. Ganesan. 1994. Isolation and characterisation of an epithelial specific receptor kinase

from an ovarian cancer cell line. *Cell Growth & Diff.* **5:** 1173–1183.

MacLeod, J.N., S.A. Liebhaber, M.H. MacGillivray, and N.E. Cooke. 1991. Identification of a splice-site mutation in the human growth hormone-variant gene. *Am. J. Hum. Genet.* **48:** 1168–1174.

Mori, S., L. Ronnstarnd, L. Claesson-Walsh, and C.H. Heldin. 1993. A tyrosine residue in the juxta-membrame domain is critical for ligand mediated endocytosis. *J. Biol. Chem.* **269:** 4917–4921.

Pawson, T. 1995. Protein modules and signaling networks. *Nature* **373:** 573–579.

Perez, J.L., X. Shen, S. Finkernagel, L. Sciorra, N. Jenkins, D.J. Gilbert, N.G. Copeland, and T.W. Wong. 1994. Indentification and chromosomal mapping of a receptor tyrosine kinase with a putative phospholipid binding sequence in its ectodomain. *Oncogene* **9:** 211–221.

Poole, S., R.A. Firtel, E. Lamar, and W. Rowekamp. 1981. Sequence and expression of the discoidin I gene family in Dictyostelium discoideum. *J. Mol. Biol.* **153:** 273–289.

Rosen, S.D., J.A. Kafka, D.L. Simpson, and W.H. Barondes. 1973. Developmentally regulated carbohydrate binding protein in Dictyostelium discoideum. *Proc. Natl. Acad. Sci.* **70:** 2554–2557.

Schlessinger, J. and A. Ullrich. 1992. Growth factor signaling by receptor tyrosine kinases. *Neuron* **9:** 383–391.

Shelling, A.N., R. Butler, T. Jones, S. Laval, J.M. Boyle, and T.S. Ganesan. 1995. Localisation of an epithelial-specific receptor kinase (EDDR1) to chromosome 6q16. *Genomics* **25:** 584–587.

Smith, C.W.J., E.B. Porro, J.G. Patton, and B. Nadal-Ginard. 1989. Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. *Nature* **342:** 243–247.

Springer, W.R., D.N.W. Cooper, and S.H. Barondes. 1984. Discoidin I is implicated in cell-substratum attachment and ordered cell migration of dictyostelium discoideum and resembles fibronectin. *Cell* **39:** 557–564.

Stubbs, J.D., C. Lekutis, K.L. Singer, B. Anhthu, D. Yuzuki, U. Srinivasan, and G. Parry. 1990. cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of epidermal growth factor-like domains linked to factor VIII-like sequences. *Proc. Natl. Acad. Sci.* **87:** 8417–8421.

Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science* **244:** 48–52.

Zuker, M. and A.B. Jacobson. 1995. "Well-determined" regions in RNA secondary structure prediction: analysis of small subunit ribosomal RNA. *Nucleic Acids Res.* **23:** 2791–2798.

*Received February 23, 1996; accepted in revised form May 10, 1996.*