Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von Willebrand factor A domains

Delphine GILGÈS*, Marie-Antoinette VINIT*, Isabelle CALLEBAUT†, Laure COULOMBEL, Valère CACHEUX‡, Paul-Henri ROMEO* and Isabelle VIGON*[1]

*INSERM U474, maternité Port-Royal, 123 Bld de Port-Royal 75014 Paris, France, †LMCP, CNRS UMRC 7590, 75005 Paris, France, and ‡INSERM, U468, Hopital Henri-Mondor, 94010 Créteil, France

To identify extracellular proteins with epidermal growth factor (EGF) domains that are potentially involved in the control of haemopoiesis, we performed degenerate reverse-transcriptase-mediated PCR on the murine bone-marrow stromal cell line MS-5 and isolated a new partial cDNA encoding EGF-like domains related to those in the Notch proteins. Cloning and sequencing of the full-length cDNA showed that it encoded a new extracellular multi-domain protein that we named polydom. This 387 kDa mosaic protein contained a signal peptide followed by a new association of eight different protein domains, including a pentraxin domain and a von Willebrand factor type A domain, ten EGF domains, and 34 complement control protein modules. The human polydom mRNA is strongly expressed in placenta, its expression in the other tissues being weak or undetectable. The particular multidomain structure of the encoded protein suggests an important biological role in cellular adhesion and/or in the immune system.

Key words: haemopoiesis, stroma, placenta, pentraxin domain.

INTRODUCTION

Pluripotent haemopoietic stem cells are functionally defined by their capacity for self-renewal and their ability to generate cells of all haemopoietic lineages throughout the life. The fate of haemopoietic stem cells is strikingly dependent on their interactions with the bone marrow microenvironment composed of heterogeneous stromal cells and extracellular matrix components. Significant advances have been made in the knowledge of proteins expressed by the bone marrow microenvironment and implicated in the biology of haemopoietic stem cells. These include the identification of many cytokines and their receptors, adhesion molecules and their bound cell ligands and constituents of the extracellular matrix. Nevertheless, studies *in vitro* and *in vivo* demonstrate that the control of haemopoiesis is still dependent on as yet uncharacterized proteins.

One unanswered question concerns the determination of lateral inhibition as a regulator of the haemopoietic stem cell fate. This signalling process is mediated in the neural system by receptors of the Notch family and their ligands [1]. Recent investigations have shown the expression of Notch proteins in $CD34^+$ haemopoietic progenitor cells and in bone-marrow stromal cells and have revealed a role of Notch protein in T-cell commitment [2–5]. To analyse further the role of Notch proteins in the control of haemopoiesis, we decided to study the expression of Notch genes in the bone-marrow stromal cell line MS-5, which is known to support human and murine primitive haemopoiesis [6–8]. As the extracellular domain of all Notch proteins contains epidermal growth factor (EGF) domains, which are also found in other proteins essential for neuronal development [9–12], we performed reverse-transcriptase-mediated PCR (RT–PCR) with degenerate primers located in regions encoding these EGF repeats and identified a new partial cDNA with EGF repeats sharing strong similarities with the EGF domains found in the Notch family. Cloning of the full-length coding sequence revealed that it encodes a very large protein containing a putative signal peptide, no transmembrane region and a completely new association of eight different protein domains. The structure of this protein, which we have named polydom, its expression pattern and the presence of 34 complement control protein (CCP) modules, a pentraxin (PTX) domain, a von Willebrand factor type A (vWF-A) domain and multiple EGF-like repeats suggest that it might have a role in cell adhesion and/or act in the immune system.

MATERIALS AND METHODS

Partial cDNA cloning of genes encoding proteins with EGF-like motifs

Total RNA was prepared from the MS-5 cell line by using TRIzol reagent (Life Technologies) and poly(A)$^+$ RNA was obtained with the MPG mRNA purification kit (Quantum Biotechnologies). A cDNA primed with oligo(dT)$_{12–18}$ was synthesized from 1 μg of poly(A)$^+$ RNA from MS-5 cells with Superscript II reverse transcriptase (RT) (Life Technologies); one-tenth of this cDNA was used as template for PCR. Degenerate oligonucleotides were designed on the basis of conserved amino acid sequences localized in the EGF-like domains of Notch ligands and receptors: the sense primer, 5′-TGYCA-(I)AAYGG(I)GSIAC(I)TGY-3′, corresponded to the amino acid sequence C(Q/H)NGGTC (single-letter codes) and the antisense primer, 5′-RCA(I)GG(I)YY(I)GA(I)AGRCAYTC-3′ to

Abbreviations used: CCP, complement control protein; CRP, complement-reactive protein; EGF, epidermal growth factor; EGFca, EGF-like domain bearing Ca^{2+}-binding site; FISH, fluorescence *in situ* hybridization; HPRT, hypoxanthine phosphoribosyl transferase; HYR, hyalin repeat; PTX, pentraxin; RT, reverse transcriptase; SAP, serum amyloid protein; STT2R, similar to thyroglobulin type 2 repeats; TNF-R, tumour necrosis factor receptor; vWF-A, von Willebrand factor type A.

[1] To whom correspondence should be addressed (e-mail vigon@cochin.inserm.fr).

The nucleotide and protein sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AF206329.

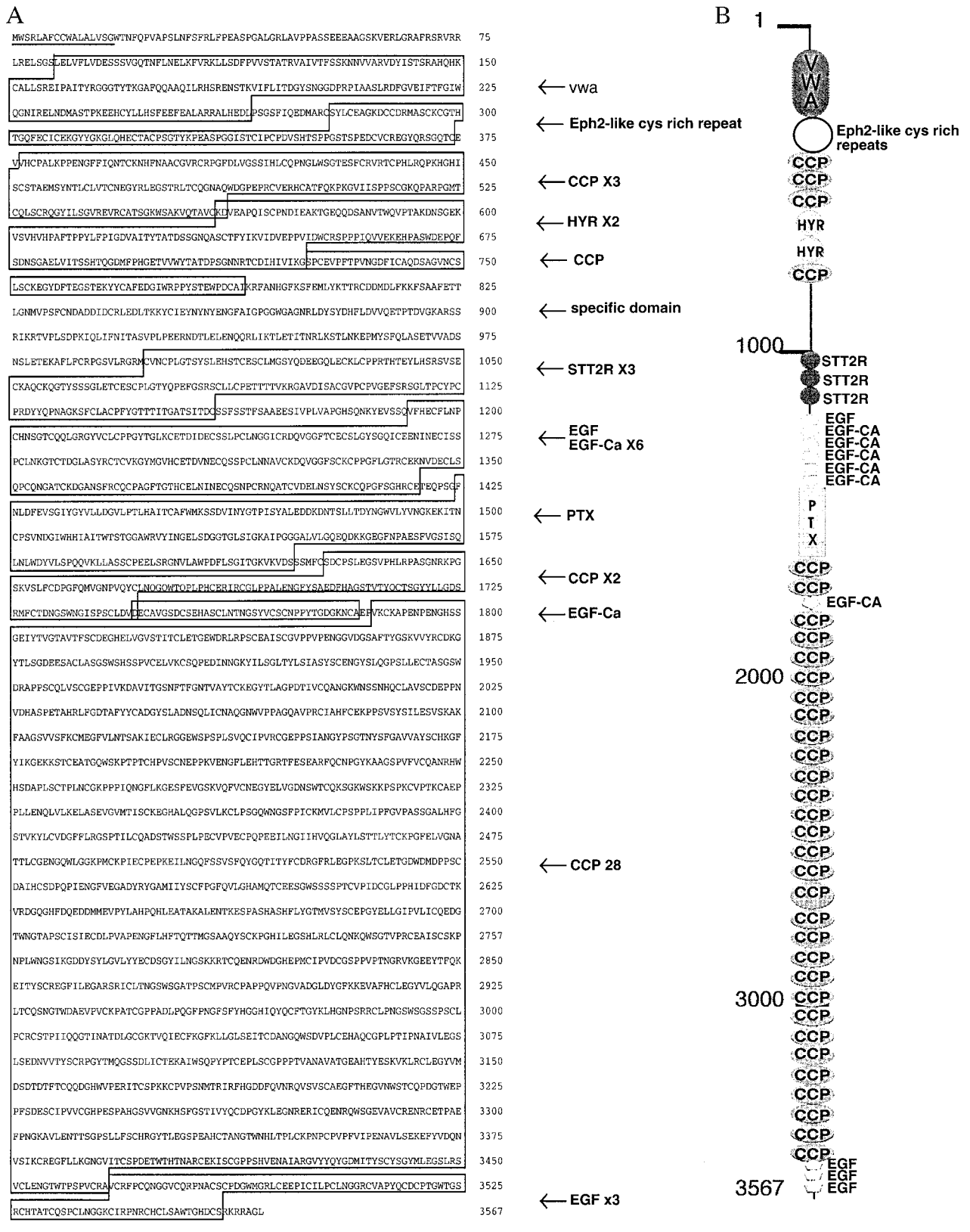

Figure 1 Amino acid sequence and modular organization of polydom protein

(**A**) Predicted amino acid sequence of polydom. Amino acid positions are indicated at the right of the sequence; the signal peptide is underlined. The various domains are boxed and identified by labelled arrows. (**B**) Schematic representation of the protein. The protein is drawn to scale. Abbreviations: EGF, EGF-like domain; Eph2, ephrin 2; vwa, von Willebrand factor type A.

the amino acid sequence ECLS(N/G)PC. PCR reactions were performed in a DNA thermal Cycler (Perkin-Elmer/Cetus) for five cycles under the following conditions: denaturation at 94 °C for 15 s, annealing at 40 °C for 30 s and elongation at 72 °C for 1 min followed by 30 cycles of denaturation at 94 °C for 15 s, annealing at 50 °C for 30 s and elongation at 72 °C for 1 min, in 50 μl with 1 × *Taq* polymerase buffer (Perkin-Elmer/Cetus), 1.5 mM $MgCl_2$, each dNTP at 0.2 mM, each primer at 10 pM and 2.5 units of *Taq* polymerase (Perkin-Elmer/Cetus). PCR products were then subcloned in the TA cloning vector (Stratagene) and sequenced with an ABI 373 DNA automated sequencer (Applied Biosystems).

Isolation of the murine polydom cDNA

Oligo$(dT)_{12–18}$ or random hexamer-primed cDNA libraries were constructed from 2 μg of poly(A)$^+$ RNA from MS-5 cells with the Superscript Choice System kit (Life Technologies). The fractions corresponding to the longest cDNA species were cloned in Lambda ZapII vector (Stratagene Cloning Systems); 10^6 clones of these libraries were first screened with an α-^{32}P-random-labelled probe corresponding to the PCR product containing EGF domains, then libraries were successively screened with the positive clones obtained.

Sequence analysis

The protein sequence was analysed with multiple approaches. Using the polydom sequence (or domains included in it), we searched the National Center of Biological Information non-redundant database (NR) for similarities by using the BLAST2 and Ψ BLAST programs [13]; various domain databases such as Pfam [14], Prosite [15] and SMART [16] were also scanned for matches with the polydom sequence. Hydrophobic cluster analysis [17] was also used to identify protein domain repeats. This two-dimensional method combined primary structure (one-dimensional) comparison with prediction of secondary structure (two-dimensional) and was therefore particularly useful in detecting similar folds sharing very low levels of sequence identity (below 20–25 % identity).

Expression pattern analysis

Clontech Northern blots (Clontech Laboratories) were hybridized with randomly primed α-^{32}P-labelled probes (Roche Molecular Biochemicals) with Expresshyb™ solution (Clontech Laboratories). Clontech Northern blots contain approx. 1 μg (human) or 2 μg (murine) of poly(A)$^+$ RNA from each tissue and the quantity of poly(A)$^+$ RNA was adjusted to reflect approximately equal amounts of actin mRNA.

Semi-quantitative RT–PCR was performed with cDNA species primed with random primers. These cDNA species were synthesized from 1 μg of total RNA with the Superscript II RT; and one-tenth of the cDNA was used as a template for PCR with the murine polydom forward primer (5′-GGCAAGGCCAGATCGTCA-3′) and reverse primer (5′-TTCCGTGTGAGTCCTGGGGGG-3′). RT–PCRs were normalized with hypoxanthine phosphoribosyl transferase (HPRT) mRNA expression with the murine forward (5′-GCTGGTGAAAAGGACCTCT-3′) and reverse (5′-CACAGGACTAGAACACCTGC-3′) primers.

Human chromosomal localization

Human metaphase cells for fluorescence *in situ* hybridization (FISH) were prepared from phytohaemagglutinin-stimulated lymphocytes by using the thymidine synchronization and bromodeoxyuridine release technique for banding. All the probes were biotinylated by the nick translation procedure (Life Technologies). A pool of biotin–polydom human probes cloned from human bone-marrow stroma mRNA and then sequenced was used for the FISH, which was performed as described [18] with some modifications. For the probe mixture, 20 ng of each probe was combined with 5 μg of the human Cot-1 DNA in a 10 μl hybridization solution [50 % (v/v) formamide/1 × SSC/10 % (v/v) dextran sulphate]. After denaturation, repetitive sequences were left to reanneal for 20 min at 37 °C before hybridization. Detection of labelled probe was achieved with fluorescein-conjugated avidin followed by one round of immunological signal amplification with a biotinylated anti-avidin antibody. Chromosomes were counterstained and banded by 4,6-diamidino-2-phenylindole in Antifade (0.2 μg/μl; Vector Laboratories); signals were detected with a Leitz microcoscope and captured on CCD computer.

Generation of murine polyclonal polydom antibody

A cDNA corresponding to the specific polydom domain (see Figure 1A, positions 789–997) was subcloned in-frame into the expression vector pGex4T-1 (Amersham Pharmacia Biotech). Recombinant polydom was produced and purified with the glutathione S-transferase gene fusion system (Amersham Pharmacia Biotech). The purified glutathione S-transferase–polydom was then used to immunize rabbits (Eurogentec). The crude antiserum specifically recognized TNT polydom products *in vitro* containing this specific domain (results not shown).

Cell line transfections and Western-blot analysis

Full-length polydom cDNA was subcloned in frame into the pcDNA3-1/Myc-His (Invitrogen) expression vector; Cos-7 cells were transiently transfected with vector alone or polydom construct with the LipofectAMINE reagent (Life Technologies). Supernatants of Cos-7 cells were collected 48 h after the transfections and the proteins were separated by SDS/PAGE [4–15 % (w/v) gel] (Bio-Rad) followed by electrophoretic transfer to PVDF nitrocellulose filters (Amersham Pharmacia Biotech). Proteins were detected with monoclonal anti-Myc clone 9E10 (Roche), which recognizes the epitope EGKLISEEDL sequence derived from the human c-Myc protein, or with the polyclonal polydom antibody. Labelled proteins were enhanced by chemiluminescence (ECL®; Amersham).

RESULTS

Characterization of an mRNA encoding a protein with multiple protein modules

Because the Delta/Notch family of proteins seems to be involved in the biology of haemopoietic stem cells, we looked for the expression of these proteins in the murine stromal cell line MS-5, which efficiently sustains primitive murine and human haemopoiesis [6–8]. We performed RT–PCR on MS-5 mRNA by using degenerate primers corresponding to conserved amino acid sequences in the EGF-like domains shared by the extracellular domains of the Notch family members. Cloning and sequencing of PCR products demonstrated that mRNA species of the murine Notch-1 and Notch-2 and of the EGF-like homoeotic molecule (Dlk) were expressed in MS-5 cells (results not shown). We also identified an unknown 188 bp cDNA fragment that

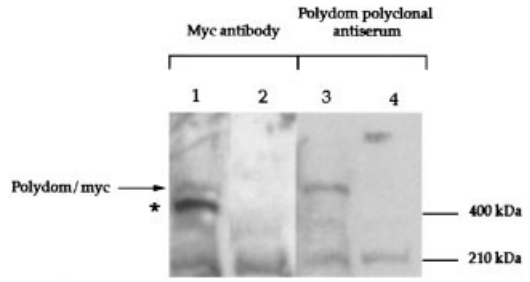

Figure 2 Secretion of polydom by Cos-7 cells

Cos-7 cells were transiently transfected with a full-length Myc-tagged polydom construct or with the vector alone. Cos-7 supernatants were collected 48 h after transfection and supernatant proteins were analysed by Western blotting with the 9E10 monoclonal Myc antibody (lanes 1 and 2) and with polydom polyclonal antiserum (lanes 3 and 4). Lanes 1 and 3, supernatant of Cos-7 cells transfected with the Myc-tagged polydom construct; lanes 2 and 4, supernatant of Cos-7 cells transfected with the vector alone. The sizes of the higher two murine laminin subunits (400 and 210 kDa) are shown; the arrow indicates a protein whose molecular mass is more than 400 kDa, specifically expressed by Cos-7 cells transfected with the Myc-tagged polydom construct. The star indicates a protein recognized by Myc antibody only in the supernatant of Cos-7 cells transfected with the Myc-tagged polydom construct.

encodes a polypeptide with one EGF-like domain related to those of the Notch proteins. To clone the corresponding full-length cDNA, we used this PCR product as a probe to screen MS-5 cDNA libraries. Numerous positive clones were isolated; after successive hybridizations, we obtained several overlapping clones covering up to 12 kb and containing a 10703 bp open reading frame. This open reading frame started at nt 221 with a potential initiator methionine codon preceded by three in-phase stop codons and finished with a stop codon at nt 10927. We did not find any classical polyadenylation signal AATAAA in the 3′ untranslated region but an ATTAAA sequence is present in several clones 18 bp before a poly(A)-rich sequence. The full-length mRNA encodes a polypeptide of 3567 residues with a molecular mass of 387 kDa (Figure 1A).

This protein, which we have named polydom, is very rich in cysteine residues (7.2% of the total residues) and exhibits several features of a secreted protein, for example: (1) a signal peptide sequence with a cleavage site between residue 17 and 18 is present after the methionine initiator codon; (2) it contains 29 potential N-glycosylation sites and 23 potential O-glycosylation sites; (3) it does not contain a hydrophobic transmembrane-spanning sequence or a sequence for a glycosaminoglycan attachment site.

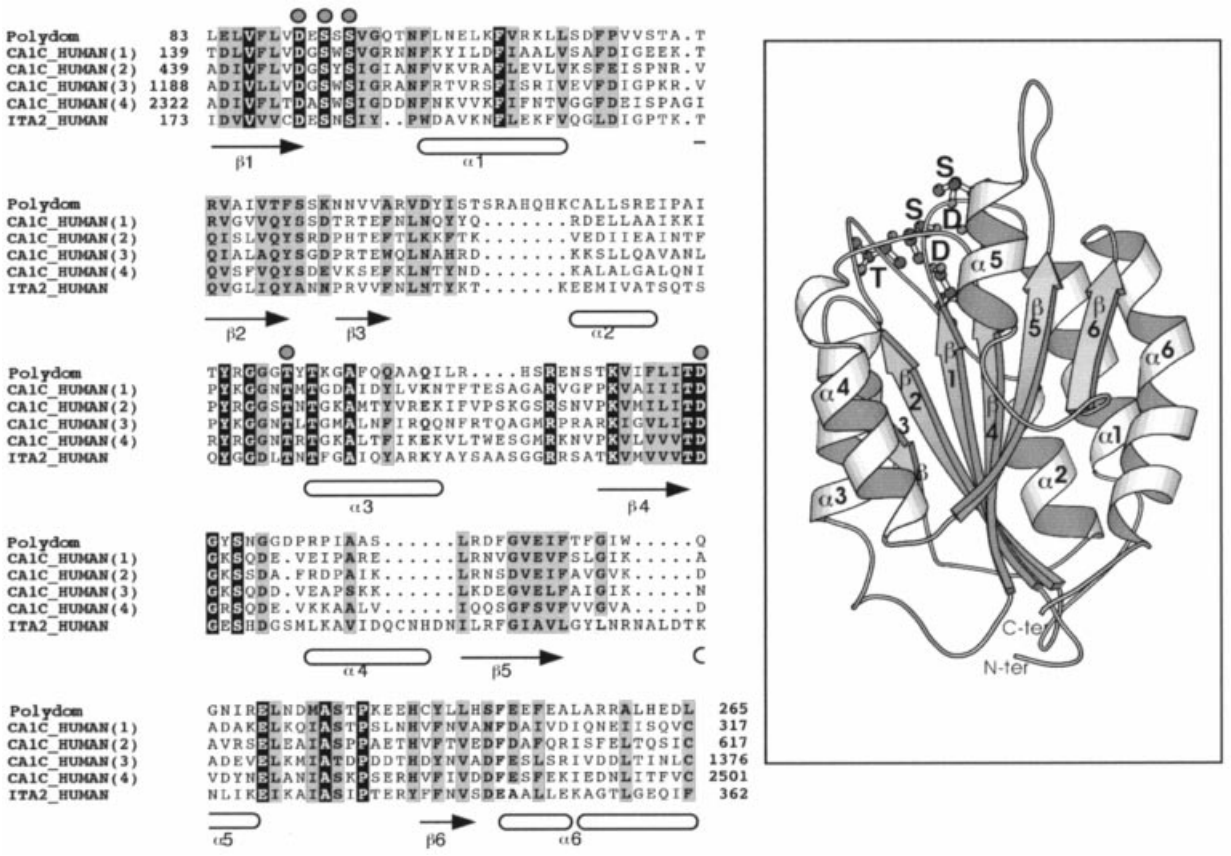

Figure 3 Alignment of the polydom vWF-A domain with other known vWF-A domains

The polydom vWF-A domains were aligned with the four vWF-A domains of human collagen α1(XII) chain (CA1C_HUMAN, SwissProt accession number Q99715) and with that of the I domain from the α subunit of human integrin α1-β2 (ITA2-HUMAN, SwissProt accession number P17301). The secondary structure (two-dimensional) and three-dimensional representations are shown respectively under the alignment and at the right. Residues that co-ordinate metal binding are shown with a filled circle in the alignment. Identical residues are shown in white on a black background; similar residues are boxed. Positions occupied by amino acids with similar properties are shaded grey. The Pfam database *E*-value against the vWF-A domain was 6.8×10^{-19}. Abbreviations: C-ter, C-terminus; N-ter, N-terminus.

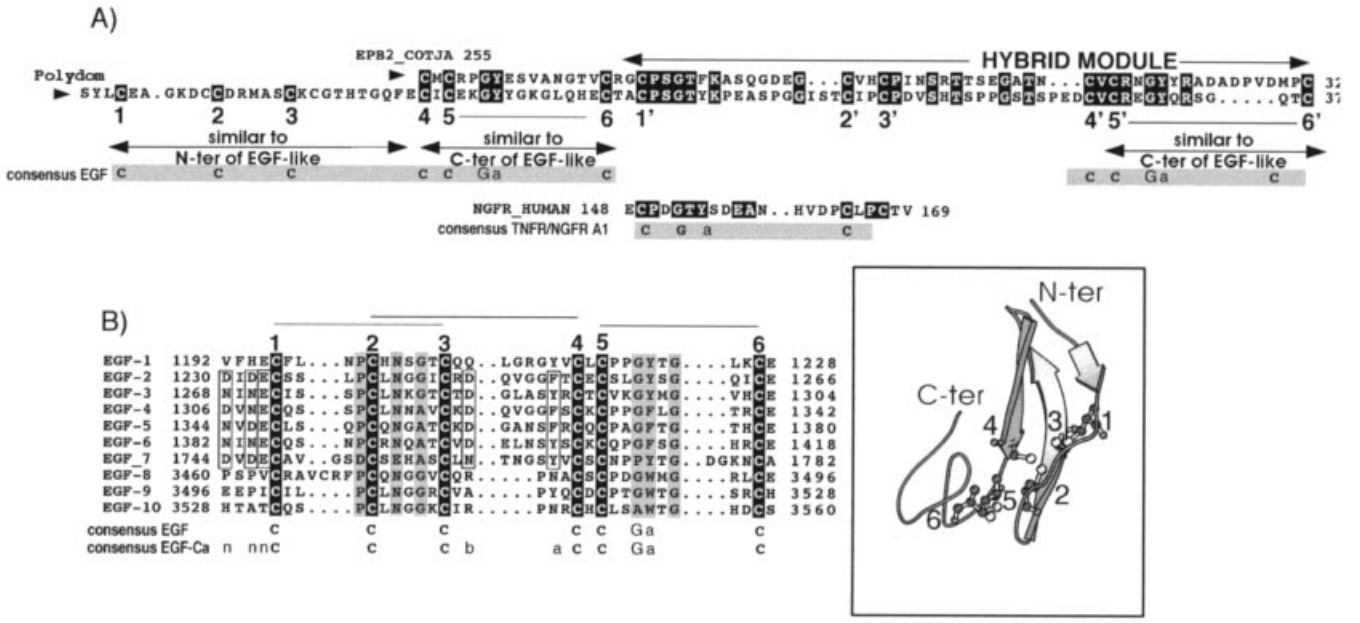

Figure 4 Alignment of the ephrin-2-like cysteine-rich repeats and EGF-like domains of polydom

(**A**) Alignment of the ephrin-2-like cysteine-rich repeat of polydom. Two cysteine-rich modules share significant similarities to the cysteine-rich region of the ephrin family of receptor protein-tyrosine kinases [the BLAST2 *E*-value with QEK5 (EPB2_COTJA, SwissProt accession number Q90344) was 8×10^{-7}; 40% identity over 65 residues]. Identical residues are white on a black background; similar residues are boxed. The C-terminal parts of these modules are related to the C-terminal subdomains of EGF domains (see consensus EGF). (**B**) Multiple alignments of the polydom EGF-like domains. Ten modules are significantly related to the EGF-like domain model [Pfam database *E*-values ranged from 2.2×10^{-9} (EGF-1 and EGF-6) to 0.14 (EGF-8)]. The three-dimensional structure of EGF (PDB identifier EGF) is shown at the right. Disulphide bond patterns are indicated up to the alignments. Identical residues are white on a black background, similar residues are boxed. Positions occupied by amino acids with similar properties are shaded grey. The consensus line for EGF-like domains (PROSITE PDOC 00021) is shown under the alignment [a, aromatic amino-acid (Phe, Tyr, Trp)]. An additional consensus line is shown for EGFca domains [PROSITE PDOC 00913; n, negatively charged or polar residues (Asp, Glu, Gln, Asn); b, possibly β-hydroxylated residue (Asp, Asn); a, aromatic residues]. Abbreviations: C-ter, C-terminus; N-ter, N-terminus.

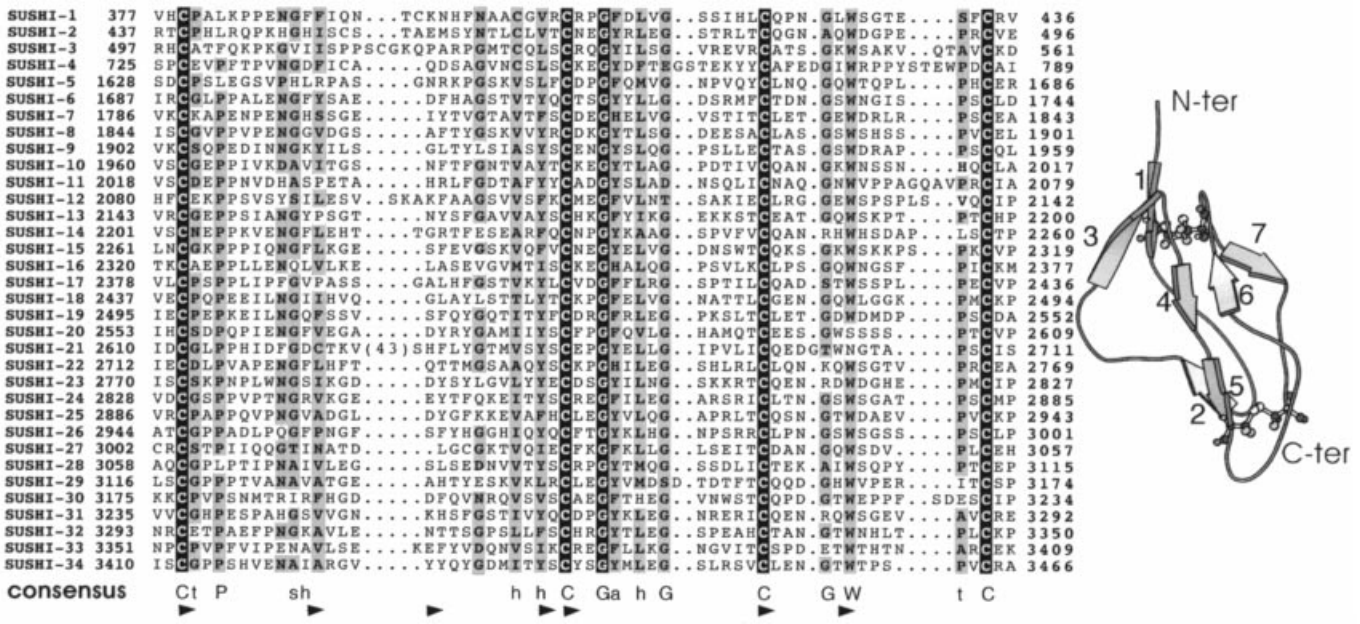

Figure 5 Multiple alignments of the 34 CCP domains in polydom

Pfam database *E*-values against the CCP/SUSHI domain model ranged from 2.7×10^{-18} (CCP-12) to 0.037 (CCP-3). Only the more divergent fourth CCP domain was not detected against the Pfam database. The 21st CCP domain is characterized by a large insertion of 43 residues between the first and second cysteine residues, corresponding to a non-globular region (large loop). Secondary structures, as deduced from the experimental structure of the 15th CCP module of human factor H (PDB identifier 1HFI), which is shown at the right, are indicated below the alignment, as well as the consensus line [over 80%; h, hydrophobic residue (Val, Ile, Leu, Met, Phe, Tyr, Trp); a, aromatic residue (Phe, Tyr, Trp, His); s, small residue; t, turn residue (Pro, Gly, Asp, Asn, Ser)]. Identical residues are white on a black background, similar residues are boxed. Positions occupied by amino acids with similar properties are shaded grey. Abbreviations: C-ter, C-terminus; N-ter, N-terminus.

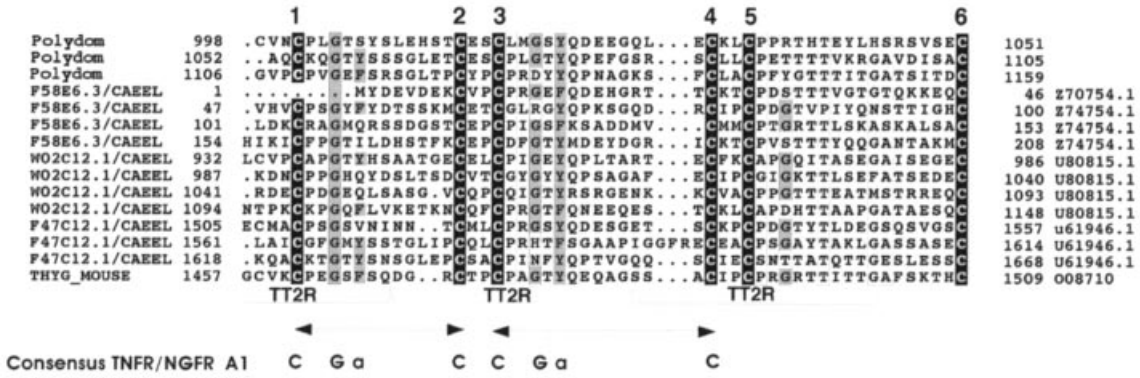

Figure 6 Multiple alignments of the three STT2R repeated domains in polydom

Three polydom cysteine-rich repeated domains are aligned with domains of *C. elegans* hypothetical proteins F47C12.1 (F47C12.1/CAEEL), F58E6.3 (F58E6.3/CAEEL), W02C12.1, (W02C12.1/CAEEL) and F55H12.3 (F55H12.3/CAEEL) and a domain of human thyroglobulin (THYG_HUMAN). BLAST values using the three polydom repeats as query [BLAST 2.0.6; non-redundant (NR) database at the National Center of Biological Information (NCBI; 334797 sequences) ranged between 8×10^{-29} (F47C12.1/CAEEL) and 1×10^{-5} (THYG_HUMAN)]. The positions of STT2R within thyroglobulin are indicated below the sequence. Accession numbers are given at the right. The first two repeats possess hallmarks of the TNF-R/nerve growth factor receptor A1 module, which are indicated below the alignment. Positions occupied by amino acids with similar properties are shaded grey.

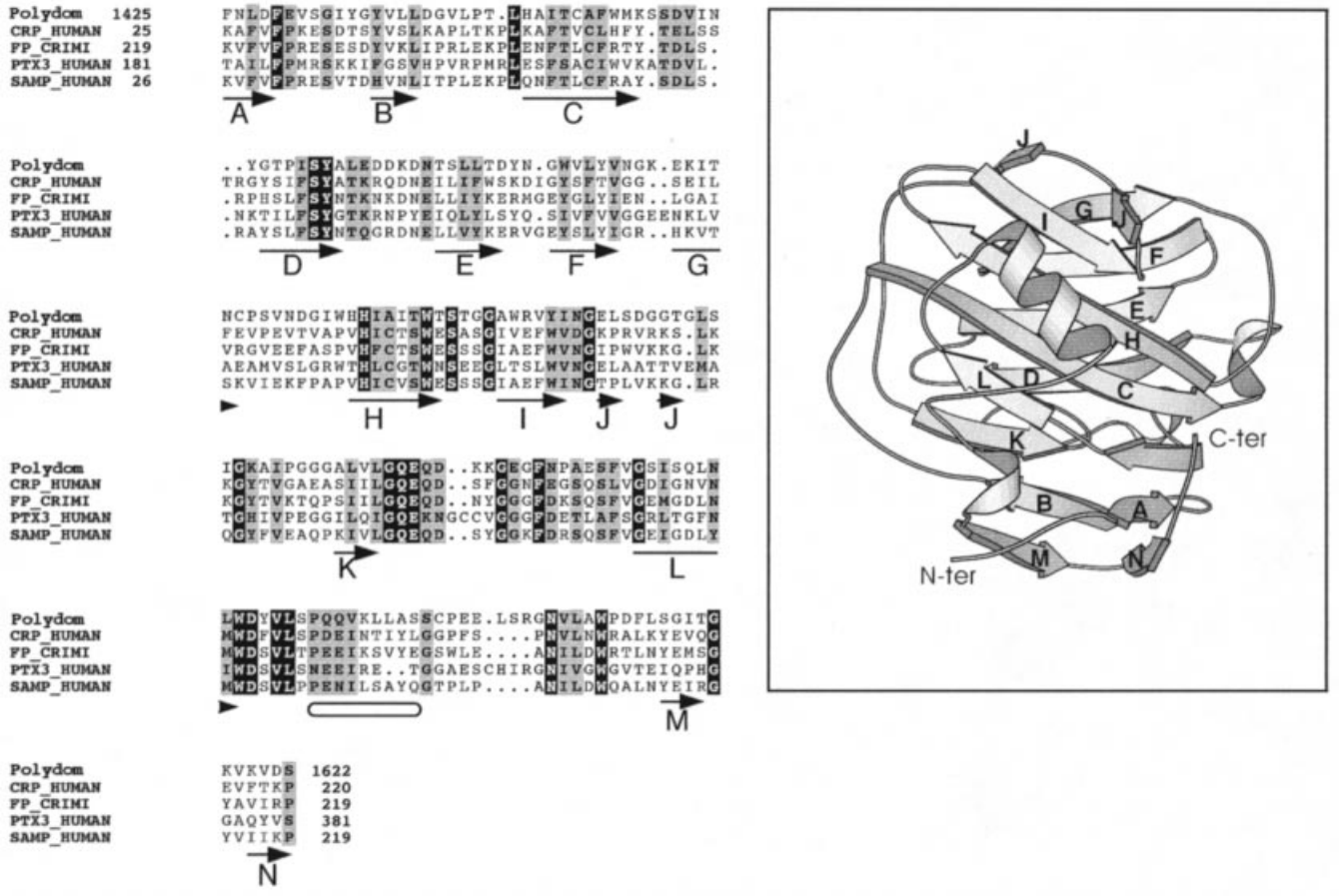

Figure 7 Alignment of the polydom PTX domain

(CRP_HUMAN, SwissProt accession number P02741), *Cricetulus migratorius* female protein (FP_CRIMI, SwissProt accession number P15697), human PTX_related protein ptx3 (PTX3_HUMAN, SwissProt accession number P26022) and human SAP (SAMP_HUMAN, SwissProt accession number P02743). The Pfam database *E*-value against the PENTRAXIN domain model was 9.5×10^{-18}. Secondary structures as predicted from the experimental structure of the human C reactive protein (PBD identifier 1GNH), shown at the right, are indicated below the alignment. Positions occupied by amino acids with similar properties are shaded grey. Abbreviations: C-ter, C-terminus; N-ter, N-terminus.

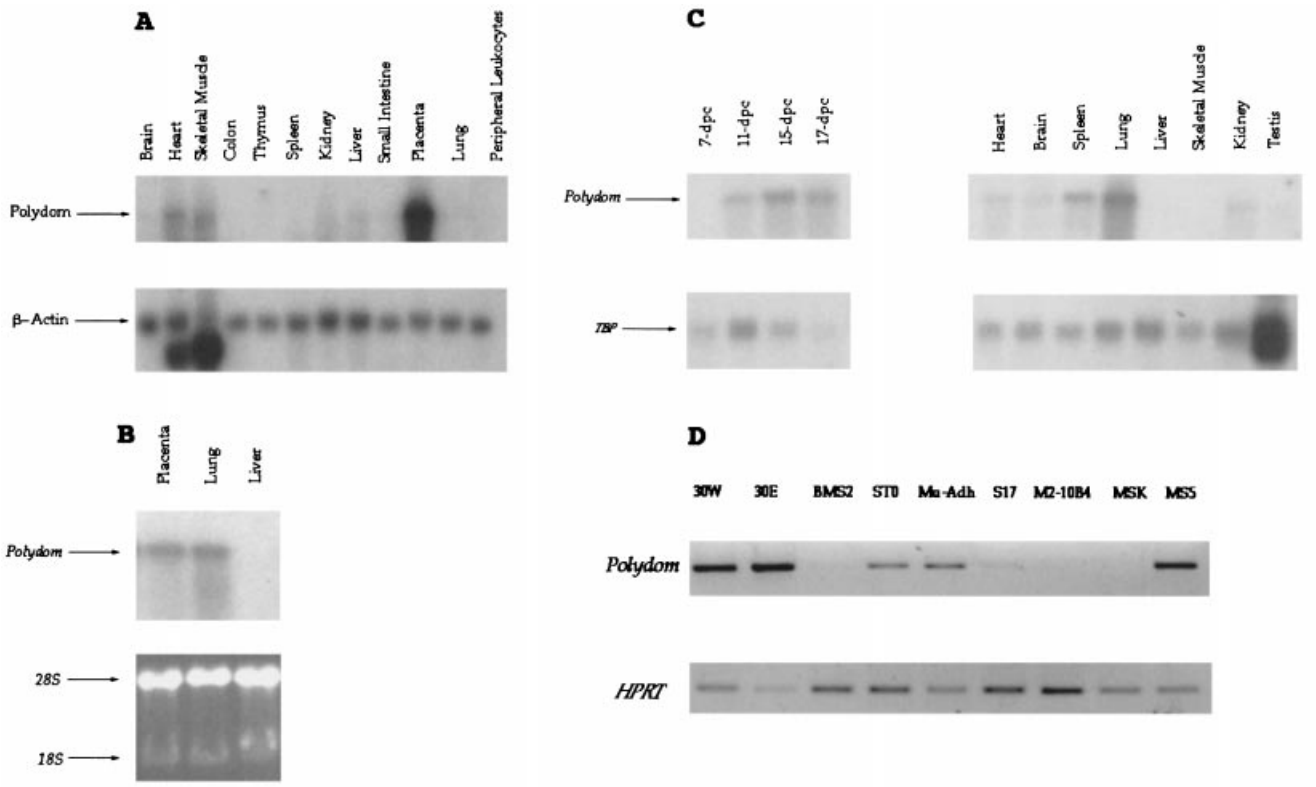

Figure 8 Expression pattern of polydom mRNA

(**A**) Clontech poly(A)$^+$ human multiple tissue blot was hybridized with a human probe homologue to the mouse polydom probe and with a β-actin probe. (**B**) Total RNA (10 μg per lane) was hybridized with a murine cDNA polydom fragment spanning nt 3558–6247. The ethidium-bromide-stained membrane is shown. (**C**) Clontech poly(A)$^+$ mouse embryo multiple tissue blot and mouse multiple tissue blot were hybridized with the same probe as that described above, then stripped and rehybridized with a TATA-box binding protein (TBP) probe as control. Abbreviation: dpc, days post-coitus. (**D**) RT–PCR was performed with total RNA isolated from stromal cells (MS-5, 30E, 30W, MS-K, M2-10B4, BMS2 and S17), murine fibroblasts (STO) and adherent layers of long-term murine cultures (Mu-Adh). As a control the amplification of RT products with HPRT primers is shown. PCR products were 443 bp for murine polydom and 248 bp for HPRT.

To assess the molecular mass of polydom, the full-length Myc-tagged polydom cDNA was transiently transfected in Cos-7 cells, which do not express polydom mRNA. At 48 h after transfection, a protein with a molecular mass of more than 400 kDa was detected with specific antibodies against polydom and Myc in the supernatant of Cos-7 cells transfected with the Myc-tagged polydom construct (Figure 2) This protein was not present in the supernatant of cells transfected with the vector alone. The size of this protein was larger than that predicted from the amino acid sequence (387 kDa). This could be explained by glycosylation of polydom, in agreement with the presence of many consensus sequences for N-glycosylation in the primary sequence of polydom. In the supernatant of Cos-7 cells transfected with the Myc-tagged polydom construct, a protein with a lower molecular mass was detected only with anti-Myc antibody (Figure 2). This protein might correspond to a form still containing the c-Myc epitope located at the C-terminus but deleted from a large region in its N-terminal part. In fact, the polydom domain used to produce polyclonal polydom antibodies is located in residues 789–798. These results suggest that polydom cDNA contains all the necessary elements for the production and secretion of polydom by Cos-7 cells.

Complementary searches for similarities in protein databases revealed that polydom is composed of many known domains (Figure 1B) but did not reveal the presence of a Delta/Serrate/Lag2 (DSL) domain, an ankyrin motif or a cysteine-rich module typical of Notch proteins, clearly indicating that this protein does not belong to the Delta/Notch family. Eight different modules, some of them found in multiple copies, were present in this large protein. We analysed polydom from its N-terminus to its C-terminus. A 200-residue vWF-A domain that contains a metal-ion-dependent adhesion site ('MIDAS') motif [19] was found at its N-terminus (Figure 3). This module was immediately followed by two 50-residue domains displaying similarities to a tandem of cysteine-rich regions (Figure 4A) that have been described previously only in the ephrin receptor family [20]; we named it ephrin-receptor-B-like Cys-rich repeat. The C-terminal end of these two domains possessed clear EGF-like markers; the N-terminal end of the second cysteine-rich domain presented striking features of the A1 domain of tumour necrosis factor receptor (TNF-R) (Figure 4A) [21]. The next domains were three CCP modules, also named short consensus repeat ('SCR') or sushi repeats. As shown in Figure 5, the consensus sequence of these domains was conserved. After these three CCP modules, we identified a duplicated 80-residue domain that shares similarities with repeated sequences in the hyalin protein, the major constituent of the hyalin layer surrounding the sea-urchin embryo [22]. Interestingly, we also found this motif in five other proteins: the human protein SPRX/ETX1, which is implicated in some cases of X-linked retinitis pigmentosa [23,24], the human SRPUL protein and three hypothetical *Caenorhabditis elegans* proteins (F47C12.1, W02C12.1 and F55H12.3). As this domain was first (and as yet has only been) identified in hyalin, we named it HYR for 'hyalin repeat', but further investigations demonstrated that

in fact it corresponds to highly divergent fibronectin type 3 domain ('Fn-3') [25]. After these HYR domains we found another CCP module (Figure 5), then a globular motif poor in cysteine residues, which showed no significant amino acid sequence similarities to any known protein module and thus might be specific to polydom. Next, we found three modules composed of three short repeated motifs, each sharing structural similarities with the type II repetitive domains present in thyroglobulin STT2R motif (named for 'similar to thyroglobulin type 2 repeats') [26] and in some hypothetical proteins from *C. elegans*: F47C12.1, W02C12.1, F55H12.3 and F58E6.3 (Figure 6). Interestingly, the sequence of the two first repeated motifs shared striking similarities with the TNF-R A1 module [21]. After these repeats came six EGF-like domains that presented a characteristic pattern of six cysteine residues organized in three disulphide bonds (1 to 3, 2 to 4 and 5 to 6) (Figure 4B). These six domains were present in tandem repeats, five of which also harboured the consensus signature for Ca^{2+} binding (EGFca; $D/NXD/NE/QX_nD/NX_nY/F$). Because two residues separated these EGFca domains, our protein might belong to the class II family of proteins containing multiple EGFca domains [27]. After these EGF-like domains we found a 200-residue domain that displayed more than 30% identity with the complete sequences of serum amyloid protein (SAP) and complement-reactive protein (CRP) and with the C-terminal end of the newly identified long PTX proteins. As shown in Figure 7, this domain is organized into 14 anti-parallel β strands as defined by the experimental structure of the SAP/PTX domain and harbours amino acid residues identified as constituting the PTX signature (HXCXS/TWXS/T) except for the cysteine residue, which is replaced by an alanine in polydom. It is interesting to note that the cysteine residue of the PTX consensus located in the βC strand is involved in disulphide bonding with another cysteine residue located in strand βF. This second cysteine residue is also missing from the polydom sequence, also replaced by an alanine residue. Amino acid residues important for Ca^{2+}-dependent ligand binding by PTX are also conserved. After the PTX domain, two CCP domains (Figure 5) and one EGF-like repeat that also presented a Ca^{2+}-binding consensus sequence were found (Figure 4B). After these motifs the protein was composed of 28 CCP domains organized in a tandem repeat of 14 CCP domains (Figure 5). With 34 copies, the CCP module represents by far the most abundant domain of polydom. Finally, three EGF-like domains were present at the C-terminus of the protein.

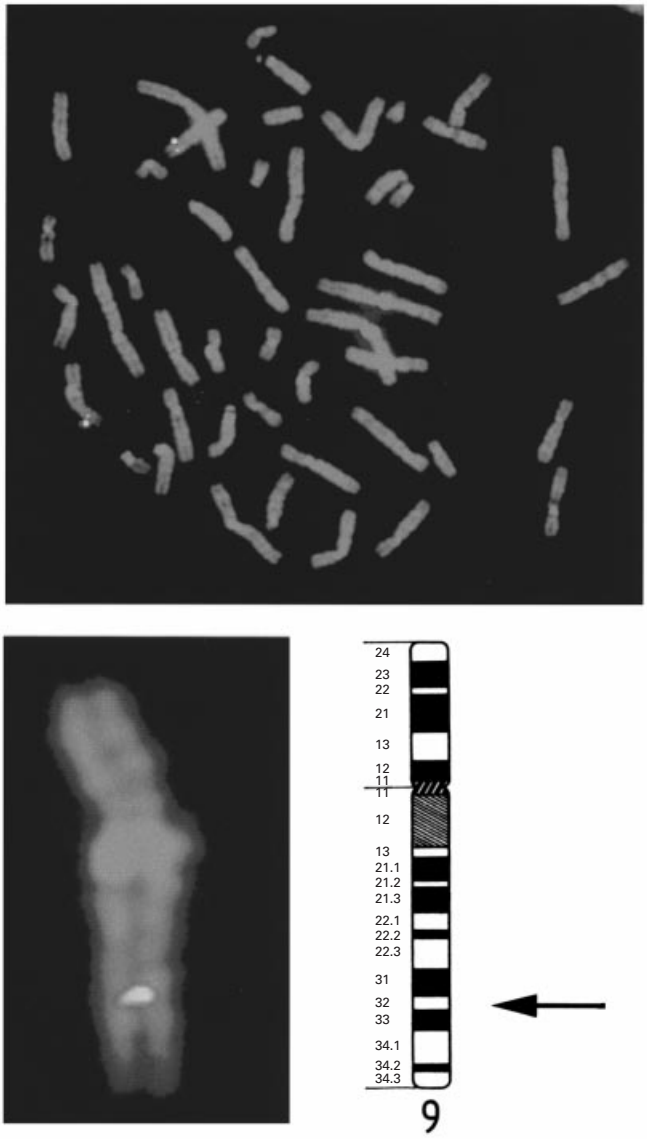

Figure 9 Chromosomal localization of the human *polydom* gene

Upper panel: complete human metaphase after FISH with the cocktail probe. Specific yellow spots are localized on the two chromosomes 9 counterstained with 4,6-diamidino-2-phenylindole. Lower panel: chromosome assignment of the human probes. The arrow indicates a double-dot signal on chromosome 9q32.

Expression pattern of the polydom mRNA in human and mouse

To determine the *polydom* gene expression pattern, we hybridized Northern blots of human and murine tissues. With regard to β-actin normalization, a strong signal was detected in human placenta, whereas in other human tissues the expression of polydom was undetectable or appeared weakly (Figure 8A). In mouse, polydom transcripts were found in lung and placenta and weakly detectable in other tissues (Figures 8B and 8C). During development, polydom mRNA was detectable from day 11 post-coitus to day 17 post-coitus (Figure 8C) in mouse.

We also wondered whether polydom mRNA was expressed in various haemopoietic and stromal cell lines. No detectable mRNA expression was observed by RT–PCR analysis in the haemopoietic cell lines studied (MEL, BaF3, 175B, Jurkat, HEL, Raji, UT7 and K562) and in activated or quiescent B and T lymphocytes (results not shown). Because polydom cDNA was cloned from MS-5 stromal cells, we analysed its expression in other stromal cells by RT–PCR. Polydom mRNA was highly expressed in the stroma adherent layer of murine bone marrow, in murine fibroblasts and in the stromal cell lines 30W, 30E and MS-5. In contrast, it was weakly or not expressed in the BMS2, M2-10B4, MS-K and S17 cell lines (Figure 8D). This expression pattern in stromal cells might be related to the heterogeneity and complexity of bone-marrow stroma.

Chromosomal localization of the *polydom* gene

To determine whether genetic disorders might be associated with the *polydom* gene, we performed chromosomal mapping of the human and murine genes by FISH. The *polydom* gene mapped to the long arm of human chromosome 9 at the 9q32 band (Figure 9) and to murine chromosome 4 at 4B-4C2, a region syntenic with human 9q32 (results not shown). The human localization indicated that the *polydom* gene was located close to

loci linked to some human diseases: acrofacial dysostosis-1, Nager type [28], familial dysautonomia [29] and muscular dystrophy limb-girdle, type 2H [30]. Furthermore, rearrangements or deletions of the 9q32 region occur in human haemopoietic malignancies, including acute myeloid leukemias and non-Hodgkin lymphomas [Cancer chromosome aberration project (NCBI): http://www.ncbi.nlm.nih.gov/ccap/, recurrent aberration results].

DISCUSSION

Using a degenerated RT–PCR approach with oligonucleotides corresponding to sequences localized in the conserved EGF-like domains shared by all Notch proteins, we have identified, in the bone-marrow stromal cell line MS-5, a new transcript encoding a very large mosaic protein that we named polydom. Polydom contains a signal peptide that allows it to be secreted and is composed of eight different domains; some of these, the EGF-like domains, the CCP modules, the vWF-A and the PTX domains, are found in a large variety of extracellular mosaic proteins [31,32].

The EGF-like module is a widely distributed domain conserved between species (vertebrates and invertebrates) and found in many extracellular molecules with various biological functions such as blood coagulation, control of cell fate, cell adhesion, activation of complement and fibrinolysis [32,33]. Genetic mutations in EGF domains are involved in hereditary human disorders, underlining the crucial role of these domains [34,35], which are involved in protein–protein interactions, directly or in association with other EGF domains or other structural motifs. Six of the ten EGF domains of polydom contain a Ca^{2+}-binding site that stabilizes protein–protein interactions and might be important for structural integrity [36]. In polydom, an array of six EGF modules separates a PTX domain and three domains made of repeated motifs sharing similarities with thyroglobulin type II repeats or TNF-R A1 modules. Thus, in polydom, the EGF-like modules might be present as spacers between these two functional modules, and/or they might be directly involved in ligand binding. Interestingly, in polydom, the EGF modules are associated with a PTX domain, an association never described before. Classical PTXs are CRP and the SAP component; these two small proteins, which assemble into pentameric structures, are acute-phase proteins implicated in innate immunity. They can bind a wide variety of ligands; these bindings are mediated by the PTX domain [37]. Several larger proteins have been identified and constitute the long PTX family. The C-terminal end of these proteins consist of a PTX-like domain, whereas their N-terminal domains show no similarity with any known protein. The PTX3/TSG14 protein, which belongs to the long PTX family, seems to be implicated in innate immunity, like CRP and SAP [38], but other identified long PTX proteins have very different functions [39]. Amino acid residues that are important for the PTX Ca^{2+}-dependent ligand-binding function are conserved in polydom and, as with the EGFca domain, Ca^{2+} binding could stabilize the conformation of the protein domain or the structure of the complete polydom, or both. Multimeric conformations of classical PTXs and PTX3 have been shown to have a role in the ligand-binding function. It will therefore be important to determine whether polydom can form multimers. It will also be of interest to consider whether, as with the classical PTXs SAP and CRP, polydom exhibits Ca^{2+}-dependent lectin-like properties and thus is able to bind agar or other classical PTX ligands such as the C1q protein. Thus we can postulate that, as well as EGFca domains, the PTX domain might be important for the structural integrity and function of polydom. We report here the characterization of the largest known molecule with a PTX domain and we describe the first example of an association of this module with other known protein domains such as EGF-like domain, CCP module and vWF-A module. Thus polydom might be the prototype of a new PTX sub-family.

The vWF-A domain is known to be distributed in more than 20 different human proteins with different organizations and is implicated in the immune and haemostatic systems, cell adhesion or matrix assembly [40]. The binding ligand function of vWF, complement proteins (factor B and C2) or the α chain of integrins is mediated by this domain; the vWF-A domain might also mediate ligand binding in other proteins whose function remains unknown, as is that of polydom. Interestingly, the vWF-A domain is the third protein domain of polydom exhibiting residues that co-ordinate the binding of bivalent metal ions. For this domain, a bivalent cation could contribute to the ligand binding activity of polydom, as has been noted for other vWF-A domains [19]. The last polydom domain often found in a wide variety of extracellular mosaic proteins is the CCP module. Prototype members of this protein family are molecules that regulate the complement system; CCP modules have been identified in most components and regulatory proteins of the complement cascade [41–43]. They are also present in the selectin family of adhesion molecules [44] and in coagulation proteins such as the β chain of factor XIII or *Limulus* factor C [45] and in β_2 glycoprotein [46]. In some of these proteins CCP modules are found as multiple copies and might either have a structural role as spacial spacers between other domains or be directly implicated in ligand binding. Polydom harbours two highly similar tandems of 14 CCP modules separating EGF domains (Figure 5). It will be interesting to investigate the role of these various CCP modules in polydom.

The other four domains found in polydom are less widely distributed in tissues and might be responsible for some biological specificity of this protein. The first is a cysteine-rich domain with repeated motifs related to the type II repetitive regions encountered in thyroglobulin [26]. The basal unit of these modules (here termed the STT2R motif) shares sequence similarities with the A1 module found in the cysteine-rich domains of TNF-R family members and in some other proteins [47–49]. The cysteine-rich domains in TNF-R are formed from two distinct types of subdomain to form the A1-C2 or the A1-B2 domains that are involved in the specificity of ligand recognition [21]. Little is known about the specific functional activity of the A1 domain itself and about the type II repetitive region of thyroglobulin. Repeated STT2R motifs are present in three hypothetical proteins from *C. elegans* (F47C12.1, W02C12.1 and F55H12.3), an evolutionary conservation that suggests its biological relevance. Another domain found in polydom, less widely distributed but conserved in some *C. elegans* proteins, is the HYR domain. This domain has been partly identified in an adhesive protein of the hyalin layer of the sea-urchin embryo and seems to be involved in cellular adhesion [22]. The third domain that does not belong to a well-defined family of domains is a cysteine-rich motif that has previously been described only in the ephrin receptor family of proteins. The ephrin receptor family is the largest subfamily of tyrosine kinase receptors and is thought to be important during development [20]. It has been shown that this cysteine-rich motif found in the extracellular parts of these receptors is dispensable for ligand binding and receptor signalling, but might affect the stability of the ligand–receptor complexes [50]. The identification of a protein that is not an ephrin receptor but contains such a domain suggests the existence of a new superfamily of proteins sharing this common structural motif. The last domain found in polydom exhibits no obvious similarity to other protein module;

this globular domain might be involved in the specificity of polydom activity.

In conclusion, most of the domains found in polydom are known to be involved in protein–protein interactions or in cellular adhesion and are found in various extracellular proteins of biological importance. It is as yet unknown whether polydom can interact with one or several types of molecules, whether its biological activity is supported by a specific protein module or by a particular module organization and finally whether the specificity of polydom could be related to the specific assembly of domains. The association in the same molecule of CCP, PTX, EGF and vWF-A domains could suggest a role of polydom in natural defence (the complement system or the acute-phase response to tissue infection or injury) and/or in cell adhesion. Finally, it is interesting to notice that two proteins of *C. elegans* (F55H12.3 and F47C12.1) share the following order of modules with polydom: EGF-like motifs, three or four STT2R motifs, a specific globular region, one or two CCP modules, and one or two HYR repeats. The fact that this specific association of five different domains is conserved in non-mammalian species suggests that this region might be important for the biological function of polydom.

We thank Dr S. Fichelson for providing stromal cell lines, B. Izac for her technical assistance with MS-5 cultures, and Dr Marie-Genevieve Mattéi for performing the murine chromosomal localization. This work was supported by Institut National de la Santé et de la Recherche Médicale (INSERM), Ministère de l'Education Nationale de la Recherche et de la Technologie, Fondation pour la Recherche Médicale, and Ligue Nationale Contre le Cancer, l'Association de la Recherche contre le Cancer. D. G. is supported by la Fondation pour la Recherche Médicale.

REFERENCES

1 Artavanis-Tsakonas, S., Matsuno, K. and Fortini, M. E. (1995) Notch signaling. Science **268**, 225–232
2 Milner, L. A., Bigas, A., Kopan, R., Brashem-Stein, C., Bernstein, I. D. and Martin, D. I. (1996) Inhibition of granulocytic differentiation by mNotch1. Proc. Natl. Acad. Sci. U.S.A. **93**, 13014–13019
3 Washburn, T., Schweighoffer, E., Gridley, T., Chang, D., Fowlkes, B. J., Cado, D. and Robey, E. (1997) Notch activity influences the $\alpha\beta$ versus $\gamma\delta$ T cell lineage decision. Cell **88**, 833–843
4 Varnum-Finney, B., Purton, L. E., Yu, M., Brashem-Stein, C., Flowers, D., Staats, S., Moore, K. A., Le Roux, I., Mann, R. and Gray, G. et al. (1998) The Notch ligand, Jagged-1, influences the development of primitive hematopoietic precursor cells. Blood **91**, 4084–4091
5 Li, L., Milner, L. A., Deng, Y., Iwata, M., Banta, A., Graf, L., Marcovina, S., Friedman, C., Trask, B. J., Hood, L. and Torok-Storb, B. (1998) The human homolog of rat Jagged1 expressed by marrow stroma inhibits differentiation of 32D cells through interaction with Notch1. Immunity **8**, 43–55
6 Itoh, K., Tezuka, H., Sakoda, H., Konno, M., Nagata, K., Uchiyama, T., Uchino, H. and Mori, K. J. (1989) Reproducible establishment of hemopoietic supportive stromal cell lines from murine bone marrow. Exp. Hematol. **17**, 145–153
7 Issaad, C., Croisille, L., Katz, A., Vainchenker, W. and Coulombel, L. (1993) A murine stromal cell line allows the proliferation of very primitive human CD34 + +/CD38- progenitor cells in long-term cultures and semisolid assays. Blood **81**, 2916–2924
8 Kobari, L., Dubart, A., Le Pesteur, F., Vainchenker, W. and Sainteny, F. (1995) Hematopoietic-promoting activity of the murine stromal cell line MS-5 is not related to the expression of the major hematopoietic cytokines. J. Cell. Physiol. **163**, 295–304
9 Lieber, T., Wesley, C. S., Alcamo, E., Hassel, B., Krane, J. F., Campos-Ortega, J. A. and Young, M. W. (1992) Single amino acid substitutions in EGF-like elements of Notch and Delta modify *Drosophila* development and affect cell adhesion in vitro. Neuron **9**, 847–859
10 D'Arcangelo, G., Miao, G. G., Chen, S. C., Soares, H. D., Morgan, J. I. and Curran, T. (1995) A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. Nature (London) **374**, 719–723
11 Rupp, F., Payan, D. G., Magill-Solc, C., Cowan, D. M. and Scheller, R. H. (1991) Structure and expression of a rat agrin. Neuron **6**, 811–823
12 Weller, A., Beck, S. and Ekblom, P. (1991) Amino acid sequence of mouse tenascin and differential expression of two tenascin isoforms during embryogenesis. J. Cell Biol. **112**, 355–362
13 Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402
14 Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D. and Sonnhammer, E. L. (1999) Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic Acids Res. **27**, 260–262
15 Hofmann, K., Bucher, P., Falquet, L. and Bairoch, A. (1999) The PROSITE database, its status in 1999. Nucleic Acids Res. **27**, 215–219
16 Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. U.S.A. **95**, 5857–5864
17 Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. and Mornon, J. P. (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell. Mol. Life Sci. **53**, 621–645
18 Lichter, P., Ledbetter, S. A., Ledbetter, D. H. and Ward, D. C. (1990) Fluorescence *in situ* hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. Proc. Natl. Acad. Sci. U.S.A. **87**, 6634–6638
19 Lee, J. O., Rieu, P., Arnaout, M. A. and Liddington, R. (1995) Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). Cell **80**, 631–638
20 Flanagan, J. G. and Vanderhaeghen, P. (1998) The ephrins and Eph receptors in neural development. Annu. Rev. Neurosci. **21**, 309–345
21 Naismith, J. H. and Sprang, S. R. (1998) Modularity in the TNF-receptor family. Trends Biochem Sci. **23**, 74–79
22 Wessel, G. M., Berg, L., Adelson, D. L., Cannon, G. and McClay, D. R. (1998) A molecular analysis of hyalin – a substrate for cell adhesion in the hyaline layer of the sea urchin embryo. Dev. Biol. **193**, 115–126
23 Dry, K. L., Aldred, M. A., Edgar, A. J., Brown, J., Manson, F. D., Ho, M. F., Prosser, J., Harwick, L. J., Lennon, A. A. and Thomson, K. (1995) Identification of a novel gene, ETX1 from Xp21.1, a candidate gene for X-linked retinitis pigmentosa (RP3). Hum. Mol. Genet. **4**, 2347–2353
24 Meindl, A., Carvalho, M. R., Herrmann, K., Lorenz, B., Achatz, H., Lorenz, B., Apfelstedt,-Sylla, E., Wittwer, B., Ross, M. and Meitinger, T. (1995) A gene (SRPX) encoding a sushi-repeat-containing protein is deleted in patients with X-linked retinitis pigmentosa. Hum. Mol. Genet. **4**, 2339–2346
25 Callebaut, I., Gilgès, D., Vigon, I. and Mornon, J.-P. (2000) HYR, an extracellular module involved in cellular adhesion and related to the immunoglobulin-like fold. Protein Sci. **7**, 1382–1390
26 Malthiery, Y. and Lissitzky, S. (1987) Primary structure of human thyroglobulin deduced from the sequence of its 8448-base complementary DNA. Eur. J. Biochem. **165**, 491–498
27 Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D. and Handford, P. A. (1996) Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. Cell **85**, 597–605
28 Zori, R. T., Gray, B. A., Bent-Williams, A., Driscoll, D. J., Williams, C. A. and Zackowski, J. L. (1993) Preaxial acrofacial dysostosis (Nager syndrome) associated with an inherited and apparently balanced X; 9 translocation: prenatal and postnatal late replication studies. Am. J. Med. Genet. **46**, 379–383
29 Eng, C. M., Slaugenhaupt, S. A., Blumenfeld, A., Axelrod, F. B., Gusella, J. F. and Desnick, R. J. (1995) Prenatal diagnosis of familial dysautonomia by analysis of linked CA-repeat polymorphisms on chromosome 9q31–q33. Am. J. Med. Genet. **59**, 349–355
30 Weiler, T., Greenberg, C. R., Zelinski, T., Nylen, E., Coghlan, G., Crumley, M. J., Fujiwara, T. M., Morgan, K. and Wrogemann, K. (1998) A gene for autosomal recessive limb-girdle muscular dystrophy in Manitoba Hutterites maps to chromosome region 9q31–q33: evidence for another limb-girdle muscular dystrophy locus. Am. J. Hum. Genet. **63**, 140–147
31 Doolittle, R. F. (1995) The multiplicity of domains in proteins. Annu. Rev. Biochem. **64**, 287–314
32 Bork, P., Downing, A. K., Kieffer, B. and Campbell, I. D. (1996) Structure and distribution of modules in extracellular proteins. Q. Rev. Biophys. **29**, 119–167
33 Davis, C. G. (1990) The many faces of epidermal growth factor repeats. New Biol. **2**, 410–419
34 Dietz, H. C. and Pyeritz, R. E. (1995) Mutations in the human gene for fibrillin-1 (FBN1) in the Marfan syndrome and related disorders. Hum. Mol. Genet. **4** Spec No: 1799–1809
35 Giannelli, F., Green, P. M., Sommer, S. S., Poon, M. C., Ludwig, M., Schwaab, R., Reitsma, P. H., Goossens, M., Yoshioka, A. and Brownlee, G. G. (1996) Haemophilia B (sixth edition): a database of point mutations and short additions and deletions. Nucleic Acids Res. **24**, 103–118

36 Kielty, C. M. and Shuttleworth, C. A. (1993) The role of calcium in the organization of fibrillin microfibrils. FEBS Lett. **336**, 323–326
37 Gewurz, H., Zhang, X. H. and Lint, T. F. (1995) Structure and function of the pentraxins. Curr. Opin. Immunol. **7**, 54–64
38 Lee, G. W., Goodman, A. R., Lee, T. H. and Vilcek, J. (1994) Relationship of TSG-14 protein to the pentraxin family of major acute phase proteins. J. Immunol. **153**, 3700–3707
39 Goodman, A. R., Cardozo, T., Abagyan, R., Altmeyer, A., Wisniewski, H. G. and Vilcek, J. (1996) Long pentraxins: an emerging group of proteins with diverse functions. Cytokine Growth Factor Rev. **7**, 191–202
40 Colombatti, A., Bonaldo, P. and Doliana, R. (1993) Type A modules: interacting domains found in several non-fibrillar collagens and in other extracellular matrix proteins. Matrix **13**, 297–306
41 Campbell, R. D., Law, S. K., Reid, K. B. and Sim, R. B. (1988) Structure, organization, and regulation of the complement genes. Annu. Rev. Immunol. **6**, 161–195
42 Sim, R. B. and Perkins, S. J. (1990) Molecular modelling of C3 and its ligands. Curr. Top. Microbiol. Immunol. **153**, 209–222
43 Reid K. B. and Day A. J. (1989) Structure–function relationships of the complement components. Immunol. Today **10**, 177–180 [erratum, Immunol. Today (1989) **10** following 216]
44 Kansas, G. S. (1996) Selectins and their ligands: current concepts and controversies. Blood **88**, 3259–3287
45 Bottenus, R. E., Ichinose, A. and Davie, E. W. (1990) Nucleotide sequence of the gene for the β subunit of human factor XIII. Biochemistry **29**, 11195–11209
46 Kato, H. and Enjyoji, K. (1991) Amino acid sequence and location of the disulfide bonds in bovine beta 2 glycoprotein I: the presence of five Sushi domains. Biochemistry **30**, 11687–11694
47 Mallett, S. and Barclay, A. N. (1991) A new superfamily of cell surface proteins related to the nerve growth factor receptor. Immunol. Today. **12**, 220–223
48 Armitage, R. J. (1994) Tumor necrosis factor receptor superfamily members and their ligands. Curr. Opin. Immunol. **6**, 407–413
49 Baker, S. J. and Reddy, E. P. (1998) Modulation of life and death by the TNF receptor superfamily. Oncogene **17**, 3261–3270
50 Labrador, J. P., Brambilla, R. and Klein, R. (1997) The N-terminal globular domain of Eph receptors is sufficient for ligand binding and receptor signaling. EMBO J. **16**, 3889–3897

Received 21 July 2000; accepted 21 August 2000