Primary structure and tissue distribution of two novel proline-rich \(\gamma\)-carboxyglutamic acid proteins \\
(cDNA cloning/\(\gamma\)-carboxyglutamic acid/vitamin K)

JOHN D. KULMAN*, JEFF E. HARRIS*, BETTY A. HALDEMAN†, AND EARL W. DAVIE*‡

*Department of Biochemistry, University of Washington, Seattle, WA 98195; and †ZymoGenetics, Inc., Seattle, WA 98102

Contributed by Earl W. Davie, June 23, 1997

ABSTRACT Two human cDNAs that encode novel vitamin K-dependent proteins have been cloned and sequenced. The predicted amino acid sequences suggest that both are single-pass transmembrane proteins with amino-terminal \(\gamma\)-carboxyglutamic acid-containing domains preceded by the typical propeptide sequences required for posttranslational \(\gamma\)-carboxylation of glutamic acid residues. The polypeptides, with deduced molecular masses of 23 and 17 kDa, are proline-rich within their putative cytoplasmic domains and contain several copies of the sequences PPXY and PXXP, motifs found in a variety of signaling and cytoskeletal proteins. Accordingly, these two proteins have been called proline-rich Gla proteins (PRGP1 and PRGP2). Unlike the \(\gamma\)-carboxyglutamic acid domain-containing proteins of the blood coagulation cascade, the two PRGPs are expressed in a variety of extracellular tissues, with PRGP1 and PRGP2 most abundantly expressed in the spinal cord and thyroid, respectively, among those tissues tested. Thus, these observations suggest a novel physiological role for these two new members of the vitamin K-dependent family of proteins.

Since the existence of a lipid-soluble antihemorrhagic vitamin was reported by Dam in 1935 (1), the molecular basis for the function of vitamin K in hemostasis has come to be understood in great detail. The identification of \(\gamma\)-carboxyglutamic acid (Gla) residues in bovine prothrombin (2–4) but not in the prothrombin of animals treated with the anticoagulant and vitamin K antagonist dicoumarol (2) clarified the role of vitamin K as a cofactor in the posttranslational \(\gamma\)-carboxylation of selected prothrombin glutamyl residues. Homologous domains with 9–12 Gla residues within the amino-terminal 48 residues have been identified in a number of circulating plasma glycoproteins. The coordination of \(Ca^{2+}\) by several of these Gla residues is required for proper conformation of the Gla domain (5–7) and allows \(Ca^{2+}\)-dependent binding of the coagulation factors to anionic phospholipid membrane surfaces at sites of vascular injury (for reviews, see refs. 8–10).

In addition to the four classical vitamin K-dependent coagulation factors, namely, prothrombin and factors VII, IX and X, this protein family includes the anticoagulant factors, proteins C and S, as well as protein Z, a plasma glycoprotein of unknown function. A recent addition to this family is Gas6, a protein expressed in response to serum starvation of cultured cells (11, 12). This vitamin K-dependent protein has been alternatively described as a growth-potentiating factor (13, 14), a cell survival factor (15), or both (16). Unlike other Gla proteins, Gas6 is expressed in a variety of extrahypothalamic tissues (12) and no role in coagulation has yet been ascribed to it. Rather, Gas6 has been identified as a ligand for the receptor tyrosine kinases Axl (17–19), Rse (alternatively, Sky, Tyro3, Bt, or Tif) (18–20), and Mer (19).

Specific glutamic acid residues within the Gla domains of these proteins are posttranslationally modified by a vitamin K-dependent \(\gamma\)-carboxylase located in the rough endoplasmic reticulum. This reaction requires a \(\gamma\)-carboxylation recognition sequence contained within a propeptide that is flanked by a signal peptide and the amino-terminal domain of the mature protein where the \(\gamma\)-carboxylation occurs.

Given the functional importance of Gla domains in the coagulation factors and gas6, we attempted to identify cDNAs encoding novel Gla domain-containing proteins by searching the dbEST database (21) with a protein query sequence designed from an alignment of all known Gla domain sequences. Herein, we report the cloning of two cDNAs encoding novel Gla domain-containing proteins. Analysis of the deduced amino acid sequence suggests that these proteins are integral membrane proteins with proline-rich cytoplasmic regions. These proline-rich regions contain the potential WW domain-binding motif, PPXY (22). The WW domain (for reviews, see refs. 23 and 24) is a recent addition to a family of protein modules that include Src homology (SH) 2, SH3, and pleckstrin homology (PH) domains (for reviews, see refs. 25 and 26). Moreover, these proline-rich regions contain several copies of the sequence PXXP, an SH3 domain-binding motif. The established importance of SH3 domain interactions and the emerging significance of WW domain interactions in various cytoskeletal components and signaling molecules suggests potential roles for these two newly identified Gla proteins.

MATERIALS AND METHODS

Expressed Sequence Tag (EST) Database Searches. The EST database dbEST (21) at the National Center for Biotechnology Information was searched by using the specialized BLAST algorithm (27) TBLASTN. The amino acid query sequence LEEXXXXXLERECXEEXCXXEEARE was derived by alignment of all known human Gla domain sequences.

Cloning of Proline-Rich Gla Proteins PRGP1 and PRGP2 cDNAs. PCR (28) was performed with a Perkin–Elmer/Cetus DNA thermal cycler and the thermostable DNA polymerases Taq (Boehringer Mannheim), KlenTaq (CLONTECH), or Tth (CLONTECH) according to the manufacturer’s instructions. Oligonucleotide primers were designed based on the nucleotide sequences of ESTs. These were used to amplify cDNA fragments from either a human hepatoma cell (HepG2) cDNA library (29) with a nested PCR strategy. The cDNAs were inserted into the pcDNA3 expression vector (30) and transfected into a variety of cell lines. The expression of these proteins was assessed by immunoblotting and immunofluorescence microscopy.
RESULTS AND DISCUSSION

Cloning of PRGP1 cDNA. A search of the dbEST database with an amino acid query sequence derived from a highly conserved region of all known Gla domains identified EST T25157, which was derived from a human colorectal tumor library (32). Identification of this EST allowed the subsequent cloning of the PRGP1 cDNA (Fig. 1). An oligonucleotide primer based on the EST sequence was used to amplify the 5′ end of the PRGP1 cDNA from a HepG2 cell library by conventional PCR methodology. The 3′ end of the cDNA was amplified by PCR using the technique of RACE from a brain poly(A)+ cDNA library, yielding a contiguous sequence of 1.8 kb with a 3′ polyadenylation consensus sequence followed by a poly(A) tail. Human multiple-tissue Northern blots (see below) using a radiolabeled probe derived from this sequence hybridized with transcripts of approximately 4.6 kb, consider-

![Diagram: Cloning of PRGP1 and PRGP2 cDNAs]

**Fig. 1.** Cloning of PRGP1 (Upper) and PRGP2 (Lower) cDNAs. Overlapping fragments of the cDNAs were generated by the indicated oligonucleotide primers by using PCR. Relative positions of ESTs and their respective accession numbers are indicated. Coding sequences, open bars; 5′ and 3′ noncoding sequences, shaded bars; sequence derived from the plasmid pPC86, solid bars. Origins of each cDNA fragment are indicated below the fragment.

**Table 1.** Oligonucleotide primers used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSP1</td>
<td>5′-TGGACGTGACCACCATGAAGCTGNGGTGAGCC-3′</td>
</tr>
<tr>
<td>G1P1</td>
<td>5′-TAGTCTGGTTTAAATGGAATTGCGTTTTTTCGCTCGGAG-3′</td>
</tr>
<tr>
<td>G1P2</td>
<td>5′-ACGGAGAAAAACCTCAATCC-3′</td>
</tr>
<tr>
<td>AP1</td>
<td>5′-CCAGCAATCTTGATAGCTCTAGATAGAGGC-3′</td>
</tr>
<tr>
<td>G1P3</td>
<td>5′-GCAATTCTCATTAAAAAGCTAC-3′</td>
</tr>
<tr>
<td>AP2</td>
<td>5′-ACCTCAATGGGACGGCCGG-3′</td>
</tr>
<tr>
<td>G1P4</td>
<td>5′-TGATGTCACCTCACTTAATGAGAAG-3′</td>
</tr>
<tr>
<td>G1P5</td>
<td>5′-CTTAAGGCTGTTAAAACCTTGAGGAG-3′</td>
</tr>
<tr>
<td>G1P6</td>
<td>5′-TGATGATATGTGCTGATTAAATAGGAGACAG-3′</td>
</tr>
<tr>
<td>G1P7</td>
<td>5′-CCATTACTCCTTCTTATCCTCTTACCTGTTGAAATAATATCC-3′</td>
</tr>
<tr>
<td>G1P8</td>
<td>5′-TGGACGTGACCACCATGAAGCTGNGGTGAGCC-3′</td>
</tr>
<tr>
<td>G2P1</td>
<td>5′-CTTGATATATGTGCTGATTAAATAGGAGACAG-3′</td>
</tr>
<tr>
<td>G2P2</td>
<td>5′-TGATGATATGTGCTGATTAAATAGGAGACAG-3′</td>
</tr>
<tr>
<td>G2P3</td>
<td>5′-CTTGATATATGTGCTGATTAAATAGGAGACAG-3′</td>
</tr>
<tr>
<td>G2P4</td>
<td>5′-GTGAAATTATAGGGGACCAC-3′</td>
</tr>
<tr>
<td>G2P5</td>
<td>5′-TTTCTTATTGGGGACAGCTAACCCTCC-3′</td>
</tr>
</tbody>
</table>

Primers are listed in the order in which they appear in Fig. 1. G1 primers were used to amplify PRGP1. G2 primers were used for PRGP2. VSP1 is specific for the plasmid pPC86. AP1 and AP2 are adaptor-specific primers used in RACE reactions.
ably larger than the expected size of 1.8 kb. This discrepancy in size was clarified by additional 3' RACE reactions using primers derived from the region of the original cDNA proximal to the polyadenylation site and a skeletal muscle poly(A)+ cDNA library. These PCRs yielded an additional 2.8 kb of 3' untranslated sequence with the expected polyadenylation consensus sequence followed by a poly(A) tail. A contiguous 4.5-kb cDNA sequence encompassing all previously identified PRGP1 cDNA fragments was amplified from HepG2 and skeletal muscle cDNA libraries and the nucleotide sequence was determined (Fig. 2). Since the 1.8-kb cDNA sequence was identical with the corresponding region of the 4.5-kb cDNA, it was concluded that the different cDNA lengths result from differential usage of polyadenylation sites rather than from differential mRNA splicing. Moreover, since no transcript corresponding to the 1.8-kb band was observed on Northern

**Fig. 2.** Nucleotide and predicted amino acid sequence of PRGP1 and PRGP2. Proposed protease cleavage sites are indicated by △. Gla domains are outlined. Putative transmembrane regions are shown in boldface type. Potential WW domain interaction motifs (PPXY) are shaded in gray. Potential SH3 domain interaction motifs (PXXP) are underlined. In-frame stop codons are denoted with asterisks. Polyadenylation signals are shown in boldface type and are boxed.
Cloning of PRGP2 cDNA. The dbEST database search identified a second potential Gla protein sequence, EST H58421, derived from a human fetal liver/spleen library. Oligonucleotide primers based on this sequence were used to clone the PRGP2 cDNA (Fig. 1). The 5’ and 3’ ends of the cDNA were amplified from HepG2 and thyroid cDNA libraries, respectively, and a contiguous 1.2-kb sequence was subsequently amplified from HepG2, kidney, and thyroid cDNA libraries and sequenced (Fig. 2). As in the case of PRGP1, the 3’ RACE product was amplified from a poly(A)+ cDNA library and contained both a polyadenylation sequence and a poly(A) tail.

Sequence Analysis of PRGP1 and PRGP2. The PRGP1 cDNA contained a 165-bp 5’ untranslated region (UTR), a 657-bp coding sequence, and an unusually long 3.7-kb 3’ UTR. The mature protein of 198 amino acid residues has a deduced molecular mass of 23 kDa after cleavage of the 20-amino acid propeptide. This cleavage site was predicted by alignment of propeptide arginine residues with the −1 and −4 arginine residues of the other Gla domain-containing proteins (Fig. 3). The first methionine codon is encountered at nucleotide 166, immediately proximal to the propeptide-encoding sequence, and is preceded by an in-frame stop codon at nucleotide 49. Therefore, unlike all other Gla domain-containing proteins, PRGP1 lacks a discernible signal peptide. A putative transmembrane region (residues 56 through 83) within the mature protein could, therefore, act as a “signal anchor” to direct the nascent polypeptide to the endoplasmic reticulum lumen. The orientation of the protein within the membrane is largely determined by the relative charges of the residues flanking the transmembrane segment, with the cytoplasmic sequence generally carrying the greater positive charge (33). PRGP1 has relatively more positive charges on the carboxyl-terminal side of the putative transmembrane region and would, therefore, be expected to orient with the carboxyl terminus within the cytoplasm. Such an orientation is consistent with the fact that glutamic acid residues within amino-terminal Gla domains are γ-carboxylated by the endoplasmic reticulum-resident carboxylase. In addition, a disulfide loop within the Gla domain would be expected to form correctly within the oxidizing environment of the endoplasmic reticulum but not within the reducing milieu of the cytoplasm.

The PRGP2 cDNA possesses a short (9 bp) 5’ UTR, a 609-bp coding sequence, and a 548-bp 3’ UTR. The mature protein of 153 amino acid residues has a deduced molecular mass of 17 kDa after cleavage of the 49 residue signal peptide. The presence of an amino-terminal signal peptide and a putative transmembrane region defines PRGP2 as a type I single-pass transmembrane protein with the carboxyl terminus oriented to the cytoplasm. As is the case with PRGP1, the PRGP2 propeptide cleavage site can be predicted by alignment of arginine residues within the propeptide with arginine residues within known cleavage sites of the other Gla domain-containing proteins (Fig. 3).

The putative cytoplasmic regions of PRGP1 and PRGP2 are uncharacteristically proline-rich. The proline-containing motifs PPXY and PXXP are found in proteins that interact with WW domains and SH3 domains, respectively. These are modules common to proteins involved in signal transduction and cytoskeletal interactions (22, 25, 26). The amino acid sequence of PRGP1 contains 2 PPXY motifs and 3 PXXP motifs and that of PRGP2 contains 1 PPXY motif and 12 PXXP motifs, many of which overlap (Fig. 2).

Tissue Distribution of PRGP1 and PRGP2. Northern blot analysis of PRGP1 and PRGP2 revealed that both have a broad tissue distribution with PRGP1 showing the highest expression in the spinal cord and PRGP2 showing the highest expression in the thyroid among those tissues tested (Fig. 4). This stands in marked contrast to the Gla domain-containing proteins, which are usually expressed in specific tissue groups. The dbEST database search identified a second potential Gla protein sequence, EST H58421, derived from a human fetal liver/spleen library and contained both a polyadenylation sequence and a poly(A) tail.

Potential Functions of PRGP1 and PRGP2. The discovery of PRGP1 and PRGP2 represents the identification of a novel modular context for Gla domains. In these molecules, the Gla domain is followed by a single putative transmembrane stretch and a small proline-rich cytoplasmic region, rather than by epidermal growth factor-like domains, kringle domains, or a disulfide loop, as is the case with all other known Gla domain proteins. It remains to be determined whether either of these two proteins actually interacts within the cytoplasm with WW
domain- or SH3 domain-containing proteins and, if so, whether they interact with the cytoskeleton or signaltransduction machinery. In addition, Gla domains have long been known to interact with surfaces rich in anionic phospholipids, particularly phosphatidylserine (9, 10), and such surfaces are attractive candidates for the extracellular target of PRGP1 and PRGP2. However, the possibility of a nonphospholipid ligand cannot be ruled out. Identification of a physiological extracellular target remains an additional challenge in the functional characterization of PRGP1 and PRGP2.

The identification of potential transmembrane sequences in PRGP1 and PRGP2 raises the possibility that these proteins are members of a larger family of transmembrane Gla proteins. A search of the nonredundant DNA sequence database at the National Center for Biotechnology Information has revealed an additional Gla domain-encoding sequence, complete with the expected propeptide-encoding sequence, but lacking that of the signal peptide. The sequence identified was that of the rabbit aortic cyclic nucleotide-gated channel, a multipass transmembrane protein (37). However, the portion of the cDNA that encoded the Gla domain was in a different reading frame than that which encoded the transmembrane regions and the cyclic nucleotide-binding domain. It therefore remains to be demonstrated whether this is truly a novel membrane Gla protein and, if so, how many other members of this class exist.

We thank David Yee, Patrick O’Hara, Don Foster, Frank Grant, and Dominic Chung for their advice and helpful discussions. This work was supported, in part, by Research Grant HL-16919 from the National Institutes of Health. J.D.K. was supported by a National Science Foundation Graduate Fellowship.