# Insights into How CUB Domains Can Exert Specific Functions while Sharing a Common Fold

## CONSERVED AND SPECIFIC FEATURES OF THE CUB1 DOMAIN CONTRIBUTE TO THE MOLECULAR BASIS OF PROCOLLAGEN C-PROTEINASE ENHANCER-1 ACTIVITY\*S

Received for publication, February 23, 2007, and in revised form, April 12, 2007 Published, JBC Papers in Press, April 19, 2007, DOI 10.1074/jbc.M701610200

Guillaume Blanc, Bernard Font, Denise Eichenberger, Christophe Moreau<sup>1</sup>, Sylvie Ricard-Blum, David J. S. Hulmes<sup>2</sup>, and Catherine Moali

From the Institut de Biologie et Chimie des Protéines, CNRS/Université Claude Bernard Lyon 1, Unité Mixte de Recherche 5086, Institut Fédératif de Recherche 128 Biosciences Lyon Gerland, 7 Passage du Vercors, 69367 Lyon cedex 7, France

Procollagen C-proteinase enhancers (PCPE-1 and -2) are extracellular glycoproteins that can stimulate the C-terminal processing of fibrillar procollagens by tolloid proteinases such as bone morphogenetic protein-1. They consist of two CUB domains (CUB1 and -2) that alone account for PCPE-enhancing activity and one C-terminal NTR domain. CUB domains are found in several extracellular and plasma membrane-associated proteins, many of which are proteases. We have modeled the structure of the CUB1 domain of PCPE-1 based on known threedimensional structures of CUB-containing proteins. Sequence alignment shows conserved amino acids, notably two acidic residues (Asp-68 and Asp-109) involved in a putative surface-located calcium binding site, as well as a conserved tyrosine residue (Tyr-67). In addition, three residues (Glu-26, Thr-89, and Phe-90) are found only in PCPE CUB1 domains, in putative surface-exposed loops. Among the conserved residues, it was found that mutations of Asp-68 and Asp-109 to alanine almost completely abolished PCPE-1 stimulating activity, whereas mutation of Tyr-67 led to a smaller reduction of activity. Among residues specific to PCPEs, mutation of Glu-26 and Thr-89 had little effect, whereas mutation of Phe-90 dramatically decreased the activity. Changes in activity were paralleled by changes in binding of different PCPE-1 mutants to a mini-procollagen III substrate, as shown by surface plasmon resonance. We conclude that PCPE-stimulating activity requires a calcium binding motif in the CUB1 domain that is highly conserved among CUB-containing proteins but also that PCPEs contain specific sites that could become targets for the development of novel anti-fibrotic therapies.

CUB<sup>3</sup> domains are widely occurring structural motifs, found almost exclusively in extracellular and plasma membrane-associated proteins. These proteins are involved in a wide range of biological functions, including complement activation (1, 2), developmental patterning (3, 4), tissue repair (5), axon guidance and angiogenesis (6, 7), cell signaling (8), fertilization (9), hemostasis (10), inflammation (11), neurotransmission (12), receptor-mediated endocytosis (13, 14), and tumor suppression (15, 16). Many CUB domain-containing proteins are proteases (1, 2, 5, 10, 17–19). Although the roles of the CUB domains are largely unexplored, a number of them have been shown to be involved in oligomerization and/or recognition of substrates and binding partners.

The protein families from which the CUB domain derives its name are the complement serine proteases C1r, C1s, MASP-1, MASP-2, and MASP-3 and the bone morphogenetic protein-1/ tolloid metalloproteases BMP-1, mTLD, mTLL-1, and mTLL-2 (or their counterparts xolloids, tolloids, and SpAN/BP10 in Xenopus, Drosophila, and sea urchin, respectively). Each consists of a catalytic domain (either N-terminal in tolloids or C-terminal in complement proteases) together with several CUB domains interspersed by calcium-binding EGF domains. In the case of the complement proteases, the CUB domains mediate dimerization and binding to the collagen-like regions of C1q (for C1r/C1s) and MBL or L- and H-ficolin (for MASP-2). The three-dimensional structures of the CUB1-EGF fragment of C1s, its homologue in MASP-2 (which also occurs as the alternatively spliced product MAp19) as well as the CUB1-EGF-CUB2 fragment of MASP-2 have been determined (20-22). These studies have revealed the presence of a calcium binding site in the CUB1 domains of both C1s and MAp19, within the loops distal to the EGF domain. In the case of MAp19, site-directed mutagenesis has shown the involvement of this region in binding to MBL and L-ficolin (21).

The BMP-1/tolloid-related metalloproteinases (BMP-1, mTLD, mTLL-1, and mTLL-2 in mammals) are members of the astacin family (M12A) of the metzincin subclan MA(M) (23, 24). Following the pro-region, which is removed by a furin-like

<sup>\*</sup> This work was supported by the Région Rhône-Alpes, the European Commission (Contract No. NMP2-CT-2003-504017), the Ligue Contre le Cancer, Engelhard (Lyon), the Centre National de la Recherche Scientifique and the Université Claude Bernard Lyon 1. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

International Stress Stress

<sup>&</sup>lt;sup>1</sup> Present address: Commissariat à l'Energie Atomique-Grenoble/Département Réponse et Dynamique Cellulaires/Laboratoire de Biophysique Moléculaire et Cellulaire, CNRS Unité Mixte de Recherche 5090, 38054 Grenoble, France.

<sup>&</sup>lt;sup>2</sup> To whom correspondence should be addressed. Tel.: 33-472-722-667; Fax: 33-472-722-604; E-mail: d.hulmes@ibcp.fr.

<sup>&</sup>lt;sup>3</sup> The abbreviations used are: CUB, module found in complement subcomponents C1r/C1s, Uegf, and BMP-1; BMP-1, bone morphogenetic protein-1; mTld, mammalian tolloid; mTLL-1, mammalian tolloid-like 1; mTLL-2, mammalian tolloid-like 2; PCP, procollagen C-proteinase; PCPE, PCP enhancer.

proteinase, each variant consists of an astacin domain followed by a C-terminal region consisting of variable numbers of CUB and calcium-binding EGF domains. In the case of BMP-1, this C-terminal region has the domain organization CUB1-CUB2-EGF1-CUB3, whereas in the longer mTLD splice variant it is CUB1-CUB2-EGF1-CUB3-EGF2-CUB4-CUB5. Variants mTLL-1 and mTLL-2 share the same domain structure as mTLD but are products of separate genes. These proteinases cleave a large number of substrates, including the fibrillar procollagens I, II, III, V, and XI, the non-fibrillar procollagen VII, prolysyl oxidases, the laminin 5  $\gamma$ 2 chain, dentin matrix protein-1, precursor forms of the small leucine rich proteoglycans biglycan and osteoglycin, the heparan sulfate proteoglycan perlecan, growth factors myostatin and GDF11, the growth factor antagonist chordin (24), and latent transforming growth factor- $\beta$ -binding protein (LTBP), which controls transforming growth factor- $\beta$  activity (25).

There is increasing evidence that tolloid proteinase activity is subject to regulation by a variety of extracellular proteins. During developmental patterning, for example, it has recently been shown that the frizzled-related protein sizzled is an endogenous inhibitor of BMP-1 (3, 26), whereas complex formation involving the protein-twisted gastrulation stimulates the activity of BMP-1 on chordin and exposes a new cleavage site (27). Procollagen C-proteinase enhancers (PCPE-1 and -2) also stimulate the activities of tolloid proteinases, in a substrate-specific manner (28). PCPE-1 has no effect on cleavage of several BMP-1 substrates (28), including chordin (29), prolysyl oxidase, osteoglycin, the laminin 5  $\gamma$ 2 chain, procollagen VII, and the N-propeptide region of procollagen V. In contrast, both PCPE-1 and -2 have been shown to stimulate the activities of tolloid proteinases during cleavage of the C-propeptide regions of the major fibrillar procollagens (types I, II, and III) (28, 30, 31). Targeted deletion of the PCPE-1 gene has recently been shown to lead to aberrant collagen fibril formation and impaired biomechanical properties of bone tissue (32).

PCPEs are extracellular glycoproteins, devoid of intrinsic proteolytic activity, consisting of two CUB domains followed by a C-terminal NTR domain (31). Enzymatic removal of the NTR domain has no effect on enhancement of tolloid proteinase activity, showing that this is a property of the CUB domain region (33, 34). Although the mechanism of action of PCPEs is not well understood, this probably involves binding to the substrate, because maximum enhancing activity is found for PCPE: procollagen molar ratio of at least 1:1. BIAcore studies have shown that PCPE-1 binds to both the C-propeptide region as well as elsewhere in the procollagen molecule (35). Recent studies using a miniprocollagen III substrate have shown that enhancing activity is unaffected by removal of essentially all of the triple-helical region using highly purified bacterial collagenase (28). Taken together, these data suggest that PCPE-1 binds to the non-triple-helical telopeptide region N-terminal to the BMP-1 cleavage site, in addition to the C-propeptide region, thereby inducing a conformational change that facilitates the action of BMP-1. It is also possible that PCPE-1 binds to tolloid proteinases, as recently shown for mTLL-1 (36). To further explore the mechanism of action of PCPE-1, it is important to identify recognition sites within the CUB domains

#### TABLE 1

Sequences of forward primers used to generate the PCPE-1 mutants Reverse primers had the complementary sequences. Mutated codons are underlined.

E26A	5'-GTG GCA AGT <u>GCG</u> GGG TTC CCC-3'
Y67A	5'-GCC TGC CGC <u>GCC</u> GAT GCT CTG-3'
D68A	5'-GCC TGC CGC TAC <u>GCT</u> GCT CTG GAG G-3'
T89A	5'-TTT TGT GGG <u>GCC</u> TTC CGG CCT G-3'
F90A	5'-TGT GGG ACC <u>GCC</u> CGG CCT GCG-3'
D109A	5'-AGG ATG ACG ACG $\underline{\text{GCT}}$ GAG GGC ACA GGA G-3'

involved in enhancing activity. Here, we show that these involve a putative calcium binding site in the CUB1 domain, as well as a site specific to the CUB1 domain of PCPEs.

#### **EXPERIMENTAL PROCEDURES**

*Molecular Modeling*—The three-dimensional model of the CUB1 domain of human PCPE-1 (residues 12–124, numbered from the N terminus of the mature protein, *i.e.* after cleavage of the signal peptide) was built with the comparative molecular modeling program Geno3D (37), using the CUB domain structure with the highest sequence homology, namely the CUB2 domain of rat MASP-2 (22). Ten three-dimensional structures were generated and superimposed with the ANTHEPROT three-dimensional package (38). The quality of the models was assessed using PROCHECK, and the most representative model, with the lowest energy, was selected.

*Recombinant Proteins*—The BMP-1 FLAG construct was obtained by PCR using the BMP-1 cDNA inserted in pCEP4 as a template (28). The FLAG sequence was inserted immediately contiguous to the BMP-1 sequence before the STOP codon. After sequencing and transfection in 293-EBNA cells, a clone was selected that gave the highest BMP-1 level of expression. This clone was used for further amplification and production of the protein.

For site-directed mutagenesis and production of PCPE mutants, an 8-histidine tag was inserted by PCR into the human PCPE-1 cDNA (28), before the STOP codon, and the PCR product was cloned into pCR-Blunt II TOPO (Invitrogen). This construct was used as a template to generate all the PCPE-1 mutants using the QuikChange<sup>TM</sup> site-directed mutagenesis kit from Stratagene, according to the manufacturer's protocol. Overlapping oligonucleotides (Table 1) were purchased from MWG-BIOTECH (Courtaboeuf, France). The mutated inserts were then subcloned into pCEP4 (Invitrogen) using the KpnI and BamHI sites for semi-stable transfections into 293-EBNA cells (28). The sequences of all mutants were confirmed by double-stranded DNA sequencing (Genome Express, France).

*Protein Purification*—For BMP-1-FLAG, the first purification step (Reactive Green) was as previously described for BMP-1 (28). BMP-1-FLAG-containing fractions (according to SDS-PAGE) were dialyzed against Buffer A (20 mM Hepes, pH 7.4, 2.5 mM CaCl<sub>2</sub>, 0.02% octyl- $\beta$ -D-glucopyranoside) plus 0.25 M NaCl, and loaded onto an Anti-FLAG M2-agarose affinity column (Sigma) equilibrated with the same buffer. After washing with Buffer A plus 0.5 M NaCl, bound BMP-1-FLAG was eluted using 0.25 mg/ml FLAG peptide dissolved in Buffer A plus 0.5 M NaCl, then extensively dialyzed against the same buffer without peptide. We checked that the purified protein was devoid of endogenous PCPEs.

#### CUB Domain Specificity in PCPE-1

For His-tagged PCPE-1 and its mutants, conditioned medium was centrifuged to remove cell debris and incubated with nickel-agarose equilibrated with 50 mM sodium phosphate, pH 8.0, 0.3 M NaCl (buffer B). The resin was then packed into a column and washed with buffer B containing 10 mM imidazole. The protein was eluted with 0.25 M imidazole in buffer B, and extensively dialyzed against phosphate-buffered saline.

*Enzymatic Activities*—BMP-1 activity was measured as previously described (28), using either [<sup>3</sup>H]procollagen I isolated from chick embryo fibroblasts or mini-procollagen III produced in 293 EBNA cells, as specified in the figure legends.

*Circular Dichroism*—Far-UV (190–250 nm) CD measurements were carried out using thermostatted 0.2-mm path length quartz cells in a Jobin-Yvon CD6 instrument, calibrated with aqueous d-10-camphorsulfonic acid. Proteins (0.3–1 mg/ml) were analyzed at 25 °C in 20 mM KH<sub>2</sub>PO<sub>4</sub>/NaOH, 150 mM NaF, pH 7.2. Spectra were measured with a wavelength increment of 0.2 nm, integration time of 1 s, and bandpass of 2 nm. Protein concentrations were determined by absorbance at 280 nm (using absorbances calculated for each mutant based on the amino acid sequence) as well as by the Bradford assay, normalizing the latter to a known concentration of wild-type PCPE-1. Secondary structure analysis was carried out on the DICHROWEB server (39) using the CDSSTR program.

*Fluorescence*—Intrinsic fluorescence was measured at 25 °C using a Photon Technology International instrument with a 10-mm path length quartz cell. Emission was recorded between 295 and 400 nm after excitation at 280 nm (excitation slits: 8 nm; emission slits: 4 nm) with a scanning rate of 1 nm/s. PCPE-1 was dialyzed against 20 mM HEPES, pH 7.4, 0.15 M NaCl and diluted to 2.5  $\mu$ M in a total volume of 500  $\mu$ l of the same buffer. EGTA and CaCl<sub>2</sub> were added in 1- $\mu$ l volumes. Fluorescence spectra were recorded 5 min after each addition and corrected for fluorescence of the buffer (containing or not EGTA and/or calcium). The effect of EGTA and calcium was also measured on a control solution of 5.4  $\mu$ M *N*-acetyltryptophan amide (NATA) to check that changes in fluorescence were not due to absorption by added compounds (40).

Surface Plasmon Resonance-Surface plasmon resonance analyses were carried out on a BIAcore 3000 instrument. Miniprocollagen III was covalently coupled to a CM4 sensor chip after activation with N-hydroxysulfosuccinimide and 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (amine coupling) and injection in 10 mM sodium acetate, pH 5. Unreacted groups on the surface were neutralized with ethanolamine hydrochloride. The whole procedure was performed at a flow rate of 5  $\mu$ l/min. A control flow cell was prepared similarly except that the mini-procollagen III solution was replaced by 10 mM sodium acetate, pH 5. The signal recorded on the control flow cell was automatically subtracted from those of the other flow cells. Prior to analysis, ligands were dialyzed against HBS-P (10 mM HEPES, pH 7.4, 0.15 M NaCl, and 0.005% P20 surfactant), and binding was monitored at 25 °C in the same buffer (containing or not 5 mM CaCl<sub>2</sub>) at a flow rate of 60  $\mu$ l/min. Regeneration was carried out with 2 M guanidinium chloride when experiments were run in the absence of added calcium or with sequential injections of 0.25 M EDTA and 2 M guanidinium



FIGURE 1. **Structural model of the CUB1 domain of human PCPE-1.** The model was generated using the known structure of the rat MASP-2 CUB2 domain as a template. Residues either highly conserved in CUB domains (Tyr-67, Asp-68, and Asp-109) or specific to the CUB1 domains of PCPEs (Glu-26, Thr-89, and Phe-90) are indicated. Also shown are loops L3, L5, L7, and L9 previously described in the CUB1 domain structures of human C1s and MAp19, as well as the N and C termini. The figure was prepared using PyMOL.

chloride in the presence of 5 mM calcium. Kinetic data were analyzed using BIAevaluation 4.1 software.

#### RESULTS

Identification of Target Residues—As sequence alignment with CUB domains of known three-dimensional structure (20– 22, 41) showed the strongest homology (33% sequence identity) with the CUB2 domain of rat MASP-2 (22), this structure was used for molecular modeling of the human PCPE-1 CUB1 domain. As shown in Fig. 1, the predicted structure is a  $\beta$ -sandwich, where the root mean square deviation is 1.12 Å with respect to the template. Fig. 2 shows a sequence alignment of selected CUB domains in PCPEs and related human proteins, compared with the positions of sheets and loops in known structures, as illustrated by the MAp19 CUB domain (21). Four large loops are present on the upper face of the structure (*loops L3, L5, L7,* and *L9,* Fig. 1), as being exposed to solvent and accessible for ligand binding.

Because previous studies on MAp19 indicated the importance of the loop regions in CUB domain interactions with MBL and L-ficolin (21), we sought to identify residues involved in the interaction of the PCPE-1 CUB1 domain with procollagens by analysis of conserved and specific residues in the loops in the model shown in Fig. 1. As shown in Fig. 2, and as previously pointed out by Gregory *et al.* (20), CUB domains from several proteins, including the CUB1 domain of PCPE-1, contain three highly conserved acidic residues (Glu-45, Asp-53, and Asp-98 in C1s; Glu-52, Asp-60, and Asp-105 in MAp19; and Glu-60, Asp-68, and Asp-109 in PCPE-1) that have been shown to be

		β3	L3		β4	L4	β5		L5	
				-		_		$\rightarrow$		_
	12	21	31	-	41		51		61	
PCPE1_1	CGGDVK.C	JESG YVI	ASEGFPN.L	YPPNKE	IWTITV	P.EGQT	VSLSF	RVFDL	ELHPACR	YDAL
TSG6	CGGVFTDE	KRIFI	KSPGFPN.E	YEDNOI	YWHIRL	K.YGOR	IHLSF	LDFDL	EDDPGCL	ADYV
CBLN_9	CGGELSG2	ATGSF	SSPGFPN.R	Y PPNKE C	IWYIRT	D.PGSS	IQLTI	HDFDV	EYHSRCN	FDVL
PCPE1_2	GGRLEK	AQGTL	TTPNWPESD	YPPGISC WDDCW77	SWHIIA	P.PDQV	IALTE	EKFDL	PDTYCR	YDSV
TLL2 5	CGGRLKAR	VOTKELY	SHAOFGDNN	YPSEAR	DWVIVA	E. DGYG	VELTE	RTFEV	REEAD.	YDYM
MTLD_5	GGQVRAI	VKTKDLY	SHAQFGDNN	YPGGVD	EWVIVA	E.EGYG	VELVF	QTFEV	EETDCG	YDYM
TLL1_5	<b>C</b> GGRLKAE	SKPRDLYS	SHAQFGDNN	Y P G Q V D C	EWLLVS	E.RGSR	LELSF	QTFEV	EEADCG	YDYV
TLL2_1 MTLD 1	CGETLQ.I	OTTGNF:	SAPGFPN.G SSPEVPN G	YPSYSHC YSAHMHC	VWRISV	T.PGEK	IVLNF	TSLDL	FKSRLCW	YDYV
TLL1_1	CGETLQ. H	SNGN L	SSPGFPN.G	YPSYTH	IWRVSV	T.PGEK	IVLNF		YKSSLCW	YDYI
NRPLN1_1	<b>C</b> GDTIKIE	SPGYL	TSPGYPH.S	Y HP SEK	EWLIQA	PDPYQR	IMINF	NPHFD	LEDRDCK	YDYV
TLL2_2 MTLD 2	CGGDMN . H	KDAGQI	QSPNYPD.D	YRPSKE	VWRITV	S.EGFH	VGLTF	QAFEI		YDYL
TLL1 2	CGGEIR.H	KNEGQI	OSPNYPD.D	YRPMKE	VWKITV	S.ESYH	VGLTF	QSFEI	ERHDNCA	YDYL
TLL2_3	C <mark>GGFIT.</mark>	KLNGT I	ISPGWPK.E	Y P T N K N <mark>C</mark>	VWQVVA	P.AQYR	ISLQF	EVFEL	EGNDVCK	YDFV
MTLD_3	CGGFLT.H	KLNGSI	ISPGWPK.E	Y P P N K N C	IWQLVA	P. TQYR	ISLQF	DFFET	GNDVCK	YDFV
TLL2 4	CAHKIS.S	SVEGT.L	ASPNWPD.K	YPSRREC	TWNISS	T.AGHR	VKLTF	NEFEI	BOHOECA	YDHL
MTLD_4	CDHKVT.S	STSGTI	ISPNWPD.K	Y PSKKE	TWAISS	T.PGHR	VKLTF	MEMDI	ESQPECA	YDHL
TLL1_4	CEQKIH.S	SPSGLI	ISPNWPD.K	YPSRKE C	TWEISA	T.PGHR	IKLAF	SEFEI	EQHQECA	YDHL
C1S 2	CSGDVFT.	ALIGE I	ASPGFPE.K	YPENSEC YPENSEC	EYOIRL	P.KMSE E.KGFO	VVVTL	RREDFDV	EPDSNPPGGMFCR	DSLV
MASP2_2	CSGQVFT	RSGE L	SSPEYPR.P	YPKLSS	TYSISL	E.EGFS	VILDE	.VESFDV	ETHPETLCP	YDFL
C1R_2	CSSELYTE	EASGYIS	SSLEYPR.S	YPPDLRC	NYSIRVI	E.RGLT	LHLKF	. LEPFDI	DDHQQVH <mark>C</mark> P	YDQL
MASP1_2 MASP1 1	CSDNLFT(	QRTGVI	ISPDFPN.P	YPKSSE VPSDSEV	LYTIEL	E.EGFM P DCFP	TKLYF	.EDIFDI	QDHPEVPCP	YDYI
MASP2_1	. LGPKWPH	EP.VFGRL	ASPGFPG.E	ANDQEF	RWTLTA	P.PGYR	LRLYF	THFDL	ELSHLCE	YDFV
C15_1	LLAWVYAR	EPTMYGEI	LSPNYPQ.A	Y PSEVER	SWDIEV	P.EGYG	IHLYF	THLDI	ELSENCA	YDSV
C1R_1	AGGSIPIE	PQKLFGEV	ISPLFPK.P	YPNNFEI	TTVITVI	P.TGYR	VKLVF	QQFDL	PSEGCF	YDYV
	β6 L	6 β7	L	7	β8 L	8 β9	9	L9	β10	
	β6 L	6 β7	, L ▶ 01	7	β8 La	8 β9	→	L9	β10	
	β6 71	6 β7 81	→ L 91	7	β8 La	8 β9 101	→ →	L9	β10 121	
PCPE1_1 PCPE2_1	$\beta 6$ L 71 EVFAG.SC	6 β7 81 STSGQRLGI	91	7 ₽		8 β 101 SNOVTL		L9 111 Egtggr.	β10 121 GFLLWYS	GR
PCPE1_1 PCPE2_1 TSG6	β6 	6 β7 81 GTSGQRLGB AN.GQRIGB DVHGFVGB	91 RFCGTFRPA RFCGTFRPG RYCGDELPD	7 P A	β8 L LVAPO LVSSO LVSSO LISTO	8 β 101 SNQVTL SNKMMV SNVMTL	RMTTDI QMISDI KFLSD	L9 111 egtggr . antagn . asvtag .	β10 121 gfllwys gfmamfs gfqikyv	GR AA AM
PCPE1_1 PCPE2_1 TSG6 CBLN_9	β6 71 EVFAG.SC DVYNG.HJ EIYDS.YI EIYGG.PI	6 β7 81 STSGQRLGH AN.GQRIGH ODVHGFVGH OFHSPRIAG	91 RFCGTFRPA RFCGTFRPA RFCGTFRPG RYCGDELPD 2LCTQRSPE	7 P	β8 LVAPO LVSSO IISTO MQVSSTO	8 β 101 GNQVTL GNVMTL GNUMTL GNELAI	RMTTD QMISD KFLSD RFTD	L9 111 egtggr. antagn. asvtag. lsingr.	β10 121 gfllwys gfmamfs gfoikyv gfnaswo	GR AA AM AV
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2	β6 71 EVFAG.SC DVYNG.HJ EIYDS.YI EIYGG.PI SVFNGCE	6 β7 81 STSGQRLGI DVHGFVGI DVHGFVGI SDDSRLGI	91 RFCGTFRPA RFCGTFRPA RFCGTFRPA QLCTQRSPE XFCGDAVPG	7 A D NP S	β8 LVAP LVSS IIST MQVSST ISSE	8 β 101 GNQVTL GNKMMV GNVMTL GNELLV GNELLV	RMTTER RMTTER KFLSD RFKTE QFVSD OFLSD	L9 111 egtggr. antagn. asvtag. lsingr. lsvtad.	β10 121 	GR AA AM AV TL FR
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI	6 β1 STSGQRLGI STSGQRLGI OVHGFVGI OFHSPRIA SDDSRRLGI OSSAPRLGI	91 RFCGTFRPA RFCGTFRPA RFCGDELPD 2LCTQRSPE KFCGDAVPG KYCGDSPPA RFCGSGPLE	7 <b>A</b> D NP P	β8 LVAP LVSS LIST MQVSST LISE LIVSE	8 β 101 SNQVTL SNKMMV SNELAI SNELLV SNELLV SNELLI SDSLMI	RMTTD QMISD RFLSD QFVSD QFVSD RFRTD	L9 111 ANTAGN . ASVTAG . LSINGR . LSUTAD . LSLTAD . DTINKK .	β10 121 	GR AA AV TL FR ST
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGGEY EAYDG.YI ELFDG.YI	6 β1 STSGQRLGI STSGQRLGI DVHGFVGI DFHSPRIA SDDSRRLGI DSSAPRLGI DSSAPRLGI	91 RFCGTFRPA RFCGTFRPA RFCGTFRPA CGDELPD QLCTQRSPE CGDSPPA RFCGSGPLE RFCGSGPLE RFCGSGPLE	7 A D NP P P E	β8 LVAPG LVSS LSS 	8 101 GNQVTL GNVMTL GNELLI GNELLI GNELLI GDSLMI GDSVLV	RMTTDI QMISD RFLSD QFVSD QFVSD QFLSD RFRTDI VKFHSD	L9 111 egtggr. antagn. asvtag. lsingr. lsutad. lsutad. dtinkk. dtitkk.	β10 121 	GR AA AM TL FR ST ST
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5 TLL1_5 TLL1_5	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGGEY EAYDG.YI ELFDG.YI ELFDG.YI	6 β1 STSGQRLGI DVHGFVGI DVHGFVGI DVHGFVGI SDDSRLGI VNDARRIGI DSTAPRLGI DSTAPRLGI DSTAVGLGI	91 RFCGTFRPA RFCGTFRPG RYCGDELPD QLCTQRSPE KYCGDSPPA RFCGSGPE RYCGSGPE	7 P D NP P E P B.	β8 LVAPO LVSSO IISTO MQVSSTO ISSEO IVSEO IVSEO VYSAO VYSAO	8 101 GNQVTL GNVTL GNELLI GNELLV RNELLI GDSLMI GDSVLV GDSVLV GDSVLV	RMTTDI QMISL KFLSD QFVSD QFLSD RFRTDI KFHSDI HFHTDI	L9 111 EGTGGR. ANTAGN. ASVTAG. LSINGR. LSITAD. DTINKK. DTINKK. DTINKK.	β10 121 GFLLWYS GFMAMFS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFHLRYT GFHLRYT	GR AA AM TL FR ST ST ST ST
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5 TLL2_1 MTLD_1	β6 71 EVFAG.SC DVYNG.HJ EIYGG.PI SVFNGAVS AVFNGGEV EAYDG.YI ELFDG.LI EVRDG.YV EVRDG.FV	6 β1 βTSGQRLGH AN.GQRIGH ODVHGFVGH ODVHGFVGH ODVHGFVGH ODVHGFRIGH SDDSRRLGH SSAPRLGH SSAPRLGH STAVGLGH WRKAPLLGH	91 RFCGTFRPA RFCGTFRPA RFCGDELPD QLCTQRSPP RFCGSGPLE RFCGSGPLE RFCGSGPPE RFCGSKLPE RFCGSKLPE	7 A D S P E P P	β8 LVAPC LVSSC IS	8 β 101 GNVTL GNVTL GNVTL GNVTL GNVLL GNELLV RNELLI GDSLMI GDSVLV GDSVLV DSRLWV DSRLWV	RMTTD QMISD KFLSD RFKTD QFLSD RFRTD KFHSD HFHTD EFRSS EFRSS	L9 111 EGTGGR. ANTAGR. LSINGR. LSINGR. LSITAD. DTINKK. DTINKK. DTINKK. SNILGK. SNWVGK.	β10 121 GFLLWYS GFLLWYS GFLLWYS GFLLWYS GFLASWQ GFLASWQ GFLASWC GFLRYT GFFLRYT GFFAAYE GFFAAYE	GR AA AAV TLL FR STI SII AI
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5 TLL1_5 TLL1_5 TLL2_1 MTLD_1 TLL1_1	β6 71 EVFAG.SC DVYNG.HA EIYGG.PI SVFNGAVS AVFNGGEV EAYDG.YI ELFDG.YI EVRDG.YV EVRDG.YV	6 β1 βTSGQRLGI GTSGQRLGI DDVHGFVGI DDV	91 RFCGTFRPA RFCGTFRPG RFCGTFRPG RFCGDELPD QLCTQRSPPA RFCGSGPLE RFCGSGPPE RFCGSGPPE RFCGSKLPE RFCGSKLPE RFCGSKLPE	7 A D S P E P P P V	β8 LVAPC LVSSC ISSC ISSC ISSC IVSC IVSC IVSC LVST LVST LTST	8 β 101 GNQVTL GNVTL GDSLMI GNVL GNSLL M GNSLL M GNSLL M GNSLL M GNVL M GNSLL M M GNSLL M M M M M M M M M M M M M	RMTTD QMISD RFLSD RFKTD QFLSD RFRTD KFHSD HFHTD EFRSS EFRSS EFRSS	L9 111 EGTGGR. ANTAGN. ASVTAG. LSINGR. LSITAD. DTINKK. DTINKK. DTINKK. SNILGK. SNWVGK.	β10 121 GFLLWYS GFMAMFS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHARYT GFHIRYK GFFAAYE GFFAAYE GFAAVYE	GR AAM AAV TLL FR ST SI AII AI
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5 TLL1_5 TLL2_1 MTLD_1 TLL2_1 NRPLN1_1	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEV EAYDG.YI ELFDG.LI EVRDG.YV EVRDG.YV EVRDG.YV EVRDG.YV	6 β1 STSGQRLGI STSGQRLGI DDVHGFVGI DDVHGFVGI DDVHGFVGI SSAPRLGI SSAPRLGI NCKAPLLGI WRKAPLLGI WRKSPLLGI	91 RFCGTFRPA RFCGTFRPA RFCGTFRPA RFCGDELPD QLCTQRSPE KFCGDAVPG KFCGSGPLE RFCGSGPLE RFCGSGPE RFCGSKLPE RFCGCKLPE RFCGCKLPE	7 P NP S P E P P P P P	β8 LVAPC LVSSC ISSC ISSC ISSC IVSC IVSC LVST LVST LVST LVST LVST	8 β 101 GNVTL GNVTL GNELAI GNELLV GNELLI GDSLMI GDSVLV DSRLWV DSRLWV DSRLWV DSRLWV	RMTTD QMISD RFKTD QFVSD QFLSD RFRTD CFLSD RFRTD HFHTD EFRSS EFRSS EFRSS	L9 111 EGTGGR. ANTAGN. ASVTAG. LSINGR. LSUTAD. DTINKK. DTINKK. DTINKK. SNIUGK. SNWVGK. SNWVGK.	β10 121 GFLLWYS GFMAMFS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFFAXYE GFFAXYE GFFAXYE GFFAXYE GFSIRYE	GRAAM AAV TLL FRT SI AII AII IF
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 TLL2_5 TLL2_1 MTLD_1 TLL2_1 MTLD_1 TLL1_1 TLL2_2 MTLD_2	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEV EAYDG.YI ELFDG.LI EVRDG.YI EVRDG.YI EVRDG.FI EVRDG.PI	6 β1 STSGQRLGI STSGQRLGI DVHGFVGI DVHGFVGI DSTAPRLGI DSTAPRLGI WRKAPLLGI WRKAPLLGI WRKAPLLGI WRKSPLLGI VNDARRIGI SESSTLGI	91 RFCGTFRPA RFCGTFRPA RFCGTFRPA RFCGDELPD 2LCTQRSPE KFCGDAVPG KFCGDAVPG KFCGSGPPE RFCGSGPPE RFCGSKLPE RFCGSKLPE RFCGCKLPE RFCGKLPE RFCGKLPE RFCGKLPE RFCGKLPE RFCGKLPE	7 A NP S P E P P P P P P P D	β8 LVAPC LVSSC ISSC ISSC ISSC IVSE IV	8 101 SNKMMV SNKMMV SNELAI SNELLV SNELLI SDSLLV SDSLLV SSRLWV SNRLWV SNRLWV SNRLWV	RMTTD QMISD KFLSD RFKTD QFVSD QFLSD RFRTD CFRSS EFRSS EFRSS KFVSD KFVSD	L9 111 EGTGGR. ANTAGN. ASVTAG. LSINGR. LSUTAD. LSLTAD. DTINKK. SNILGK. SNILGK. SNWVGK. SNWVGK. SNWVGK. GSINKA	β10 121 GFLLWYS GFMAMFS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFFAAYE GFFAAYE GFFAAYE GFAAVE GFAANFF GFAANFF	GRAAM TLR STISTISTISTISTISTISTISTISTISTISTISTISTIS
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5 TLL2_1 MTLD_1 TLL2_1 NRPLN1_1 TLL2_2 MTLD_2 TLL1_2	β6 71 EVFAG.SC DVYNG.HJ EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI ELFDG.LI EVRDG.YI EVRDG.YI EVRDG.FI EVRDG.HS EVRDG.HS	6 β7 81 STSGQRLGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI WRKAPLLGI WRKAPLLGI WRKSPLLGI SESSILGI SENSPLIGI	91 RFCGTFRPA RFCGTFRPA RFCGDELPD QLCTQRSPE KFCGDAVPG KYCGSGPPE RFCGSGPPE RFCGSKLPE RFCGSKLPE RFCGSKLPE RFCGCKLPE RFCGYEKPE RFCGYEKPE RFCGYEKPE RFCGYEKPE RFCGYEKPE	7 P D P P P P P P D D	β8 LVAPG LVSSG ISST MQVSSTG ISSEG IVSEI	8 101 GNKMMV GNVMTL GNELLV GNELLI GNELLV GDSVLI OSRLWV OSRLWV OSRLWV SNRLWM SNRLWM	RMTTE RMTTE RMTSE RFKTE QMISE RFKTE QFVSE QFVSE RFRTE EFRSS EFRSS EFRSS EFRSS EFRSS EFRSS EFRSS EFRSS EFRSS EFRSS EKFVSE KFVSE	L9 111 EGTGGR . ANTAGN . ASVTAG . LSINGR . LSUTAD . LSLTAD . DTINKK . DTINKK . SNILGK . SNWVGK . SNWVGK . SNWVGK . GSINKA . GSINKA .	β10 121 GFLLWYS GFMAMFS GFLWYS GFNASWQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFFARYE GFFAAYE GFFAAYE GFFAVYE GFAAVFF GFAANFF GFAANFF	GRAAM AAV TIR STI AAI IFE KE KE
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5 TLL1_5 TLL2_1 MTLD_1 TLL1_1 NRPLN1_1 TLL2_2 MTLD_2 TLL1_2 TLL1_2 TLL2_3	β6 71 EVFAG.SC DVYNG.H2 EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI ELFDG.LI EVRDG.YV EVRDG.FV EVRDG.FV EVRDG.FY EVRDG.H3 EVRDG.H3 EVRDG.SC	6 β1 STSGQRLGI AN.GQRIGI DDVHGFVGI DDVHGFVGI DSSAPRLGI DSSAPRLGI DSSAPRLGI DSSAPLLGI WRKAPLLGI WRKAPLLGI WRKAPLLGI SESSTLIGI SESSTLIGI SENSPLIGI	91 RFC GTF RPAG RFC GTF RPAG RFC GDELPD QLC TQRSPE KFC GDAVPG KYC GDSPPA RFC GSGPPE RFC GSGPPE RFC GSKLPE RFC GSKLPE RFC GDKLPE RFC GYEKPE RFC GYEKPE RFC GYEKPE RFC GYEKPE RFC GYEKPE RFC GYEKPE	7 A D S P E P P P P D D V	β8 LVAPG LVSSG LVSSG ISSE ISSE IVSE I	8 101 GNVMTL GNVMTL GNELAI GNELLV GNELLI GDSLMI GDSVLV GDSVLV DSRLWV DSRLWV SNRLWN SNRLWM SNRLWM	RMTTE RMTTE RMTSD KFLSD RFKTE QFVSD QFLSD RFFTD HFHTD EFRSS KFVSD KFVSD KFVSD KFVSD KFVSD	L9 111 EGTGGR. ANTAGN. ASVTAG. LSINGR. LSVTAD. LSINGR. DTINKK. DTINKK. SNILGK. SNWVGK. SNWVGK. SNWVGK. GSINKA. GSINKA. GTVNKA.	β10 121 GFLLWYS GFMAMFS GFLWYS GFNASWQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFFARYE GFFAVE GFFAVE GFAANFF GFAANFF GFAANFF GFAANFF GFAANFF	GAAAM AAV TIR SSI AAI IFE KKE SD
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5 TLL2_1 MTLD_1 TLL1_1 NRPLN1_1 TLL2_2 MTLD_2 TLL1_2 TLL2_3 MTLD_3 TLL1_3	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI ELFDG.LI EVRDG.YV EVRDG.FV EVRDG.FV EVRDG.FY EVRDG.FY EVRDG.ST EVRDG.SS EVRSG.LS EVRSG.LS	6 β1 STSGQRLGI AN.GQRIGI DDVHGFVGI DDVHGFVGI DSTAPRIGI DSTAPRIGI DSTAPRIGI WRKAPLLGI WRKAPLLGI WRKAPLLGI SESSTLIGI SESSTLIGI SENSPLIGI SESSTLIGI SESSTLIGI SESSTLIGI SESSTLIGI SESSTLIGI	91 RFC GTF RPAG RFC GTF RPAG RFC GDELPD QLC TQRSPE KFC GDAVPG KYC GDSPPA RFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSGLE RFC GSKLPE RFC GYEKPE RFC GYEKPE RFC GYEKPE RFC GYEKPE RFC GSETPE RFC GSETPE	7 A D S P E P P P P D D V V V	β8 LVAPG LVSS ISSC ISSC ISSC IVSE IVS	8 101 SNVMTL SNVMTL SNELLV SNELLV SDSLMI SDSVLV SSRLWV SSRLWV SSRLWV SSRLWN SSRLWL SSRLWM SSRLW	RMTTE RMTTE RMTSD KFLSD KFLSD RFKTE CQFVSD QFLSD RFFTD KFVSD KFVSD KFVSD KFVSD EFKSD EFKSD	L9 111 EGTGGR.A ANTAGN.A ASVTAG.LSINGR.LSINGR.LSVTAD. LSITAD. DTINKK.S DTINKK.S SNWVGK.S SNWVGK.S SNWVGK.GSINKA.GSINKA.GSINKA. GSINKA.S GTVNKA.NTVSKK.	β10 121 	GAAMVLL FSTTITE SSITIE SSDDD SDDD
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TLL2_5 MTLD_5 TLL2_1 MTLD_1 TLL1_1 NRPLN1_1 TLL2_2 MTLD_2 TLL2_2 MTLD_2 TLL2_3 MTLD_3 TLL2_4	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI ELFDG.II EVRDG.YI EVRDG.YI EVRDG.FI EVRDG.FI EVRDG.FI EVRDG.SS EVRSG.LS EVRSG.LS EMYDG.PI	6 β1 STSGQRLGI ODVHGFVGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI SESSTLIGI SESSTLIGI SESSTLIGI SESSTLIGI SESSKLHGI OSLAPILGI	91 RFCGTFRPAG RFCGTFRPAG RFCGDELPD 2LCTQRSPE KFCGDAVPG KFCGSGPLE RFCGSGPLE RFCGSGPLE RFCGSGPLE RFCGSGPLE RFCGSGPLE RFCGSGPLE RFCGSKLPE RFCGSKLPE RFCGSEKPE RFCGSEKPE RFCGSEKPE RFCGSEKPE RFCGSKKPD	7 A D S P E P P D D V V V P D P	β8 LVAPG LVSS ISSE ISSE ISSE IVSE	8 101 SNVMTL SNKMMV SNVMTL SNELLV SNELLV SDSLMI SDSVLV SDSVLV SSRLWV SSRLWV SSRLWV SSRLWM SSRLWM SSRLWL SSRLWM SSRLW	RMTTE QMISD KFLSD RFKTD QFLSD RFFKD EFRSS EFRSS KFVSD KFVSD EFKSD EFKSD EFKSD EFKSD	L9 111 ANTAGN . ASVTAG . LSINGR . LSINGR . LSITAD . DTINKK . DTINKK . SNWVGK . SNWVGK . SNWVGK . GSINKA . GSINKA . GTVNKA . NTVSKK . NTVSKK . ASVQRK .	β10 121 GFLLWYS GFMAMFS GFLWYS GFAASVQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFFAAYE GFFAAYE GFFAAYE GFFAAYE GFAANFF GFAANFF GFAANFF GFAANFF GFAAHFF GFKAHFF GFKAHFF GFKAHFF GFKAHFF	GAAMVLL FSTTITEEEEDDDD SSDTE
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TIL2_5 MTLD_5 TLL2_5 MTLD_5 TLL2_1 MTLD_1 TLL2_1 MTLD_1 TLL2_2 MTLD_2 TLL2_3 MTLD_2 TLL2_3 MTLD_3 TLL2_4 MTLD_4 MTLD_4	β6 71 EVFAG.SC DVYNG.HX EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI ELFDG.II EVRDG.YY EVRDG.FY EVRDG.EY EVRDG.EY EVRDG.HS EVRCG.SC EVRSG.LS EVRSG.LS EMYDG.PI EVFDG.RI	6 β1 81 STSGQRLGI ODVHGFVGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI SSTAVGLGI WRKAPLLGI WRKAPLLGI WRKAPLLGI SESSTLIGI SSESSTLIGI SSESSTLIGI SSESSLI	91 RFC GTFRPA RFC GTFRPA RFC GDELPD 2LC TQRSPE KFC GDAVPG KFC GDAVPG KFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSKLPE RFC GSKLPE RFC GYEKPE RFC GYEKPE RFC GSETPE KFC GSEKPE RFC GSKKPE RFC GSKKPE	7 P D S P E P P D D V V P D P	β8 	8 101 SN VTL SN VTL SN VTL SN VTL SN VTL SN VTL SN VL SN SN S	RMTTE QMISD KFLSD RFKTD QFVSD QFLSD RFRTD RFRTD HFHTE EFRSS KFVSD KFVSD EFKSD EFKSD EFKSD RFYSD	L9 111 ANTAGN . ASVTAG . LSINGR . LSINGR . LSINGR . DTINKK . DTINKK . SNILGK . SNWVGK . SNWVGK . GSINKA . GSINKA . GSINKA . GSINKA . MTVSKK . NTVSKK . NTVSKK . NTVSKK .	β10 121 GFLLWYS GFLWYS GFLWYS GFLWYS GFASYK GFASYK GFFASYK GFFARYT GFFARYT GFFARYT GFFARYE GFFAAYE GFFAAYE GFFANFF GFAANFF GFAANFF GFRAHFF GFKAHFF GFQAVHS GFQASHA	GAAMVL FRTSTIITE SSITIE SSDE SSDE
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TIL2_5 MTLD_5 TIL1_5 TIL1_5 TIL1_1 NRPLN1_1 TIL1_1 NRPLN1_1 TIL2_2 MTLD_2 TIL2_3 MTLD_3 TIL1_3 TIL1_3 TIL1_4 MTLD_4 TIL1_4 NRPLN1_2	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI ELFDG.YI ELFDG.LI EVRDG.YV EVRDG.FY EVRDG.FY EVRDG.FY EVRDG.FY EVRDG.SY EVRDS.SY EVRSS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRSS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRSS.SY EVRSS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRSS.SY EVRDS.SY EVRDS.SY EVRSS.SY EVRSS.SY EVRDS.SY EVRDS.SY EVRDS.SY EVRSS.S	6 β1 STSGQRLGI AN.GQRIGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSTAPRLGI OSTAPRLGI OSTAPRLGI OSTAVGLGI WRKAPLLGI WRKAPLLGI SESSTLIGI SEST SES	91 RFC GTFRPA RFC GTFRPA RFC GDELPD 2LC TQRSPE KFC GDAVPG KFC GSGPLE RFC GSGPLE RFC GSGPLE RFC GSGPLE RFC GSGPLE RFC GSGPLE RFC GSKLPE RFC GYLKPE RFC GYLKPE RFC GSEKPE RFC GSEKPE RFC GSKKPE RFC GSKKPE RFC GSKKPE	7 P D P P E P P D D V P D P	β8 	8 101 SNVMTL SNKMMV SNVMTL SNELLI SNELI SNELLI	RMTTE QMISD KFLSD RFKTD QFVSD RFFKD RFFKSD HFHTD EFRSS KFVSD KFVSD EFKSD EFKSD RFYSD RFYSD RFYSD	L9 111 ANTAGR . ASVTAG . LSINGR . LSINGR . LSINGR . DTINKK . DTINKK . SNILGK . SNILGK . SNWVGK . SNWVGK . GSINKA . GSINKA . GSINKA . GSINKA . MTVSKK . NTVSKK . NTVSKK . NTVSKK . ASVQRK . ASVQRK .	β10 121 GFLLWYS GFMAMFS GFLWYS GFMAMFS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFFAAYE GFFAAYE GFFAAYE GFFAAYE GFAANFF GFAANFF GFAANFF GFAANFF GFKAHFF GFKAHFF GFQAYHS GFQASHA GFQATHS	GAAMVL FRTSTIITE SSITIEE SSDEEE TTTTT
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TIL2_5 MTLD_5 TIL1_5 TIL1_1 NTPLN1_1 TIL1_1 NTPLN1_1 TIL2_2 MTLD_2 TIL1_2 TIL2_3 MTLD_3 TIL1_3 TIL1_3 TIL1_4 NTPLN1_2 CIS 2	β6 71 EVFAG.SC DVYNG.HJ EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEV EAYDG.YI ELFDG.II EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.SI EVRDG.SI EVRSG.LS EVRSG.LS ENYDG.RI EVFDG.RI EVFDG.FI EIWGG.FI FVAGD.	6 β1 STSGQRLGI AN.GQRIGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSTAPRLGI OSTAPRLGI OSTAPRLGI OSTAVGLGI WRKAPLLGI WRKAPLLGI SESSTLIGI SESTLIGI SESSTLIGI SESTLIGI SESTLIGI SESTLIGI SESTLIGI SESTLIGI	91 RFCGTFRPA RFCGTFRPA RFCGTFRPA RFCGDELPD 2LCTQRSPE KFCGDAVPG KFCGSGPLE RFCGGSGPLE RFCGGSGF RFC	7 P D S P E E P D D V V V P R P	β8  L   LVAPG LVSS   LVSS ISSG    MQVSSTG ISSG   ISSG IVSE   IVSC IVSE   IVSC IVSE   IVSC IVSE   IVSC IVSE   IVST IVST   IVSS IVSE   IVSE IVAT   INSE INSE   INSE INE	8 101 SNKMMV GNVMTL GNVMTL GNELAI GNELLI GNELLI GDSVLV GDSVLV GDSVLV GDSVLV GDSVLV SNRLWV SNRLWV SNRLWN SNRLWM SNRLLMM	RMTTD QMISD KFLSD RFKTD QFLSD RFFKD RFFKSD EFRSS KFVSD KFVSD EFFKSD RFYSD RFYSD RFYSD RFYSD RFYSD RFYSD RFYSD RFYSD RFYSD RFYSD RFYSD	L9 111 ANTAGR . ASVTAG . LSINGR . LSINGR . LSINGR . LSINGR . DTINKK . DTINKK . SNILGK . SNWVGK . SNWVGK . GSINKA . GSINKA . GSINKA . GSINKA . MTVSKK . NTVSKK . NTVSKK . NTVSKK . NTVSKK . NTVSKK . NTVSKK . NTVSKK . NTVSKK . LSVQRK . ASVQRK . SAIAKE . LTGOKK .	β10 121 GFLLWYS GFMAMFS GFLWYS GFLWYS GFOIKYV GFNASWQ GFSASYK GFIGHYI GFHARYT GFHARYT GFFAAYE GFFAAYE GFFAAYE GFFAAYE GFFAAYE GFFAAYF GFAANFF GFAANFF GFKAHFF GFKAHFF GFKAHFF GFKAHFF GFXAF GFXAFF	GAAMVL RTATIFTSSITIEEEDDDU SSTEEELDD GAAMVL GAAMVL GAAMVL FSTSITIEEEDDDU GAAMVL FSTSITIEEEDDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDDU GAAMVL FSTSITIEEEDU GAAMVL FSTSITIEEEDU GAAMVL FSTSITIEEEDU GAAMVL FSTSITIEEEDU GAAMVL FSTSITIEEEDU GAAMVL FSTSITIEEUU FSTSITIEEUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TIL2_5 MTLD_5 TIL1_5 TIL1_1 NTPLN1_1 TIL1_1 NTPLN1_1 TIL2_2 MTLD_2 TIL1_2 TIL2_3 MTLD_3 TIL1_3 TIL2_4 MTLD_4 TIL1_4 NRPLN1_2 CIS_2 MASP2_2	β6 71 EVFAG.SC DVYNG.HJ EIYDG.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI ELFDG.LI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.SC EVRDG.SC EVRDG.SC EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.SC EVRSC.SC EVRSC.S	6 β1 STSGQRLGI AN.GQRIGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI SSTAVGLGI WRKAPLLGI WRKAPLLGI WRKAPLLGI SSESSTLIGI	91 RFC GTFRPA RFC GTFRPA RFC GDELPD 2LC TQRSPE CGDAVPG KFC GDAVPG KFC GSGPLE RFC GSCKPE RFC GSKKPE RFC GSKFE RFC SF RFC	7 P D S P E E P D D V V P	β8  L   LVAPG LVSS   LVSS ISSE   ISSE ISSE   ITSQ ITSQ   ITSSE ITSSE   ITSSE ITSSE   ITSQ ITSSE	8 101 SNKMMV GNVMTL GNVMTL GNVMTL GNELLI GNELLI GDSVLV GDSVLV GDSVLV GDSVLV GDSVLV GDSVLV SNRLWV SNRLWV SNNMRV SNNMV SNNMV SNNMV SNNMV SNNMV SNNMV SNNMV SNNMV SNNMV SNNV SN	RMTTD QMISD KFLSD RFKTD QFLSD RFFTD RFFTD EFRSS EFRSS KFVSD EFFKSD RFVSD RFYSD RFYSD RFYSD RFYSD RFYSD RFYSD TFVTD	L9 111 ANTAGR . ASVTAG . LSINGR . LSINGR . LSINGR . LSINGR . DTINKK . DTINKK . SNILGK . SNILGK . SNWVGK . SNWVGK . GSINKA . GSINKA . GSINKA . GSINKA . NTVSKK . NTVSKK . NTVSKK . NTVSKK . ASVQRK . ASVQRK . SAIAKE . LTGQKK . ESGDHT .	β10 121 GFLLWYS GFMAMFS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHARYT GFHLRYT GFHLRYT GFFAAYE GFFAAYE GFFAVYE GFSANFF GFAANFF GFAANFF GFAANFF GFKAHFF GFKAHFF GFKAHFF GFKAHFF GFKAHFF GFXANFS GFQASHA GFQASHA GFSANYS GWKLRYH GWKIHYT	GAAMVL RTAAVL SSITIEEEDDDEEELDT VGST
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TIL2_5 MTLD_5 TIL1_5 TIL1_1 NTPLN1_1 TIL1_1 NTPLN1_1 TIL2_2 MTLD_2 TIL2_3 MTLD_3 TIL1_3 TIL2_4 MTLD_3 TIL2_4 MTLD_4 TIL1_4 NRPLN1_2 CIS_2 MASP2_2 CIR_2 CIR_2	β6 71 EVFAG.SC DVYNG.HJ EIYDG.YI EIYGG.PI SVFNGAVS AVFNGGEY EAYDG.YI ELFDG.LI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.SC EVRDG.SC EVRDG.SC EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.SC EVRSC.SC EVRSC.S	6 β1 STSGQRLGI AN.GQRIGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI SSESSTLIGI SENSPLIGI SENSPLIGI SSESSTLIGI	91 RFC GTFRPA RFC GTFRPA RFC GDELPD 2LC TQRSPE CGDAVPG CGDSPPA RFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSCKPE RFC GSKLPE RFC GSEKPE RFC GSEKPE RFC GSEKPE RFC GSEKPE RFC GSKKPD RFC GSK RFC GSK	7 P D S P E E D D V D V P	β8  L   LVAPG LVSS   LVSS ISSE    MQVSSTG ISSE   ISSE ISSE   IVSE ISSE   IVSE IVSE   IVSE ITSQ   IVAT IVAT   IRSE INETE   INISE INISE   INISE INISE   INISE INISE	8 101 SNVMTL GNVMTL GNVMTL GNVMTL GNVMTL GNVMTL GNVMTL GDSVLV GDSVL	RMTTD QMISD KFLSD RFKTD QFLSD RFFTD KFHSD HFHTD EFRSS EFRSS KFVSD EFKSD RFVSD RFYSD RFYSD RFYSD RFYSD RFYSD RFYSD TFVTD TFVTD	L9 111 ANTAGR . ASVTAG . LSINGR . LSINGR . LSINGR . LSINGR . DTINKK . DTINKK . DTINKK . SNILGK . SNWVGK . SNWVGK . GSINKA . GSINKA . GSINKA . GSINKA . MTVSKK . NTVSKK . NTVSKK . ASVQRK . ASVQRK . SAIAKE . ESGDER .	β10 121 GFLLWYS GFMAMFS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHARYT GFHLRYT GFHLRYT GFFAAYE GFFAAYE GFFAVYE GFSANFF GFAANFF GFAANFF GFAANFF GFAANFF GFKAHFF GFKAHFF GFKAHFF GFKAHFF GFXANFS GFQASHA GFQASHA GFSANYS GWKLRYT GWKLRYT	GAAMVLRTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TIL2_5 MTLD_5 TIL1_5 TIL1_1 NTPLN1_1 TIL1_1 NTPLN1_1 TIL1_2 TIL2_2 MTLD_2 TIL1_2 TIL2_3 TIL2_3 TIL2_3 TIL2_3 TIL1_3 TIL2_4 MTLD_4 TIL1_4 NRPLN1_2 C1S_2 MASP2_2 C1R_2 MASP1_1	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEV EVFNGAVS EVFNG.YI ELFDG.LI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.SI EVRDG.SI EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.SI EVFDG.RI EVFDG.RI EVFDG.FI FVAGD.SI EIWSG.SI EIWSG.SI EIWSG.SI EVRSG.S	6 β1 STSGQRLGI AN.GQRIGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI SSESSTLIGI SENSPLIGI SENSPLIGI SESSLLGI SSESSTLIGI S	91 RFC GTFRPA RFC GTFRPA RFC GDELPD 2LC TQRSPE CGDAVPG CGDSPPA RFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSCKPE RFC GSKLPE RFC GSKLPE RFC GSEKPE RFC GSEKPE RFC GSEKPE RFC GSKKPD RFC GSK RFC GSK RFC GSK RFC S RFC	7 P D P P E P	β8  L   LVAPG LVSS   LVSS ISSE   ISSE ISSE   ISSE ISSE   ISSE ISSE   IVSE ISSE   IVSE IVSE   ITSQ VLATC   IVATE INSE   INSE INESE   INSE INESE   INE INE   INE INE   INE INE   INE INE	8 101 SNKMMV GNVMTL GNVMTL GNELLI GNELLI GDSVLV GDSVL	RMTTP QMISD KFLSD KFKTD QFLSD KFFK5D KFFK5D KFVS	L9 111 ANTAGR . ASVTAG . LSINGR . LSINGR . LSINGR . LSINGR . DTINKK . DTINKK . DTINKK . SNILGK . SNWVGK . SNWVGK . SNWVGK . SNWVGK . SNWVGK . SNWVGK . SNWVGK . SNWVGK . SNVVRK . ASVQRK . ASVQRK . ASVQRK . ASVQRK . SAIAKE	β10 121 GFLLWYS GFLWYS GFLWYS GFLWYS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFFAAYE GFFAAYE GFFAAYE GFFAAYE GFAANFF	GAAMVLRTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TIL2_5 MTLD_5 TLL1_5 TIL1_5 TLL2_1 MTLD_1 TLL1_1 NRPLN1_1 TLL2_2 MTLD_2 TLL2_3 MTLD_3 TLL2_3 MTLD_3 TLL2_4 MTLD_4 TLL1_4 NRPLN1_2 C1S_2 MASP2_2 C1R_2 MASP1_1 MASP2_1	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEV EVFNG.YI ELFDG.LI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.SI EVRDG.SI EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.SI EVFDG.RI EVFDG.RI EVFDG.SI EVFDG.SI EVFDG.SI EVRSG.S	6 β1 STSGQRLGI AN.GQRIGI ODVHGFVGI OFHSPRIAG SDDSRRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI OSSAPRLGI SSESSTLIGI SENSPLIGI SENSPLIGI SESST SESST	91 RFC GTFRPA RFC GTFRPA RFC GDELPD 2LC TQRSPE CGDAVPG CGDSPPA RFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSGPPE RFC GSCKPE RFC GSKLPE RFC GSEKPE RFC GSEKPE RFC GSEKPE RFC GSKKPD RFC GSK RFC SC RFC SC R	7 P D P E E P D P	β8  L   LVAPG LVSS   LVSS ISSE    MQVSSTG ISSE   ISSE ISSE   IVSE ISSE   IVSE IVSE   IVATE INSE   IVATE INSE   IVATE INSE   INSE INSE	8 101 SNKMMV GNVMTL GNELAI GNELLI GNELLI GDSVLV GDSVL	RMTTD QMISD KFLSD KFKTD QFLSD KFFK5D KFFK5D KFVSD KFVSD KFVSD KFVSD KFVSD KFVSD KFVSD KFVSD KFVSD KFVSD KFVSD TFFTD LFFTD LFFTD TFRSD	L9 111 GTGGR. ASVTAG. LSINGR. LSINGR. LSINGR. LSINGR. SINKA. DTINKK. DTINKK. DTINKK. SNILGK. SNWVGK. SNWVGK. GSINKA. GSINKA. GSINKA. GSINKA. SNVVGK. NTVSKK. NTVSKK. ASVQRK. ASVQRK. ASVQRK. LSQDHT. ESGDSR. FSNEERP.	β10 121 GFLLWYS GFLWYS GFLWYS GFLWYS GFQIKYV GFNASWQ GFSASYK GFIGHYI GFHLRYT GFHLRYT GFHLRYT GFFAAYE GFFAVYE GFFAVYE GFAANFF	GAAMVLRTTTSSSITIEEEDDDEEELDTEAAVE
PCPE1_1 PCPE2_1 TSG6 CBLN_9 PCPE1_2 PCPE2_2 TILL2_5 MTLD_5 TLL1_5 TLL2_1 MTLD_1 TLL1_1 NRPLN1_1 TLL2_2 MTLD_2 TLL2_3 MTLD_3 TLL1_3 TLL2_3 MTLD_3 TLL2_3 MTLD_3 TLL1_3 TLL2_4 MTLD_4 TLL1_4 NRPLN1_2 C1S_2 MASP2_2 C1R_2 MASP1_2 MASP1_1 MASP2_1 C1S_1 C1S_1	β6 71 EVFAG.SC DVYNG.HA EIYDS.YI EIYGG.PI SVFNGAVS AVFNGGEV AVFNGGEV EVFDG.YI ELFDG.LI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.YI EVRDG.BI EVRDG.SI EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.LS EVRSG.S EV	6 β7 81 STSGQRLGI AN.GQRIGI ODVHGFVGI ODVHGFVGI OSTAPRIGI OSTAPRIGI OSTAPRIGI OSTAPRIGI OSTAPRIGI OSTAPRIGI SESSTLIGI	91 RFC GTFRPA RFC GTFRPA RFC GDELPD 2LC TQRSPE CGDAVPG KFC GDSCPPA RFC GSCPPE RFC GSCPPE RFC GSCPPE RFC GSCPPE RFC GSCPPE RFC GSCPPE RFC GSCPE RFC GSC RFC GSCPE RFC SSCPE RFC SSCPE	7 P D P E E E P P P P P P P P P P TEQTPGC TERAPGF NPHSPIV	β8  L   LVAP LVSS   LVSS LVSS   ISSE LVSS   ISSE LVSS   IVSE LVST   LVST LVST   LVAT LVAT   LVAT LVAT   LVST LVAT   LVST LVAT   LVAT LVAT   LVAT LVAT	8 101 SNKMMV GNVMTL GNVMTL GNVMTL GNVMTL GNVMTL GNVMTL GNVMTL GNVVTL GNVVT SNTLWM SNTLWM SNNMRV SNNMV SNNMRV SNNMV	RMTTD QMISD KFLSD CKFLSD CKFLSD CFVSD CFFSS CFFS	L9 111 ANTAGR . ASVTAG . LSINGR . LSINGR . LSINGR . LSINGR . SNILGK . SNILGK . SNILGK . SNILGK . SNWVGK . SNWK . SNWVGK .	β10 121 GFLLWYS GFLWYS GFLWYS GFLWYS GFQIKYV GFSASYK GFIGHYI GFHLRYT GFHLRYT GFHLRYT GFFAAYE GFFAAYE GFFAAYE GFAANFF GFAA	GAAMVLRTTFSTITIKKEEDDDEEELDTEAAVET

FIGURE 2. **Alignment of CUB domains.** Amino acid sequences of selected extracellular human proteins were aligned using ClustalW. Conserved cysteines are highlighted in *pink*, tyrosines in *blue*, and acidic residues involved in calcium coordination (based on known three-dimensional structures) in *red*. Residues in surface loops specific to PCPE CUB1 domains are highlighted in *green*. Secondary structure elements (β sheets and loops) are also shown based on the structure of the MAp19 CUB domain (equivalent to MASP2-CUB1). Residue numbering refers to the CUB1 domain of mature PCPE-1 (*i.e.* after cleavage of the signal peptide). Individual CUB domains in multi-CUB domain proteins are distinguished by numbers starting at the N terminus. *TSG6*, tumor necrosis factor-stimulated gene *6; CBLN*, cubulin; *NRPLN1*, neuropilin-1. The figure was prepared using ESPript (53).

#### CUB Domain Specificity in PCPE-1

involved in the coordination of a calcium atom in the threedimensional structures of C1s (20) and MAp19 (21). This calcium binding site is formed by the loops located on the upper face of the  $\beta$ -sandwich structure (Fig. 1). Furthermore, a tyrosine residue involved in stabilizing this calcium binding site through formation of an H-bonding network (Tyr-17 in C1s, Tyr-24 in MAp19) is also conserved in PCPE-1 (Tyr-32). These sequence identities therefore point to a putative calcium binding site in PCPE-1. In addition to Glu-60, Asp-68, and Asp-109, Tyr-67 in PCPE-1 CUB1 (Tyr-52 in C1s and Tyr-59 in MAp19) is also conserved and is predicted to occur in the vicinity of this putative calcium binding site, with its side chain exposed to solvent (Fig. 1).

Calcium coordination involves one side-chain oxygen of Glu-45 in C1s or its homologue Glu-52 in MAp19. In contrast, while Asp-53 and Asp-98 in C1s, and Asp-60 and Asp-105 in MAp19, both contribute to calcium coordination, there are differences in the actual atoms involved, involving both side-chain and carbonyl oxygens (20, 21). For these reasons Glu-60 and Tyr-32 in PCPE-1 CUB1 are likely to be intimately involved in calcium binding, whereas Tyr-67, Asp-68, and Asp-109, although strongly conserved in CUB domains and also likely to be involved in calcium binding, might also be available for interactions with procollagens.

To identify residues specific to PCPE CUB1 domains that might be involved in binding to procollagens, we carried out a sequence alignment with CUB domains most closely related, in the data base of human proteins, to those of PCPEs. As shown in Fig. 2, only three surface-exposed residues specific to the CUB1 domains of PCPE-1 and -2 were found within loops L3, L5, L7, and L9: Glu-26, Thr-89, and Phe-90. We therefore identified six residues potentially involved in interactions with procollagens, including three conserved residues (Tyr-67, Asp-68, and Asp-109) and three specific to PCPE CUB1 domains (Glu-26, Thr-89, and Phe-90).

Production and Characterization of Site-directed Mutants-To determine the roles of the residues identified by the molecular modeling, we produced a series of recombinant site-directed mutants and tested their effects on PCPE-1 enhancement of procollagen processing by BMP-1. Alanine mutations were systematically introduced into recombinant full-length human PCPE-1 followed by expression in 293-EBNA cells. All constructs contained a polyhistidine tag at the C terminus of the NTR domain, both for ease of purification and also to separate the recombinant protein from traces of endogenous wild-type PCPE-1 produced by these cells (28). The presence of the His tag had no effect on the tolloid proteinase-enhancing activity of PCPE-1 (data not shown). As shown in supplemental Fig. S1, all mutants were expressed and efficiently purified by nickel-agarose chromatography. In addition, CD analysis showed no major changes in secondary structure as a result of the mutations (supplemental Fig. S2 and Table S1).

*Residues Involved in PCPE-enhancing Activity*—Enhancing activity was initially tested using the previously described miniprocollagen III substrate (28), which consists of the C-propeptide region, the C-telopeptide region, and the 33 most C-terminal (Gly-Xaa-Yaa) triplets of human procollagen III, in the presence of recombinant, human BMP-1. As shown in Fig. 3,



FIGURE 3. Effects of PCPE-1 mutations on BMP-1-enhancing activity using a mini-procollagen III substrate. *A*, mini-procollagen III (19 pmol) was incubated for 2 h at 37 °C in the presence or absence of 19 pmol of wild-type His-tagged PCPE-1 or its mutants, in 50 µl of assay buffer, with or without 0.4 pmol of BMP-1, followed by analysis in 10% acrylamide gels under nonreducing conditions and staining with Coomassie Blue. *B*, densitometric analysis of the gel shown in *A*. Results are expressed as percentages of the total substrate (mini-procollagen III) plus product (C-propeptide III). *MiniIII*, mini-procollagen III; *CPIII*, C-propeptide III.

the presence of PCPE-1, at an equimolar ratio with regard to the substrate, enhanced the activity of BMP-1 on mini-procollagen III, as previously reported (28). As judged both qualitatively and quantitatively from the Coomassie Blue-stained gels (Fig. 3), the degree of enhancement appeared similar to wild type for the PCPE CUB1-specific E26A and T89A mutants. In contrast, the PCPE CUB1-specific F90A mutation led to an almost total loss of PCPE-enhancing activity. Mutation of the conserved residues Asp-68 and Asp-109, likely to be involved with the putative calcium binding site, led to a marked drop in enhancing



FIGURE 4. Effects of PCPE-1 mutations on BMP-1-enhancing activity using a radioactive procollagen 1 substrate. *A*, [<sup>3</sup>H]tryptophan-labeled procollagen 1 (12.5 pmol) was incubated in 100  $\mu$ l of assay buffer for 1 h and 30 min at 37 °C, with or without 12.5 pmol of wild-type PCPE-1 or its mutants, in the presence of 0.24 pmol of BMP-1-FLAG, followed by scintillation counting of released C-propeptides. The stimulation factor is the ratio of BMP-1 activity in the absence and presence of PCPE-1. Means  $\pm$  S.D.s are indicated, for three independent measurements. *B*, BMP-1 activity (dpm) as a function of the molar ratio PCPE-1:[<sup>3</sup>H]procollagen 1 for wild-type PCPE-1 and for the F90A and D109A mutants, with fixed amounts of enzyme and substrate. The conditions were the same as in *A* except that the amounts of PCPE-1 and mutants were varied. Means  $\pm$  S.D.s are indicated, for three independent measurements.

activity. The Y67A mutation was associated with a relatively small drop in enhancing activity.

To obtain additional quantitative data on the effects of the mutations, PCPE-1-enhancing activity was measured using a  $[^{3}H]$ tryptophan-labeled procollagen I substrate, as described (27, 42), except that BMP-1 FLAG was used instead of BMP-1. The presence of the C-terminal FLAG had no effect on BMP-1 activity (43). The extent of enhancement, measured as the ratio of BMP-1 activity in the presence and absence of PCPE-1, is shown in Fig. 4*A*. The results confirm those obtained with the mini-procollagen III substrate. The E26A and T89A mutations had little effect on enhancing activity, although E26A appeared to be slightly more active (+14%) and T89A slightly less active (-11%) than the control. In contrast, the Y67A mutation led to a pronounced decrease (-29%), whereas mutations D68A, F90A, and D109A almost completely abolished enhancing activity (-85%, -88%, and -90%, respectively).

To determine whether the loss in enhancing activity was due to a change in the optimum enhancer:substrate stoichiometry, the enhancing activities of two of the most affected mutants (F90A and D109A) were measured as a function of the ratio of

#### CUB Domain Specificity in PCPE-1

enhancer to [<sup>3</sup>H]procollagen I, in comparison with wild-type PCPE-1. As expected (30, 44), maximum enhancement with wild-type PCPE-1 was obtained at an enhancer:substrate ratio of 1:1. In contrast, for the mutants F90A and D109A, increasing the enhancer:substrate ratio up to 5:1 led only to a gradual increase in enhancing activity, with no sign of a plateau, and enhancing activity remained considerably less than that of the control (Fig. 4*B*).

Effects of Mutations on PCPE-1 Binding to Procollagen—To determine whether the observed changes in PCPE-1-enhancing activity of the mutants were associated with changes in binding affinity for the substrate, the interactions of wild-type and mutant forms of PCPE-1 with immobilized mini-procollagen III were measured using surface plasmon resonance technology (BIAcore). As expected, injection of wild-type PCPE-1 over immobilized mini-procollagen III resulted in concentrationdependent specific binding to the surface (data not shown). As shown in Fig. 5, the binding of PCPE-1 to the immobilized miniprocollagen III was initially fast and then slowed down without reaching a plateau, even when the injection time was doubled. This sensorgram is reminiscent of the sensorgrams previously obtained when the C-propeptides of procollagen I or III were injected over immobilized PCPE-1 (35). The various PCPE-1 mutants were then sequentially injected over the same regenerated mini-procollagen III surface at a fixed concentration of 100 nm, in the absence or presence of 5 mm  $Ca^{2+}$  (Fig. 5). At first glance, the most striking effect of calcium was the shape of the sensorgrams: association was even faster with calcium, whereas dissociation was much slower especially for wild-type PCPE-1 and E26A. Moreover, absolute responses were significantly higher for all the proteins, even for the mutants that gave a very low signal in the absence of calcium (D68A and F90A). Altogether, these results suggest that the interaction between PCPE-1 and mini-procollagen III is much tighter in the presence of calcium, and this remains true for mutants of the putative calcium binding site (D68A and D109A).

With regard to mutations having little effect on PCPE-enhancing activity, binding of E26A to mini-procollagen III was similar to control in both conditions but binding of T89A in the absence of calcium seemed significantly impaired. The binding was partially restored in the presence of calcium to reach a level (-11% at the end of injection compared with wild type), which correlates well with the small drop in enhancing activity observed with this mutant. Concerning the mutations having major effects on enhancing activity, as shown in Fig. 5, the extent of binding of the Y67A mutant was less than wild-type, in agreement with the drop in enhancing activity (Figs. 3 and 4), whereas binding of mutants D68A, F90A, and D109A to miniprocollagen III was even lower (although partially rescued by calcium addition), consistent with the virtual abolition of enhancing activity (Figs. 3 and 4). Thus, in general, changes in binding affinity to mini-procollagen III mirrored the changes in enhancing activity introduced by the mutations.

Surface plasmon resonance was also used to determine dissociation constants for PCPE-1 and its mutants. On the same sensor chip (with a low level of immobilization to avoid mass transfer limitations), the concentrations of injected PCPE-1 and mutants were varied to obtain kinetic and affinity con-





FIGURE 5. BIAcore analysis of the binding of PCPE-1 and its mutant forms to immobilized mini-procollagen III in the absence (A) and presence (B) of 5 mm calcium. Mini-procollagen III was covalently coupled to one flow cell of a CM4 sensor chip (305 resonance units). A control flow cell was similarly prepared but in the absence of added protein. His-tagged, recombinant human PCPE-1 in HBS-P (A) or HBS-p plus 5 mm CaCl<sub>2</sub> (B) buffer was injected simultaneously over both flow cells (flow rate 60  $\mu$ I/min) and the response (resonance units) automatically calculated from the difference in signals. Dissociation was measured by injection of buffer alone. Wild-type PCPE-1 or its mutant forms were sequentially injected, at a fixed concentration of 100 nm, over the same mini-procollagen III surface, after regeneration with 2 m guanidinium chloride (A) or 0.25 m EDTA followed by 2 m guanidinium chloride (B).

stants. Analysis of the data using the BIAevaluation 4.1 software showed a significant improvement of the quality of the fit for PCPE-1 when the "heterogeneous ligand" model was used instead of the basic 1:1 binding model. The reasons for this ligand heterogeneity (mini-procollagen III in this case) are not completely clear, but plausible explanations are: 1) some critical lysine in the ligand may be involved in cross-link formation with the sensor chip surface leading to a decreased activity for a part of the mini-procollagen III population; 2) there are two populations in the initial mini-procollagen III preparation (due for example to incomplete loading with calcium); and 3) there are two independent binding sites on the ligand. To test the first hypothesis, we immobilized an anti-c-Myc antibody on the sensor chip, captured the mini-procollagen III which is c-Myc-

#### TABLE 2

### Dissociation constants of PCPE-1 and its mutants for mini-procollagen III

Constants were determined after analysis of the curves obtained by surface plasmon resonance with at least six different concentrations of PCPE-1 (and mutants) after immobilization of 305 resonance units of mini-procollagen III. The data were obtained in the absence of calcium and analyzed with BIAevaluation 4.1 software (BIAcore). Two models ("1:1 Binding with mass transfer" and "Two-state reaction") could be excluded based on preliminary controls performed according to the recommendations of the manufacturer. Other models were used to fit the data and the best fits were always obtained with the "Heterogeneous Ligand" model.

				-	-			
	PCPE-1	E26A	Y67A	D68A	T89A	F90A	D109A	
<i>K</i> <sub>D1</sub> (пм)	1.8	1.8	12	$ND^{a}$	46	ND	66	
$K_{D2}$ (nm)	48	118	188	ND	274	ND	689	
Chi2	0.814	2.19	4.84		4.19		1.66	

<sup>a</sup> ND, not determined.

tagged (28) and injected PCPE-1 as before. Capture was very stable (three c-Myc tags are present on one mini-procollagen III molecule), but even in this case, the "heterogeneous ligand" model was the most appropriate to fit the kinetic data. Consequently, the first hypothesis can be ruled out, but the last two hypotheses remain valid as the collagen triple helix has already been suspected to bind PCPEs (31) and mini-procollagen III could possibly offer several binding sites to PCPE-1.

The dissociation constants obtained for the interaction between PCPE-1 and mini-procollagen III with the "heterogeneous ligand" model were 1.8 and 48 nM (Table 2). The first constant is in agreement with the value obtained for the interaction between procollagen I and PCPE-1 (1.1 nm (35)), although the kinetic parameters differ ( $k_{on}$  and  $k_{off}$  are, respectively, 16-fold and 26-fold higher). The same analysis was performed with the various mutants except for D68A and F90A, for which the kinetic constants were too high to be accurately fitted. Reliable fits were obtained for E26A (no change in  $K_{D1}$ ; slight increase in  $K_{D2}$ ) and D109A (36- and 15-fold increase in  $K_{D1}$  and  $K_{D2}$ , respectively, compared with wild type). For Y67A and T89A, statistical parameters were not as good as for the previous mutants but in agreement with previous results, deduced  $K_D$  values were significantly higher than for wild type and lower than for D109A. Determination of the dissociation constants in the presence of calcium gave less reliable data due to the strength of the interaction that hindered complete regeneration of the surface between injections.

Although calcium is clearly a key partner in the interaction between mini-procollagen III and PCPE-1, it is impossible to infer from previous data if calcium binds to mini-procollagen III, PCPE-1, or both. We have already shown that the C-propeptide of procollagen III binds calcium (35) and thus, mini-procollagen III like PCPE-1 is expected to bind calcium. To explore further the role played by calcium in PCPE-1, we monitored the effect of calcium binding on the intrinsic fluorescence of PCPE-1. As can be seen in Fig. 6A, the emission spectrum obtained with wild-type PCPE-1 has a maximum  $\sim$  335 nm. Upon addition of a 40-fold excess of EGTA, the spectrum was not strongly modified, either because the protein does not contain calcium or because calcium cannot be removed by this treatment. Interestingly, however, when calcium is added to the EGTA-treated PCPE-1 (free concentration around 1.9 mM), the emission spectrum was blue-shifted and the peak area significantly decreased. The same type of behavior has already



FIGURE 6. Changes in intrinsic fluorescence observed upon calcium binding to PCPE-1 (A) and mutant D68A (B). Solutions of protein (2.5  $\mu$ M in 20 mM HEPES, pH 7.4, 0.15 M NaCl) were excited at 280 nm, and emission spectra were recorded between 295 and 400 nm (*continuous line*). EGTA (100  $\mu$ M final concentration; *dotted line*) and CaCl<sub>2</sub> (1.9 mM final concentration; *dashed and dotted line*) were sequentially added, and spectra recorded 5 min after each addition. In the case of the wild-type (WT), the same experiment was also performed with MgCl<sub>2</sub> instead of CaCl<sub>2</sub> (*dashed line*; curve indistinguishable from EGTA treated). Total dilution in each case was less than 1%. In *B*, the spectra recorded with *N*-acetyl tryptophan amide (NATA) alone (*continuous line*), NATA plus 100  $\mu$ M EGTA (*dotted line*) and NATA plus 100  $\mu$ M EGTA plus 1.9 mM CaCl<sub>2</sub> (*dashed line*; curve indistinguishable from NATA plus EGTA) are also shown for comparison.

been observed for C1s, another CUB-containing protein for which calcium binding has clearly been demonstrated (20, 45). When the experiment was performed with  $MgCl_2$  instead of  $CaCl_2$ , no change in fluorescence was observed (Fig. 6A) indicating that the changes observed with calcium were specific to this ion.

In contrast to wild-type PCPE-1, D68A seemed to be sensitive to both EGTA and  $Ca^{2+}$  (Fig. 6*B*), because EGTA caused an increase in intrinsic fluorescence, whereas  $CaCl_2$  caused a large diminution of the peak area. This probably indicates that calcium is more labile in D68A compared with wild-type PCPE-1, but even the wild-type does not seem to be completely loaded with calcium in our conditions (no calcium present during the purification). Interestingly, other mutants supposedly involved in calcium binding or close to the calcium site (Y67A and D109A) displayed spectra similar to D68A, whereas mutants of the specific residues T89A and F90A behaved like wild type (data not shown).

#### DISCUSSION

We recently demonstrated that PCPEs are more specific than the tolloid proteases they activate (28), and because their action is solely on fibrillar procollagens, we proposed that PCPEs could be interesting targets to develop new anti-fibrotic therapies. This will require a better knowledge of the structural features that allow PCPEs to bind to procollagens, and the work described here is aimed at identifying residues in the CUB1 domain of PCPEs involved in the interaction with procollagens I–III.

Identification of regions involved in procollagen binding was hindered by the lack of high resolution structural data for PCPE-1. It is known that tolloid protease-enhancing activity is a property of the CUB domain-containing region (34, 46). Furthermore, data from our laboratory<sup>4</sup> have shown that proteolytic cleavage within the CUB1 domain leads to loss of PCPE function. In addition, others have shown the importance of surface-exposed loops connecting  $\beta$ -strands in the interactions of CUB domains in neuropilin-1 (7) and MAp19 (21). In the light of these observations, we decided to focus on residues in connecting loops that were either conserved in CUB domains or specific to the CUB1 domains in PCPEs.

Among the conserved residues, the most prominent effects were obtained when the acidic residues Asp-68 and Asp-109, two putative calcium ligands, were individually mutated to neutral alanines. Moreover, we demonstrate here that calcium affects the local structure of PCPE-1 and that this effect cannot be reproduced with another ion sharing similar properties such as magnesium. This supports the previous hypothesis (20) that PCPE CUB domains belong to the subset of CUB domains that bind calcium and that this ion is necessary to maintain their structural integrity. Although the putative calcium binding site is found in about two-thirds of the CUB domains, this ability to bind calcium was only discovered recently with the crystal structure of the CUB1-EGF fragment of C1s (20). Other such CUB domains are found in the complement proteases C1s, C1r and MASPs, the tolloid enzymes, the inflammatory GAG-binding protein tumor necrosis factor-stimulated gene 6 (TSG-6) and in receptors such as cubilin and neuropilins (Fig. 2).

Crystal structures have revealed that the exact nature of the calcium-binding residues and the involvement of side-chain and/or main-chain carbonyls are subject to variations from one CUB domain to another. For example, the aspartate homologous to PCPE Asp-68 contributes one (MAp19) or two (C1s) ligands, and the aspartate homologous to Asp-109 can be involved through its side chain (MAp19) or through its main chain (C1s). This raises the question of whether calcium ligands can also develop electrostatic contacts with the interacting partner of the CUB domain. This question was discussed by Gregory and colleagues (21) who concluded in favor of a disruption in structural integrity rather than of a key interaction of the CUB domain with the partner to explain the drops in affinity observed upon mutation of some calcium ligands. The iden-

<sup>&</sup>lt;sup>4</sup> B. Font, C. Moali, D. Eichenberger, and D. J. S. Hulmes, unpublished observations.

#### CUB Domain Specificity in PCPE-1

tification of the exact nature and role of the calcium ligands in the CUB1 module of PCPEs will certainly require more investigation, because the full-length PCPEs are not the most appropriate tools to distinguish between the role of the calcium ion in CUB1 *versus* CUB2. Production of the CUB1 domain alone is in progress and should help to dissect the structure-function relationships in this particular domain. What is already clear, however, is the major role played by calcium ions in CUB domains. This is illustrated by the occurrence of several mutations in calcium ligands in CUB2, CUB3, and CUB4 of *Drosophila* tolloid, leading to disruptions in dorso-ventral patterning (47, 48), as well as in CUB1 of human MASP-2, resulting in chronic inflammatory disease (49). These mutations have been shown to impede tolloid function (43, 50) and formation of the MASP-2/mannan binding lectin complex (21, 49), respectively.

We also probed the role of tyrosine 67, a well conserved residue immediately preceding Asp-68. Mutation of the homologous residue in MAp19 was shown to yield the most pronounced effect on dissociation constants for MBL and L-ficolin among a series of 12 mutants (21) and to be directly involved in the interaction with the binding partners. In our case, the effect of replacing Tyr-67 by alanine was significant and reproducible but not as severe as in the case of MAp19. This result seems to indicate that, in some cases, conserved residues can influence differently the structure and activity of CUB domains even though they maintain a conserved framework of structural features. In the CUB1 domain of PCPE-1, the interaction mediated by Tyr-67 may not be as strong as in MAp19.

One of the most striking results in this study was the demonstration that phenylalanine 90 is a major determinant of PCPEenhancing activity, and of the PCPE-1/procollagen III interaction, probably through hydrophobic contacts. We also note that mutation of threonine 89 led to a reduction in binding, and a small drop in enhancing activity. It is possible that this residue is involved in a hydrogen-bonding network that stabilizes the PCPE interaction site. We note that the sequence CGTFRP, including Thr-89 and Phe-90, is completely conserved in fish, frog, rodent, and human PCPEs (not shown), suggesting an important functional role. In contrast, PCPE-1 mutant E26A was found to stimulate BMP-1 activity at least as much as wild type. Because Glu-26 is diametrically opposite to Phe-90 in the three-dimensional structure, this may help to delineate the surface most probably involved in procollagen binding. We note that Glu-26 is adjacent in the three-dimensional structure to the N-terminal sequence of PCPE-1.

The agreement between the observed changes in PCPE-enhancing activity and binding to the mini-procollagen III substrate is striking. The mutants can be ranked in the same order in both assays. These data support the notion that the mechanism of action of PCPE involves binding to the BMP-1 substrate, although binding to the enzyme cannot be excluded (36).

Interestingly, mutations within adjacent loops in the neuropilin-1 CUB1 domain also resulted in dramatic decreases in the affinity of this receptor for class 3 semaphorins (7). These mutations are also located on the upper face of the CUB domain (see Fig. 1), in loops L3, L5, and L9, whereas a mutation in loop L2 on the lower face had no effect. Phenylalanine 90 in PCPE-1 CUB1 is also located on the upper face, but in loop L7. Similarly, residues involved in CUB domain interactions in *Drosophila* tolloid (50) and MAp19 (21) are located on the upper face, containing the calcium binding site. Based on these observations, we propose that the CUB L3, L5, L7, and L9 loops first defined in spermadhesins (41), and subsequently found in the calciumbinding CUB domains of C1s (20) and MAp19 (21), are the principal regions involved in ligand binding.

Although the loop regions in CUB domains seem to be required for ligand binding, the less mobile  $\beta$ -sheets appear to be involved in homo- (20, 21) and hetero-dimerization (41). In contrast to complement proteases and spermadhesins, however, PCPEs do not seem to form dimeric structures, as revealed by small angle x-ray scattering, even at the high concentrations required for data collection (51). The ability of CUB domains to form multimers may therefore be a versatile property of CUBcontaining proteins that depends on the nature of the CUB module and on available partners for multimerization in a given physiological context. Thus, even though binding modes seem to be conserved among calcium-containing CUB domains, specific functions may result from different sets of residues in loops and  $\beta$ -strands and through a choice of functional modules located on either side of the CUB domain (EGF, CUB, NTR, and others).

We have previously shown that PCPE-1 binds to procollagen I with a  $K_d$  of 1.1 nm (35), but the one or more binding sites in procollagens are not known. Interestingly, PCPE-1 also binds both cleavage products released by BMP-1, pN-collagen, and the C-propeptide trimer. Because we have recently demonstrated that most of the triple helix can be removed without affecting PCPE activity (28), it is tempting to speculate that PCPEs bind to both the C-propeptide and the C-telopeptide, thereby inducing a conformational change in the procollagen molecule. The natural occurrence of shorter fragments of PCPE-1 that retain their enhancing activity despite the loss of the NTR domain allowed the demonstration that CUB domains alone are sufficient to enhance PCP activity (33, 34). Although the requirement for both CUB1 and CUB2 domains has never been challenged, this study showing that a single mutation in CUB1 can severely reduce the activity of PCPE-1, without affecting the overall structure of the protein, clearly demonstrates that CUB1 is required for enhancing activity. Even though the role of the CUB2 domain remains completely unexplored, we hypothesize that one PCPE CUB domain could bind to the C-propeptide and the other to the C-telopeptide in agreement with the size and shape of the various partners (51, 52). Future work will focus on which CUB domain binds to which side of the BMP-1 cleavage site and where in the C-telopeptide/C-propeptide the interaction sites are located. In the light of the results described here, especially the necessity for the specific residue Phe-90 in maintaining enhancing activity, it seems unlikely that CUB1 can be exchanged with another CUB domain, for example PCPE-1 CUB2, without loss of activity, but this remains to be demonstrated.

Acknowledgments—We are grateful to Alexandre Pozza and Attilio di Pietro for their help during the fluorescence experiments and to Gérard Arlaud and Nicole Thielens for helpful discussions.

#### REFERENCES

- Gaboriaud, C., Thielens, N. M., Gregory, L. A., Rossi, V., Fontecilla-Camps, J. C., and Arlaud, G. J. (2004) *Trends Immunol.* 25, 368–373
- Sorensen, R., Thiel, S., and Jensenius, J. C. (2005) Springer Semin. Immunopathol. 27, 299–319
- Lee, H. X., Ambrosio, A. L., Reversade, B., and De Robertis, E. M. (2006) Cell 124, 147–159
- Davidson, G., Mao, B., del Barco Barrantes, I., and Niehrs, C. (2002) Development 129, 5587–5596
- 5. Greenspan, D. S. (2005) Top. Curr. Chem. 247, 149-183
- Neufeld, G., Cohen, T., Shraga, N., Lange, T., Kessler, O., and Herzog, Y. (2002) Trends Cardiovasc. Med. 12, 13–19
- Gu, C., Limberg, B. J., Whitaker, G. B., Perman, B., Leahy, D. J., Rosenbaum, J. S., Ginty, D. D., and Kolodkin, A. L. (2002) *J. Biol. Chem.* 277, 18069–18076
- Reigstad, L. J., Varhaug, J. E., and Lillehaug, J. R. (2005) FEBS J. 272, 5723–5741
- 9. Topfer-Petersen, E., Romero, A., Varela, P. F., Ekhlasi-Hundrieser, M., Dostalova, Z., Sanz, L., and Calvete, J. J. (1998) *Andrologia* **30**, 217–224
- Tao, Z., Peng, Y., Nolasco, L., Cal, S., Lopez-Otin, C., Li, R., Moake, J. L., Lopez, J. A., and Dong, J. F. (2005) *Blood* 106, 4139–4145
- 11. Milner, C. M., and Day, A. J. (2003) J. Cell Sci. 116, 1863–1873
- Zheng, Y., Mellem, J. E., Brockie, P. J., Madsen, D. M., and Maricq, A. V. (2004) *Nature* 427, 451–457
- Sugiyama, T., Kumagai, H., Morikawa, Y., Wada, Y., Sugiyama, A., Yasuda, K., Yokoi, N., Tamura, S., Kojima, T., Nosaka, T., Senba, E., Kimura, S., Kadowaki, T., Kodama, T., and Kitamura, T. (2000) *Biochemistry* 39, 15817–15825
- 14. Christensen, E. I., and Gburek, J. (2004) Pediatr. Nephrol. 19, 714-721
- Hooi, C. F., Blancher, C., Qiu, W., Revet, I. M., Williams, L. H., Ciavarella, M. L., Anderson, R. L., Thompson, E. W., Connor, A., Phillips, W. A., and Campbell, I. G. (2006) *Oncogene* 25, 3924–3933
- 16. Kang, W., and Reid, K. B. (2003) FEBS Lett. 540, 21-25
- Duke-Cohan, J. S., Tang, W., and Schlossman, S. F. (2000) Adv. Exp. Med. Biol. 477, 173–185
- Ge, W., Hu, H., Ding, K., Sun, L., and Zheng, S. (2006) J. Biol. Chem. 281, 7406–7412
- 19. Wu, Q. (2003) Curr. Top. Dev. Biol. 54, 167-206
- Gregory, L. A., Thielens, N. M., Arlaud, G. J., Fontecilla-Camps, J. C., and Gaboriaud, C. (2003) *J. Biol. Chem.* 278, 32157–32164
- Gregory, L. A., Thielens, N. M., Matsushita, M., Sorensen, R., Arlaud, G. J., Fontecilla-Camps, J. C., and Gaboriaud, C. (2004) *J. Biol. Chem.* 279, 29391–29397
- Feinberg, H., Uitdehaag, J. C., Davies, J. M., Wallis, R., Drickamer, K., and Weis, W. I. (2003) *EMBO J.* 22, 2348–2359
- 23. Bond, J. S., and Beynon, R. J. (1995) Protein Sci. 4, 1247-1261
- 24. Ge, G., and Greenspan, D. S. (2006) *Birth Defects Res. C. Embryo. Today* **78,** 47–68
- 25. Ge, G., and Greenspan, D. S. (2006) J. Cell Biol. 175, 111-120
- Muraoka, O., Shimizu, T., Yabe, T., Nojima, H., Bae, Y. K., Hashimoto, H., and Hibi, M. (2006) *Nat. Cell Biol.* 8, 329–338
- Scott, I. C., Blitz, I. L., Pappano, W. N., Maas, S. A., Cho, K. W., and Greenspan, D. S. (2001) *Nature* 410, 475–478
- Moali, C., Font, B., Ruggiero, F., Eichenberger, D., Rousselle, P., Francois, V., Oldberg, A., Bruckner-Tuderman, L., and Hulmes, D. J. S. (2005) J. Biol.

Chem. 280, 24188-24194

- Petropoulou, V., Garrigue-Antar, L., and Kadler, K. E. (2005) J. Biol. Chem. 280, 22616–22623
- 30. Adar, R., Kessler, E., and Goldberg, B. (1986) Collagen Rel. Res. 6, 267-277

CUB Domain Specificity in PCPE-1

- Steiglitz, B. M., Keene, D. R., and Greenspan, D. S. (2002) J. Biol. Chem. 277, 49820-49830
- Steiglitz, B. M., Kreider, J. M., Frankenburg, E. P., Pappano, W. N., Hoffman, G. G., Meganck, J. A., Liang, X., Hook, M., Birk, D. E., Goldstein, S. A., and Greenspan, D. S. (2006) *Mol. Cell Biol.* 26, 238–249
- 33. Kessler, E., and Adar, R. (1989) Eur. J. Biochem. 186, 115-121
- Hulmes, D. J. S., Mould, A. P., and Kessler, E. (1997) Matrix Biol. 16, 41–45
- Ricard-Blum, S., Bernocco, S., Font, B., Moali, C., Eichenberger, D., Farjanel, J., Burchardt, E. R., van der Rest, M., Kessler, E., and Hulmes, D. J. S. (2002) *J. Biol. Chem.* 277, 33864–33869
- Ge, G., Zhang, Y., Steiglitz, B. M., and Greenspan, D. S. (2006) J. Biol. Chem. 281, 10786–10798
- Combet, C., Jambon, M., Deleage, G., and Geourjon, C. (2002) *Bioinformatics* 18, 213–214
- Deleage, G., Combet, C., Blanchet, C., and Geourjon, C. (2001) Comput. Biol. Med. 31, 259–267
- Whitmore, L., and Wallace, B. A. (2004) Nucleic Acids Res. 32, W668-W673
- Divita, G., Di Pietro, A., Roux, B., and Gautheron, D. C. (1992) *Biochemistry* 31, 5791–5798
- Romero, A., Romao, M. J., Varela, P. F., Kolln, I., Dias, J. M., Carvalho, A. L., Sanz, L., Topfer-Petersen, E., and Calvete, J. J. (1997) *Nat. Struct. Biol.* 4, 783–788
- 42. Kessler, E., and Goldberg, B. (1978) Anal. Biochem. 86, 463-469
- Hartigan, N., Garrigue-Antar, L., and Kadler, K. E. (2003) J. Biol. Chem. 278, 18045–18049
- Moschcovich, L., Bernocco, S., Font, B., Rivkin, H., Eichenberger, D., Chejanovsky, N., Hulmes, D. J. S., and Kessler, E. (2001) *Eur. J. Biochem.* 268, 2991–2996
- Thielens, N. M., Enrie, K., Lacroix, M., Jaquinod, M., Hernandez, J. F., Esser, A. F., and Arlaud, G. J. (1999) *J. Biol. Chem.* 274, 9149–9159
- 46. Takahara, K., Kessler, E., Biniaminov, L., Brusel, M., Eddy, R. L., Janisait, S., Shows, T. B., and Greenspan, D. S. (1994) *J. Biol. Chem.* **269**, 26280 – 26285
- 47. Childs, S. R., and O'Connor, M. B. (1994) Dev. Biol. 162, 209-220
- Finelli, A. L., Bossie, C. A., Xie, T., and Padgett, R. W. (1994) *Development* 120, 861–870
- Stengaard-Pedersen, K., Thiel, S., Gadjeva, M., Moller-Kristensen, M., Sorensen, R., Jensen, L. T., Sjoholm, A. G., Fugger, L., and Jensenius, J. C. (2003) N. Engl. J. Med. 349, 554–560
- Canty, E. G., Garrigue-Antar, L., and Kadler, K. E. (2006) J. Biol. Chem. 281, 13258–13267
- Bernocco, S., Steiglitz, B. M., Svergun, D. I., Petoukhov, M. V., Ruggiero, F., Ricard-Blum, S., Ebel, C., Geourjon, C., Deleage, G., Font, B., Eichenberger, D., Greenspan, D. S., and Hulmes, D. J. S. (2003) *J. Biol. Chem.* 278, 7199–7205
- Bernocco, S., Finet, S., Ebel, C., Eichenberger, D., Mazzorana, M., Farjanel, J., and Hulmes, D. J. S. (2001) *J. Biol. Chem.* 276, 48930–48936
- Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* 15, 305–308

#### Insights into How CUB Domains Can Exert Specific Functions while Sharing a Common Fold: CONSERVED AND SPECIFIC FEATURES OF THE CUB1 DOMAIN CONTRIBUTE TO THE MOLECULAR BASIS OF PROCOLLAGEN C-PROTEINASE ENHANCER-1 ACTIVITY

Guillaume Blanc, Bernard Font, Denise Eichenberger, Christophe Moreau, Sylvie Ricard-Blum, David J. S. Hulmes and Catherine Moali

J. Biol. Chem. 2007, 282:16924-16933. doi: 10.1074/jbc.M701610200 originally published online April 19, 2007

Access the most updated version of this article at doi: 10.1074/jbc.M701610200

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:

http://www.jbc.org/content/suppl/2007/04/20/M701610200.DC1

This article cites 53 references, 21 of which can be accessed free at http://www.jbc.org/content/282/23/16924.full.html#ref-list-1