Available online at www.sciencedirect.com
SCIENCE DIRECT
Protein Expression and Purification 48 (2006) 261–273
Protein Expression & Purification
www.elsevier.com/locate/yprep

Definition, expression, and characterization of a protein domain in the N-terminus of pregnancy-associated plasma protein-A distantly related to the family of laminin G-like modules

Henning B. Boldt, Simon Glerup, Michael T. Overgaard, Lars Sottrup-Jensen, Claus Oxvig *

Department of Molecular Biology, Science Park, University of Aarhus, Gustav Wieds Vej 10 C, DK-8000 Aarhus C, Denmark

Received 19 December 2005, and in revised form 23 January 2006
Available online 23 February 2006

Abstract

Although pregnancy-associated plasma protein-A (PAPP-A), a modulator of insulin-like growth factor (IGF) activity through its cleavage of IGF-binding protein (IGFBP)-4 and -5, has been known for more than two decades, knowledge about its domain architecture is still incomplete. Using position-specific iterative BLAST, we have identified distant relatives of the PAPP-A N-terminal sequence stretch of 250 residues. We present evidence that a protein domain with weak similarity to known laminin G-like (LG) modules is contained within this region, and we propose that PAPP-A and PAPP-A2 are new and unique members in the group of LG proteins as the pappalysins represent the first examples where LG modules are associated with proteinases. Fourteen β-strands characteristic for the LG structure were tentatively located within the PAPP-A LG (PA-LG) module using secondary structure prediction and sequence alignment. Upon mammalian expression of PAPP-A truncation mutants, we defined domain boundaries showing that PA-LG is an autonomously folding unit, which spans the first 243 residues. We were unable to express PAPP-A variants which lack the PA-LG module, suggesting a possible role in stabilization of the proteolytic domain. To obtain larger amounts of protein for functional and structural analysis, the defined PA-LG domain was expressed in bacteria and folded in vitro. In addition, the availability of recombinant PA-LG module may potentially improve diagnostic assays based on the measurement of PAPP-A antigen, and also facilitate the study of PAPP-A in animal model systems.

Keywords: Laminin G-like domain; LNS domain; Domain boundaries; Pappalysin; Proteolysis; Recombinant expression

Pregnancy-associated plasma protein-A (PAPP-A)[1] is a 400 kDa insulin-like growth factor-binding protein (IGFBP) proteinase composed of two identical subunits of 1546 residues linked by a disulfide bond [1]. The 300-residue proteolytic domain is located in the N-terminal half of the subunit (Fig. 1), flanked by uncharacterized sequence stretches, and PAPP-A is the founding member of a fifth family within the metzincin superfamily, the pappalysins [2]. PAPP-A cleaves IGFBP-4 only in the presence of IGF [3]. It also cleaves IGFBP-5, [4], a characteristic shared with PAPP-A2, the only known homologue of PAPP-A [5]. The proteolytic activities of IGFBP proteinases such as PAPP-A and PAPP-A2 counteract the antagonizing effect of IGFBPs and thus regulate IGF bioavailability. Clinically, circulating PAPP-A is used as a marker of Down's syndrome [6], and is also a promising marker of acute coronary syndromes [7].

The group of proteins harboring a so-called laminin G-like (LG) domain, present in the α chain of laminin [8], in neurexins [9], and in sex hormone-binding globulin (SHBG) [10], has collectively been termed the LNS proteins [11]. The family has expanded as the LG module was identified in

* Corresponding author. Fax: +45 8942 5068.
E-mail address: co@mb.au.dk (C. Oxvig).

[1] *Abbreviations used:* PAPP-A, pregnancy-associated plasma protein-A; PAPP-A2, pregnancy-associated plasma protein-A2; IGF, insulin-like growth factor; IGFBP, IGF-binding protein; LG, laminin G-like; PA-LG, PAPP-A laminin G-like; LNS, laminin, neurexin and sex hormone-binding globulin; CD, circular dichroism.

1046-5928/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2006.01.024

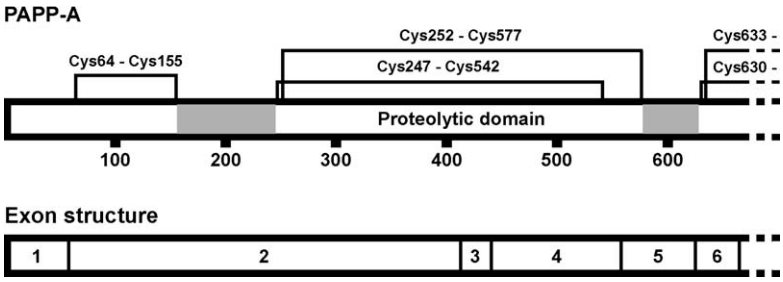

Fig. 1. Schematic representation of the disulfide bridges, the putative domain structure, and the exon structure of PAPP-A. For simplicity, only the N-terminal region of PAPP-A is included and only two of the disulfide bonds in the proteolytic domain are shown. The boundaries between putative domains are expected to be within the shaded areas [1].

several other proteins, also referred to as LG proteins [12]. Despite a rather low sequence similarity of ≈20%, structure determination of three representative LG modules from murine laminin α2 chain (LG5) [13], neurexin Iβ [14], and SHBG [15], respectively, revealed a common fold of the LG modules, comprising ≈200 residues.

The LG domain is composed of 14 β-strands forming a sandwich of two β-sheets, each sheet having 7 β-strands arranged in an anti-parallel manner. The overall fold contains a concave and a convex side with protruding loops at the rims on both ends of the sheet, partially filling the groove formed at the concave site. Part of the structure, comprising the four N-terminal as well as the four C-terminal strands, is recognized as a jelly-roll folding motif [13–15]. A similar fold, which is related to the fold observed in lectins [12], has been identified in the homologous pentraxins (10–15% sequence similarity) [16,17]. The structure is furthermore reminiscent of the lectin-like domains found in the tetanus neurotoxin [18] and also in the sialidases of various pathogenic bacteria and parasites [19].

In this study, we have analyzed the first ≈250 residues of PAPP-A. We propose that PAPP-A has a single LG module located N-terminal to its proteolytic domain, and that PAPP-A therefore is a member of the LG/LNS group of proteins. The availability of large amounts of recombinant LG module of PAPP-A will allow structural and functional studies to be carried out. In addition, it will facilitate the generation of antibodies towards PAPP-A of other species, as full-length protein has proven difficult to express and no other modules of PAPP-A has been expressed in an isolated form.

Materials and methods

Database searching, secondary structure predictions, and alignment

The N-terminal sequence of PAPP-A spanning 300 residues was used to search for distant structural relatives of PAPP-A and PAPP-A2 using position-specific iterative blast (PSI-BLAST) [20] in the non-redundant database at NCBI. Positions of secondary structure elements in the N-termini of PAPP-A and PAPP-A2 were predicted using the set of algorithms included in Jpred2 [21]. The first 360 residues of mature PAPP-A or residues 234–612 of PAPP-A2 were submitted to the server. The N-terminal Glu of mature PAPP-A is residue 1 in accordance with [22], while for PAPP-A2, numbering starts with the initial Met of the signal peptide as residue 1 [5]. The sequences of the different proteins containing either pentraxin, lectin-like or LG modules of known structure as identified by PSI-BLAST were aligned with the N-terminal sequences of PAPP-A and PAPP-A2 using ClustalW [23], and subsequently modified manually according to secondary structure prediction and sequence. Because the lectin-like fold of TeNT and TcTS show deviations in the topology of the β-sheets compared to that found in the pentraxin and LG/LNS families, the sequences corresponding to strand A, M and N in these two proteins were not included in the alignment. The extension of each module defined by alignment was used to calculate the sequence identity between PAPP-A and the representatives of the pentraxin, lectin-like and LG/LNS families by pairwise alignment of the individual sequences using ClustalW [23].

PAPP-A plasmid construction and mutagenesis for mammalian expression

An expression construct, encoding residues 1–310[2] of human PAPP-A was constructed using pPA-*Bsp*EI [2] as the template in a PCR. Primers were 5′-GACGCTAAGCTTATGAAGGATTCCTGC-3′ (5′ end; *Hin*dIII site underlined) and 5′-GCACTGTCTAGA**CTA**G TTGTAT TGCTTGAAGGCCTC-3′ (3′ end; *Xba*I site underlined; stop codon in bold). The resulting PCR fragment was cloned into the *Hin*dIII/*Xba*I sites of pcDNA3.1+ (Invitrogen) to generate pPA1–310. Plasmids encoding a set of five C-terminally truncated variants (pPA1–185, pPA1–197, pPA1–225, pPA1–243, and pPA1–266) was constructed in which the codons encoding Arg-186, His-198, Pro-226,

[2] The numbering of the 1546-residue mature PAPP-A polypeptide is used in this paper [22]. Glu-1 of mature PAPP-A is at position 81 of preproPAPP-A (Accession No. AAC50543). The absence of Val-27 (Val-107 in AAC50543) observed in this paper, and substantiated by database searching, is in conflict with the published amino acid sequence of PAPP-A [22].