Journal of Proteomics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

The intriguing heterogeneity of human salivary proline-rich proteins
Short title: Salivary proline-rich protein species

Barbara Manconi [a], Massimo Castagnola [b,c,1], Tiziana Cabras [a], Alessandra Olianas [a], Alberto Vitali [c], Claudia Desiderio [c], Maria Teresa Sanna [a], Irene Messana [a,*]

[a] Department of Life and Environmental Sciences, Biomedical section, University of Cagliari, Cagliari Italy
[b] Biochemistry and Clinical Biochemistry Institute, Medicine Faculty, Catholic University of Rome, Roma Italy
[c] Institute of Chemistry of the Molecular Recognition, CNR, Roma Italy

ARTICLE INFO

Article history:
Received 29 July 2015
Received in revised form 1 September 2015
Accepted 7 September 2015
Available online xxxx

Keywords:
Acidic proline-rich proteins
Basic proline-rich proteins
N-/*O*-glycosylated proline-rich proteins
Phosphorylation
N-pyroglutaminylation

ABSTRACT

The most heterogeneous family of human salivary proteins is represented by proline-rich proteins (PRPs) divided in acidic, basic, and basic glycosylated (aPRPs, bPRPs, gPRPs). They are encoded by six genes, clustered on chromosome 12p13.2: *PRH1-2* encode aPRPs, *PRB1-4* encode bPRPs and gPRPs. Each gene exists in different allelic forms: two for *PRH2*, three for *PRH1*, *PRB2*, and *PRB4*, four for *PRB1*, and *PRB3*. During granule maturation, PRP proproteins undergo proteolysis by the action of convertases and carboxypeptidases. Differently from bPRPs, proteolysis of aPRPs is not complete, and, besides fragments, entire protein species are also secreted. Maturation process generates ten aPRPs (PRP-1, PRP-2, PIF-s, Db-s, Pa, PRP-3, PRP-4, PIF-f, Db-f, P-C), and at least 18 bPRPs (II-2, P-E, IB-6, Ps-1, Ps-2, IB-1, P-J, IB-8a, P-F, P-H, P-D, II-1, protein glycosylated A, CD-IIg, and Gl1-4). In addition, single nucleotide and length polymorphisms, and differentially spliced transcripts originate several natural variants. Phosphorylation, N-pyroglutaminylation, dimerization, and *N*-/*O*-glycosylation also occur during maturation, enlarging the number of protein species, further increased by proteolytic events governed by carboxy- and endopeptidases during and after secretion, and giving rise to a huge number of small peptides. The PRP functional role is still poorly understood.

Significance: The high polymorphism of PRPs gives an important contribution to the high heterogeneity and inter-individual variability of the human salivary proteome. The products of six genes clustered on chromosome 12p13.2 comprise a mixture of entire, truncated, phosphorylated, glycosylated and dimerized protein/peptide species, sharing large part of their sequences, and possibly involved in different biological activities. Whatever the role of PRP species is, it should be crucial, given that PRPs are the most conserved oral salivary proteins among mammals.

1. Introduction

Proline is unique among the common 20 protein amino acids because the cyclization of the side-chain onto the nitrogen atom of the backbone limits its conformation (Φ angle about 65°) and restricts the arrangements of the preceding residue with a prominent preference for β-sheet conformation [1]. As a consequence, polyproline peptides are prone to adopt a structure called polyproline (PP) II helix, distinctive for a continuous hydrophobic strip around the surface of the helix. PP II helices are very common motifs in globular proteins, where they are generally solvent exposed [2]. The relative rigidity connected to the low conformational entropy upon binding affords weak and very fast on- and off-rates for binding with other widespread motifs, such as SH3 and WW domains [3]. Weak, but specific, interactions in intracellular signaling pathways are at the basis of the preponderance of proline-based recognition motifs in the eukaryotic proteome [4]. Indeed, proline-rich sequences are largely represented in proteins/peptides involved in processes requiring fast recruitment or interchange with cognate proteins, such as cytoskeletal rearrangements, signaling cascade or transcription initiation, and thus they are very common both in the animal and vegetable kingdoms [5]. In humans, proline-rich proteins (PRPs) or proline-rich domains are very common and participate to almost all biological processes, such as RNA splicing and processing [6], modulation of signaling pathways [7], misfolded protein binding [8], collagen fibril organization and chondrocyte development [9], to cite a few. Proteins involved in the formation of the epithelial cornified envelope (cornifins, involucrins and small PRPs) [10], and salivary PRPs are the largest and most heterogeneous families of human PRPs. Indeed, PRPs represent more than 20–30% (w/w) of total proteins in whole human saliva and more than 50–60% (w/w) of proteins of parotid saliva [11]. They are commonly divided in three classes: acidic PRPs (aPRPs),

* Corresponding author at: Department of Life and Environmental Sciences University of Cagliari, Monserrato Campus, Monserrato CA 09042 Italy.
E-mail address: imessana@unica.it (I. Messana).
[1] Manconi, Barbara and Castagnola, Massimo contributed equally to this manuscript.

http://dx.doi.org/10.1016/j.jprot.2015.09.009
1874-3919/

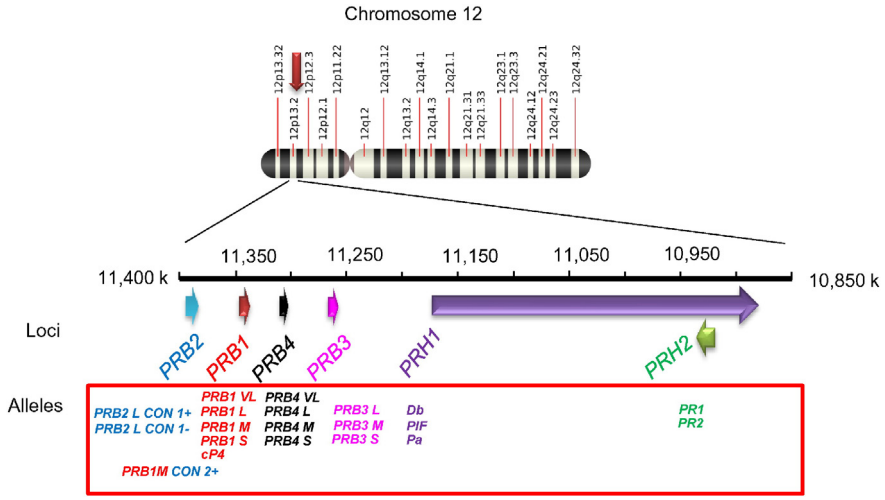

Fig. 1. Schematic representation of the human PRP gene cluster. The six genes of PRPs (*PRB2*, ID: 653247; *PRB1*, ID: 5542; *PRB4*, ID: 5545; *PRB3*, ID: 5544; *PRH1*, ID: 5554; *PRH2*, ID: 5555) are contained within an ~0.5 Mb segment of the chromosome 12p13.2. The red box reports the main alleles found in Caucasian population.

basic PRPs (bPRPs), and glycosylated (basic) PRPs (gPRPs) [12]. The six genes encoding salivary PRPs are localized in a cluster on chromosome 12p13.2 (Fig. 1), two (*PRH1* and *PRH2*) encoding for aPRPs and four (from *PRB1* to *PRB4*) for bPRPs. Proline is the predominant amino acid in salivary PRP sequences (25–40% of all amino acids), but Gly and Gln are also highly represented, and globally these three amino acids account from 70 to 88% of all the residues [13]. Salivary PRPs are unique among the PRP families for the complete absence of hydroxyproline, hydroxylysine, and aromatic amino acids. As it will be better described in the following sections, the major aPRPs are 150 residue-long and the acidic portion is restricted to the first 30 residues for the presence of many Asp and Glu residues. The remaining part of the sequence shows high similarities with bPRPs and is highly repetitive, although aPRP repeats differ slightly from bPRP repeats. Due to these structural features, bPRPs and aPRPs elute as distinct chromatographic clusters in RP-HPLC separations (Fig. 2). While aPRPs are secreted by both parotid and submandibular/sublingual glands (in different percentages), bPRPs are secreted only by parotid glands. A further distinction between aPRPs and bPRPs is that while aPRPs can be found in saliva both as intact and truncated proteoforms, bPRPs encoded by *PRB1*, *PRB2* and *PRB4* genes are detectable in saliva only as fragments of the bigger proproteins. bPRPs are more polymorphic than aPRPs. Alleles of bPRPs exhibit tandem repeat length variation, and single nucleotide polymorphisms (SNP) in the coding region as well as polymorphic cleavage sites. Polymorphic stop codons and alternative splicing result in multiple transcript variants encoding distinct proteins [14–17]. In addition to proteolytic fragmentation, other PTMs, such as glycosylation, phosphorylation, and N-terminal pyroglutamic acid (pGlu) formation, occur before, during, and after secretion on these proteins. As a result, PRPs are characterized by a very high heterogeneity and inter-individual variability that have made their identification and characterization demanding. Top-down proteomics represents the solely approach able to investigate the naturally occurring protein species present in a biological sample [18], including the human salivary PRPs. In fact, in bottom-up proteomics, the presence of a protein in a complex sample is inferred by LC–MS/MS detection of proteotypic peptides generated by enzymatic digestions of the sample before the analysis. This approach reveals its limits when the goal is to evidence peptides or proteins carrying multiple PTMs in sites separated by the digestion. Moreover, it is inapplicable to investigate naturally occurring proteolytic fragmentations. Conversely, the top-down approach, based on the analysis of the "intact" proteome, represents an exclusive strategy for the comprehensive observation of protein species, but suffers from several limits, the major linked to the detection of high molecular size and/or poorly soluble proteins, limits which, however, do not apply to salivary PRPs.

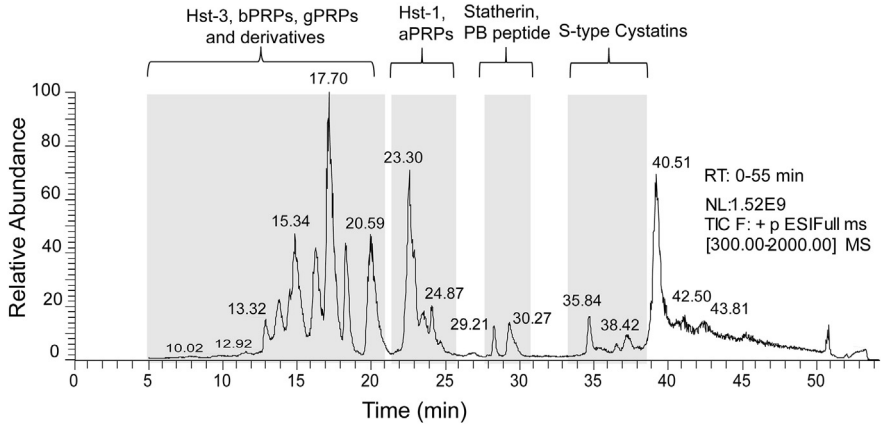

Fig. 2. Typical TIC HPLC–ESI–MS profile of the acidic soluble fraction of adult human whole saliva. The main families of secretory salivary peptides and proteins, i.e. aPRPs, bPRPs and gPRPs, histatins, statherin and PB peptides, S-type cystatins, due to the structural similarity of their members, elute in well-defined chromatographic clusters (only histatins are partly overlapped to PRPs). The chromatographic column was a Vydac (Hesperia, CA) C8 column with 5 μm particle diameter (150 × 2.1 mm); 0.056% (v/v) aqueous TFA (eluent A), and 0.05% (v/v) TFA in acetonitrile–water 80/20 (eluent B). The gradient applied was linear from 0 to 55% of B in 40 min, and from 55% to 100% of B in 10 min, at a flow rate of 0.30 ml/min.

Please cite this article as: B. Manconi, et al., The intriguing heterogeneity of human salivary proline-rich proteins, J Prot (2015), http://dx.doi.org/10.1016/j.jprot.2015.09.009