Available online at www.sciencedirect.com

ScienceDirect

journal homepage: http://www.elsevier.com/locate/euprot

Little things make big things happen: A summary of micropeptide encoding genes

*Jeroen Crappé**, *Wim Van Criekinge*, *Gerben Menschaert**

Lab of Bioinformatics and Computational Genomics (BioBix), Department of Mathematical Modelling, Statistics and Bioinformatics, Faculty of Bioscience Engineering, Ghent University, 9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 31 October 2013
Received in revised form
6 December 2013
Accepted 14 February 2014
Available online 6 March 2014

Keywords:
Micropeptide
sORF
Ribosome profiling
Peptidomics
(l)ncRNA
Bioactive peptide

ABSTRACT

Classical bioactive peptides are cleaved from larger precursor proteins and are targeted toward the secretory pathway by means of an N-terminal signaling sequence. In contrast, micropeptides encoded from small open reading frames, lack such signaling sequence and are immediately released in the cytoplasm after translation. Over the past few years many such non-canonical genes (including open reading frames, ORFs smaller than 100 AAs) have been discovered and functionally characterized in different eukaryotic organisms. Furthermore, *in silico* approaches enabled the prediction of the existence of many more putatively coding small ORFs in the genomes of *Sacharomyces cerevisiae*, *Arabidopsis thaliana*, *Drosophila melanogaster* and *Mus musculus*. However, questions remain as to what the functional role of this new class of eukaryotic genes might be, and how widespread they are. In the future, approaches integrating *in silico*, conservation-based prediction and a combination of genomic, proteomic and functional validation methods will prove to be indispensable to answer these open questions.

1. Introduction

It is a well-known fact that small peptides play important roles in all kinds of biological processes [1]. The largest and most extensively studied class of small peptides comprises classical bioactive peptides. These are enzymatically cleaved from longer protein precursors containing an N-terminal signal sequence, hence directing the translation product toward the secretory pathway (see Fig. 1). Once released from the secretory vesicles, most of these peptides act as ligands of membrane receptors (mostly G protein-coupled receptors) and exert their extra-cellular biological signaling function in a autocrine, paracrine or endocrine way [2]. Examples are neuropeptides, peptide hormones and growth factors [3–6]. Other secreted peptides exercise their function in host defense systems having antimicrobial or toxic properties or show antihypertensive, antithrombotic or antioxidative activity [7,8]. Recently, other (non-classical) peptides – encoded by small open reading frames (sORFs) – have been discovered, presumably defining a new eukaryotic gene family [9–16]. These so-called micropeptides are translated from sORFs shorter

* *Corresponding authors at:* Lab of Bioinformatics and Computational Genomics (BioBix), Department of Mathematical Modelling, Statistics and Bioinformatics, FBE, Ghent University, Coupure Links 653, 9000 Ghent, Belgium. Tel.: +32 9 264 99 22.
E-mail addresses: Jeroen.Crappe@Ugent.be (J. Crappé), Gerben.Menschaert@Ugent.be, Gerben.Menschaert@gmail.com (G. Menschaert).

http://dx.doi.org/10.1016/j.euprot.2014.02.006

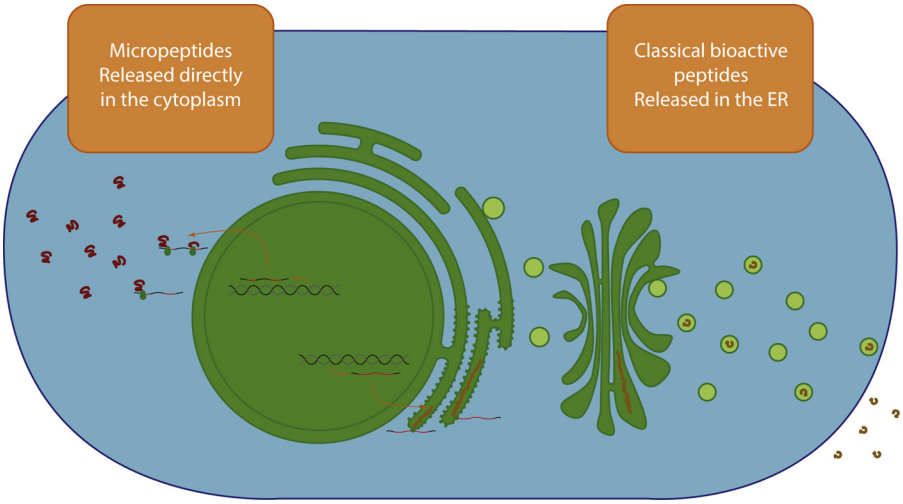

Fig. 1 – Localization of classical bioactive peptides and micropeptides. Classical bioactive peptides contain an N-terminal signal sequence directing the translation product toward the secretory pathway. As a consequence, these biologically active peptides exert an extra-cellular function. In contrast, micropeptides lack an N-terminal signaling sequence, and are consequently released in the cytoplasm immediately after translation.

than 100 AAs [14,17]. Sometimes these are also referred to as polycistronic peptides in the case were they are translated from a polycistronic mRNA [13] or as short open reading frame (sORF)-encoded polypeptides (SEPs) [18]. In contrast to other bioactive peptides, micropeptides are not cleaved from a larger precursor protein and lack an N-terminal signaling sequence. As such they are in principle released in the cytoplasm immediately after translation. This review focuses on this new class of peptides (see Fig. 1).

In the past, many molecules have been overlooked because of various biases and/or simplifications introduced in the performed discovery strategy. For example, it was only in 1993 that the first microRNA (lin-4) was discovered in *Caenorhabditis elegans* [19]. In the 2 subsequent decades more than 2500 microRNAs were identified in human alone [20]. Since micropeptides came into the limelight, ever more research is conducted to this new type of biomolecules, providing increasing evidence that this type of biomolecules is possibly also long overlooked [17]. It was assumed, especially for comprehensive cDNA annotation studies, that protein-coding genes do not code for translation products shorter than 100 AAs [21]. This arbitrarily chosen minimum length reduces the likelihood of false-positive detection by gene-prediction software and genome annotation algorithms, but at the same time vastly underestimates the true number of (atypical) small proteins [22,23]. This generalization is also noticeable in (manually curated) protein databases such as SwissProt-KB, where at the time of writing only 680 (3.4%) out of a total of 20,271 reviewed human proteins have a length shorter than 100 AAs. Although micropeptide research is not yet widespread and much remains to be learned about their abundance, functional activity and localization, a handful of these peptides have recently been functionally annotated in different eukaryotic organisms (see next paragraph for an extensive overview or Table 1 for a brief summary). Though important to an argument of the general conservation and function across all kingdoms of life, sORFs in bacteria and viruses [24–28], will not be covered in this review. The above-mentioned references can serve as a brief overview of putatively coding sORF (pcsORF) detection in lower organisms.

2. Overview of functionally annotated micropeptides

The first eukaryotic micropeptide was only described in 1996. While investigating the function of *early nodulin 40* (*Enod40*), formerly annotated as a ncRNA gene in legumes, van de Sande et al. transformed tobacco plants with a soybean GmENOD40-2 construct [29], proving that this construct was active in the non-legume tobacco, modulating the action of auxin. Sequence comparison of the tobacco and legume *Enod40* clones revealed a highly conserved sORF coding for a 10 (tobacco) or 12 (soybean) AAs long peptide [29]. Later on, a second overlapping coding sORF of 24 AA was identified in soybean, categorizing *Enod40* as a polycistronic mRNA. *Enod40* is a well-known factor that functions in root nodule organogenesis in legumes and also displays a high sequence conservation among other plant species including monocots, suggesting a more general biological function [9]. In addition, *Enod40* shows a highly conserved secondary topology, giving it the characteristics of a structural RNA [30]. The presence of peptide encoding sORFs and of structured RNA, both playing a role in developmental processes, indicates that *Enod40* acts as a bi-functional or dual RNA [31,32]. Since the discovery of this first micropeptide in plants, others have been functionally annotated. In Arabidopsis, the *POLARIS* (*PLS*) gene, identified as a promoter trap transgenic line predominantly showing expression in the embryonic basal region, affects root growth and vascular development [10]. Mutation analysis has shown that the 36 AAs peptide encoded by *PLS* interacts with PIN proteins, forming a network that plays an important role in the hormonal crosstalk between auxin, ethylene and cytokinin [33,34]. In maize, the recessive mutation of *Brick1* (*Brk1*) leads