Journal of Proteomics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Proteomics

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

Unveiling antimicrobial peptide–generating human proteases using PROTEASIX

Paulo Bastos [a,b,*], Fábio Trindade [a,c], Rita Ferreira [b], Mercedes Arguello Casteleiro [d], Robert Stevens [d], Julie Klein [e,f,1], Rui Vitorino [a,c,*,1]

[a] Institute of Biomedicine - iBiMED, Department of Medical Sciences, University of Aveiro, Portugal
[b] QOPNA, Mass Spectrometry Center, Department of Chemistry, University of Aveiro, Portugal
[c] Department of Physiology and Cardiothoracic Surgery, Faculty of Medicine, University of Porto, Porto, Portugal
[d] School of Computer Science, University of Manchester, Kilburn Building, Oxford Road, Manchester, UK
[e] INSERM U1048, Institute of Cardiovascular and Metabolic Disease, Toulouse, France
[f] Université Toulouse III Paul-Sabatier, Toulouse, France

ARTICLE INFO

Article history:
Received 20 September 2016
Received in revised form 20 February 2017
Accepted 22 February 2017
Available online xxxx

Keywords:
Bioinformatics
Antimicrobial peptides
Proteases
Knowledgebase
Pathophysiology

ABSTRACT

Extracting information from peptidomics data is a major current challenge, as endogenous peptides can result from the activity of multiple enzymes. Proteolytic enzymes can display overlapping or complementary specificity. The activity spectrum of human endogenous peptide-generating proteases is not fully known. Hence, the indirect study of proteolytic enzymes through the analysis of its substrates is largely hampered. Antimicrobial peptides (AMPs) represent a primordial set of immune defense molecules generated by proteolytic cleavage of precursor proteins. These peptides can be modulated by host and microorganismal stimuli, which both dictate proteolytic enzymes' expression and activity. Peptidomics is an attractive approach to identify peptides with a biological role and to assess proteolytic activity. However, bioinformatics tools to deal with peptidomics data are lacking. PROTEASIX is an excellent choice for the prediction of AMPs-generating proteases based on the reconstitution of a substrate's cleavage sites and the crossing of such information with known proteases' specificity retrieved by several publicly available databases. Therefore, the focus of the present tutorial is to explore the potential of PROTEASIX when gather information concerning proteases involved in the generation of human AMPs and to teach the user how to make the most out of peptidomics results using PROTEASIX.
Significance: This tutorial provides a step-by-step guide on to use PROTEASIX for making sense and unveiling the biological implications of peptidomics data. In addition to its educational focus, this work also reveals which proteases contribute the most to the formation of human antimicrobial peptides distributed throughout body fluids in human defense barriers.

1. Introduction

Making sense of peptidomics results is a major current challenge, as endogenous peptides can result from the activity of multiple enzymes. Such proteolytic enzymes can display overlapping or complementary specificity. Furthermore, the activity spectra of human endogenous peptide-generating proteolytic enzymes are not fully known. Hence, the indirect study of proteolytic enzymes via alterations at their substrate level is largely hampered.

Antimicrobial peptides (AMPs) constitute a family of defense molecules extraordinarily diverse with a continuum of activity spectra, ranging from two to more than 150 amino acids, acting either intracellularly or extracellularly, directly over microorganisms or through activation of other immune mediators. Most AMPs are found extracellularly, primarily at the defense barriers where they are thought to be generated by proteolytic cleavage of precursor proteins. Accordingly, thousands of fragmentation peptides can be found in the extracellular milieu [1–3]. Among these, a growing number of AMPs has been identified and the corresponding antimicrobial activity discerned [4–7]. For instance, the salivary short form of Thymic stromal lymphopoietin (isoform 2) results from proteolytic cleavage of Thymic stromal lymphopoietin and displays antimicrobial properties [4,8]. Similarly, eotaxin-3/CCL26 is generated by mast cell protease-mediated cleavage during allergic inflammatory responses to bacterial infections of the airways and is active against several airway pathogens, including *Streptococcus pneumoniae* and *Staphyloccous aureus* [5]. However, with few exceptions [9–11], little is known concerning how variations at the peptidome level may be associated with human diseases and mediated or triggered by pathogenic microorganisms. Therefore, the identification and prediction of proteases implicated in AMPs generation might

* Corresponding authors at: Institute of Biomedicine – iBiMED, Department of Medical Sciences, University of Aveiro, Aveiro 3810-193, Portugal.
E-mail addresses: pauloandrediasbastos@ua.pt (P. Bastos), rvitorino@ua.pt (R. Vitorino).
[1] Senior authors.

http://dx.doi.org/10.1016/j.jprot.2017.02.016
1874-3919/

help to understand some disease mechanisms and the role of disease-associated as well as pathogen-associated peptides as well as to pinpoint potential therapeutic targets.

With the flourishing of proteomics, many tools for protein-protein interactions' annotation and prediction (e.g. STRING [12]) as well as curated knowledgebase (e.g. UniProt [13]) have emerged. In contrast, in the field of peptidomics there is a lack of supporting bioinformatics tools, which may stem from the larger complexity of the human peptidome over its proteome counterpart and from the fact that most protein-centered studies employ enzymatic treatment (e.g. tryptic digestion), which obliterates any information provided by endogenous peptide fingerprints. In order to address this challenge, PROTEASIX [14] can be of tremendous utility, since it allows the accurate elucidation of cleavage site from peptide sequence inputs, thus uncovering the proteases potentially implicated in endogenous peptide formation. Still, its correct exploitation requires users to be fully elucidated on how to use this tool and to know what it can provide them.

With the application of PROTEASIX [15] to a set of proteolysis products, the user can find the answers to the following questions: What are the known proteases and their target cleavage sites (observed and predicted)? For a given peptide and the respective precursor protein, what are the cleavage sites that led to its production? Are these peptides the product of observed or predicted proteolysis? What are the functions and cellular locations of proteases and their substrate proteins in the species in question? Which are the specific cleavage sites for a given protease? Therefore, the focus of the present tutorial is to explore the potential of PROTEASIX in gathering information concerning proteases involved in the generation of human AMPs and to teach the user how to make the most out of peptidomics data using PROTEASIX.

2. PROTEASIX: a peptide centric tool

PROTEASIX [14] is an open-source, updated, peptide-centric knowledgebase and tool for in silico prediction or retrieval of proteases known to be involved in native peptide generation. It can be used for both small and large-scale investigations in an automatic fashion (for the updated version, please refer to the currently available at http://proteasix.cs.manchester.ac.uk). In order to organize data from a variety of public sources, PROTEASIX employs its own ontology (PROTEASIX Ontology) [15], which re-uses parts of other Protein, Gene, Chemical and Biological Ontologies. By doing so, PROTEASIX knowledgebase supports the PROTEASIX tool, allowing the linkage of peptide fragments to their corresponding proteases as well as the uncovering of possible disease and innate response mechanisms.

Proteolytic cleavage requires the recognition of short amino acid sequences, also called motifs or cleavage sequences (CS). In turn, proteases (both endopeptidase and exopeptidase) exhibit varying binding specificity/affinity for the recognition of such motifs, spanning from being strictly restricted to one/few critical amino acids in specific positions, to being unspecific for generic amino acids or groups thereof [16,17]. In PROTEASIX, each protease is firstly associated with their corresponding CS. By aligning CS with the scissile bond (a covalent chemical bond susceptible to enzymatic cleavage such as proteolytic hydrolysis) positions on the peptide sequence provided by the user, PROTEASIX attempts to determine which enzymes generate such peptides; entries that cannot be aligned are, thus, discarded.

Since co-location of proteases and respective substrates is imperative to proteolysis, in PROTEASIX all proteases and associated substrates are annotated with the Gene Ontology Cellular Component in order to verify whether a common location can be found and thus increase the confidence of the cleavage prediction. Moreover, because this binding specificity/affinity depends on the amino acid sequence, mutations may lead to alterations in proteolytic activity and explain why some proteases/fragmentation peptides appear as deregulated across human disorders.

3. PROTEASIX step-by-step

1. The information required for each peptide consists of a peptide identifier (anything will do - e.g. a group identifier, an ordinal attribute, a random number) as well as the UniProt Accession Number or identifier of the parent protein from which the peptide is

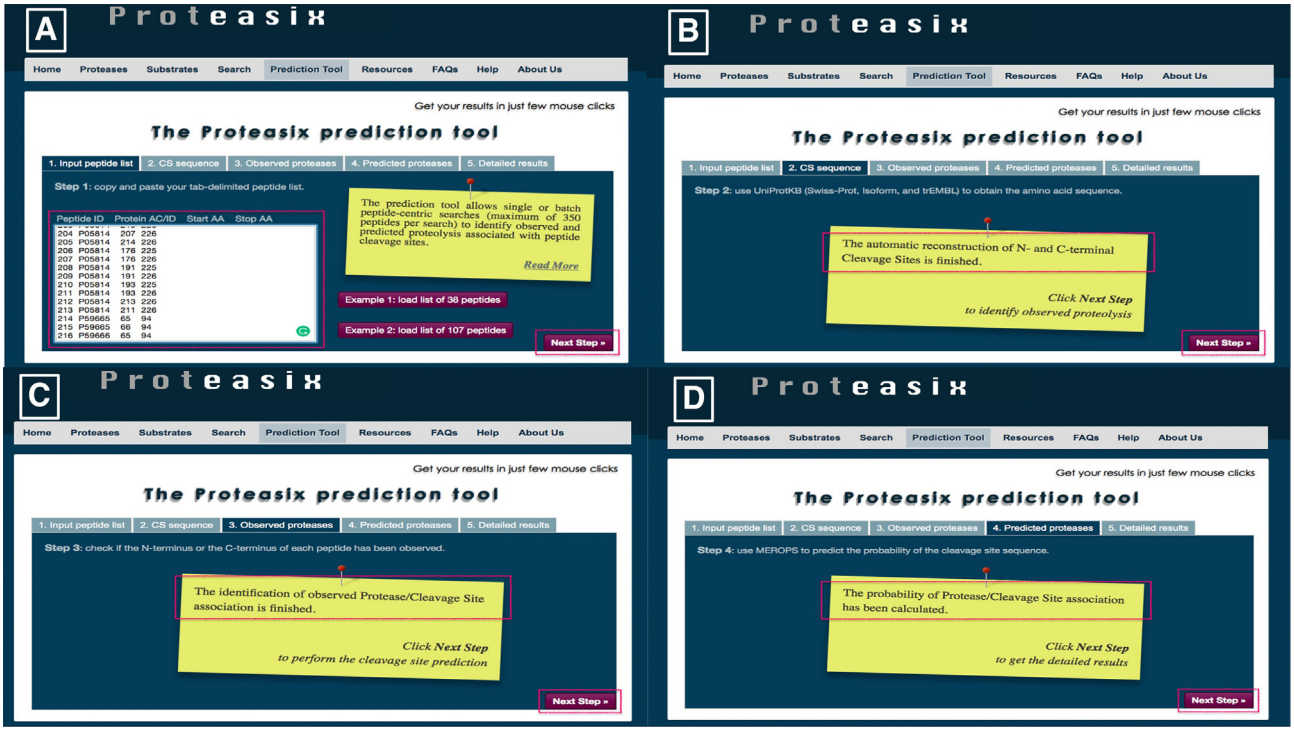

Fig. 1. Screenshots of the step-by-step tutorial on how to use PROTEASIX from input data to download phase: A) upper-left, step 3; B) upper-right, step 5; C) lower-left, step 6; D) lower-right, step 8.

Please cite this article as: P. Bastos, et al., Unveiling antimicrobial peptide–generating human proteases using PROTEASIX, J Prot (2017), http://dx.doi.org/10.1016/j.jprot.2017.02.016