BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 1 2012, pages 130–131
doi:10.1093/bioinformatics/btr604

Sequence analysis Advance Access Publication November 3, 2011

AMPA: an automated web server for prediction of protein antimicrobial regions

Marc Torrent[1,2,*], Paolo Di Tommaso[3], David Pulido[2], M. Victòria Nogués[2], Cedric Notredame[3], Ester Boix[2] and David Andreu[1,*]

[1]Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona Biomedical Research Park, Dr Aiguader 88, [2]Department of Biochemistry and Molecular Biology, Universitat Autònoma de Barcelona, Biosciences Faculty, Cerdanyola del Vallès and [3]Bioinformatics and Genomics program, Centre for Genomic Regulation (CRG) and UPF, Dr Aiguader 88, 08003, Barcelona, Spain

Associate Editor: Burkhard Rost

ABSTRACT

Summary: AMPA is a web application for assessing the antimicrobial domains of proteins, with a focus on the design on new antimicrobial drugs. The application provides fast discovery of antimicrobial patterns in proteins that can be used to develop new peptide-based drugs against pathogens. Results are shown in a user-friendly graphical interface and can be downloaded as raw data for later examination.

Availability: AMPA is freely available on the web at http://tcoffee.crg.cat/apps/ampa. The source code is also available in the web.

Contact: marc.torrent@upf.edu; david.andreu@upf.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on June 14, 2011; revised on October 7, 2011; accepted on October 26, 2011

1 INTRODUCTION

The widespread use of antibiotics has promoted the emergence of antibiotic-resistant strains that often pose serious health problems. Many efforts have been directed toward finding alternative antibiotics unaffected by resistance mechanisms. Antimicrobial peptides (AMPs) are regarded as one of the most promising alternatives in this regard (Yeung *et al.*, 2011).

AMPs usually act through membrane permeation, receptor-independent mechanisms that greatly reduce the risk of inducing resistance in microorganisms (Hilpert, 2011). Moreover, many AMPs are active against a wide diversity of pathogens, including Gram-negative and Gram-positive bacteria, fungi and parasites. All these features make AMPs attractive candidates for fighting microbial infection. A growing interest, therefore, exists for identifying natural AMP leads that can be developed into novel antibiotics.

An increasing number of proteins and peptides with antimicrobial effects have been identified in many organisms, many of such proteins belonging to the innate defense system but others are unrelated to immune response. The design of new AMPs based on these proteins is hampered by the fact that the mechanisms of their antimicrobial action are not yet fully elucidated. It is commonly accepted that an adequate combination of hydrophobic and cationic amino acid residues underlies the action of antimicrobial regions, generally by conferring them an amphipathic secondary structure that favors interaction with microbial membranes, that induces cell damage which kills the microorganism (Wimley, 2010). The antimicrobial activity is usually confined to discrete stretches of the protein rather than spread over the entire sequence. Such stretches are often located at the N- or C-termini of proteins found in the secretion granules of immune cells such as neutrophils and eosinophils, from which they are released by limited proteolysis upon infection (Ramanathan *et al.*, 2002).

Due to the growing interest in AMPs, algorithms for predicting antimicrobial activity have been developed (Lata *et al.*, 2007; Torrent *et al.*, 2011; Wang *et al.*, 2011) and used to propose short AMP sequences (Fjell *et al.*, 2009) that after high-throughput synthesis and screening can be validated as novel AMP leads (Hadley and Hancock, 2010). In contrast, the *in silico* identification of protein sequence stretches with potential AMP activity has not to our knowledge been successfully addressed. Here, we present a method for efficient screening of protein sequences in search of antimicrobial stretches that can provide novel AMP leads.

2 METHODS

2.1 Description

The AMPA algorithm is based on an antimicrobial propensity scale derived from high-throughput screening results (Torrent *et al.*, 2009b) from the AMP bactenecin 2A for which antimicrobial IC_{50} values for all amino acid replacements at each position have been determined (Hilpert *et al.*, 2005). From these data, an antimicrobial index (AI) can be calculated providing a fair assessment of the tendency for such amino acids to be present in an AMP sequence. As low IC_{50} values correspond to high activity, amino acids with a low index are the most favored to be part of an AMP. By using a sliding-window method the algorithm draws an antimicrobial profile. The regions (>12 amino acids length) located below the threshold are considered putative antimicrobial domains.

The AMPA server has been integrated within the new T-Coffee web server framework (Notredame *et al.*, 2000), developed in Perl and Java. To submit a job, the query sequences must be provided in FASTA format. Also the window size and the threshold values must be specified (default values are provided). Once launched, the job is processed and the output results page is displayed and sent to the email address provided. On the top of the

*To whom correspondence should be addressed.

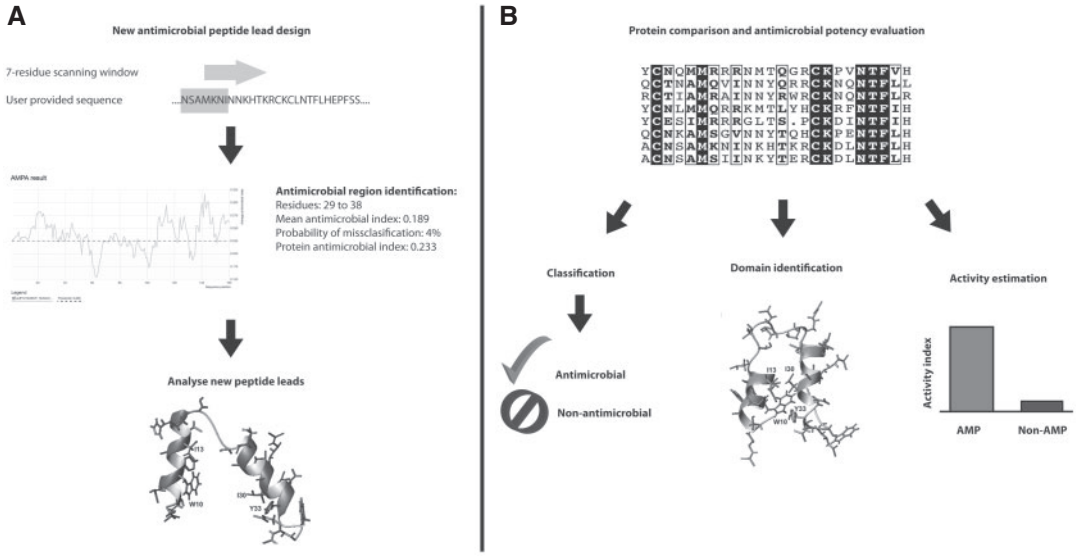

Fig. 1. The AMPA algorithm can automatically scan protein sequences to identify potential antimicrobial regions guiding lead discovery (**A**). AMPA can also be used to evaluate antimicrobial activity and guide antimicrobial domain identification (**B**).

results page, an interactive antimicrobial profile of the query sequences is provided. Antimicrobial profiles can be switched on and off for a better visual inspection of the results. The information on the antimicrobial stretches of the query protein (i.e. the number and location of the antimicrobial regions and the mean AI of the protein together with a statistical evaluation of the results) is displayed on the top of the webpage and can be downloaded for later examination.

AMPA is intended for multiple purposes. While its main application is the fast automatic detection of antimicrobial regions in proteins that can serve as new templates for AMP design, as shown in Figure 1, AMPA-derived AI values can be used to automatically classify proteins or domains thereof as either antimicrobial or non-antimicrobial, and to compare different protein sequences in this regard. When used in conjunction with the T-coffee alignment tool, antimicrobial regions can be checked to identify potentially conserved antimicrobial domains.

The AMPA screening system has been optimized to work with the default window size and threshold values provided. Both search parameters, however, can be tuned for specific purposes such as locating antimicrobial regions in very large—or very short—proteins, or to analyze sequences using more stringent than usual criteria. For instance, long antimicrobial domains in large (e.g. >1000 residue) proteins may be better defined by increasing the window size from 7 to 9–19 amino acids. Larger window sizes, however, are not advised, as the averaging process may miss small antimicrobial motifs. On the other hand, for very short sequences (<25 residues), small window sizes may provide slightly better defined antimicrobial profiles. Threshold values, on their part, while not altering the antimicrobial profile, can be adjusted to highlight the highest-scoring stretches in proteins with more than one antimicrobial region.

2.2 Validation

The algorithm has been extensively validated *in silico* and found to correctly identify 80–90% of the antimicrobial proteins and properly predict their antimicrobial domain (Torrent *et al.*, 2009b). Besides, the system has successfully allowed predicting the N-terminal domain of eosinophil cationic protein as the region embodying the antimicrobial activity of the entire protein (Torrent *et al.*, 2009a). To experimentally test the automatic detection of antimicrobial domains by the AMPA algorithm, we have analyzed six proteins reported as antimicrobial but without any identified domain. Peptides derived from the regions identified by AMPA were found to retain antimicrobial activity in a micromolar range (Supplementary Material). Hence, we propose the use of AMPA as a valuable resource to accurately detect antimicrobial domains in proteins.

Funding: M.T. is a recipient of an Alianza Cuatro Universidades fellowship. Work at UPF funded by HEALTH-F3-2008-223414 (LEISHDRUG) project from the EU; at UAB funded by MICINN (BFU2009-09371) and Generalitat de Catalunya (2009 SGR 795).

Conflict of Interest: none declared.

REFERENCES

Fjell,C.D. *et al.* (2009) Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem.*, **52**, 2006–2015.

Hadley,E.B. and Hancock,R.E. (2010) Strategies for the discovery and advancement of novel cationic antimicrobial peptides. *Curr. Top Med. Chem.*, **10**, 1872–1881.

Hilpert,K. (2011) Identifying novel antimicrobial peptides with therapeutic potential against multidrug-resistant bacteria by using the SPOT synthesis. *Mini Rev. Org. Chem.*, **8**, 157–163.

Hilpert,K. *et al.* (2005) High-throughput generation of small antibacterial peptides with improved activity. *Nat. Biotechnol.*, **23**, 1008–1012.

Lata,S. *et al.* (2007) Analysis and prediction of antibacterial peptides. *BMC Bioinformatics*, **8**, 263.

Notredame,C. *et al.* (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.

Ramanathan,B. *et al.* (2002) Cathelicidins: microbicidal activity, mechanisms of action, and roles in innate immunity. *Microbes Infect.*, **4**, 361–372.

Torrent,M. *et al.* (2009a) Bactericidal and membrane disruption activities of the eosinophil cationic protein are largely retained in an N-terminal fragment. *Biochem. J.*, **421**, 425–434.

Torrent,M. *et al.* (2009b) A theoretical approach to spot active regions in antimicrobial proteins. *BMC Bioinformatics*, **10**, 373.

Torrent,M. *et al.* (2011) Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. *PLoS One*, **6**, e16968.

Wang,P. *et al.* (2011) Prediction of antimicrobial peptides based on sequence alignment and feature selection methods. *PLoS One*, **6**, e18476.

Wimley,W.C. (2010) Describing the mechanism of antimicrobial peptide action with the interfacial activity model. *ACS Chem. Biol.*, **5**, 905–917.

Yeung,A.T. *et al.* (2011) Multifunctional cationic host defence peptides and their clinical applications. *Cell Mol. Life Sci.*, **68**, 2161–2176.