Synthetic and Systems Biotechnology 1 (2016) 89–94

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: keaipublishing.com/synbio

Norine: A powerful resource for novel nonribosomal peptide discovery

M. Pupin [a,b], Q. Esmaeel [c], A. Flissi [a,b], Y. Dufresne [a,b], P. Jacques [c,d], V. Leclère [a,b,c,*]

[a] Univ Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France
[b] Inria-Lille Nord Europe, Villeneuve d'Ascq Cedex F-59655, France
[c] Univ Lille, INRA, ISA, Univ Artois, Univ Littoral Côte d'Opale, EA 7394 - ICV - Institut Charles Viollette, F-59000 Lille, France
[d] Bioindustry Unit, Gembloux Agro-Bio Tech-University of Liege, Passage des Déportés, 2, Gembloux 5030, Belgium

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form 26 November 2015
Accepted 26 November 2015
Available online

Keywords:
Norine
Nonribosomal peptides
Secondary metabolites
Database

ABSTRACT

Since its first release in 2008, Norine remains the unique resource completely devoted to nonribosomal peptides (NRPs). They are very attractive microbial secondary metabolites, displaying a remarkable diversity of structure and functions. Norine (http://bioinfo.lifl.fr/NRP) includes a database now containing more than 1160 annotated peptides and user-friendly interfaces enabling the querying of the database, through the annotations or the structure of the peptides. Dedicated tools are associated for structural comparison of the compounds and prediction of their biological activities. In this paper, we start by describing the knowledgebase and the dedicated tools. We then present some user cases to show how useful Norine is for the discovery of novel nonribosomal peptides.

1. Introduction

Nonribosomal peptides (NRPs) are attractive natural compounds because of their numerous biological activities potentially exploited by industries in diverse areas such as phytosanitary sector, cosmetics or health. They are produced by microorganisms (including bacteria and fungi) through specialized biosynthetic pathways. NRPs are biosynthesized by enzymatic modular complexes called NonRibosomal Peptide Synthetases (NRPSs) working as multidomain assembly lines.[1] The mode of synthesis leads to the production of compounds displaying a broad range of structures. Indeed, if some of them look like classical peptides because they are linear, most of them are more complex, including one or more cycles and branches. Moreover, those peptides are composed of monomers that are not limited to the 20 proteinogenic amino acids. Up to now, we have identified more than 530 building blocks composing the different NRPs. The structural biodiversity is also due to the monomer modifications occurring during the synthesis made by the NRPSs themselves or performed post synthesis by accessory enzymes (also named tailoring or decorating enzymes). Famous examples for NRPs are the antibiotics penicillin,[2] bacitracin and vancomycin,[3] or the immunosuppressor cyclosporine.[4] In addition, some NRPs show antitumor activity such as Dactinomycin.[5] A current worrying public health issue is to find and develop new drugs to overcome multi-resistant pathogens. Therefore, it is important to develop bioinformatics tools for secondary metabolite discovery, such as antiSMASH[6] and tools especially dedicated to NRPSs and NRPs, such as Florine,[7] NaPDos,[8] and Norine.[9,10] The development of Norine was first motivated by the availability of computational tools allowing structure comparison of all NRPs,[11] in spite of their complexity. For this purpose, we needed a database gathering all known NRPs, annotated according to their monomeric structure (i.e. monomer composition and 2D topology). Until now, the Norine team screened the literature to enter new peptides and annotated them manually. To get a more complete database, we have recently opened it to crowdsourcing through an easy-to-use web-based application.[10] Moreover, a semi-automatic process to extract data from external sources is currently under development.

2. The Norine database

2.1. Description and querying

Norine is a platform that includes the unique database dedicated to NRPs, associated with computational tools for their analysis. It has gained an international recognition thanks to high quality and manually curated annotations. Containing about 700 annotated NRPs for its first release in 2008, Norine database now contains more than 1160 NRPs that are clustered into 214 families, and composed of

* Corresponding author. Univ Lille, ICV, Charles Viollette Institute, PrBioGEM team, Polytech'Lille, Avenue Langevin, Villeneuve d'Ascq Cedex F-59655, France. Tel.: +33 320 43 46 68.
E-mail address: valerie.leclere@univ-lille1.fr (V. Leclère).
Peer review under responsibility of KeAi Communications Co., Ltd.

http://dx.doi.org/10.1016/j.synbio.2015.11.001

at least 530 distinct monomers. Among the peptides, 73.5% are tagged with the "curated" status, which means that their nonribosomal origin is supported experimentally (for example due to identified NRPS) while 26.5% are annotated with "putative" status due to only presumed nonribosomal origin (often based on structural features). Two thirds of the peptides are cyclic or partially cyclic or contain at least one cycle. The sizes of the peptides range from 2 to 26 monomers, if polytheonamide is excluded, which was described as being the biggest NRP with 49 monomers for a long time but recently was identified to be an RiPP (Ribosomally synthesized and post-translationally modified peptide).[12] Thus, in the near future, a third category will be created to tag all deprecated peptides when the hypothetical NRPS origin is finally excluded.

Each peptide page includes a comprehensive description of the peptide with the name, activities and structural atomic and monomeric details. The monomeric structure can be automatically obtained through the integrated smiles2monomers tool (s2m) when SMILES are available.[13] When identified, links to UniProt (for synthetases), PDB and PubChem (for structural data on the peptides) are provided. Moreover, a direct link to the NRPS gene clusters annotated in MIBiG[14] will be added soon.

Norine is queried from all over the world by biologists and biochemists to further analyze the nonribosomal peptides they study. For example, Desriac et al.[15] queried Norine to predict the antibacterial activity of a putative NRP produced by *Pseudoalteromonas*, while Bills et al.[16] used Norine to investigate the structural differences between bacterial and fungal NRPs. Indeed, for this purpose, the Norine platform provides visualization and editing applets for monomeric structure as well as tools to compare monomeric structures. Currently, Norine can be queried either by annotations (through "general search" tab) or by structural information (through "structure search" tab) of the peptides.

2.1.1. General search

Norine provides a basic interface that enables to query the database and search for peptides by combining multiple criteria, such as the name of the peptide, the Norine ID, the biological activities, the structure type, the producing organism, or the title or authors of references associated to the NRP. The main advantage of this interface is that it allows users to extract data and get statistics according to different criteria. For example, one can query for all siderophores produced by "any bacteria" (check "siderophore" in the "activity" field and enter "bacteria" in the "organism search" field), or all peptides with a linear structure, or simply search for all NRPs produced by the genus "*Pseudomonas*" (enter "pseudomonas" in the "organism search" field) (see the results in Fig. 1). The first output is a list of all the peptides corresponding to the criteria selected, classified by families. A click on a peptide name directs to the peptide page containing all details on the compound. Moreover, a click on the pie chart icon located above the list of results provides graphical output (Fig. 1). Pie charts and diagrams enable to filter the obtained results in order to refine them, by clicking on a slice, for example by structure type or monomers size.

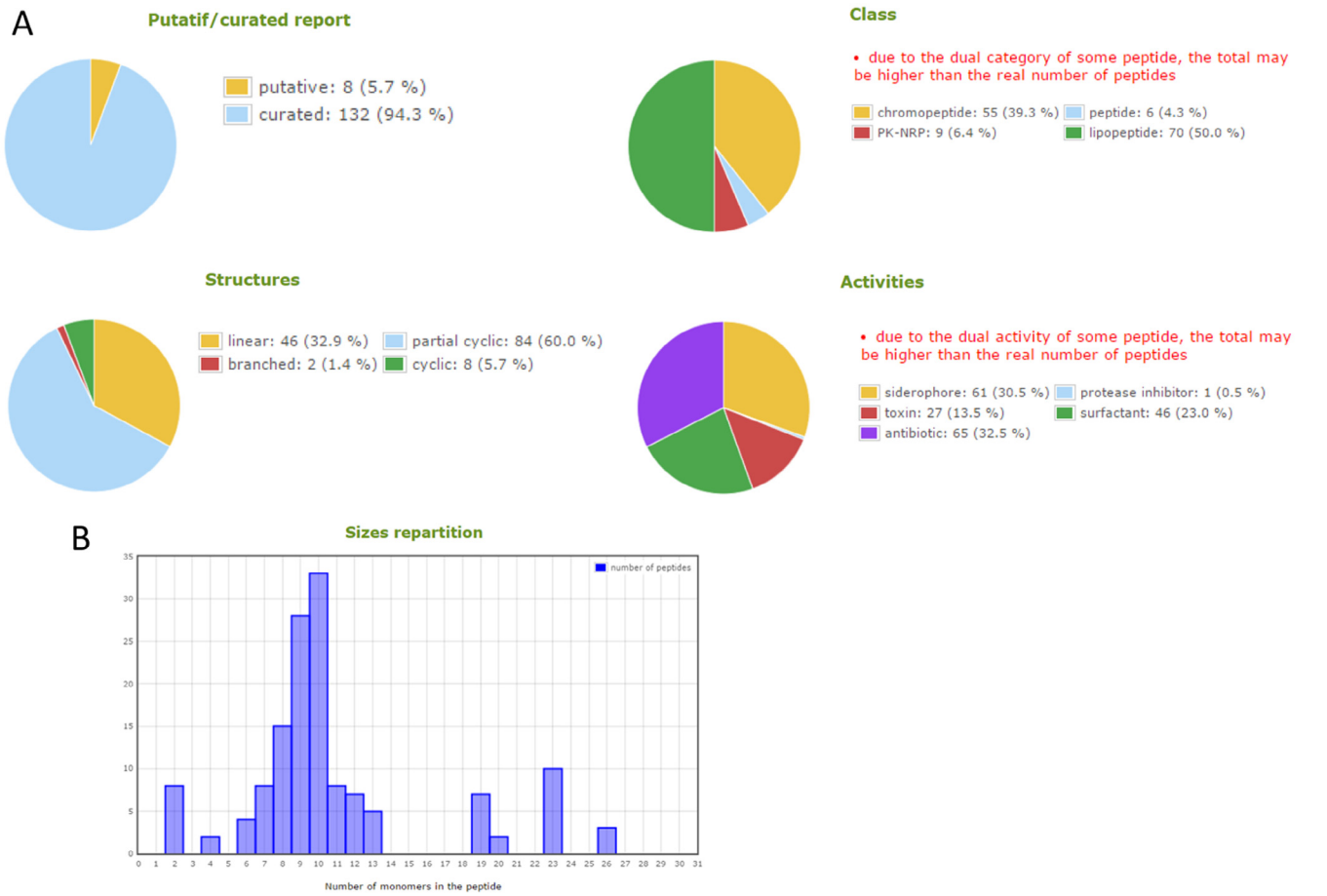

Fig. 1. Graphical output provided with *Pseudomonas* query in "organism search" form. (A) Pie charts representing the percentages of the nonribosomal peptides produced by *Pseudomonas*, according to their status, their class, their structure types and their activities. (B) Histogram representing size distribution of the peptides produced by *Pseudomonas*. For lipopeptides, the fatty acid is considered as one monomer.