



**HAL**  
open science

# Norine: Bioinformatics Methods and Tools for the Characterization of Newly Discovered Nonribosomal Peptides

Areski Flissi, Matthieu Duban, Philippe Jacques, Valérie Leclère, Maude Pupin

► **To cite this version:**

Areski Flissi, Matthieu Duban, Philippe Jacques, Valérie Leclère, Maude Pupin. Norine: Bioinformatics Methods and Tools for the Characterization of Newly Discovered Nonribosomal Peptides. Michael Burkart; Fumihiko Ishikawa. Non-Ribosomal Peptide Biosynthesis and Engineering. Methods and Protocols, 2670, Springer US, pp.303-318, 2023, Methods in Molecular Biology, 978-1-0716-3214-7. 10.1007/978-1-0716-3214-7\_16 . hal-04119620

**HAL Id: hal-04119620**

**<https://hal.science/hal-04119620>**

Submitted on 24 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Norine: Bioinformatics Methods and Tools for the Characterization of Newly Discovered Nonribosomal Peptides

Areski Flissi<sup>1</sup>, Matthieu Duban<sup>2</sup>, Philippe Jacques<sup>3</sup>, Valérie Leclère<sup>2</sup> and Maude Pupin<sup>1</sup>

<sup>1</sup>Université de Lille, CNRS, Centrale Lille, UMR 9189 CRISAL, Lille, France

<sup>2</sup>Université de Lille, UMRt BioEcoAgro 1158-INRAe, Métabolites secondaires d'origine microbienne, Charles Viollette Institute, Lille, France

<sup>3</sup>Université de Liège, UMRt BioEcoAgro 1158-INRAe, Métabolites secondaires d'origine microbienne, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, Gembloux, Belgium

## Abstract

In this chapter, we present Norine (<https://norine.univ-lille.fr/norine>), the unique resource dedicated to nonribosomal peptides. First, the content of the knowledgebase and the related tools are described. Then, a study case shows how to query Norine by annotations or structure and how to interpret the obtained results.

## 1 Introduction

Norine (<https://norine.univ-lille.fr/norine>) is the unique resource dedicated to computational biology analysis of nonribosomal peptides (NRPs) that represent a huge reservoir of microbial secondary metabolites with applications in human, animal, and environmental health. First released in 2006 [1], this resource is composed of a database together with tools useful for NRP analysis including visualization and structure comparison. A specificity of the database is that it can be queried according to both annotation and peptide structures. This project results from a University Lille (France) collaboration between computer science researchers from CRISAL (Centre de Recherche en Informatique et Automatique de Lille) and microbiologists from ICV (Charles Viollette Institute) and BioEcoAgro crossborder joint research unit.

## 2. Materials

### 2.1. Norine

The name Norine was firstly attributed to the database part of the resource. Currently (May 2022), Norine contains about 1750 entries describing NRPs with various annotations classified in different tabs (Fig. 1):

- **Peptide:** This tab gives general information about the NRP such as the Norine ID (e.g., NOR01984), the peptide name, family, synonyms, known biological activities or properties (antimicrobial, surfactant, antitumoral, siderophores, etc.), category (peptide, glycopeptide, lipopeptide, etc.), chemical formula, mono-isotopic mass, and so on. Each validated NRP in Norine is associated with a DOI (Digital Object Identifier) like <https://doi.org/10.26097/nor01984>.
- **Structure:** A specific notation, named monomeric structure, was designed to represent the complex structures of NRPs that can contain cycles and/or branches and that are composed of various monomers such as the 20 proteogenic amino acids, nonproteogenic amino acids like Ornithine (Orn), aminobutyric acid (Aib), and diaminobutyric acid (Dab) or more

original building blocs as vancosamine (Van) and dihydroxybenzoic acid (diOh-Bz), chromophores and a variety of fatty acids. This notation includes the representation of the monomers composing an NRP and the chemical bonds between them. It allows to describe the structure of any NRP and determine its number of monomers and its structure type (linear, cyclic, branched, etc.). Three visualizations of NRPs are provided, the monomeric structure related to the biological mode of synthesis and two chemical (atomic) structures constructed by s2m and rBAN tools from the SMILES notation where the atoms of a monomer are in the same color to highlight the monomeric composition.

- **Organisms** that produce the NRP are represented in a taxonomic tree. The list comes from literature and is not exhaustive.
- **References** from which entry data were extracted (authors, title of the publication, DOI, and PubMed ID). Intentionally, not more than one or two references are proposed. They correspond to the articles describing for the first time the peptide with its structure and/or the synthetase, which proves it is indeed an NRP.
- **Links** to other resources in which the NRP is described (PubChem or ChEMBL for the atomic structure, UniProt or MiBIG for the synthetases, BIRD or PDB for 3D structure, etc.)

The screenshot shows the Norine database entry for 'milkisin A'. On the left, there is a list of properties: Norine ID: NOR01359, DOI: 10.26097/nor01359, Family: amphisin, Synonym(s):, Activity: antimicrobial, surfactant, Category: lipopeptide, Monoisotopic mass: 1381.0 g/mol, Comment: Isomery of each monomer is only predictive, not experimentally amphisin family, Source: mynorine, Contributor (creation): **Schlusshuber Margot** [Université de Caen Norm], Entry information: status: curated - change status to non-NRP (Norine validators only), and a link to view all entry history. On the right, a cyclic peptide structure is shown with monomers: D-Ser, Leu, D-Glu, D-Ile, D-Leu, C80-CH(2), D-Leu, D-Asp, D-Thr, Leu, and Gln. The 'Contributor' field in the left sidebar is highlighted with a red circle.

Fig. 1 : An example of a peptide page corresponding to milkisin A. Contributors to the page are mentioned here highlighted surrounded in red

Norine database also contains annotations about all the monomers identified in at least one NRP including the full usual name, IUPAC names, or SMILES.

Since its creation, Norine's entries were manually extracted from scientific literature. Several articles are often exploited to collect comprehensive information for one peptide. To increase the amount of data while maintaining the quality, we have further developed a tool called MyNorine [2] to open the database to expert sourcing. Scientists all over the world can contribute to Norine by submitting new entries or suggesting modifications or additional annotations. All submissions are further manually validated by Norine curators. Currently, more than 50 scientists are registered in Norine for their contributions and they are cited on the corresponding peptide page as illustrated in the Fig. 1 (contributor item).

In 2019, we have developed a new pipeline [3] for Norine whose purpose is to semiautomatically import massive data (Fig. 2). External databases are mined to extract NRPs that are not yet in Norine or to complete the annotations. We ensure that high quality of the Norine data are maintained using filters to validate the new annotations, and the corresponding NRPs are tagged with an unreviewed status. This pipeline led to the introduction of about 540 new NRPs.

To increase interoperability with other databases, Norine provides a free web API to access the data in various formats (HTML, JSON, XML, and CSV). This is implemented with the RESTful architectural style [4]. The Norine REST services URIs use the following general syntax: [https://.../norine/rest/<path>/<format>/\[parameter\]](https://.../norine/rest/<path>/<format>/[parameter]). The path argument determines the type of service, format can be XML or JSON, and parameter corresponds to the user query. For instance, anyone can access all the annotations of a unique NRP with an URL like <https://norine.univ-lille.fr/norine/rest/id/json/NOR1984> (JSON) or get all the NRPs with “cyclosporin” in their name <https://norine.univ-lille.fr/norine/rest/name/xml/cyclosporin> (XML).

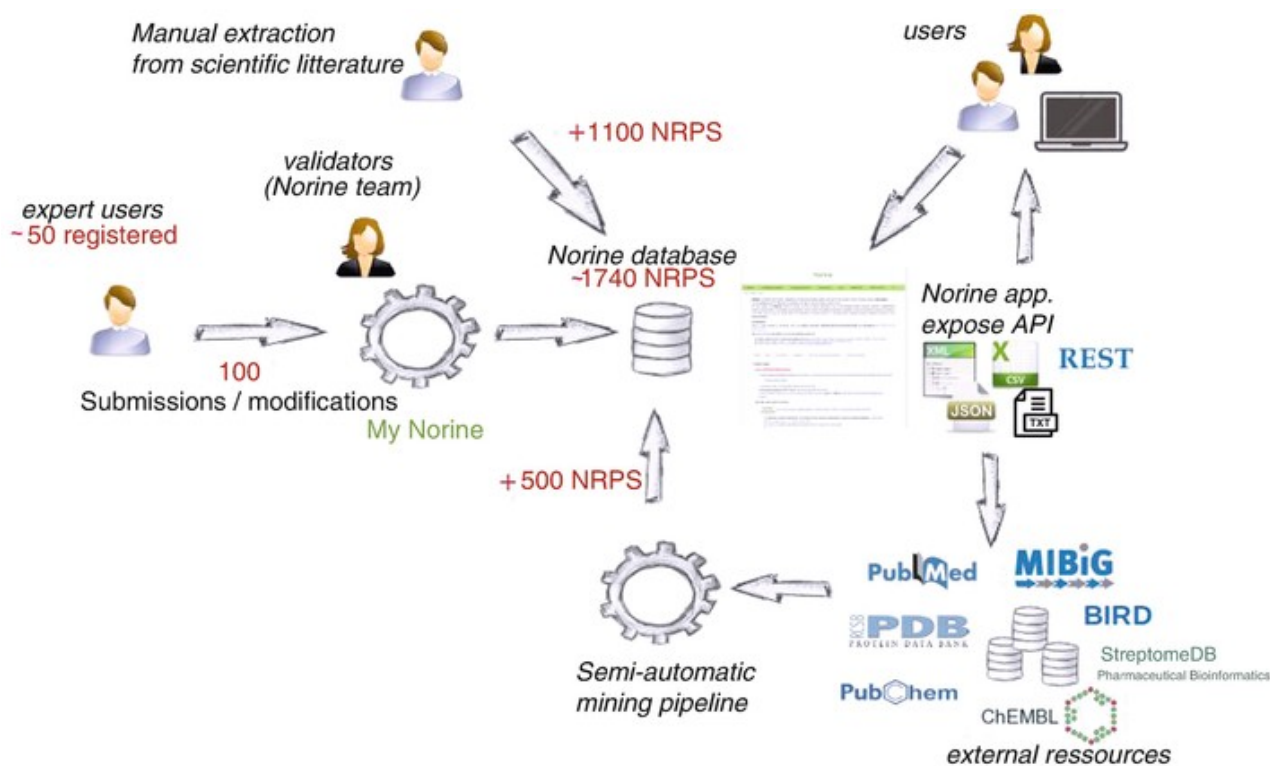


Fig. 2 : *Various processes to enrich Norine database*

Thus, thanks to these results, Norine has been selected on October 2019 by the European bioinformatics ELIXIR consortium (<https://elixir-europe.org>) as an important resource for the scientific community in life science domain (Service Delivery Plan) and is referenced in bio.tools, the registry of software metadata and description (<https://bio.tools/NORINE>).

## 2.2. Smiles2Monomers

Smiles2Monomers (s2m) [5] is a software to automatically infer monomeric structures from chemical structures represented by SMILES notation, mainly extracted from PubChem or PDB databases. s2m maps Norine monomers on the given atomic structure determine the best tiling to cover all the atoms of the NRP polymer. It outputs a graphical atomic structure, where atoms of a

monomer are represented in the same color so that one color is attributed to each monomer. This tool has allowed to complete or correct about 570 SMILES or monomeric structure of NRPs in Norine.

### 2.3. rBAN

rBAN (*retroBiosynthetic Analysis of Nonribosomal peptides*) software [6] has been integrated into Norine in 2019. As s2m, rBAN infers the monomeric structure of an NRP from a SMILES (Fig. 3). However, the strategy is different as, to simulate the retro-biosynthesis of NRPs, rBAN cuts polymers between two atoms identified as parts of bonds between two consecutive monomers such as a peptide bond or a disulfide bridge. Resulting fragments are then matched to Norine's monomers. When a monomer cannot be matched to Norine's monomers, missing substructures are searched in PubChem so as to suggest potential new ones.

Figure 3 shows an example of execution of rBAN (it displays a directed graph format highlighting the bond types between monomers) and s2m for vancomycin antibiotic.

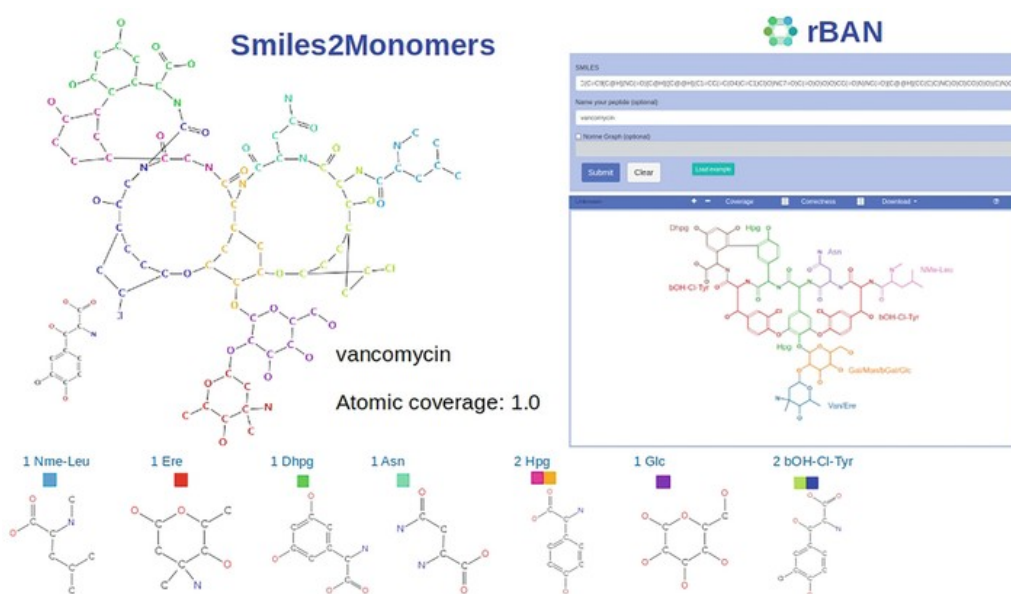


Fig. 3 : *Smiles2Monomers and rBAN results for Vancomycin*

### 2.4. MyNorine

As stated above, the MyNorine [ 2 ] tool was developed to open the Norine database to experts in NRPs, and especially the biologists and biochemists involved in the discovery of new metabolites. Indeed, they are considered as the best experts to enrich the database and improve it in terms of quality and quantity. New data are carefully validated by Norine team before they are added in Norine. In other words, MyNorine distinguishes two roles: curators and validators. While curators contribute to the database by submitting new NRPs/annotations or suggesting modifications, validators are responsible of validation steps. Contributors firstly create an account and can use different sophisticated dynamic interfaces to enrich the database, as illustrated in Fig. 4.

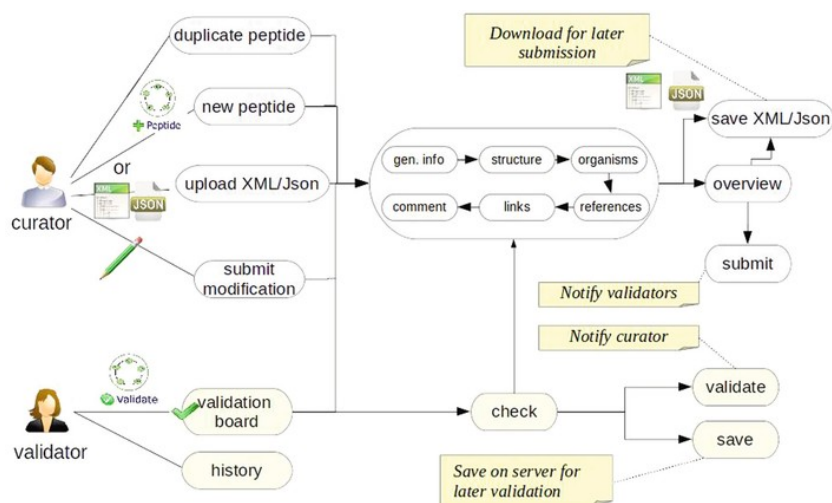


Fig. 4 : *Overview of MyNorine: roles and use cases*

### 3. Methods

#### 3.1. Querying the Norine Database

The Norine database can be queried based on two strategies: the first one is a classical querying using annotations through annotation search tab, while the second one is based on peptide structural features available through structure search tab.

##### 3.1.1. How to Query the Database with Annotations?

The *annotation search* tab contains about 20 fields available to build the queries including the taxonomy related to the producing organisms, literature references, and the peptide itself such as its name, its category (peptide, lipopeptide, chromopeptide, etc.), the presence of monomers or their derivatives, the peptide structure type (linear, cyclic, partial cyclic, branched, etc.), or biological activities. Autocomplete helps to enter existing values. While a user starts filling a field, suggestions of existing values are dynamically proposed. Complex queries can be created by combining several fields with three operators “**AND**,” “**OR**,” and “**AND NOT**”. As an example, if someone is interested in all curated peptides annotated with a SMILES, containing more than 10 monomers, introduced by the Norine team, excluding peptaibols and chromopeptides, the combined criteria will return a list of 68 peptides gathered into 25 families (Fig. 5).

# Norine

Norineannotation searchstructure searchmonomersmyREST APITerms of use

home > NRP > norine

**New!** You can now create sophisticated queries, by using this flexible and dynamic form.

Status

**AND** Smiles  Search only peptides with smiles defined

**AND** Number of monomers > 10

**AND** Source norine

**AND NOT** Category  glycopeptide  lipopeptide  peptaibol  chromopeptide  
 peptide  PK-NRP



**Search for:**  
{status = 'curated'} AND (Peptides that smiles is defined) AND (source is 'norine') AND (containing a number of monomers SUP 10) AND NOT (category is peptaibol | chromopeptide)

---

**68 peptide(s), grouped in 24 familie(s)**

Select/Unselect all

- Calcium dependent antibiotic (2)
  - CDA4b
  - CDA4a
- actinoidin (1)
  - actinoidin A
- amphisin (5)
  - lokisin
  - amphisin
  - pholipeptin
  - tensin
  - arthrofactin

Fig. 5 : Results returned after query using combined criteria

In the returned list, clicking on a peptide name opens the corresponding peptide page where all annotations are accessible and can be downloaded in XML or JSON format. The annotations of all the peptides obtained by a query can be downloaded by clicking on the corresponding icon above the list. A form allows to select the exported fields, and the file format among XML, CSV, or HTML (Fig. 6).



Search for:

(status = 'curated') AND (Peptides that smiles is defined) AND (source is 'norine') AND (containing a number of monomers SUP 10) AND NOT (category is peptaibol | chromopeptide)

Back to the form

68 peptide(s), grouped in 24 familie(s)

Select/Unselect all

Calcium dependent antibiotic (2)

- CDA4b
CDA4a

actinoidin (1)

- actinoidin A

amphisin (5)

- lokisin
amphisin
pholipeptin
tensin
arthrofactin

Peptide data Select all | Deselect all

- Norine id
Name
General name
Synonyms
Activites
Peptide class
Molecular formula
Molecular weight
Comment
Status

Structure data Select all | Deselect all

- Structure type
Number of monomers
PDB SEQRES
graph representation

Organism data Select all | Deselect all

- Organism name
Taxonomy
Gram
Synonyms
taxid

Bibliographical references Select all | Deselect all

- Title
Year
Journal
Location
Pmid
Authors

Links Select all | Deselect all

- Accession number
Database

XML text html



```
peptide
  general
    <id>409882</id>
    <name>amphisin</name>
    <formula>C21H37N7O12</formula>
    <class>amphisin</class>
    <description>amphisin is a complex of 10 peptides which differ in their fatty acid acetyl. It is used to treat complicated skin infections.
  organism
    <name>Streptomyces rimosus</name>
    <taxonomy>
      <kingdom>Bacteria</kingdom>
      <phylum>Firmicutes</phylum>
      <class>Gram positive bacteria</class>
      <order>Streptomyces</order>
      <family>Streptomyces</family>
      <genus>Streptomyces</genus>
      <species>Streptomyces rimosus</species>
  bibliographical references
    <title>amphisin</title>
    <year>1982</year>
    <journal>Journal of Antibiotics</journal>
    <location>Tokyo</location>
    <pmid>7111111</pmid>
    <authors>K. Shimizu, M. Yamashita, M. Yamashita, M. Yamashita, M. Yamashita, M. Yamashita, M. Yamashita, M. Yamashita, M. Yamashita, M. Yamashita</authors>
```

Fig. 6 : Process to download and select annotations to export in various formats

Another icon that gives access to the distribution of the NRPs list are presented as pie charts or histograms to facilitate the overview of categories, structure type, or activities (Fig. 7). A subset of peptides corresponding to a specific criterion can be accessed by clicking on an area of interest within a graphic.





Fig. 7 : Example of graphical views obtained from results of the query

### 3.1.2. Structure Search

The *structure search* tab has been created because peptides have to be considered as identical if they have the same structure (monomer composition and the same sequence including monomer isomery, branches, and cycles), whatever the peptide name or producing organisms. It is also very useful to identify if a newly characterized nonribosomal peptide is really new or is a new variant belonging to a known family. To facilitate the queries, an editor, that provides a column on the left side where the monomers are presented in a clustered list, has been developed. A structure can be easily drawn by clicking on a monomer and drag and drop it in the space available on the right side. Alternatively, a set of derivatives can also be chosen by clicking on a node of the tree.

### 3.2. Submission of a New Peptide

With the MyNorine tool, experts can submit new NRP and add or suggest modifications about existing peptides. The purpose of this module of Norine is to enhance the database by increasing the quantity of described NRPs and improve quality of annotations. Moreover, contributors are cited as the authors of the submission/modification. Users just have to create an account and can access to the tool to contribute by filling forms, as illustrated in Fig. 8.

**Peptide form**

**General information**

Name: cyclo

Family: cyclochlorotine, hydroxycyclochlorotine

Activity: cyclocinamide A, cyclocinamide B, siderophore, protease inhibitor, toxin, surfactant, immunomodulating, antiatherogenic, antiinflammatory

Class: lipopeptide

Synonym(s): A21978C, cubicin

Formula: C72H101N17O26

Molecular weight: 1620.6707

Status: curated, putative

**Next** **Clear**

**Go to final step**

**Automatically suggest name, synonyms, etc**

**Peptide**

Please enter general peptide information

Activity: to select multiple activities, use the 'Ctrl' key.

Synonyms: if exist, MyNorine will automatically suggest you existing synonyms in Norine database.

*Fields in bold are required.*

Fig. 8 : *Form filling to add a new peptide to Norine*

### 3.3. Study Case

For this use case, we created a Fictive Lipopeptide containing 6 Monomers within the peptide moiety, so-called FL6M displaying the following structure: C12:0\_D-OH-Asp\_Dab\_D-Ser\_Leu\_Ser\_Orn with a putative ring between D-Ser and Orn (where C12:0 is the fatty acid containing 12 carbons, Dab is a 2,4-diaminobutyric acid, OH-Asp is a hydroxylated aspartic acid, Orn is for ornithine, and D- specifies the D-isomery). The workflow presented below aims to identify if this peptide is original and to define its place among the NRP diversity. All the numerical results presented are those obtained after querying on May 2022.

#### 3.3.1. Overview on NRPs Sharing Traits with FL6M

The first query is for the name and, as expected, no NRP named FL6M is already annotated in the Norine database. Then, the procedure described below aims to collect a maximum of information on Norine lipopeptides to check if lipopeptides with the same size (containing the same number of monomers) or sharing common specific monomers are described, in order to identify how similar with FL6M they are. For this purpose, it is possible to query the database following different pathways. They are finally leading to same results but differs in the intermediate information available. As illustrated in Fig. 9 a, the query can start by searching for lipopeptides that are 314 gathered into 37 families. Clicking on graphical view icon directs toward various distributions. In

the size distribution bar chart, it is easy to be informed that 20 peptides include 7 monomers (6 amino acids and 1 fatty acid) and by clicking on the corresponding bar, the list of these peptides is returned. Another way to query the database is to start from the number of monomers (Fig. 9 b), 7 in this example. The database contains 176 peptides containing exactly 7 monomers, grouped into 40 families. The pie chart on categories indicates that among the 176 NRPs, 20 are lipopeptides. Both ways arrive to the same conclusion that 20 lipopeptides constituted of 7 monomers are stored in Norine. However, the first way allows to have an overview on lipopeptides, while the other way gives an overview on NRPs containing 7 monomers.



Fig. 9. : *Overview on NRPs sharing annotation with FL6M*

The presence of specific monomers can also be used as a starting point to retrieve information from Norine. In the FL6M example, the presence of a hydroxy-aspartate seems to be original. A search based on the presence of aspartate and its derivatives can be used as a query filter. Through graphical view tool, it is possible to see that among the 272 NRPs containing at least one aspartate derivative, 67 are constituted of 7 monomers. Among them, 6 are lipopeptides, all belonging to the same marinobactin family (Fig. 10 a). As shown on Fig. 10 b, the list of the 6 marinobactins can be directly obtained following a combined query with “asp derivatives” AND number of monomers =7 AND category lipopeptides. It is probably more direct but does not provide some intermediate information obtained as a courtesy of the graphical view tool.

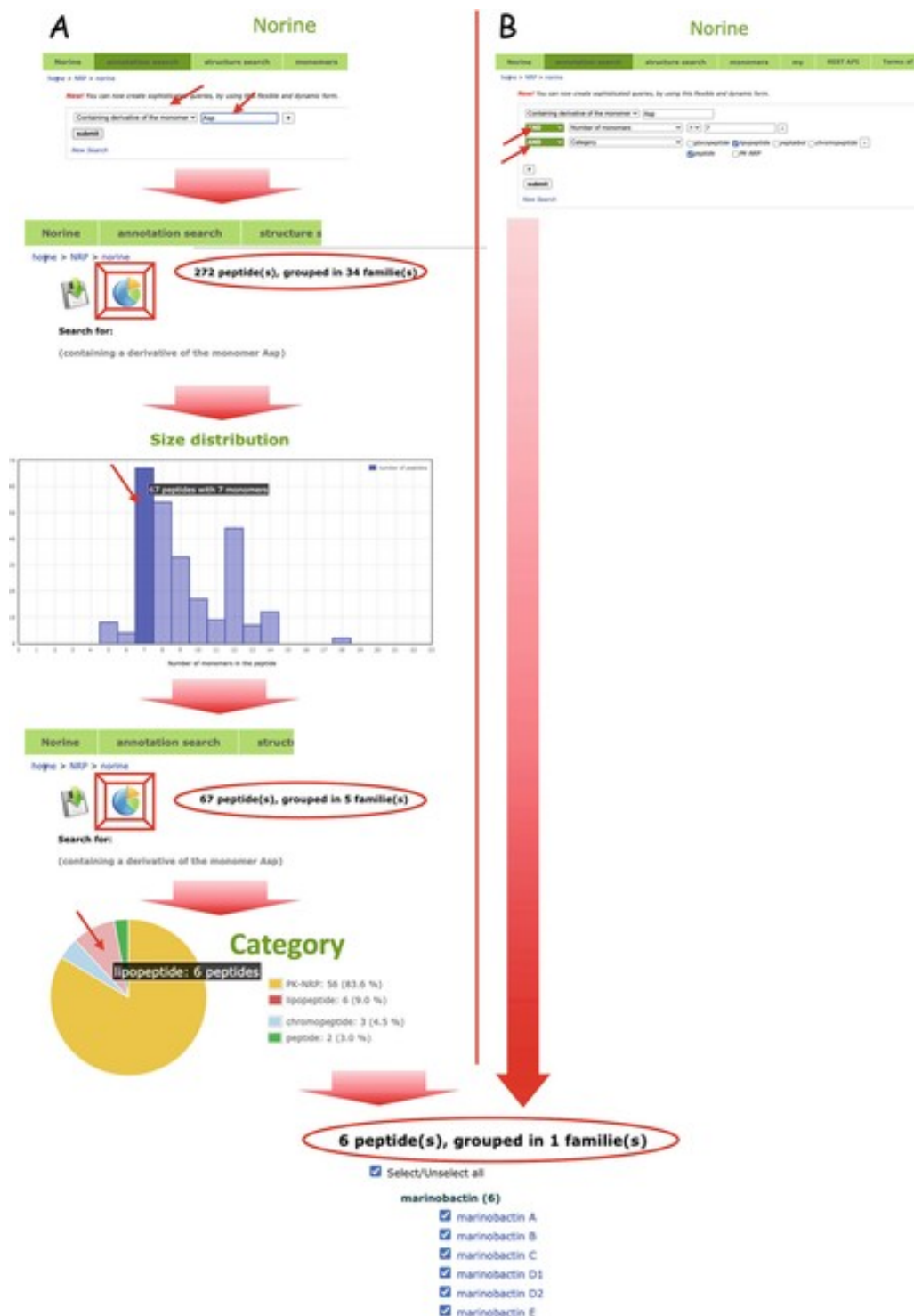


Fig. 10. : Overview on NRPs sharing monomers with FL6M

### 3.3.2. Structure Comparison with NRPs Stored in Norine

An added value of the Norine database is the possibility to search for a peptide through its structure. This allows to determine if the compound of interest shares its structure, or part of its structure, with known NRPs. A first approach can be the comparison of FL6M monomer composition to all Norine NRPs. This can be performed using the “monomer composition fingerprint search.” In the graphical editor, drag and drop each of the 7 monomers from the clustered monomer list to the drawing panel. As shown in Fig. 11, the search for “D-OHAsp” instead of “Asp\*” (meaning Asp derivatives) outputs slightly different results. This can be explained by the presence of OHAsp instead of D-OHAsp in the corrugatin, together with 3 Dab residues in this lipopeptide constituted of 9 monomers. Looking at the structures of both lipopeptides clearly indicate that they are not really similar (7 vs. 9 monomers, not the same sequence).

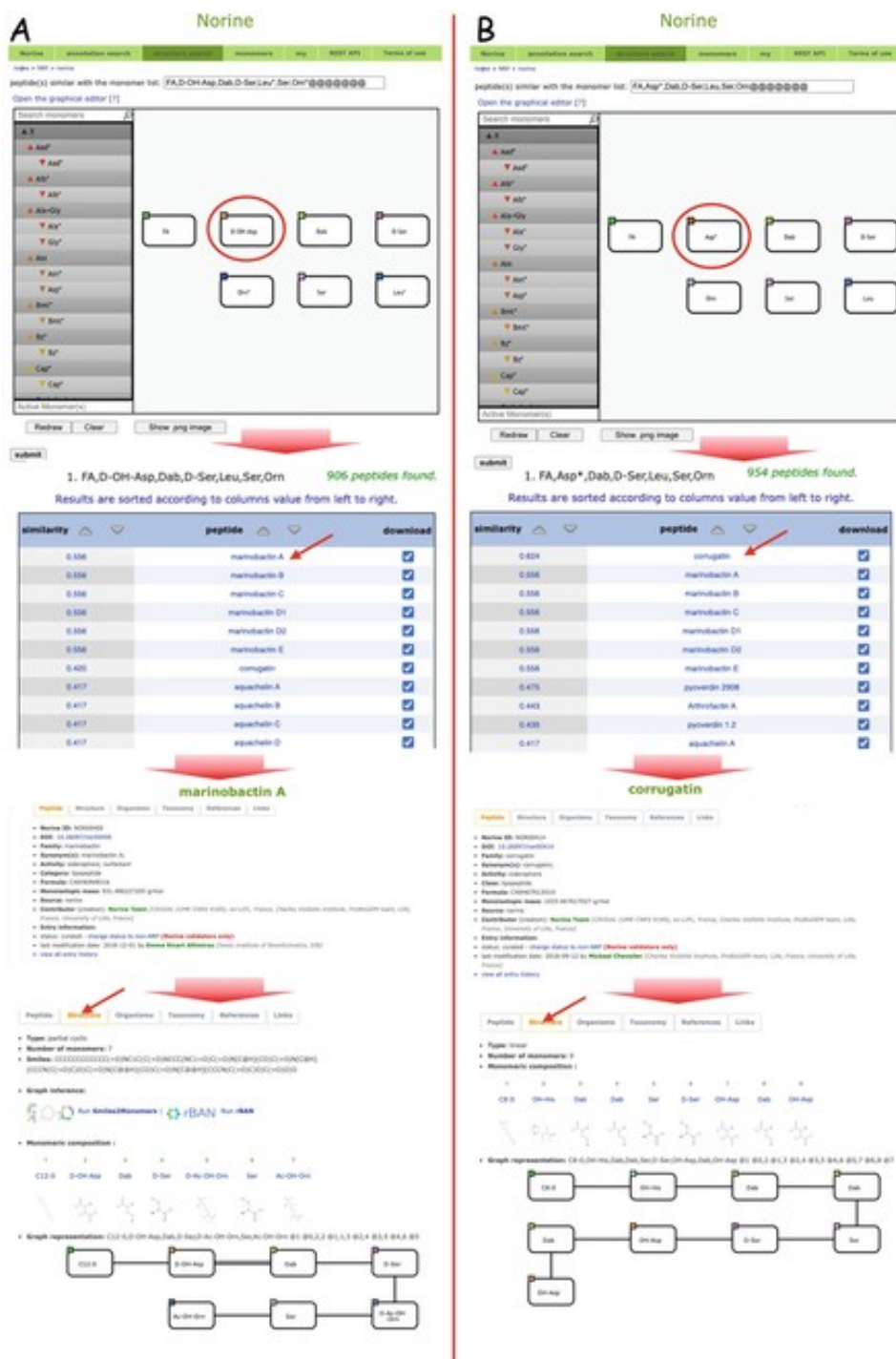


Fig. 11 : Comparison of results obtain after D-OH Asp or Asp derivative query

To further compare the structures, the “structure-based search” tries to map the given structural pattern to Norine NRPs. The graphical editor helps to design the pattern allowing to retrieve the monomers and bond them together. Figure 12 illustrates the different results obtained while searching for peptides containing the complete structural pattern of FL6M or pattern substructures with at least 2, 3, or 4 monomers.

**Structure-based search [?]**

peptide(s) containing the structural pattern: `FA,D-OH-Asp, Dab, D-Ser, Ser, Leu, Orn@1@0,2@1,3@2,5,6,`

Open the graphical editor [1]

Search monomers

- Ala
- Asn
- Asp
- Asx
- Asn-Gly
- Asx
- Gly
- Ala
- Asp
- Asn
- Asp
- Ser
- Ser
- Cys
- Cys

Active Monomer(s)

Norine

---

Check the box if you want to obtain the fingerprint distance between peptides containing the pattern and your pattern

Calculate fingerprint

containing the complete pattern

containing the pattern substructure with at least  monomers

**Add a fragment**

If there is no structure, the peptide is linear

**submit**

Norine   annotation search   **structure search**   monomers   my

home > NRP > norine

Sorry, no result matches your criteria

Query : Peptide(s) containing at least one of these fragments with no less than 0 monomer :

---

Check the box if you want to obtain the fingerprint distance between peptides containing the pattern and your pattern

Calculate fingerprint

containing the complete pattern

containing the pattern substructure with at least  monomers

**Add a fragment**

If there is no structure, the peptide is linear

**submit**

**Peptide(s) containing at least one of these fragments with no less than 2 monomers :**

1. FA,D-OH-Asp,Dab,D-Ser,Ser,Leu,Orn@1@0,2@1,3@2,5,6@5,6@3,4@3,4

peptide	download	peptide id	download
apocytin A	<input checked="" type="checkbox"/>	apocytin Q176	<input checked="" type="checkbox"/>
apocytin B	<input checked="" type="checkbox"/>	apocytin Q45	<input checked="" type="checkbox"/>
apocytin C	<input checked="" type="checkbox"/>	apocytin R	<input checked="" type="checkbox"/>
apocytin D	<input checked="" type="checkbox"/>	apocytin R296	<input checked="" type="checkbox"/>
antithrombin A	<input checked="" type="checkbox"/>	antithrombin 1	<input checked="" type="checkbox"/>
antithrombin B	<input checked="" type="checkbox"/>	Synthetic PPTA/PCPAE-14 Complex	<input checked="" type="checkbox"/>
antithrombin C	<input checked="" type="checkbox"/>	antithrombin 2	<input checked="" type="checkbox"/>
antithrombin D	<input checked="" type="checkbox"/>	antithrombin 3	<input checked="" type="checkbox"/>
antithrombin E	<input checked="" type="checkbox"/>	antithrombin 4	<input checked="" type="checkbox"/>
antithrombin F	<input checked="" type="checkbox"/>	antithrombin 5	<input checked="" type="checkbox"/>
antithrombin G	<input checked="" type="checkbox"/>	antithrombin 6	<input checked="" type="checkbox"/>
antithrombin H	<input checked="" type="checkbox"/>	antithrombin 7	<input checked="" type="checkbox"/>
antithrombin I	<input checked="" type="checkbox"/>	antithrombin 8	<input checked="" type="checkbox"/>
antithrombin J	<input checked="" type="checkbox"/>	antithrombin 9	<input checked="" type="checkbox"/>
antithrombin K	<input checked="" type="checkbox"/>	antithrombin 10	<input checked="" type="checkbox"/>
antithrombin L	<input checked="" type="checkbox"/>	antithrombin 11	<input checked="" type="checkbox"/>
antithrombin M	<input checked="" type="checkbox"/>	antithrombin 12	<input checked="" type="checkbox"/>
antithrombin N	<input checked="" type="checkbox"/>	antithrombin 13	<input checked="" type="checkbox"/>
antithrombin O	<input checked="" type="checkbox"/>	antithrombin 14	<input checked="" type="checkbox"/>
antithrombin P	<input checked="" type="checkbox"/>	antithrombin 15	<input checked="" type="checkbox"/>
antithrombin Q	<input checked="" type="checkbox"/>	antithrombin 16	<input checked="" type="checkbox"/>
antithrombin R	<input checked="" type="checkbox"/>	antithrombin 17	<input checked="" type="checkbox"/>
antithrombin S	<input checked="" type="checkbox"/>	antithrombin 18	<input checked="" type="checkbox"/>
antithrombin T	<input checked="" type="checkbox"/>	antithrombin 19	<input checked="" type="checkbox"/>
antithrombin U	<input checked="" type="checkbox"/>	antithrombin 20	<input checked="" type="checkbox"/>
antithrombin V	<input checked="" type="checkbox"/>	antithrombin 21	<input checked="" type="checkbox"/>
antithrombin W	<input checked="" type="checkbox"/>	antithrombin 22	<input checked="" type="checkbox"/>
antithrombin X	<input checked="" type="checkbox"/>	antithrombin 23	<input checked="" type="checkbox"/>
antithrombin Y	<input checked="" type="checkbox"/>	antithrombin 24	<input checked="" type="checkbox"/>
antithrombin Z	<input checked="" type="checkbox"/>	antithrombin 25	<input checked="" type="checkbox"/>

32 peptides found.

---

Check the box if you want to obtain the fingerprint distance between peptides containing the pattern and your pattern

Calculate fingerprint

containing the complete pattern

containing the pattern substructure with at least  monomers

**Add a fragment**

If there is no structure, the peptide is linear

**submit**

**Peptide(s) containing at least one of these fragments with no less than 3 monomers :**

1. FA,D-OH-Asp,Dab,D-Ser,Ser,Leu,Orn@1@0,2@1,3@2,5,6@5,6@3,4@3,4

peptide	download
antithrombin A	<input checked="" type="checkbox"/>
antithrombin B	<input checked="" type="checkbox"/>
antithrombin C	<input checked="" type="checkbox"/>
antithrombin D	<input checked="" type="checkbox"/>
antithrombin E	<input checked="" type="checkbox"/>
antithrombin F	<input checked="" type="checkbox"/>
antithrombin G	<input checked="" type="checkbox"/>

6 peptides found.

---

Check the box if you want to obtain the fingerprint distance between peptides containing the pattern and your pattern

Calculate fingerprint

containing the complete pattern

containing the pattern substructure with at least  monomers

**Add a fragment**

If there is no structure, the peptide is linear

**submit**

Sorry, no result matches your criteria

Query : Peptide(s) containing at least one of these fragments with no less than 4 monomers :

Fig. 12 : Search for peptides containing complete structural pattern or pattern substructures

To conclude if FL6M is a novel lipopeptide or belongs to a known family (new variant), it is necessary to compare carefully the structures. In the presented use case, while a score of 0.556 with marinobactin A is obtained, a first glance at the structure clearly indicates that they are different, especially regarding the position and sequence of the lactone ring. Consequently, they likely belong to different families.

A same warning point can be mentioned about the assignment of putative activities that will have to be experimentally verified in any case. It is also important to keep in mind that Norine database is not comprehensive, accordingly this limits the completeness of the results related to the annotations as well as to the structure.

## References

1. . Caboche S, Pupin M, Leclère V et al (2008) NORINE: a database of nonribosomal peptides. *Nucl Acids Res* 36:D326–D331
2. Flissi A, Dufresne Y, Michalik J et al (2016) Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing. *Nucl Acids Res* 44:D1113–D1118
3. Flissi A, Ricart E, Campart C, Chevalier M et al (2020) Norine: update of the nonribosomal peptide resource. *Nucl Acids Res* 48: D465–D469
4. Fielding RT, Taylor RN (2002) Principled design of the modern web architecture. *ACM Trans Internet Technol* 2:115–150
5. Dufresne Y, Noé L, Leclère V et al (2015) Smiles2Monomers: a link between chemical and biological structures for polymers. *J Cheminform* 7:62
6. Ricart E, Leclère V, Flissi A et al (2019) rBAN: retro-biosynthetic analysis of nonribosomal peptides. *J Cheminform* 11:13