

Database and comparison of non ribosomal peptides

Ségolène Caboche¹, Valérie Leclère², Philippe Jacques², Maude Pupin¹ and Gregory Kucherov¹

¹ Laboratoire d'Informatique Fondamentale de Lille, UMR CNRS 8022 - Université des Sciences et Technologies de Lille, Villeneuve d'Ascq, France

² Laboratoire ProBioGEM, Polytech'Lille - Université des Sciences et Technologie de Lille, Villeneuve d'Ascq, France

Main Thematics: *Sequence analysis, motifs, others*

Technical Fields: *Database, protein, algorithmics*

Keywords: Non Ribosomal Peptide Synthesis, database, non linear structure comparison

Peptides are synthesized in bacteria or fungi not only through the classical pathway from the transcription of DNA to the translation of mRNA into peptides, but also through a ribosome-independent pathway. This alternative pathway is called Non Ribosomal Peptide Synthesis (NRPS). It is performed by a huge protein complex called synthetase or NRPS (Non Ribosomal Peptide Synthetase). These are modularly organized multi-enzyme complexes which represent at the same time template and biosynthetic machinery. The molecules synthesized by NRPS are short (two to about fifty monomers), include a high density of non proteogenic amino acids and often contain amino acids connected by bounds other than peptide or disulfide bounds (for a review, see [1]). So the first dimension of the peptides synthesized is not always linear. It can forms cycles (partial or not) and it can exhibit branches.

Those peptides have interesting biological properties such as immunomodulating, iron chelating or antibiotic activities, as illustrated by the famous antibiotic penicillin which is synthesized from ACV-tripeptide precursor produced by the NRPS pathway. The construction of a database with all known NRPS synthesized peptides should represent a good tool to study these products. To our knowledge, such a database is not available, except “a knowledge based resource for analysis of Non-Ribosomal Peptide Synthetases and Polyketide Synthetases” (see [2] and [3]) which only concerns NRPS enzymatic complexes but not directly products (only about twenty NRPS produced peptides are mentioned). We created a database of non ribosomal peptides which contains the peptide's structure and various annotations like their biological activity, the organism which produces it or bibliographical references. We needed a linear nomenclature in order to represent the peptide's non linear structure and to facilitate the data input. So we created a nomenclature which is inspired by the IUPAC nomenclature to name the various amino acids and which represents cycles and branches by specific symbols. The non ribosomal peptides database currently contains several hundreds of peptides and will be soon distributed on the Web.

The database can be queried via annotations but also via peptide's structure. Indeed, existing software allows a prediction of the peptide from the synthetase proteomic sequence. It is interesting to compare those predicted peptides to those of the database. Therefore, we are interested in creating a software allowing the search for a given peptide in the database by comparing 1D non linear structures. This challenge gives a new dimension to the comparison of peptides and needs a specific strategy. Indeed, existing sequence alignment algorithms can't be directly applied to this case due to a non-linear structure of the non-ribosomal peptides. Using the linear nomenclature, we can compare two peptides as a whole or find similar parts of peptides.

This work comes from a narrow collaboration between biologists and bioinformatics teams and supplies a new dimension to the comparison of peptides due to the non linear structure of NRPS peptides.

References

- [1] SA. Sieber and MA. Marahiel, Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics, *Chemical Reviews*, Vol. 105, No. 2, 2005.
- [2] MZ. Ansari, G. Yadav, RS. Gokhale and D. Mohanty, NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthetases, *Nucleic acids research*, 32 (Web sever issue) W405-13, 2004.
- [3] <http://www.nii.res.in/nrps-pks.html>