PepFuNN: Novo Nordisk open-source toolkit to enable peptide *in silico* analysis

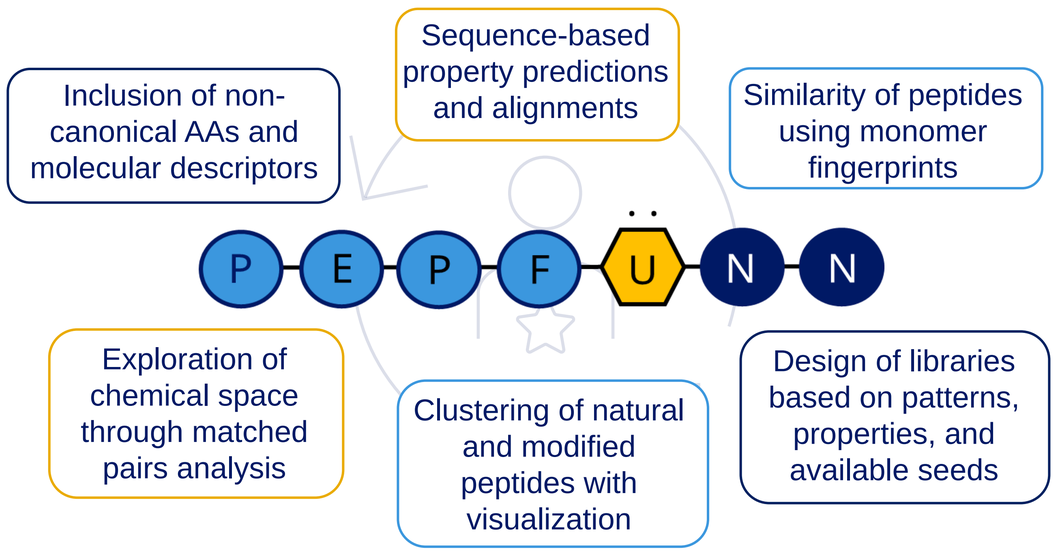
Rodrigo Ochoa[1], Kristine Deibler[2]

[1]Novo Nordisk A/S, Novo Nordisk Park, Måløv, DK-2760, Denmark

[2]Novo Nordisk Research Center Seattle, Novo Nordisk A/S, 818 Stewart St. Suite #910, Seattle, WA, 98101, USA

Source code

https://doi.org/10.17952/37EPS.2024.P1270



Open computational protocols for peptide analysis

- We present PepFuNN, a new open-source version of the PepFun package with functions to study the chemical space of peptide libraries and perform structure-activity relationship (SAR) analysis.
- PepFuNN modules allow calculating physicochemical properties, performing similarity analysis using different peptide representations, running SAR exploration studies adapted for peptides, and functions to support the design of peptide libraries based on specific requirements.

What input do you need?

Simple linear peptides can be represented using FASTA format. For non canonical elements, line notations to allow the representation of complex peptides are suggested (i.e., BILN format*).

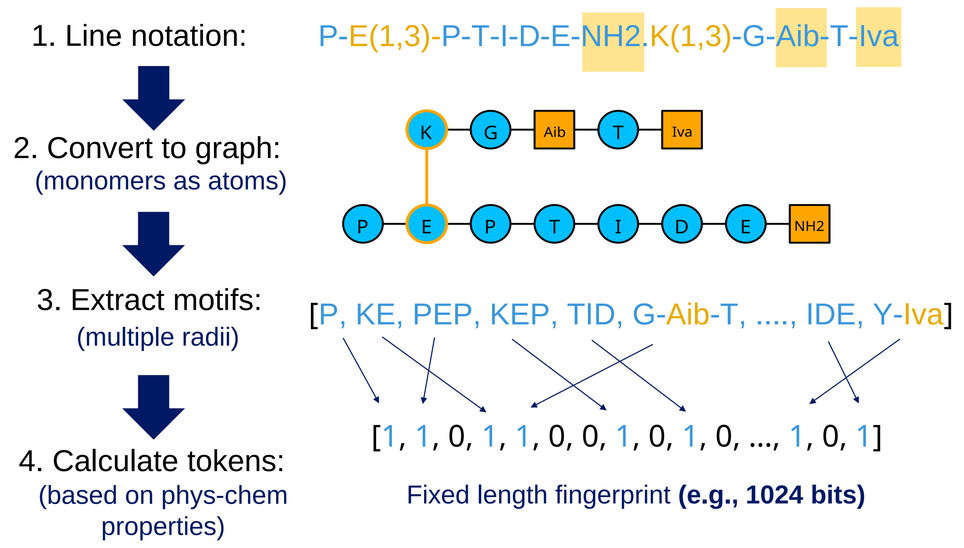
*Fox, Bieler, Haebel, Ochoa, Peters, Weber. *J. Chem. Inf. Model.* 2022

Description	FASTA	BILN
Simple linear peptide	PEPTIDE	P-E-P-T-I-D-E
Peptide with internal C-C disulfide bridge and amidation at the C-terminus	ACGAGCD	A-C(1,3)-G-A-G-C(1,3)-D-NH2
C-term to N-term cyclic peptide	CYCLIC	C(1,1)-Y-C-L-I-C(1,2)

What new methodologies are included in PepFuNN?

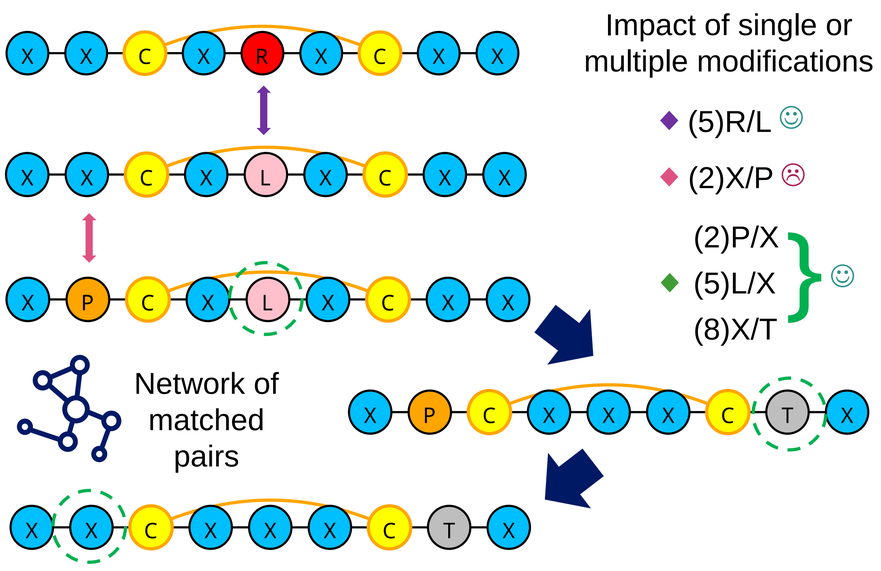
For complex peptides, it is possible to generate monomer-based fingerprints to run similarity analysis.

- A line notation sensible to complex chemical structures is used to represent the peptides.
- The complex peptide is represented as a graph, which is decomposed into fragments up to a defined radius.
- Per each combination, a numerical token is generated based on a combination of physicochemical properties per monomer.
- Each token is converted into a bit that will be part of a fixed-size fingerprint used to run similarity searches.



How to run SAR analysis?

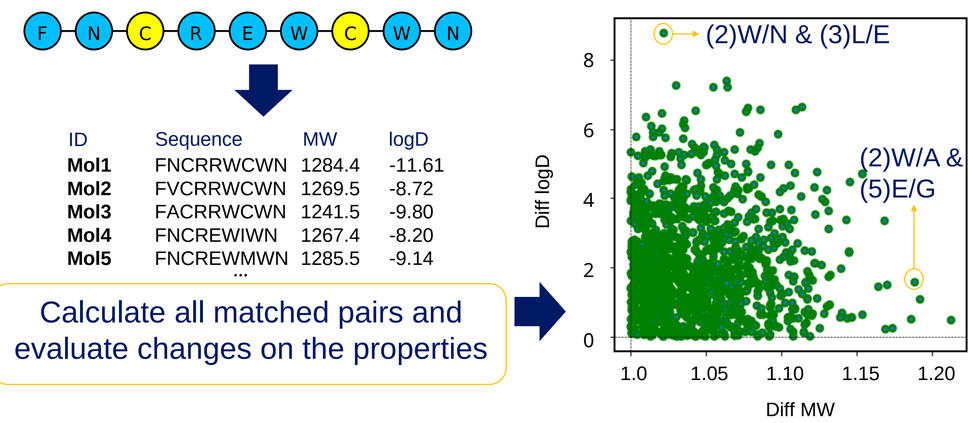
For peptides, we allow the option to calculate molecular matched pairs to identify single modifications that can have a positive or negative impact in an experimental assay.



Matched pairs analysis of a library built with PepFuNN, using a peptide scaffold as a reference

- The matched pairs methodology compares the properties of two molecules that differ by a single chemical transformation.
- The method is relevant to deeply exploring SAR data and collecting ideas for designing optimized molecules.

Library design using a scaffold



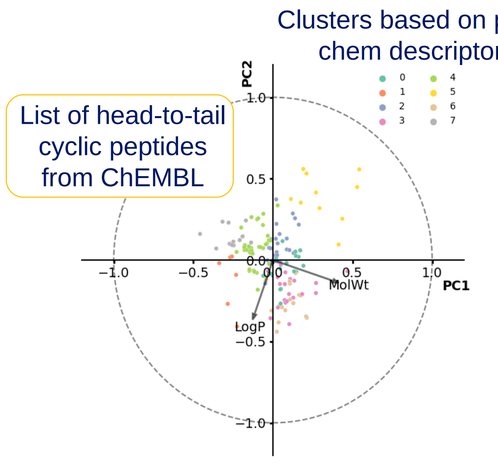
How to check the similarity between simple and/or complex peptides?

We can use different notations and representations to compare the peptides. In this case, the chemical similarity strategy can rely on the use of Morgan fingerprints (used also for small molecules) or monomer-based fingerprints. The latest are obtained from the BILN representations of the sequences.

Availability

- Project name: PepFuNN (version 1.0)
- Project home page: https://github.com/novonordisk-research/pepfunn
- Operating system(s) tested: Linux
- Programming language: Python 3.9 or higher
- Other requirements: RDKit 2020 or later; Biopython 1.7.9 recommended
- License: MIT

- Strategy to cluster and compare cyclic head-to-tail peptides from the ChEMBL database.
- We can generate a Principal Component Analysis (PCA) plot of the descriptors, with arrows showing the properties with higher importance in the dataset.



Chemical similarity

Use as input molecular structures or sequences

P(1,1)-V-N-P-F-V-**V(1,2)**

Molecular similarity (Morgan fingerprints)

Sequence similarity (Monomer fingerprints)

P(1,1)-F-I-L-P-V-**N(1,2)**
P(1,1)-V-N-P-F-V-**V(1,2)**
F(1,1)-F-W-V-L-**P(1,2)**
P(1,1)-F-V-F-L-E-**C(1,2)**
I(1,1)-F-Y-L-P-L-**P(1,2)**

Computational Drug Design • Molecular AI

raoc@novonordisk.com

Digital Science & Innovation DSI.