Computational approaches for analysis and design of peptides containing non-canonical amino acids

Rodrigo Ochoa, and many other collaborators*

Novo Nordisk A/S, Novo Nordisk Park, Måløv, DK-2760, Denmark (https://doi.org/10.17952/38EPS.2026.P251)

*See acknowledgements

Motivation

Computational methods are required to study peptides and to gain insights into rational design strategies. In the special case of peptides containing non-canonical amino acids, additional strategies and data are necessary to overcome computational challenges.

This work aims to summarize a set of tools and benchmark studies for modified peptides, using both ligand- and structure-based approaches.

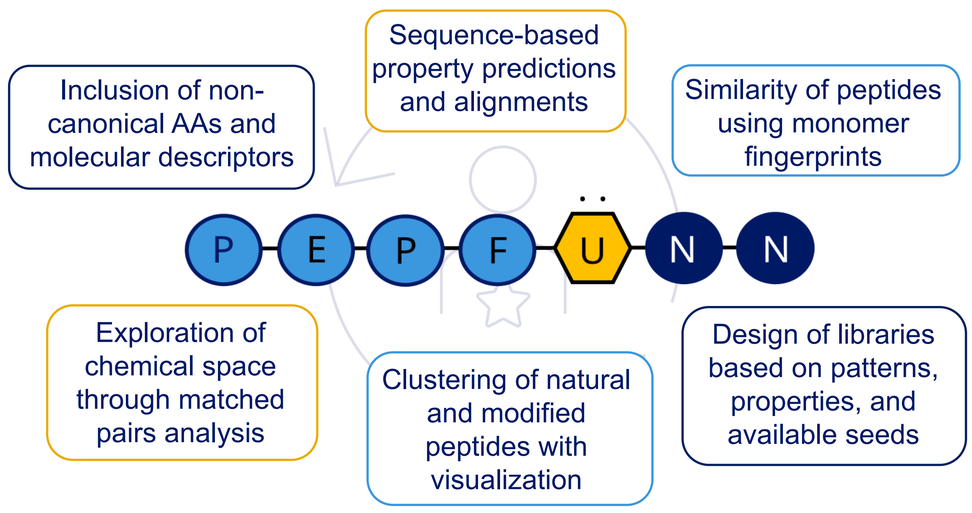
How to represent modified peptides?

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).

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-Aib-G-C(1,3)-D-NH2
C-term to N-term cyclic peptide	CYCLIC	C(1,1)-Y-C-L-I-C(1,2)

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

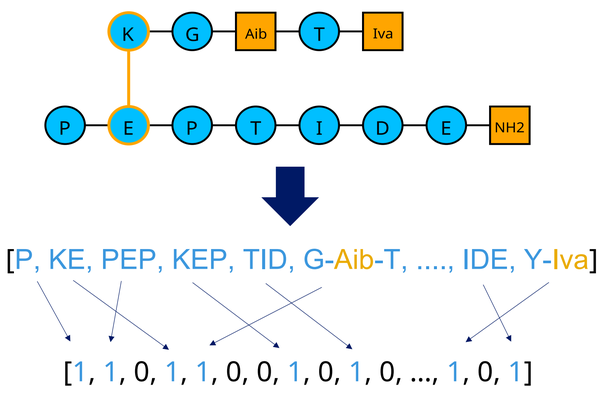
1. PepFuNN – peptide analysis toolkit from Novo Nordisk



PepFuNN modules contain multiple functions to support the analysis and design of peptides, including those modified with non-canonical components.

Special fingerprints and properties can be estimated for modified peptides, for example, to create customized machine learning models

Monomer-based fingerprints:



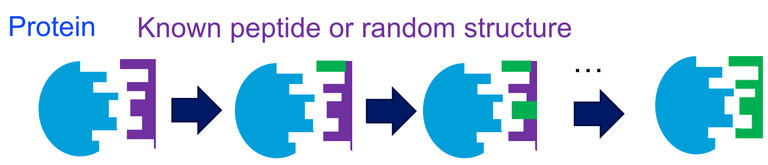
Fixed length fingerprint **(e.g., 1024 bits)**

https://github.com/novonordisk-research/pepfunn

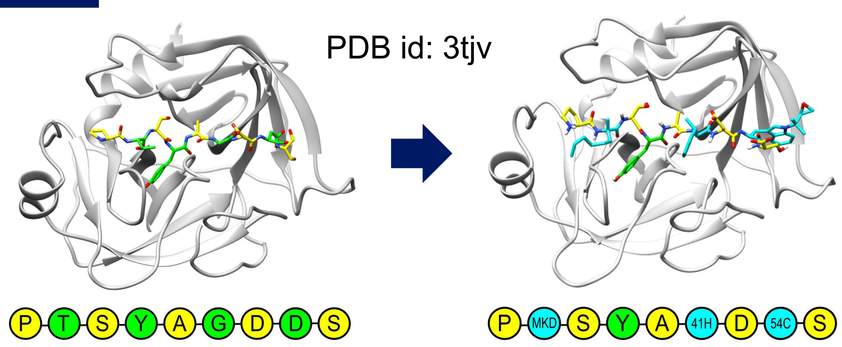
Ochoa & Deibler. *J. Pept. Sci.* 2025

2. mPARCE – protocol for the optimization of bound modified peptides

mPARCE allows adding non-canonical amino acids into peptide binders while improving their binding affinity. The workflow makes single point mutations on the peptide sequence based on a Monte Carlo approach



Non-canonical amino acids from the PDB can be incorporated

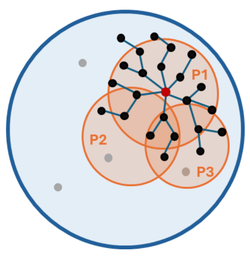


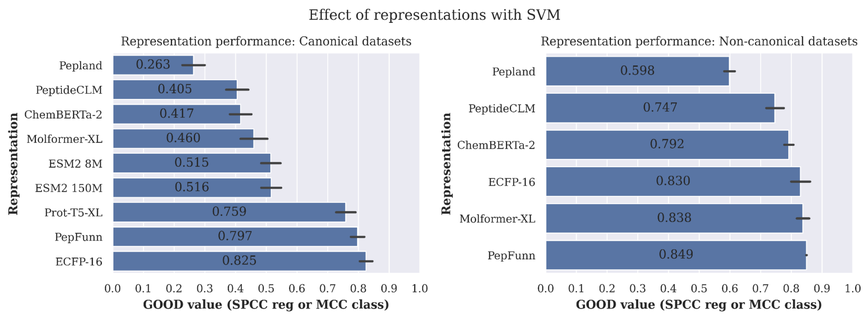
Ochoa, Cossio, Fox. *J. Comput.-Aided Mol. Des.* 2022

3. Machine learning (ML) representation for modified peptides

Traditional sequence-based approaches are often inadequate for modelling non-canonical peptides, and there is a critical need for advanced ML techniques that can effectively represent modified peptides, addressing challenges related to molecular representation and dataset partitioning.

- Molecular similarity functions can be effective for guiding dataset partitioning for both canonical and non-canonical peptides
- The findings inform the best configurations (model and representations) for enabling the routine modeling of peptides with or without chemical modifications. Example using a SVM architecture:



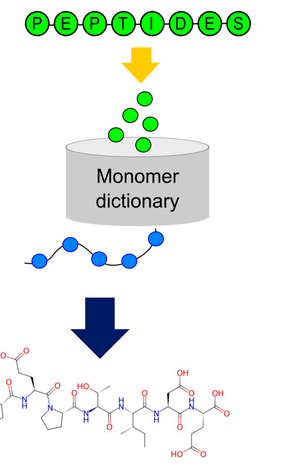


Fernandez-Diaz, Ochoa, Hoang, Lopez, Shields. *J. Cheminformatics.* 2025

4. pyPept – peptide format conversion

Python package to connect natural and non canonical residues through defined R-groups, and translate BILN and FASTA format to all-atom formats (i.e. SMILES)

- **Input:** peptide (BILN format) + Monomer dictionary with non-canonical components
- **Output:** RDKit molecular object that can be transformed to 2D and 3D formats

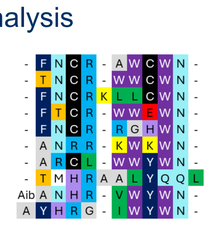


Ochoa, Brown, Fox. *J. Cheminformatics.* 2023

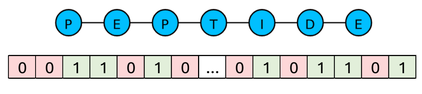
Summary

These open-source tools allow users to incorporate chemical modifications for in silico peptide drug discovery purposes

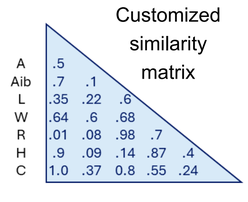
Structure activity relationship (SAR) analysis



Similarity calculations



Property prediction



Sequence design



Erckes, Abderrahmane, Jusot, Steuer, Ochoa. *Drug Discov. Today.* 2026

Acknowledgements

- Kristine Deibler (Molecular AI team, Novo Nordisk, USA)
- Jovan Damjanovic & David Rodriguez (Digital Chemistry & Design, Novo Nordisk, Denmark)
- Thomas Fox (Computational Chemistry, Boehringer Ingelheim, Germany)
- Denis Shields & Raul Fernandez (University College Dublin & IBM Research, Ireland)
- Vanessa Erckes & Christian Steuer (ETH Zurich, Switzerland)
- Maud Jusot & Massina Abderrahmane (Isomorphic Labs, UK & Iktos, France)
- Pilar Cossio (Flatiron Institute, USA)