PeptiDesCalculator 2.0: a next-generation descriptor platform for peptide informatics, sequence analysis, and QSAR modeling

Laureano E. Carpio[1], Rafael Gozalbes[1,2] & Eva Serrano-Candelas[2]

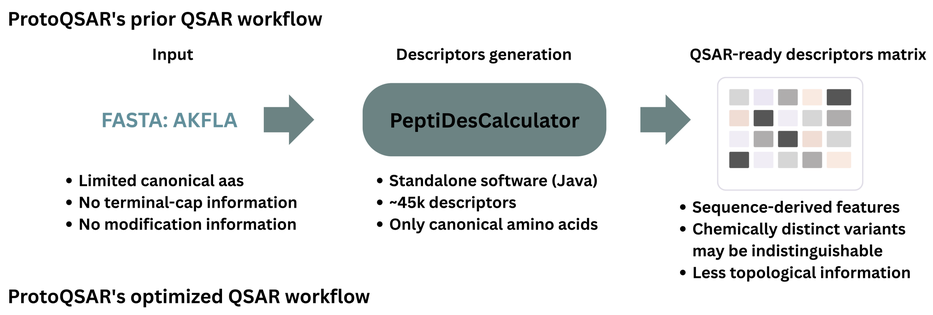
[1]*MolDrug AI Systems, Technological Park of Valencia, Spain*
[2]*ProtoQSAR, Technological Park of Valencia, Spain*

Email: lcarpio@moldrug.com

https://doi.org/10.17952/38EPS.2026.P256

Introduction

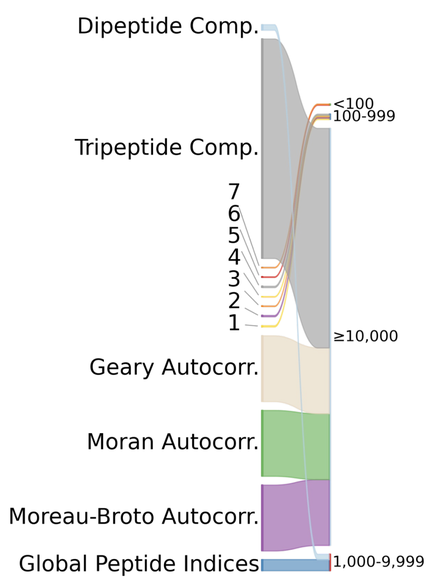
Peptides are increasingly relevant therapeutic modalities due to their high selectivity, structural diversity and broad biological applicability[1]. However, their computational representation for Quantitative Structure-Activity Relationships (QSAR) modelling and predictive purposes remains challenging because peptide properties depend not only on amino acid composition, but also on sequence order, residue-level physicochemical patterns, stereochemistry and chemical modifications. Conventional molecular descriptors do not explicitly capture many of these peptide-specific features, while existing sequence-based tools are frequently restricted to canonical amino acids[2,3]. Therefore, extensible and peptide-centered descriptor frameworks are required to support sequence analysis, chemical interpretation and machine learning-driven discovery.



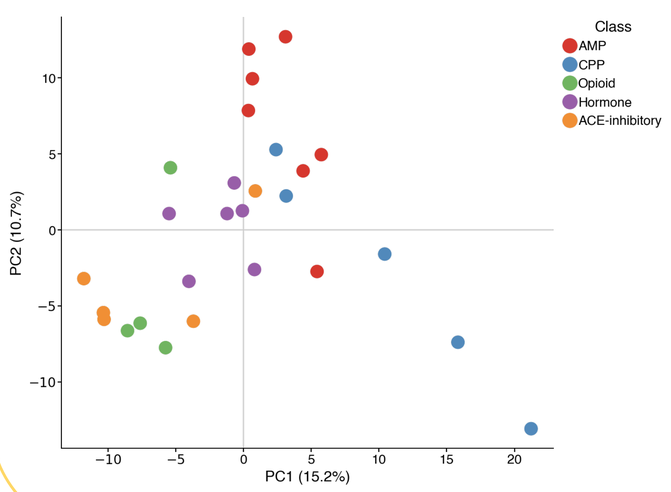
Results

1. PeptiDesCalculator 2.0 provides a modular universe of 127,840 descriptors

The Java engine was reimplemented in Python as a modular framework comprising **13 descriptor families and up to 127,840 variables**. Tripeptide composition and sequence autocorrelations account for most of the descriptor universe, while smaller families provide complementary information.



2. Multi-family descriptors capture chemically meaningful peptide diversity



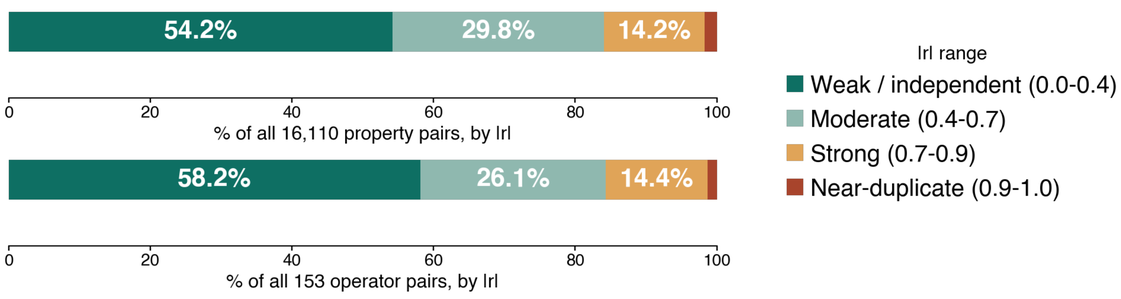
An unsupervised PCA was performed using 375 non-constant descriptors calculated using PeptiDesCalculator v2.0 for 27 published peptides from five functional classes. Patterns emerged without supervision, demonstrating that the descriptors capture peptide information and effectively represent differences among peptide classes.

3. Descriptor families respond predictably to D-residue substitutions

Descriptor group	Achatini	Deltorphin I	Dermorphin	Fulicin
Composition: AA / di / tri	Changed	Changed	Changed	Changed
CTD / Conjoint / autocorr. / GPI	Preserved	Preserved	Preserved	Preserved
PSAC / APSAC / QSO	Partial N/A	N/A	N/A	Partial N/A
Global descriptors	Partial N/A	Partial N/A	Partial N/A	Partial N/A

Across four natural D-peptides, identity-based descriptors detected the D/L substitution, physicochemical families remained invariant, and unsupported formulations returned explicit missing values.

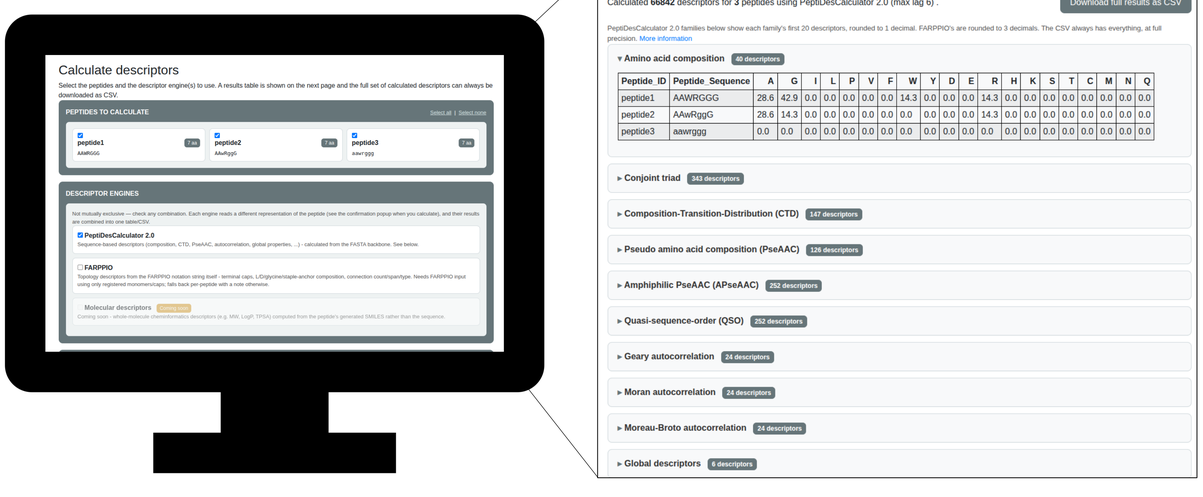
4. AAindex scales and aggregation operators show limited pairwise redundancy



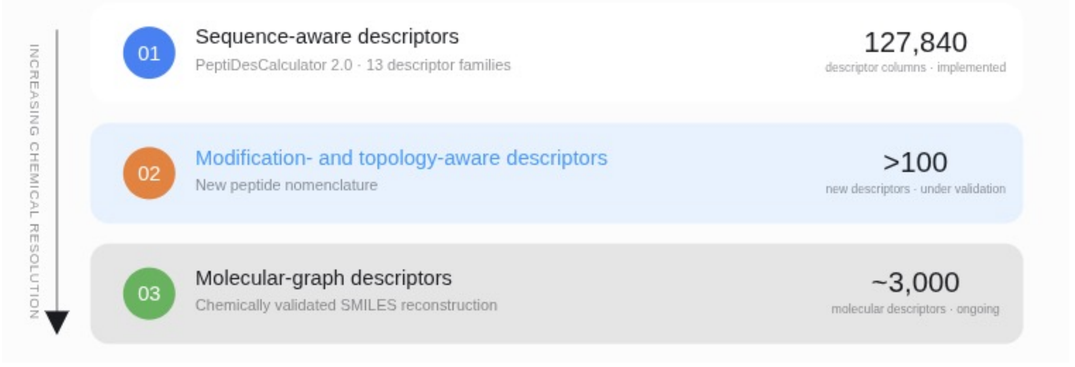
Only approximately 16% of AAindex-scale and operator pairs showed strong or near-duplicate correlations. This supports retaining broad descriptor coverage during calculation, while dataset-specific feature selection is performed later within each QSAR workflow.

Web server ProtoPEP

PeptiDesCalculator 2.0 is integrated into **ProtoPEP**, a web-based platform that calculates the selected peptide descriptors and exports the resulting machine learning ready matrix as a downloadable CSV file.



Ongoing work & conclusions



- PeptiDesCalculator 2.0 offers 13 modular descriptor families and ML-ready CSV export.
- Combined descriptors capture meaningful peptide-class variation.
- D/L and correlation analyses confirm predictable, complementary behavior.
- The new nomenclature adds modification-, topology- and SMILES-based molecular descriptors that can be employed for QSAR modelling.

References

1. L. Wang, N. Wang, W. Zhang, X. Cheng, Z. Yan, Y. Shao, G. Wang, X. Wang, R. Fu, P. Liu, Y. Wang, Q. Lei, Z. Guo and C. Fu, Signal Transduct. Target. Ther. 2022, 7, 48.
2. B. Zheng, Y. Chen, J. Li, H. Zhang and X. Liu, Int. J. Mol. Sci. 2025, 26, 5123-5148.
3. S. J. Barigye, S. Gómez-Ganau, E. Serrano-Candelas and R. Gozalbes, Proteins 2021, 89, 174-184.