



**HAL**  
open science

## Leveraging Zernike polynomials for peptide conformer identification

Diego Amaya-Ramirez, Bernard Maigret, Etienne Corman, Jean-Luc Taupin,  
Marie-Dominique Devignes

► **To cite this version:**

Diego Amaya-Ramirez, Bernard Maigret, Etienne Corman, Jean-Luc Taupin, Marie-Dominique Devignes. Leveraging Zernike polynomials for peptide conformer identification. GGMM SFCi, Sep 2021, Villeneuve d'Ascq, France. . hal-04578212

**HAL Id: hal-04578212**

**<https://hal.science/hal-04578212>**

Submitted on 16 May 2024

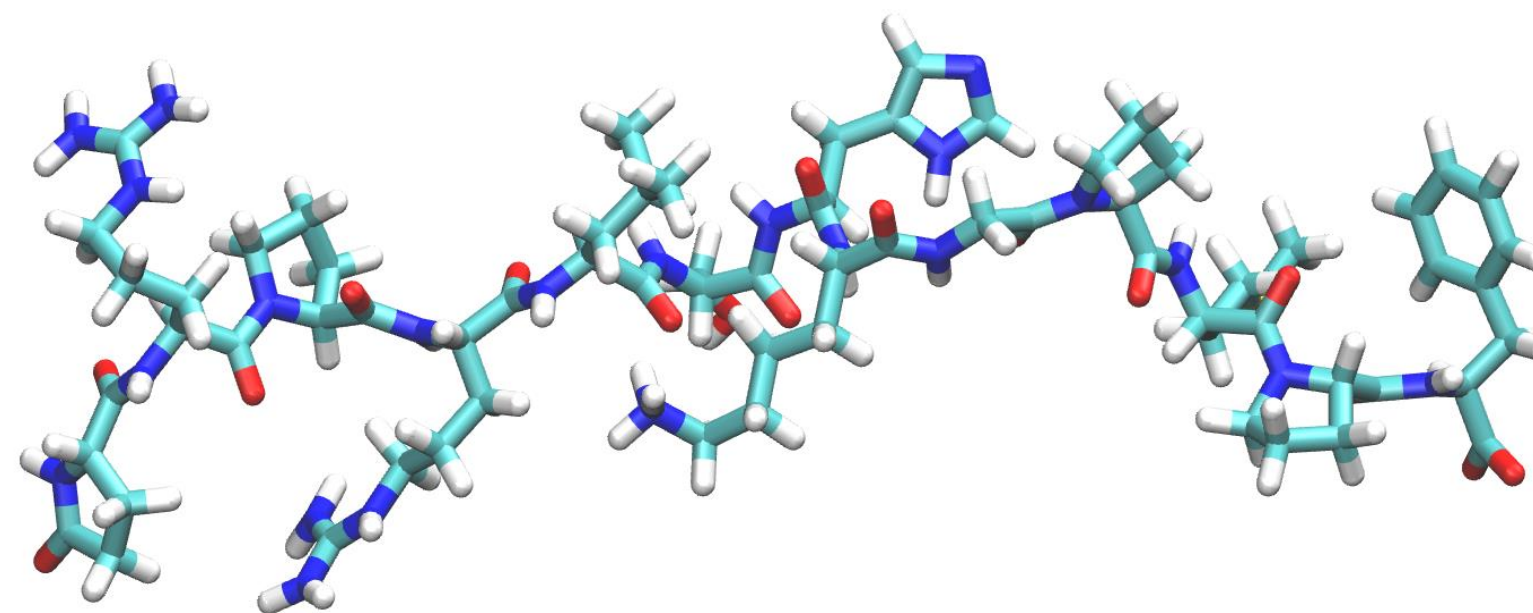
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.

## Introduction

Peptides play a key role in multiple biological processes. For this reason, peptides have gained relevance in recent years as pharmacological agents<sup>1</sup>. However, peptides are highly flexible leading to diverse three-dimensional (3D) conformers that can modify their specificity and efficacy in binding to their targets. Identifying and understanding such 3D-conformers is therefore fundamental. This can be achieved by analysis of molecular dynamics (MD) runs. However, classical methods such as RMSD maps are difficult to process automatically and manual interpretation can be subjective.

## Materials and Methods



Peptide [Pyr1]apelin-13

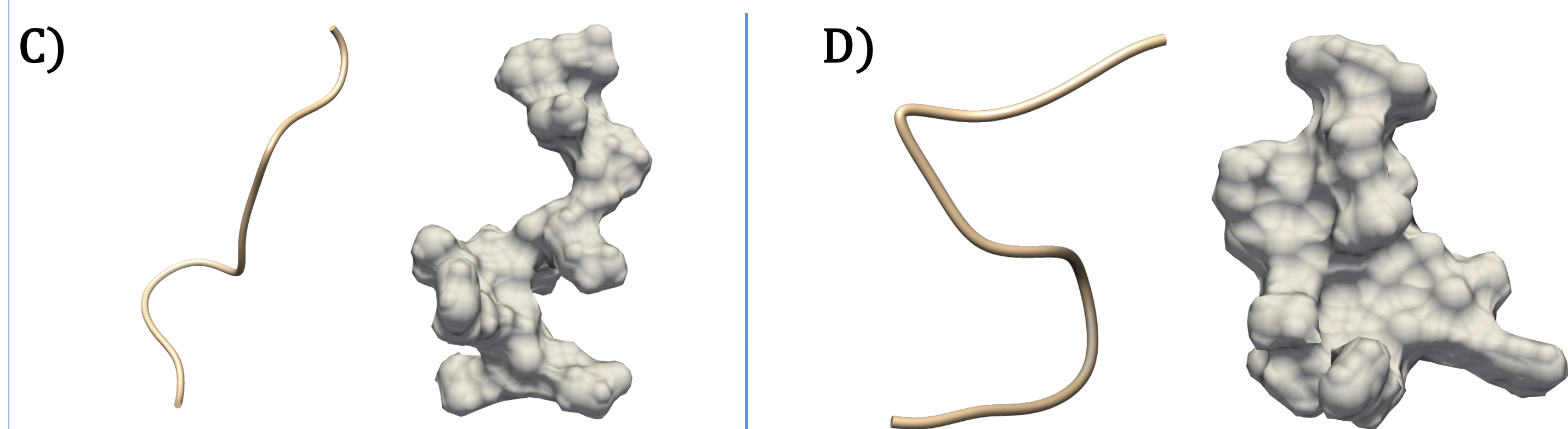
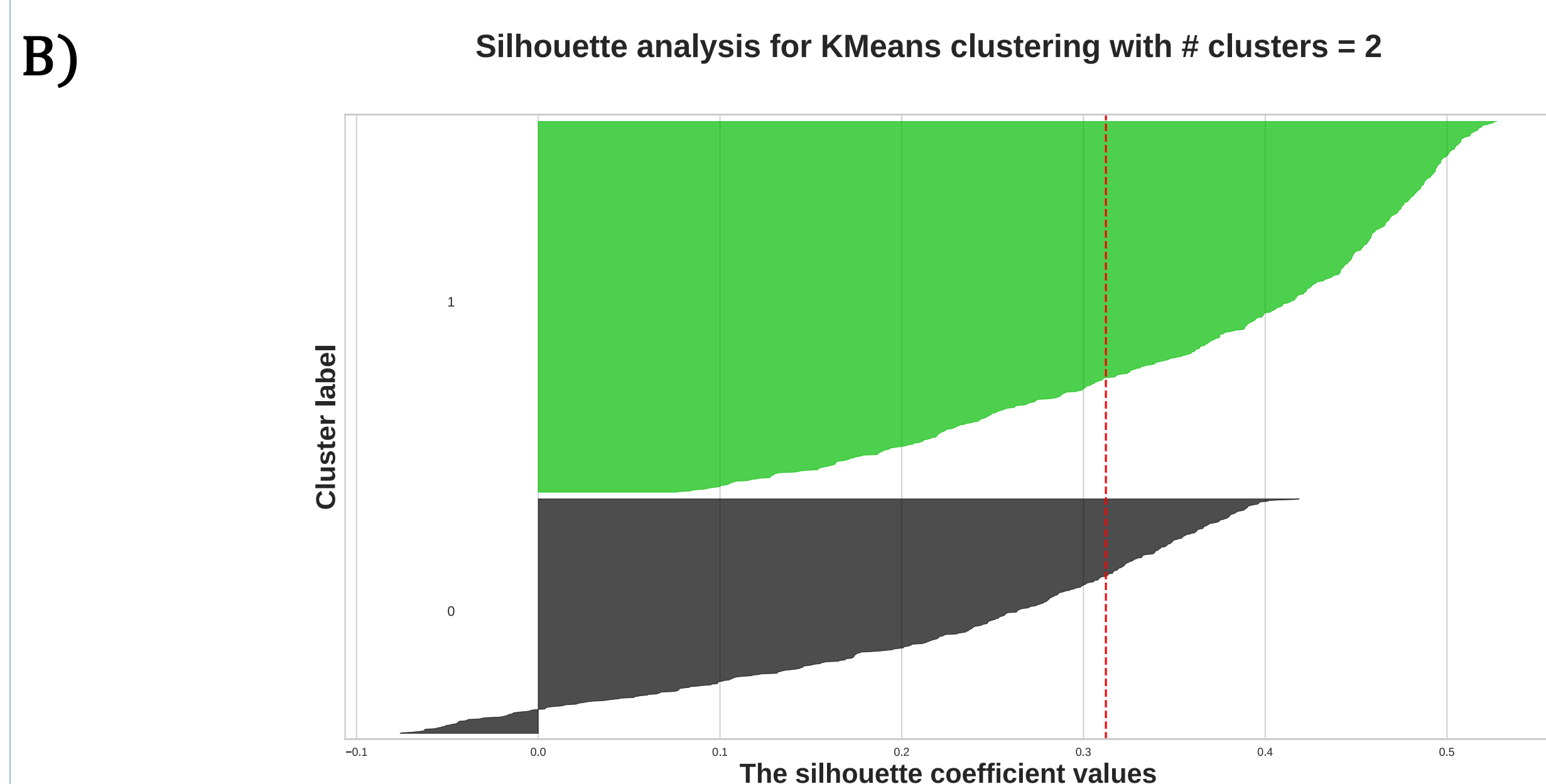
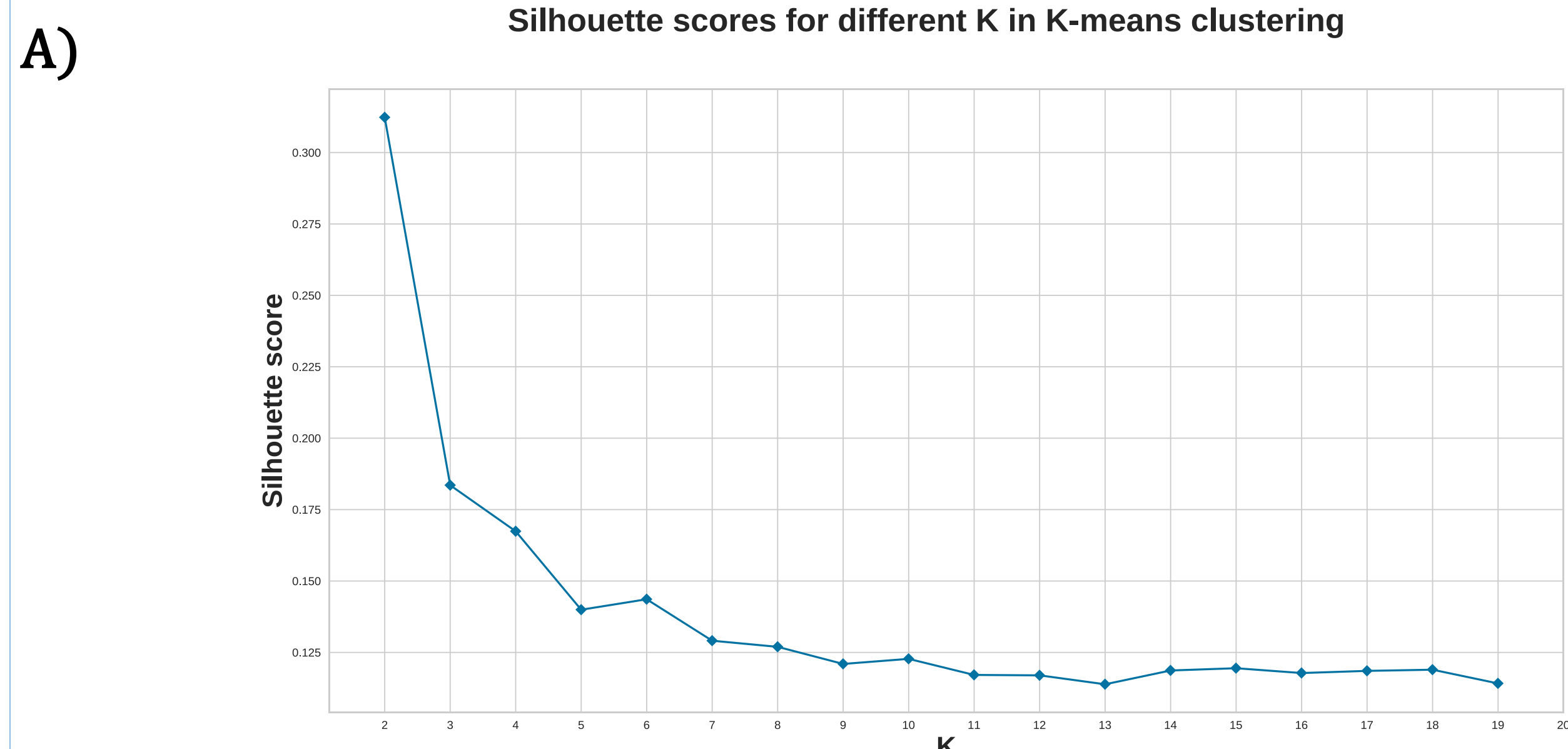
- Peptide [Pyr1]apelin-13<sup>2</sup> of sequence : pE-R-P-R-L-S-H-K-G-P-M-P-F.
- 1000 frames from a MD simulation of 100ns : using NAMD<sup>3</sup>, CHARMM36<sup>4</sup> force field and TIP3P (Transferable Intermolecular Potential with 3 Points) explicit solvent model.

For each frame, a 3D Zernike descriptor<sup>5,6</sup> is used to represent the 3D shape of the peptide as a vector of invariants representing a series of orthogonal polynomials. The 3D shape is the solvent-accessible surface, constructed from the Van der Waals radius sphere associated with each atom type and obtained using Chimera<sup>7</sup>. The MindBoggle<sup>8</sup> python library is then adapted to calculate the 3D Zernike descriptor relying on 121 invariants.

Conformer identification is performed by clustering the previously calculated 3D Zernike descriptors, using the K-means and silhouette analysis algorithms to identify the optimal number of clusters. For each cluster, the conformation closest to its centroid is chosen as a representative.

## Results

According to silhouette analysis, two main clusters are discernable. The representative conformers correspond to frames 748 and 840.



A) Silhouette scores for different number of clusters. B) Silhouette analysis for K-means clustering of 3D Zernike descriptors. The dashed line corresponds to the average silhouette score of 0,312. C) Backbone (on the left) and surface (on the right) visualization of the representative conformer for the Cluster #0 (frame 840). D) Backbone (on the left) and surface (on the right) visualization of the representative conformer for the Cluster #1 (frame 748).

## Conclusion

The 3D Zernike descriptors have already been used for protein tertiary structure comparison but to our knowledge, not to identify 3D-conformers from MD runs. The Zernike vector format allows using a variety of clustering methods to process MD frames and to automatically determine the optimal number of clusters/conformers. This strategy could be promising for other biomolecules such as proteins or nucleic acids.

## Acknowledgements

Diego Amaya-Ramirez is recipient of an Inria-Inserm doctoral fellowship.

Experiments presented in this poster were carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see <https://www.grid5000.fr>).

Experiments presented in this poster were carried out using the MBI-DS4H platform hosted by Inria/Loria funded by CPER IT2MP (Contrat Plan État Région, Innovations, Technologiques, Modélisation & Médecine Personnalisée, including a FEDER co-funding).

## References

- (1) Salmaso, V. *et al.* Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach. *Structure* 2017, 25 (4), 655-662.e2.
- (2) Yang, P. *et al.* [Pyr1]Apelin-13(1-12) Is a Biologically Active ACE2 Metabolite of the Endogenous Cardiovascular Peptide [Pyr1]Apelin-13. *Front. Neurosci.* 2017, 11.
- (3) Phillips, J. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics*, 153:044130, 2020.
- (4) Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* 2017 Jan;14(1):71-73.
- (5) Novotni, M. *et al.* Shape retrieval using 3D Zernike descriptors. *Computer-Aided Design.* 2004, 36, 1047-1062.
- (6) Di Rienzo, L. *et al.* Superposition-Free Comparison and Clustering of Antibody Binding Sites: Implications for the Prediction of the Nature of Their Antigen. *Sci. Rep.* 2017, 7 (1), 45053.
- (7) Pettersen, E. *et al.* UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 2004, 25 (13), 1605-1612.
- (8) Klein, A. *et al.* Mindboggling Morphometry of Human Brains. *PLOS Comput. Biol.* 2017, 13.