

Unsupervised

Theory (PBE+VdW).

## Mapping and classifying molecules from a high-throughput structural database



Felix Musil<sup>1,2</sup>, Sandip De<sup>1,2</sup>, Michele Ceriotti<sup>1,2</sup>

 National Center for Computational Design and Discovery of Novel Materials (MARVEL), Lausanne, Switzerland.
Laboratory of Computational Science and Modelling, Institute of Materials, Ecole Polytechnique Fédérale de Lausanne, Lausanne ne. Switzerland.



## Sketch-map Representation

machine

Sketchmaps [5] are two dimentional projections based on the structural similarity between conformers of the dataset, i.e. each point represents a conformer and their relative distance is directly related to their similarity.



The protonated lysine conformers are color-coded on both sketchmaps and dendrogram according to their conformational energy, the minimal distance between the oxygen atoms of the backbone and the nitrogen atom of the tail,  $D_{ON}$ , and the backbone dihedral angles  $\phi$ ,  $\psi$ ,  $\omega_1$  and  $\omega_2$ .

Protonated lysine conformers, perturbed by Ca<sup>2+</sup> and colored according to their stability, are projected on the sketchmap of the pure molecule shown in grey color as a reference





## **Physical Insights**

Protonated lysine conformers across this dataset are forming structurally-related groups. Investigating the representative structures of these clusters allow for a quick identification of the structural motifs explaining the layout of the conformational landscape:

• The dihedral angles  $\omega_1$  and  $\omega_2$  forming the four combinations of cis and trans states (see representative structures a-d).

• The distance D<sub>ON</sub> is directly related to the formation of a H-bond between the tail and the backbone of the amino acid (see representative structures b and c).

• The energetics of the conformers is mainly driven by the extended or bent state of the tail. The perturbed system is projected on a subset of the original conformational space showing that the  $Ca^{2+}$  ion preferably binds to the oxygen atoms of the backbone resulting in extended conformers. An analysis only focused on the Ramachandran dihedrals,  $\phi$  and  $\psi$ , would have missed one of the main features of the structural landscape that is critical to characterize the relation between structure and energetics.

## References

- aldauf, and M. Ceriotti, "Mapping and classifying n .m, "Berlin ab initio amino acid DB," 2016. [Online
- v. Bidni, Berni ab initio animo acid DB, 2010. [c nd G. Csányi, "On representing chemical environme ányi, and M. Ceriotti, "Comparing molecules and so and M. Parrinello, "Simplifying the representation of "Simplifying the representation of the second second
- ving molecules from a high-throughput structural database," J. Cheminform., vol. 9, no. 1, p. 6, Dec. 2017. Jnlinel. Available: http://aminoaciddb.rz-berlin.mgg.de/. mix, "Phys. Rev. Ds. Condens. Matter Mater. Phys., vol. 87, no. 18, p. 184115, May 2013. lids across structural and alchemical space," Phys. Chem. Chem. Phys., vol. 18, no. 20, p. 13754, 2016. of complex free-energy landscapes using sketch-map," Proc. Natl. Acad. Sci. U. S. A., vol. 108, no. 32, pp. 13023–8, Aug. 2011. Available: http://sketch



