# Exploring Esterase Diversity with Shape Descriptors

**Alena Endres[1], Pablo Cea Medina[1], Holger Gohlke[1,2]**

[1]Institute of Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-Universität Düsseldorf

[2]Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich

Esterases are hydrolytic enzymes that break ester bonds, releasing an alcohol and an acid. Their high chemo-, regio- and stereoselectivity makes them attractive for the chemoenzymatic synthesis of fine chemicals. Moreover, some esterases can cleave synthetic polymers such as polyethylene terephthalate (PET) enabling green biodegradation of plastic waste[1].

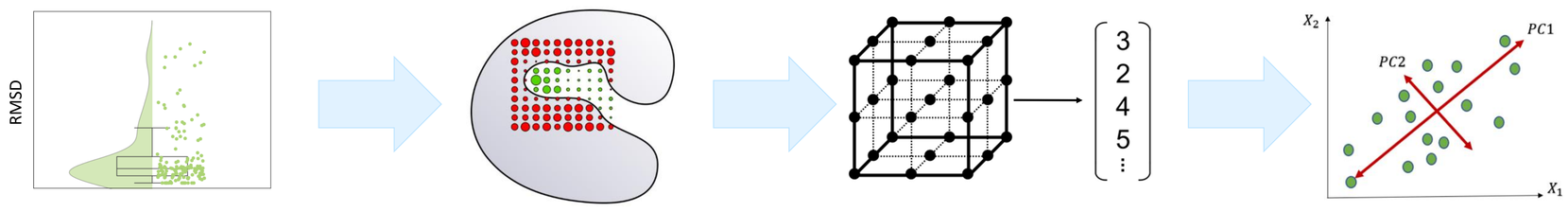
(Project-ID: lipases)

**Aim**

Development of predictive tools to search for novel biotechnologically relevant esterases

**Workflow**

Alignment to Reference Triad → Calculate Drugscore Grids → Calculate Zernike Descriptors → Dimensionality Reduction

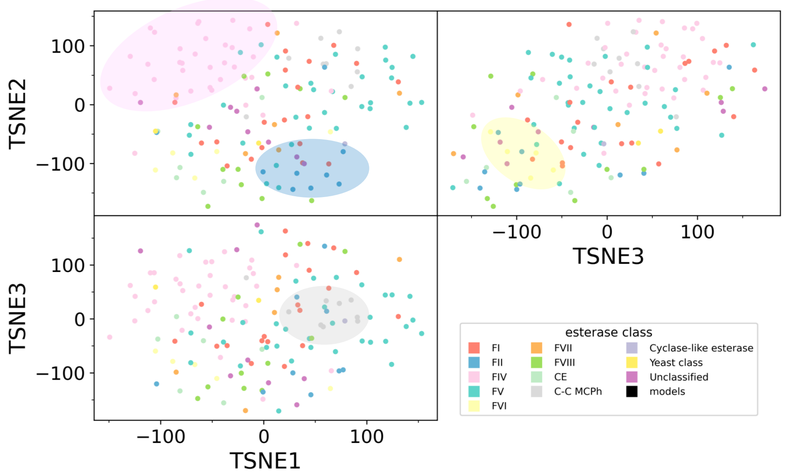



**Key concept**

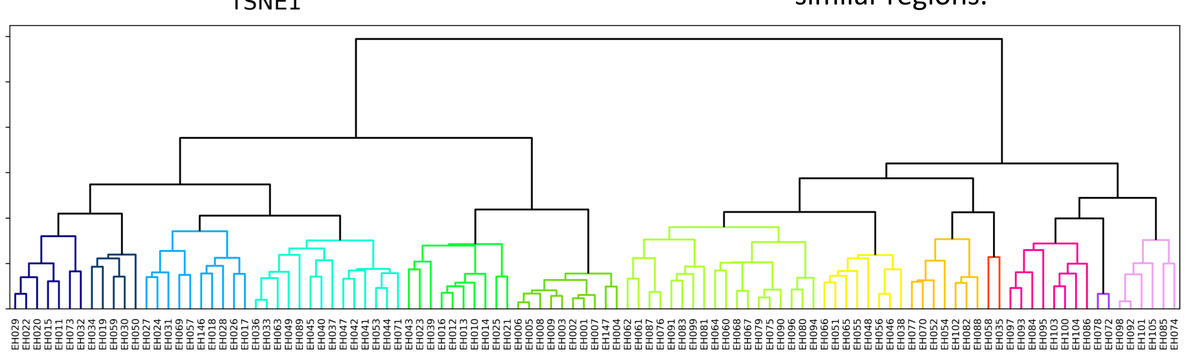
ZDs are large 1-dimensional vectors, obtained by applying a Zernike polynomial expansion on 3D molecular interaction fields[2]

## Using Zernike Descriptors to predict substrate specificity

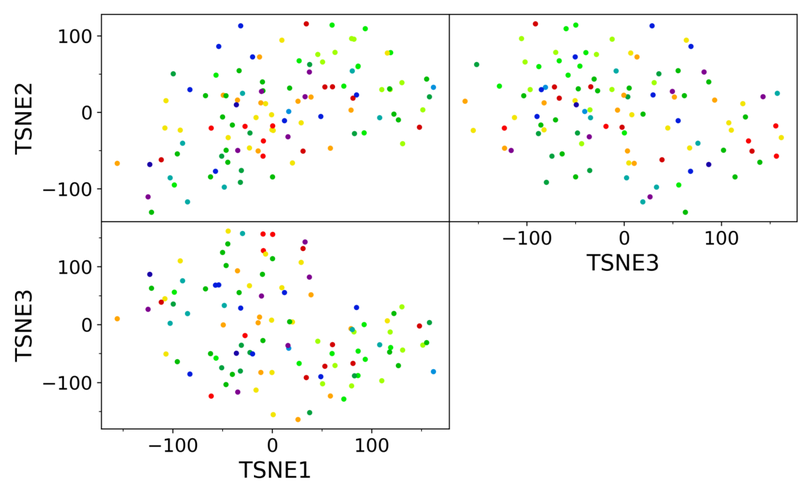
To develop a tool to predict substrate specificity we used a dataset consisting of 147 experimentally characterized esterases against 96 different substrates[3]. Using the drugscore interaction fields allows to investigate the relevance of the physicochemical properties and characteristics of the active site to substrate specificity.



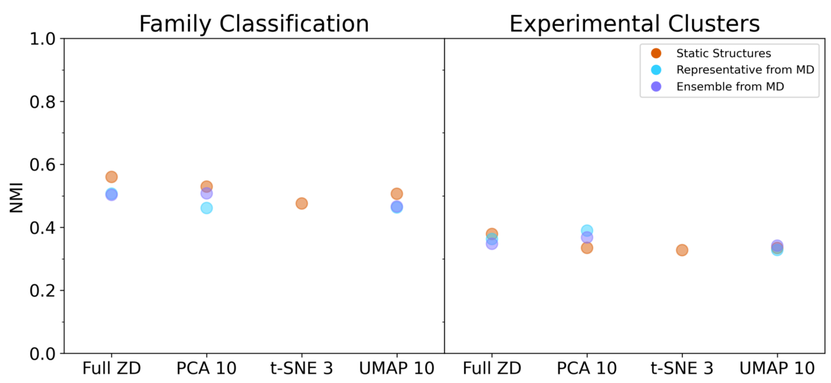

After MD simulations, an ensemble of structures was used to calculate the ZDs. The data is colored according to the sequence-based Jaeger-Arpigny classification[4]. The plot shows, highlighted by the colored ellipses, that esterases from the same class cluster in similar regions.



To assess whether Clusters of the high dimensional space based on ZD match the activity profiles of the whole dataset, we performed hierarchical clustering of the experimental data, grouping by the activity on chemically similar substrates.

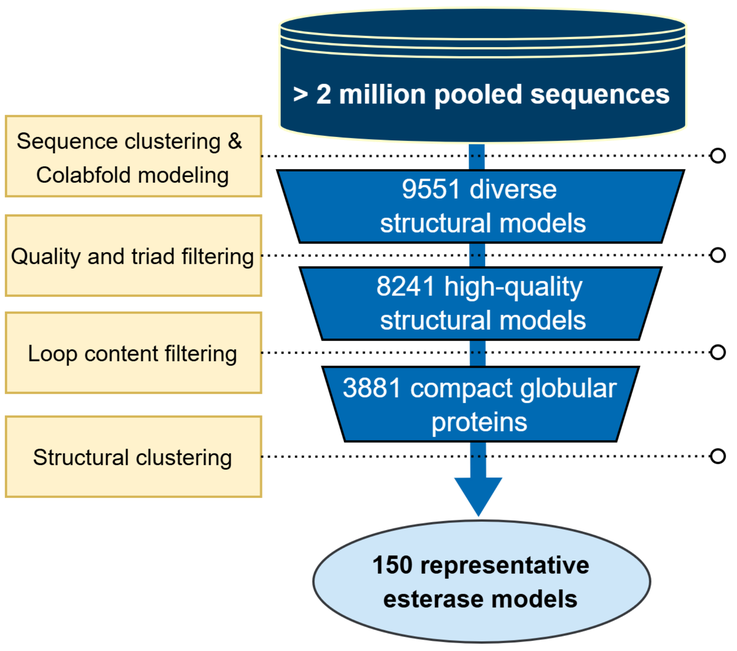



Mapping the substrate-based clusters onto the same ZD derived projection space revealed that esterases with similar substrate preference profiles do not group on the same region of the space.

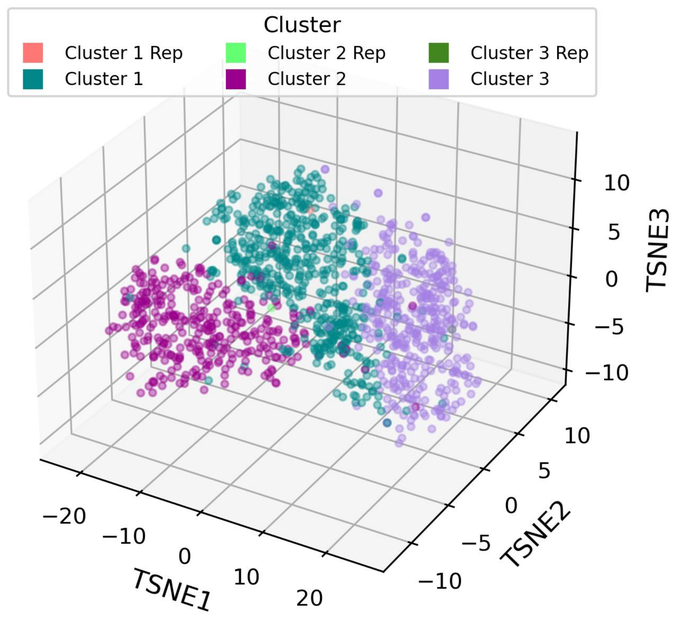



To quantify the deviation between the clusters we calculated the normalized mutual information (NMI) scores. The results show low NMI scores. This means the ZDs alone might not be able to effectively discriminate among the different types of esterases. Future research will pursue the search for a classifier, possibly using ZDs as part of a combined classification.

## Using Zernike Descriptors to identify novel Esterases

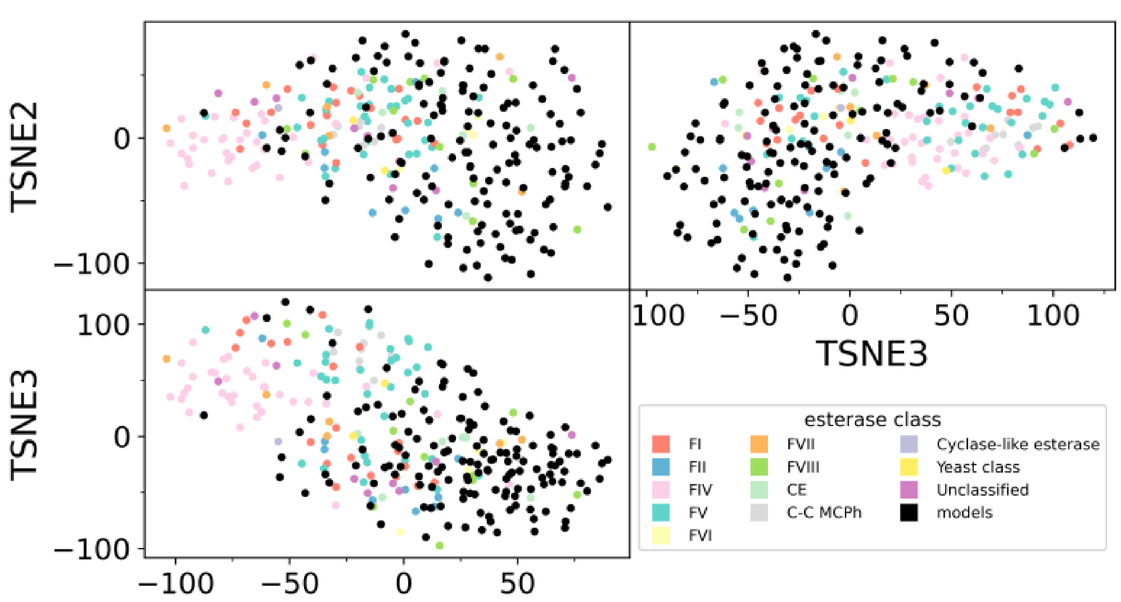



Since ZDs could be used as a tool to distinguish active sites grouped according to the Jagger-Arpigny classification scheme, they can be used to inspect the esterase space. Therefore, we constructed a larger dataset that encompasses the entirety of enzymes with putative esterase activity from the UniProtKB database.




To assess whether the models within one cluster defined by FoldSeek are not only similar in terms of the fold but also have similar ZDs, we calculated the ZDs for all the members of the three biggest clusters.

The FoldSeek clusters do form well-defined groups after dimensionality reduction and the representative lies in the center of each cluster.




Most of the ColabFold models are within the projection space of the experimental esterases. Around the borders are a few model esterases a little further away from any experimental esterase, indicating a less explored area. This could lead to the identification of a novel esterase with a novel substrate spectrum.

**CITATIONS**

[1] D. Danso, J. Chow, W.R. Streit, Appl Environ Microbiol, 85 (2019).

[2] B. Nisius, H. Gohlke, J Chem Inf Model, 52 (2012)

[3] M. Martinez-Martinez, C. Coscolin, G. Santiago, J. Chow, et al., ACS Chem Biol, 13 (2018)

[4] J. L. Arpigny, K.-E. Jaeger, Biochemical J. 343 (1999) .

**ACKNOWLEDGEMENTS**

We gratefully acknowledge the support by the German Federal Ministry od Education and Research (BMBF) through funding number 031B1342A “LipoBiocat” and the computing time granted by the John von Neumann Institute for Computing (NIC).

hhu.de