Towards automatic derivation of geometry-based descriptors as surrogates for complex structural approaches in enzyme-substrate prediction

Our main objective is to develop a methodology capable of correctly modeling reaction mechanism signatures of enzyme families using simple geometric descriptors and chemical informatics. To achieve this, we hypothesize the following:

- An extensive docking experiment can sample acceptable binding poses for an enzyme-substrate complex (ES-complex).
- Different substrates employed by a particular Enzyme Commission number (EC) class share a common chemical substructure (referred from now on as **reactive fragment)**, and such substructure is essential for the chemical reaction.
- Structural information obtained from docking experiments can be used to derive geometry-based descriptors.
- Simple geometric descriptors obtained between atoms of the reactive fragment and the catalytic residues of an enzyme can be used to discriminate if a new chemical compound can act as a substrate or not in an enzymatic reaction.

With this in mind, we will gather information about known ES-complexes from public databases and the scientific literature for three EC classes. Then, we will perform intensive and detailed docking experiments with these complexes to generate a structural database. In parallel, for each EC class, we will identify characteristic chemical groups among its known substrates and refer to them as reactive fragments, hence, all substrates in an EC class must contain within the reactive fragment. Following, the reactive fragment selected and the docked conformations will be used to measure geometric descriptors between the enzyme and the substrate. We will group the docked conformations depending of the similarity of its geometrical descriptors. Next, we will analyze all the produced groups to determine which ones could be used as a geometrical signature for a reaction mechanism, we will refer to the approved groups as candidates. The selected candidates will be tested by cross-validation to address their validity, and the approved candidates will be considered as filters for a reaction mechanism. Finally, we will perform Virtual Screening of new chemical compounds harboring the reactive fragment and test experimentally the top hits as a final validation to our protocol.

As result, we expect to obtain a methodology which will be able to obtain simple geometric rules to characterize a complex enzymatic reaction mechanism. We believe that our methodology will allow researchers to predict substrates for a defined reaction mechanism in a simple way, without the need to use more advanced techniques as Quantitative Structure Activity Relationship or Quantum Mechanics / Molecular Mechanics. This will result in a considerable speed-up and a reduction of costs in the substrate prediction process. Furthermore, we believe that the methodology developed within this project could help in the elucidation of unknown reaction mechanisms of known ES-complexes.