BONAFIDE: a Python package for calculating features for atoms and bonds in molecules

Lukas M. Sigmund[1], Michele Assante[2,3], Matthew Ball[1,4], Mikhail Kabeshov[1]

[1] *Molecular AI, Discovery Sciences, R&D, AstraZeneca Gothenburg, Pepparedsleden 1, 43183 Mölndal, Sweden.*
[2] *Innovation Centre in Digital Molecular Technologies, Department of Chemistry, University of Cambridge, Lensfield Rd, CB2 1EW Cambridge, United Kingdom.*
[3] *Compound Synthesis & Management, The Discovery Centre, Cambridge Biomedical Campus, 1 Francis Crick Avenue, AstraZeneca Cambridge, CB2 0AA Cambridge, United Kingdom.*
[4] *Laboratory of Cheminformatics, University of Strasbourg, 4 rue B. Pascal, 67081 Strasbourg, France.*

Molecular descriptors are essential for all cheminformatics projects, including the training of ML models. Several general software packages have been developed that facilitate the mapping between a molecular entity and a given descriptor, thereby granting practitioners straightforward access to hundreds of values per molecule. While global descriptors are very important, there are many scenarios in chemistry in which local information, that is, **descriptors for atoms and bonds**, is indispensable. For example, the modeling of regio- and site-selectivity requires local features. In the context of medicinal chemistry, site-of-metabolization predictions or the precise description of protein-ligand interactions can be based on local atom and bond descriptors.

Several separate software tools have been developed for the calculation of specific atom and bond features. However, a consistent interface to the individual featurization tools does not exist, which hampers easy access to respective features. **Therefore, we have developed the Bond and Atom Featurizer and Descriptor Extractor (BONAFIDE). It is a Python package that provides a uniform API to 11 featurization dependencies combined with from-scratch feature implementations, in total spanning 460 atom and 120 bond descriptors**. Data can be obtained from SMILES strings (2D representation) as well as from XYZ or SD files (3D representations). The package ensures consistent atom and bond indexing across different data formats and facilitates processing steps such as the identification of specific functional groups or the analysis of molecular symmetry. Moreover, BONAFIDE interfaces with quantum chemistry engines to extract electronic-structure-based features; previously generated electronic structure data can also be used. Lastly, the feature output format can be widely customized, and users can implement custom featurizers in a straightforward manner. BONAFIDE includes detailed online documentation, and the package's applicability is demonstrated with two case studies.

Web page: https://molecularai.github.io/atom-bond-featurizer/