Contribution ID: **15**   Type: **Poster**

TopEnzyme significantly improves classification of enzyme function by incorporating computational enzyme structures.

Tuesday 20 February 2024 10:45 (1 hour)

It is essential to understand target enzyme function for applications in biomedicine and biotechnology. A good method to predict the function of new enzymes is the classification through neural networks in combination with large structural datasets. To keep computational requirements feasible for these large systems, we created a more sophisticated representation of an enzyme than the sequence or fold while retaining local chemical information. This localized 3D enzyme descriptor improves enzyme function prediction compared to established methods in the field for GCNs.

For this project we developed TopEnzyme, a database of structural enzyme models created with TopModel and it is linked to the SWISS-MODEL and AlphaFold Protein Structure Database to provide an overview of structural coverage of the functional enzyme space for over 200,000 enzyme models. It allows the user to quickly obtain representative structural models for 60% of all known enzyme functions. We assessed the models we contributed with TopScore and found that the TopScore differs only by 0.04 on average in favor of AlphaFold2 models. We tested TopModel and AlphaFold2 for targets not seen in the respective training databases and found that both methods create qualitatively similar structures.

Testing the localized 3D descriptor on this database improves the F1-score up to 17% in enzyme classification tasks compared to fold representation methods. Furthermore, we implemented better GCNs, SchNett and DimeNetPP, for atom classification. This increases the performance by 13% and 16% on the enzyme classification task. Furthermore, we investigated the networks using GNNExplainer and found relational information more important when classifying residue-based objects, while chemical interactions are marked more important when classifying on atom objects. Our results demonstrate that a localized 3D descriptor is the better alternative to current reduced structure representations used in enzyme prediction networks.

Consent

Yes

Authors

van der Weg, Karel[a], Merdivan, Erinc[b], Piraud, Marie[b], Gohlke, Holger[a, c]

Affiliation

[a]Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany [b]Helmholtz AI Central Unit, Ingolstädter Landstraße 1, 85764 Oberschleißheim, Germany [c] Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

Primary author: VAN DER WEG, Karel (Forschungszentrum Juelich)

Co-authors: MERDIVAN, Erinc; GOHLKE, Holger (IBG-4, Forschungszentrum Jülich GmbH); PIRAUD, Marie (H.AI/ HMGU)

Presenter: VAN DER WEG, Karel (Forschungszentrum Juelich)

Session Classification: Poster session