

Checkmyblob - Machine Learning-Based Tool for Ligand Recognition and Validation

Joanna Lenkiewicz¹, Dariusz Brzezinski², Wlodek Minor¹
¹University of Virginia, ²University of Virginia, Poznan University of Technology
joanna_len@iwonka.med.virginia.edu

CheckMyBlob (CMB)^{1,2} is a software tool designed to automatically recognize and validate small molecule ligands in electron density maps of macromolecular structures. The new version of CMB uses the existing machine-learning algorithm. However, the training set was enhanced with data from the PDB-REDO server³. CMB's capabilities were increased by focusing on specific tasks. First, we have improved prediction accuracy for metal-containing ligands by developing a separate model with a custom training set.

Additionally, in case of finding metals, the CMB website redirects to the CheckMyMetal (CMM)⁴ server to automatically show the metal validation and allow the usage of CMM standard functions. Second, CMB has been trained on ligand sets corresponding to drug screening cocktails to include functionality that can suggest potential drugs. Finally, several run examples have been added to demonstrate the tool's power and help the researchers understand how to use it effectively.

References:

- {1} Kowiel M., Brzezinski D., Porebski P.J., Shabalin I.G., Jaskolski M., Minor W. Automatic recognition of ligands in electron density by machine learning. *Bioinformatics*. 2019; 35:452–461.
- {2} Brzezinski D, Porebski PJ, Kowiel M, Macnar JM, Minor W. Recognizing and validating ligands with CheckMyBlob. *Nucleic Acids Res*. 2021; 49: W86-W92.
- {3} Joosten, R.P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A.C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, {4} A.L., Deleage, G., Diarena, M., Fabbretti, R., Fettahi, G., Flegel, V., Gisel, A., Kasam, V., Kervinen, T., Korpelainen, E., Mattila, {5} K., Pagni, M., Reichstadt, M., Breton, V., Tickle, I.J., Vriend, G., and Roversi, P. PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J. Appl. Crystallogr*. 2009; 42, 376-384
- {6} Gucwa M, Lenkiewicz J, Zheng H, Cymborowski M, Cooper DR, Murzyn K, Minor W CMM-An enhanced platform for interactive validation of metal binding sites. *Protein Sci*. 2023; 32: e4525.