Please find here a brief description of submitted SI files.

**SI\_BCFtrainset.xlsx** and **SI\_BCFtestset.xlsx**

The Training and the Test set of the model. SMILES, CAS and Experimental BCFs are reported.

**SI\_ExcludedOutliers34.csv**

List of 34 excluded outliers from the training set. SMILES, CAS, Experimental BCF and model’s predicted BCF are reported. In addition, next to each outlier its closest analogue (Tanimoto similarity) is reported with its experimental BCF.

**SI\_FullDataset\_rawdata.csv**

The whole collected raw data from the mined databases. For all compounds, CAS, SMILES and experimental BCF are reported. When available, additional information are stated (e.g. literature reference, additional test conditions, species, …)