ToBvalid version 0.9.2

ToBvalid is a Python library and a program for the statistical analysis and validation of Macromolecular ADPs. It comes under MPL-2.0 license and supports Python3.

Installation

In order to install ToBvalid from source:

pip install tobvalid

Github repository (*https://github.com/ToBvalid/)* also may be used:

pip install git+https://github.com/ToBvalid/tobvalid

Dependencies

ToBvalid has dependencies on **pandas, fire, matplotlib, numpy, scipy, gemmi>=0.3.8, seaborn, statsmodels**. All dependencies will be automatically checked and installed, if necessary, during ToBvalid installation.

ToBvalid functionalities

This tool is designed for modelling of ADP distribution and their validation on both global and local levels. Internal functionalities of ToBvalid include:

- Overall statistical analysis of ADP distribution.
- Parameterisation of ADP distribution (mixture) and validation of parameters.
- Search for potentially lighter and heavier than neighbouring atoms.
- Comparison of the relative occupancy of two atoms. (only in the library, not implemented in the program yet.)
- Validation of ligands.

Usage

ToBvalid reads pdb files and implements the validation of ADPs.

It is strongly advised to re-refine the structure with any refinement software before the validation.

The input file is pdb and output files are html reports, graphs and text files with outliers' lists of both local and global analysis.

tobevalid -i {Input file} -o {output directory} -m {number of modes|auto} -t {tolerance} -hr {plot resolution in dpi} -a {'all'|'local'|'global'}

Default values:

- o - current directory
- m - 1
- t - 1e-5
- a - all

Sometimes ADP distribution may be multimodal (as a mixture of Shifted Inverse Gamma Distributions (SIGD)). The default number of modes is one. If the number of modes is needed to be defined use *-m auto*.
-hr is needed to generate high quality plots for publications.
If only local/global analysis is needed, instead of *-a all* here should be used *-a local* or *-a global* respectively.

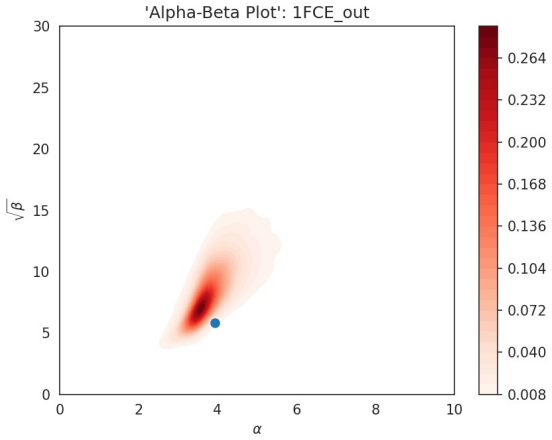
The details of the input may be printed with *help:*

tobvalid -h

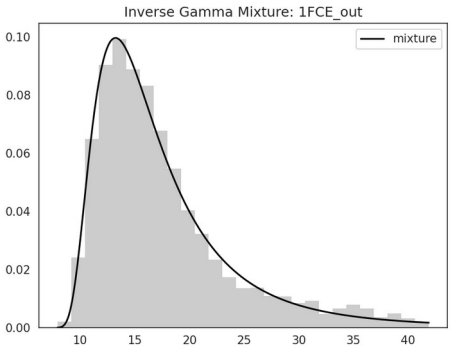
Outputs:

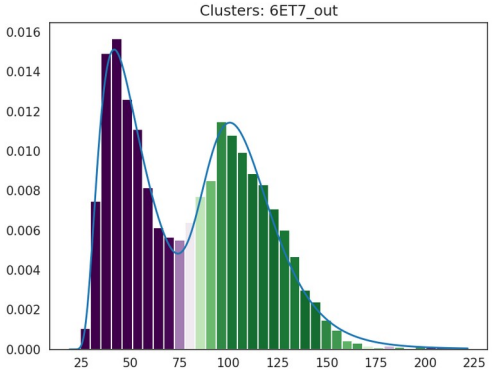
Outputs of the global analysis are listed below:

1. The results of the global analysis and overall statistics reports are given as html files.
2. alpha-beta plot. SIGD parameters (α, β) are placed on the contour plot. If these parameters are out of the contour this structure should be considered for rebuilding and re-refinement.

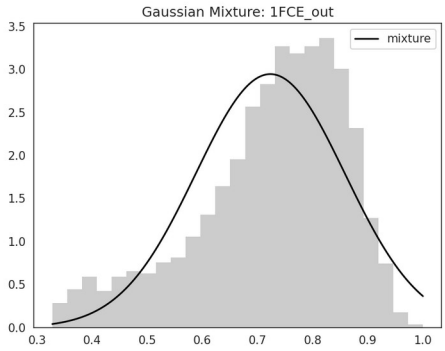


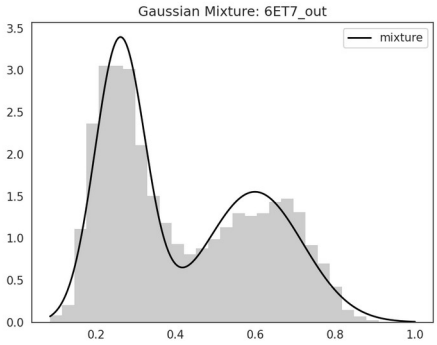
3. SIGD (mixture) plot. If there is one mode in the ADP distribution there is just one plot, but for the multimodal cases the plot of clusters is also given.



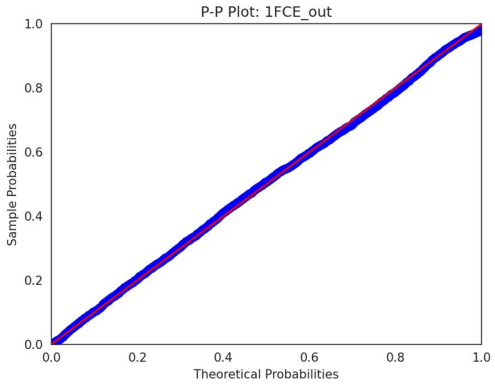


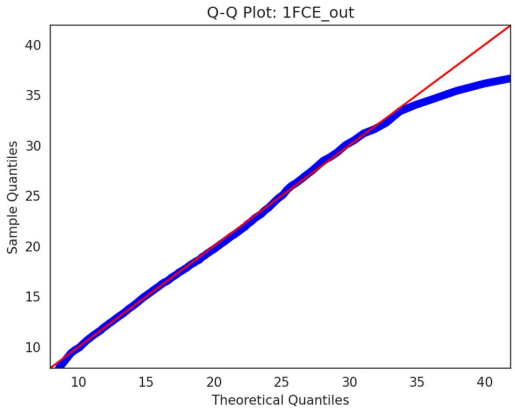
4. Peak height distribution plot. Here is the relation between B values and the resolution. In this plot several atoms in the very left/right may be the sign of light and heavy atoms respectively.





5. qq and pp plots illustrate the agreement the ADP distribution of given structure to its theoretical probabilities.





6. List of Interquartile outliers. Very small/large values of ADPs are not only considered suspicious, but also affect the parametrization. The program removes outliers before global analysis.

Outputs of the local analysis are listed below:

1. File named *{filename}_local.txt* lists the report of the overall local analysis. In this validation each atom is checked with consideration of its environment (neighbouring atoms). In this file, potential lighter/heavier atoms are listed with the calculated optimal occupancy and also with the list of neighbouring atoms and the basic statistics for the environment.
2. File named *{filename}_water.txt* lists water molecules with the number of neighbouring atom 6 and more.
3. File named *{filename}_ligand.txt* lists the results of the local analysis for each ligand.
4. *ligands_validation.txt* provides the list of suspected ligands.