1. PROGRAM AND DATA DEPOSITION

For the **original AnalogExplorer** program and **AnalogExplorer2**, two subfolders are deposited, respectively.

1.1. Under "**Programs**":

- Three batch and jar files are provided for different applications including the analysis of multiple analog series from a given compound set, analysis of an individual series, and selectivity analysis, as detailed below.
- "lib" contains jar files of external libraries, in which **the OpenEye OEChem library should also be added here by the user**.

1.2. Under "**SampleSets**":

- All compound sets analyzed and reported in the original publication of AnalogExplorer [1] and AnalogExplorer2 are deposited in respective directory.

2. REQUIRED LIBRARY AND INPUT FILES

2.1. The OpenEye OEChem library should be put into the subfolder "lib".

- In our study, OEChem version 2.0.2 is used. The corresponding jar file is "oejava-2014.Oct.2-Windows-x64.jar", which requires a license. Therefore, the library was not provided in the deposition. The user has to download the file and obtain the license from OpenEye.

2.2. SD file is the data format for input files.

- In an SD file, two fields must be included, i.e.,
  - **Compound index**: e.g., "CmpdID", "CpdID", "CompoundName", etc.
  - **Compound potency**: e.g., "pK_i", "pIC_{50}", etc.

3. HOW TO RUN AND APPLY THE PROGRAM

3.1. Depending on various applications, different batch files are selected.

3.1.1. For the analysis of a **single series** of analogs active against a given target, "**AnalogExplorer_IndividualSeries.bat**" or "**AnalogExplorer2_IndividualSeries.bat**" file is chosen.

- A pop-up panel is shown by double clicking the batch file.
- **Program settings**:

- Select an SD file: e.g., "SingleSeries_AndrogenReceptor.sdf" under "\AnalogExplorer_Original\SampleSets\".

- Select the tag for compound indices

- Select the tag for compound potency

- Select "Complete graph" if the user wants to view the complete graph. Otherwise, the reduced graph will be generated.

- Select the minimal size of the scaffold: default value is 7

- Select the minimal size of the series: default value is 10, indicating that the series must contain 10 or more analogs.

- Output file reporting subsets of compounds with varying R-groups at the same site or site combination can be specified.

- The potency range is determined on the basis of all compounds comprising the series

3.1.2. For the **selectivity analysis** of a **single series** active against two targets, "**AnalogExplorer_Selectivity.bat**" or "**AnalogExplorer2_Selectivity.bat**" file is chosen.

- **Program settings** are according to **3.1.1.** with one difference

- **Two input SD files** should be provided.

For two targets A and B, one file contains a series of analogs with activity for target A and the other file consists of the same set of analogs with reported activity for target B, e.g., "SeriesForSelectivityAnalysis1_KinaseABL.sdf" and "SeriesForSelectivityAnalysis1_KinaseSRC.sdf" under "\AnalogExplorer_Original\SampleSets\".

- The potency range is determined on the basis of all compounds within two targets, ranging from the lowest to the highest potency of a compound active against one or the other target.

3.1.3. To **systematically explore all analog series available in a compound set**, "**AnalogExplorer_MultiSeriesInCompoundSet.bat**" or "**AnalogExplorer2_MultiSeriesInCompoundSet.bat**" file is chosen.

- **Program settings** are according to **3.1.1.** with few differences

- The SD file contains all compounds available for a given target, irrespective of the number of series they belong to, e.g., "CompoundSet_KinaseSRC.sdf" under "\AnalogExplorer_Original\SampleSets\".

- When "Join subsets" is selected, structurally related analog series are combined. Otherwise, different graphs will be generated for individual series. In order to avoid redundancy, only the combined series is shown by choosing "Join subsets". Structurally distinct series are displayed by de-selecting "Join subsets".

- Different output files will be generated, depending on user's selection, i.e., all series vs. combined series.

- The potency range is determined on the basis of all compounds within the set. Therefore, SAR patterns can be compared across different series.

3.2. In the panel of generated complete or reduced graph

3.2.1. Different functions are included.

- The number of analogs contained in a given series, the minimal and maximal potency values for color spectrum are reported in panel description (top left of the panel).

- The graph can be zoomed in and out and moved to desired position by clicking the graph and dragging the mouse.

- By pressing "l", one can show or hide node labels

- By pressing "s", one can save the current view as an image in PDF

- By clicking "View MCS", the correspond MCS structure will be shown in a separate panel

- To interactively adjust the layout of the graph, the user can switch the view from "transforming" to "picking"

- To view the site or site combination, the number of analogs and the corresponding mean potency value by moving the cursor onto a given node

- By clicking "View RGroup Trees", all possible R-group trees will be generated in a separate panel

3.2.2. Multiple graphs will be created in different panels that overlay on each other if more than one series is available. One can separate them by dragging these windows.

3.3. In the panel of generated R-group trees

3.3.1. Different functions are also included.

- By pressing "l", the structures of substituents will be shown in SMILES for nodes at different levels
- By pressing "s", one can save the current view as an image in PDF
- The layout of the nodes in R-group trees could also be modified by switching the view from "transforming" to "picking"
- One can view the number of analogs, the compound index (bottom layer), the corresponding (mean) potency value and R-group(s) by moving the cursor onto a given node

3.3.2. In **AnalogExplorer2**, **stereoisomers** are explicitly considered and shown in R-group trees. Stereoisomers belonging to the same subset (i.e., compounds with different stereochemistry at the same site) are identified by a unique index. If different subsets of stereoisomers are present in an R-group tree, incremental indices are used to identify and distinguish them (i.e., "1", "2" etc.).

4. OUTPUT FILES

The de-convolution of an analog series into sub-series is reported in output files, i.e., subsets of compounds that contain substituent(s) at the given site or site combination.

4.1. In the output files of the **original AnalogExplorer** program, six fields are provided.

- BMS_ID: the index of a qualifying BM scaffold for the graph generation
- BMS_SMILES: the SMILES representation of the BM scaffold for a series
- MCS_SMILES: the SMILES representation of the MCS
- Site/SiteCombiations: the index of a substitution site or indices of a site combination
- #Cpd: the number of compounds that contain substituent(s) at the given site or site combination
- CpdList: the corresponding list of compounds that are separated by "|"

4.2. In the output files of **AnalogExplorer2**, six fields are provided.

- The first four fields are according to **4.1.**
- Cpd_ID: the index of the compound belonging to a subset
- Cpd_SMILES: the canonical SMILES of the compound with which stereochemical information can be reconveyed.

4.3. For combined series, five fields (except BMS_SMILES) are reported.

NOTE: In these two versions of AnalogExplorer, we have further increased the consistency of compound mapping to MCS considering structural symmetry. Therefore, the assignment of substitution sites might differ from those we have reported in the original publication [1]. For example, for a given MCS, the site originally identified as R1 might become R2 in current implementation. However, it doesn't affect the organization of the series and the graphical analysis.

[1] Zhang, B.; Hu, Y.; Bajorath, J. AnalogExplorer - A New Method for Graphical Analysis of Analog Series and Associated Structure-Activity Relationship Information. *J. Med. Chem.* **2014**, *57*, 9184-9194.