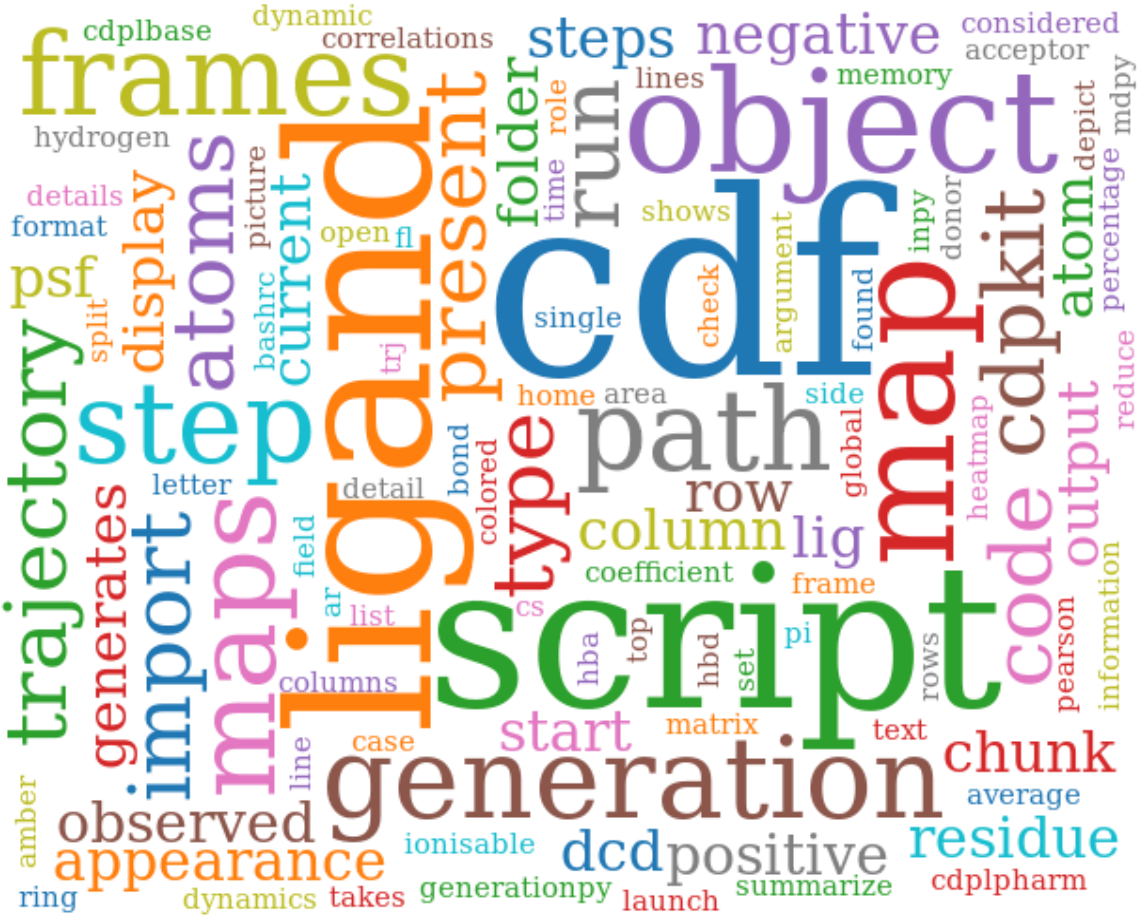
Interaction maps Tutorial

The interaction map displays the frequency of individual interactions between a ligand and its environment during a molecular dynamics simulation. The interaction map is a heatmap. Column labels contain the type of interaction and the ligand information, while the row labels contain the corresponding environment information. The color of each field of the matrix represents the frequency of an interaction as percentage. The type of interaction (like HBA-HBD) always consists of the type of feature on the ligand side followed by the environment side. Six types of pharmacophore features are considered: HBD for Hydrogen Bond Donor, HBA for Hydrogen Bond Acceptor, H for Hydrophobic interactions, PI for Positive ionisable area, NI for Negative Ionisable area and AR for Aromatic ring.



Towards Next Generation Pharmacophore Models
Arthur Garon
September 2018

universität wien

SERVIER

inte:ligand
Your partner for in-silico drug discovery.

The correlation map display the pearson correlation coefficient between all observed interactions during a molecular dynamics simulation.
The correlation map is a heatmap, where column and row labels contain the interactions information. The color of each field of the matrix represents the pearson correlation coefficient between the interaction associated with the column and the interaction associated with the row. If a value is less than -0.4, a negative correlation is observed ─ when the interaction of the column is present, it is unlikely that the interaction of the row is also present ─; if a value is more than 0.4, a positive correlation is observed ─ when the interaction of the column is present, it is likely that the interaction of the row is also present.

Interaction maps Tutorial

⬡ 1. Prerequisites

- Python version 2.7
- Python dependencies:
    - You must be able to start a python session and successfully run the `import_check.py` script

➢ If the CDPL imports are not working

- Please make sure the location of the CDPKIT installation is included in your PYTHONPATH by following these steps:
    - Open the .bashrc file, located in your home directory
    - Check that these two lines are present:
        - `PYTHONPATH=$PYTHONPATH:/path/of/the/CDPKIT/installation/Python`
        - `export PYTHONPATH`
    - In case of modification of the .bashrc file, launch another terminal

➢ Each script can be launched with the -h command to display a quick description of the script's role as well as details of the supported arguments:

- `python script_im_interested_in.py - h`

⬡ 2. Generation of interaction maps

The map generation process is divided in **3 independent steps**:

- Generate the cdf file ─ cdf, CDPKIT format ─ from a topology and a trajectory files
- Generate the gt object file from the cdf file
- Generate the Interaction and Correlation maps from the gt object file

These 3 steps are explained in detail below but can be done with a **single script**, see the **next section**.

In the following, each step is accomplished with a dedicated script, which mean that if you have already generated the cdf file, you can go directly to the second step. Same thing if you already have the gt object, you can start in the third step. The first and second steps are the most time consuming. For a system with 80 000 atoms and 2000 frames, it takes around 1h30 to generate the cdf file, and around 2h to generate the gt object. The third step takes only a few minutes.

➢ 2.1. Step 1: cdf file generation

This script generates a cdf file (CDPKIT format) from both a topology and a trajectory files.

- Run the `cdf_generation_MD.py` script
- It requires several arguments:
    - [Required] `-trj` The path of the topology file
    - [Required] `-top` The path of the trajectory file
    - [Optional] `-o` The output folder where the cdf file will be generated (Default: current directory)
    - [Optional] `-cs` The number of frames to consider per chunk (default: 500 frames)
    - [Optional] `-n` The name of the cdf file (Default: name of the dcd file)

The number of frames to consider for the generation allows to split the cdf files in chunks, to reduce the use of memory during the processing.

➢ 2.2. Step 2: gt object generation

This script generates the global trajectory object file from the cdf file.
This object is a simple representation of all pharmacophore interactions present between the ligand and the protein of the cdf file.

- Run the global_trajectory_generation.py script
- It requires several arguments:
    - [Required] -cdf The path of the cdf file
    - [Required] -lig The 3-letters code of the ligand

    - [Optional] -o The output folder where the gt object file will be generated (Default: current directory)
    - [Optional] -n The name of the gt object file (Default: name of the cdf file)

The 3-letter code of the ligand can be found in the fourth column of the topology file (psf) for any line corresponding to a ligand atom.

➢ 2.3. Step 3: Map generation

This script generates a picture of the ligand with atom numbering, 2 interaction maps, a correlation map and a text file that summarize the positive and negative correlations of the correlation map.
The ligand atom numbering is used by the maps to depict in detail which set of atoms are involved in the pharmacophore interactions. The two interaction maps display the percentage of appearance of the interactions between the ligand and its environment during the molecular dynamic simulation. One interaction map shows all found unique pharmacophore interactions, and the second shows merged interactions per environment residue type.

- Run the interaction_map_generation.py script
- It requires several arguments:
    - [Required] -cdf The path of the cdf file
    - [Required] -gt The path of the gt object file
    - [Required] -lig The 3-letters code of the ligand

    - [Optional] -fl The list of frames to consider, e.g. [frame_start, frame_end] (Default: all frames)
    - [Optional] -o The output folder where the pictures will be generated (Default: current directory)

⬡ 3. Single script for the Interaction map generation

This script generates a picture of the ligand with atom numbering, 2 interaction maps, a correlation map and a text file that summarize the positive and negative correlations of the correlation map.

- Run the get_interaction_map.py script
- It requires several arguments:
    - [Required] -trj The path of the topology file (psf)
    - [Required] -top The path of the trajectory file (dcd)
    - [Required] -lig The 3-letters code of the ligand

    - [Optional] -o The output folder where the files will be generated (Default: current directory)
    - [Optional] -cs The number of frames to consider per chunk (default: 1000)

The number of frames argument allows to split the cdf files in chunks, to reduce the use of memory during the processing.