ComplexArchitect

A python package for protein complex modeling from pairwise interactions of its subunits.
*Oscar Camacho

Index

- Description
- Download and Installation
  - Package tree
- Software Requirements
- Execution
- Documentation

Description

ComplexArchitect is a stand-alone python3 application developed by **Oscar Camacho**. It builds biological macrocomplexes taking a set of interacting pairs that can include proteins, DNA or RNA. This software could serve to study quaternary structures that are difficult to study *in vivo*.

Below is shown how to install and use this program as a stand-alone command line script (executing the script *carchitect*).

Download and Installation

You can download our package using Git with the next command:

```
$ git clone https://github.com/oscar-camacho/ComplexArchitect
$ cd ComplexArchitect
```

The directory ComplexArchitect should contain the directories and files described bellow:

Package tree The package has the following structure:

```
ComplexArchitect/
  README.md
  README.pdf
  LICENSE.txt
  setup.py
  comparch/
      __init__.py
      carchitect
      functions.py
      pdb_classes.py
      exception_classes.py
      optimize_functions.py
      utilities.py
  examples/
      1gzx/
      3t72/
      5fj8/
```

```
    6gmh/
    5nss/
doc/
    report.pdf
    tutorial.md
    tutorial.pdf
    functions.m.html
    pdb_classes.m.html
    exception_classes.m.html
    optimize_functions.m.html
    utilities.m.html
images/
unfinalized_scripts/
    CA_GUI.py
```

- README.md, README.pdf: contains information about how to install and execute the application, and a tutorial.
- setup.py: script to install the application in the Python side-packages.
- comparch: folder with the following scripts:
  - carchitect: the command-line script to launch the program.
  - functions.py: module requiered by CA_launcher.py where the classes of the application are defined.
  - pdb_classes.py: module where custom classes derived from the Biopython package are defined.
  - exception_classes.py: module where the exception classes are defined.
  - utilities.py: module where different types of required of data are stored.
  - optimize_functions.py: module where the functions to optimize the resulting model are defined.
- examples: folder with several examples stored in sub-directories that serve as input to the program.
- doc: folder with a report with the theoretical background and algorithm implemented on the program, a tutorial with examples and the documentation of each module.

Installation The user has to install the package in the python site-packages. You can install the package using pip with the next command on the terminal:

```
$ pip3 install comparch
```

Alternatively, it can also be installed with the next command:

```
$ sudo python3 setup.py install
```

Be sure to have the dependencies previously stated.

Software Requirements

In order to run this package with all its functionalities the user must have this software:

- Python 3.6
- Python modules:
  - Biopython
  - Modeller v.9.19
  - pandas
  - matplotlib
  - argparse
  - os
  - sys
  - re

  - random
  - copy

For the GUI (under development) the following ones are also necessary:

- Tkinter

Execution

The command line arguments that are needed to run ComplexArchitect are the following ones:

```
$ carchitect.py -h

usage: carchitect.py [-h] [-i INPUT] [-o OUTPUT] [-fa FASTA]
                     [-sto STOICHIOMETRY] [-m MODEL_NUM] [-C CHAINS]
                     [-clash CLASH_DIST] [-opt] [-v]

ComplexArchitect is a Python application designed to generate macrocomplex
structures from simple pair inetractions PDB files.


optional arguments:
  -h, --help            show this help message and exit
  -i INPUT              Input directory where PDB files with the pair
                        interactions are located.
  -o OUTPUT             Name of the output file, no extension is needed
  -fa FASTA             FASTA file with the sequences of the chains that will
                        conform the macrocomplex. They have to correspond to
                        the sequences of the chains from the PDB files. The
                        file should contain unique sequences; sequences don't
                        need to be repeated.
  -sto STOICHIOMETRY    Desired stoichiometry of the resulting complex. The
                        format should be like the follow example: A:2,B:4,C:1
  -m, MODEL_NUM         Maximum number of models the program is going to
                        build. The default number is 1.
  -C, CHAINS            Maximum number of chains that the resulting model
                        complex is going to have. The default number is 100.
  -clash CLASH_DIST     Minimum clash distance between 2 atoms. The default
                        minimum is 1.5 A.
  -opt OPTIMIZE         Refines the resulting model complexes and creates a
                        DOPE plot comparison between the unoptimized and the
                        optimized model.
  -v VERBOSE            Shows the progress of the program.
```

Documentation

Further information about how to use the program and a tutorial with examples are provided in the documentation section.