ACW – Amyloid coordinate wizard

ACW is a PyMol plugin for finding interfaces between amyloid beta sheets, and calculating descriptors: shape complementarity, buried surface area and surface detail index. It can be used on crystal structures containing beta sheets infinite in one direction. The plugin also serves as a database and a viewer for the amyloid oligopeptide structures downloaded from the PDB [1].

Method:

The script generates ±1 unit cells of the input peptide. Then it attempts to find H-bond-connected sheets starting in the original unit cell (reference sheets), then in the surrounding cells. After that it searches for possible interfaces of the reference sheets. Contact regions are accepted as interfaces based on criteria of the size solvent excluded surface area and average number of interacting sidechains of a peptide contributing to the interface. Valid interfaces are categorized into groups based on the interacting sidechains, and their averaged descriptor values are displayed. For more information read our accompanying article [2].

Requirements:

The plugin works with recent Windows and Linux PyMol editions with python version 2.7 and above. The calculation methods require installation of the CCP4 [3] program package (including programs AREAIMOL [4] and SC [5]).

Installation:

1. Download or clone this repository.
The folder contains the plugin file (acw.py and its module folder); a configuration file (config.txt); and the database of available PDB structures of amyloid oligopeptide crystal structures (set of folders and acw_database.json file). The database should be in the same folder to use the database mode.
2. (Optional) Edit the path of CCP4 installation folder in the configuration file (config.txt, inserting the path after the equals sign). More information about system specific syntax is in the config file. Note, for using only the “database mode” (if the plugin will be used only as a database and viewer), no need for pre-installation of CCP4 nor editing the config file.

User guide:

Input files are PyMol compatible structure files containing peptide amyloid crystal structure. Each peptide chain should have a different chain ID letter and no gaps in residue numbering. The whole peptide chain should be recognizable as protein-type molecule by PyMol. After generating the symmetry equivalent chains, each sheet is expected to contain 3 or 6 strands. The script attempts to use unit cell information for generation of neighbouring sheets (CRYST1 section), however it also works with cryo-EM or NMR data containing 3 or 6 stranded beta sheets.

To start the plugin: load the acw.py file selecting ‘file/run script…’ or ‘file/open…’ in the scroll-down menu of PyMol, then type in the PyMol command line the following: ‘acw’ to run the script.

If CCP4 is available the plugin starts in evaluation mode which serves for analysis of amyloid structures and calculation of descriptors. Otherwise the plugin will start in Database mode, or you can manually change to it. It serves for viewing of the previously analysed structures from both literature and saved user data. The plugin will print information both to the PyMol console and to the top right section of the PyMol display window.

1. Evaluation mode:

Step 0: Open an amyloid structure.
Step 1: Press '1: Set up the sheets', this will create first an output folder with the name of your PDB file, unless it is already part of the current database. Then the plugin will create and find sheets that are suitable for calculations.
Step 2: press '2: Automatic calculation', the script will calculate (using CCP4 programs) all possible contact surfaces between the sheets involving the asymmetric unit and the surrounding sheets. Interfaces that fit the criteria will be saved in the new folder generated from the file name of the structure.
Step 3: For each interface found, a dotted line is shown between the two endpoints of the interface projected in the direction of the sheets (see SDi calculation in the article for more details). If multiple interfaces are found, you can view their data by clicking 'next interface'. Resulting descriptor values are shown on screen, also these are automatically saved as ACW_output_yourfilename.txt.
Step 4: You can add the results to the database with 'save to database' to view it there later. Click on 'change to evaluation' to go back.
Step 5: Press '5: Reset' to analyse a new structure. Output files are preserved.

Under 'Viewing:' You can use the buttons in the viewing section to take pictures, hide the labels of each sheet or reorient and rotate the view of the amyloid.

Manual mode:
It is possible to test contacting surfaces that do not meet the interface criteria. Turn on 'Edit: Manual mode'. Then click on two different sheets and the script will attempt to analyse the interface between them. These interfaces will not be saved in the database, however a log file is still created for those calculations (ACW_manual_output_yourfilename_X.txt).

Database mode:

Step 1: click on 'Change to database'.
Step 2: Navigate in the database using the 'Next peptide', 'Previous peptide' buttons.
With 'Selection criteria' you can choose the class (topological class of the largest interface [6]), secondary class (for structures with multiple topological classes), peptide length and packing type of the structures [2], source of the structure (PDB or personal) and amino acid type to select a subset of the database. These work as toggle buttons (selecting/deselecting a characteristic).
You can jump directly to an existing PDB entry of the database by using the command line: type 'acwgoto X' where X is either the file name, PDB code, or its index in the list.
Step 3: You can view the different interfaces and their descriptor values by clicking 'View interface (…/…)'. These are saved in the folder of the PDB file: ACW_output_yourfilename.txt'.

Edit options:
'Manual mode' is available here too.
If the current structure was not present in the original database, it can be removed from the database by clicking 'delete from database', this will also delete its folder and it cannot be undone.
You can export information (interfaces, composition, descriptor values …) of the currently selected structures. A json file and two tabulated files will be created, one with structure information, one with interface information.
The current database data are compiled in file 'acw_personal_database.json'. If you want to reset the whole database to the database extracted from PDB, delete this file, so the script will create a new copy from 'acw_database.json' when it is run next time.

Output:

The script creates a folder for output files of each structures where it is installed. The folder names are created from the PDB file names. The folder contains PDB files containing the generated sheets, also log files from automatic and manual calculations.

The following is written into the 'ACW_output_ yourfilename.txt':

The main topological class of the amyloid and secondary class if present.
The dimensionality of packing, i.e. the dimensionality of the network created by amyloid interfaces within the crystal packing. Finally the number of interfaces found.

Then for each different interface:
- A list of sheet-pairs that form the interfaces (referring to sheet labels shown in PyMol)
- List of amino acids that interact with their side chains for each peptide chain
- Shape complementarity
- Buried surface area
- Surface detail index
- The class of the interface
- The contribution of each amino acid to the buried surface area. If there were more instances of the interface, each of these instances is listed separately.

Example output of 8ANM:

Evaluation of 8ANM
Main class is 8
Packing dimension is 3D
The number of interfaces is 1
Interface 1
Interacting sheet-pair(s): 8+1, 1+9, 1+12, 14+1
→the first interface was found 4 times: between sheets 8 and 1, between sheets 1 and 9 etc.. These are considered the same type of interface, because the contributing residues, the topology class and relative orientations of the chains are the same.
Sidechains in the interface: (1L3I,4Q6L)+(2Y,3I5W)
→the interface is formed by 4 chains, the first chain of sheet 8 forms contacts by Leu1 and Ile3; the

other chain in sheet 8 with Gln4 and Leu6. Likewise for sheet 1
Sc: 0.701
B.Area: 86.388
SDi: 1.317 → these values have been averaged for the four interfaces
Class: 8 → this is a class 8 interface

Known problems:

-Topological class might be misidentified, if the structure contains very differently bent backbones.
-This method doesn't separate LARKS and Out-of register classes, instead it assigns them to classes 1-10.
-Packing dimension may get misidentified in the following cases:
    -due to low symmetry there is not enough sheets generated. Example: 4W5Y
    -if a two dimensional packing motif is made of multiple smaller interaction. Example: 6BZP
-The script only works with 3 and 6 stranded beta sheets (after generating the neighbouring unit cells sheets should have 3 or 6 chains)
- AREAIMOL and SC may fail to calculate surface areas, if the protein includes elements other than C, N, O, S or unknown sidechains.
-Peptide terminal modifications are ignored if PyMol doesn't recognise them as protein atoms
-Nonstandard amino acids will be named X in the output as their one-letter code. Also they may cause SC and AREAIMOL to fail.

References:

[1] Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242 (2000).

[2] Soon

[3] Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Cryst D* **67**, 235–242 (2011).

[4] Lee, B. & Richards, F. M. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology* **55**, 379-IN4 (1971).

[5] Lawrence, M. C. & Colman, P. M. Shape Complementarity at Protein/Protein Interfaces. *Journal of Molecular Biology* **234**, 946–950 (1993).

[6] Eisenberg, D. S. & Sawaya, M. R. Structural Studies of Amyloid Proteins at the Molecular Level. *Annu Rev Biochem* **86**, 69–95 (2017).