ACWF – Amyloid coordinate wizard fibril

ACWF is a PyMol plugin for finding local interfaces in amyloid fibrils, and calculating their descriptors: shape complementarity (Sc), buried surface area (Ab) and surface detail index (SDi). It can be used on amyloid fibril structures determined by cryo-EM. The plugin also serves as a database and a viewer for the amyloid polypeptide structures downloaded from the PDB [1].

Method:

The plugin cleans up the input structure and generates additional chains of the filaments by fitting it to the input structure. After that the sliding window methods appoints and analyzes local interfaces similar to APR interfaces.

For the Nth residue, first the contacting atoms of every Nth residue's sidechain are found (<2.8 Å between Van der Walls radii). Other residues are considered interacting, if they have at least one atom in one chain in contact with the Nth residue. Then the residues are separated into two groups: with residue number between N-3 and N+3 together with the original residue are assigned to the same polypeptide chain and form surface-A of the local interface; residues outside the range of residue no. N-3 and N+3 or in other filaments are assigned to the counteracting surface(s) (surface-B) of the interface. The descriptor values Sc, Ab and SDi are calculated similarly as described previously [2] for surfaces A and B and assigned to the Nth residue.

Structures of different protein types were clustered using structure-based alignment of all structure-pairs for that protein type. The hierarchical clustering was carried out to minimize the RMSDs modified by the aligned peptide lengths within each cluster. These clusters be considered as different amyloid polymorphs and variations of amyloid 2D-folds.

For more information read our accompanying article [3].

Requirements:

The plugin works with recent Windows and Linux PyMol editions with python version 2.7 and above. (tested versions: 2.1 and 2.4.1) The calculation methods require installation of the CCP4 [4] program package (including programs AREAIMOL [5] and SC [6]).

Installation:

1. Download this repository, and unpack it in a separate folder. The folder contains the plugin file (acwf.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 acwf_database.json file). The database should be in the same folder as the plugin file to use the database mode otherwise only the evaluation is available.
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:

Database content and limitations of use. Because of uncertainties in generating fibrils, and requirements by the programs used for calculating Sc and Ab the following criteria were applied for the structures to be analyzed: 1) Amyloid fibril structures determined by Cryo-EM. 2) The whole polypeptide chain should be recognizable as protein-type molecule by PyMol. 3) Each polypeptide chain of the structure should have a different chain ID letter. 4) Each protofilament should contain the same number of strands, and at least 3 strands per protofilament. The program can be used in **evaluation mode** for analysis of structures not present in the current database using **input files,** that are PyMol compatible structure files containing amyloid fibril structures meeting the above criteria. Note the plugin works best with cryo-EM data (the plugin may work with NMR data including a single model, but it was not tested extensively).

To start the plugin: load the **acwf.py** file selecting 'file'→'run script' or 'file'→'open' in the scroll-down menu of PyMol and browsing to the location of your installation of ACWF. The plugin starts immediately, for reopening the plugin type in the PyMol command line the following: '**acwf**'.

The plugin starts in **Database mode**, it serves for viewing of the previously analyzed structures collected from the PDB. If CCP4 is available the plugin can be changed into the evaluation mode which serves for analysis of amyloid structures and calculation of descriptors. The plugin will print information both to the PyMol console and to the top right section of the PyMol display window. Toggle between database and evaluation modes using "Change to Evaluation/Database Mode"

1 Database mode:

Navigating

- With '**Selection criteria**' button you can choose the protein type, and number of protofilaments. These work as toggle buttons (selecting/deselecting a certain characteristic).
- Navigate in the database using the **'Next structure'**, **'Previous structure'** buttons. If protein type(s) and/or number(s) of protofilaments were selected, you can navigate in the selected subset of the database.
- After selecting one of those protein types, that have enough elements to cluster, you can also use the '**Next in cluster'**, '**Previous in cluster**', buttons to navigate according to its clustering order. (Note, this option of navigating in the order of clusters works only when exactly one protein type of the clustered ones is selected. Selection for number of protofilament is ignored during navigation by 'Next in cluster', 'Previous in cluster'.)
- You can jump directly to an existing PDB entry of the database by using the command line: type '**acwfgoto X**' where X is either the file name, PDB code, or its index in the list within the database.

Viewing options

Under '**Viewing**:' You can use the buttons in the viewing section to recolor the structure, save images, hide the labels or reorient and rotate the view of the amyloid.

'**Coloring**':

- '**Even-odd**' color the residues alternatingly green and orange according to the residue number in the sequence
- '**recolor-sc'** recolor each residue for the Sc value of its local interface on a rainbow gradient
- '**recolor-ab'** recolor each residue for the Ab value of its local interface on a rainbow gradient

- '**recolor-sdi'** recolor each residue for the SDi value of its local interface on a rainbow gradient
- '**protofilament**' colors each protofilament with a different color
- '**Color range: absolute/relative**': changes the range of the color gradient: **Absolute** means lowest value for the current structure: blue to highest value in the current structure: red. **Relative** means a predefined range that gets displayed in the console (blue: lowest to red: highest value). Relative coloring is recommended for comparison structures.
- '**Color inter/intra**': toggle between only intra-protofilament interface / intra- plus inter-protofilament interfaces.

Toggle labels and axes'

- You can also display the chain and residue labels, the cartesian axes, and the fibril axis using the '**Toggle labels and axes**' menu button.
- The Ab, Sc or SDi descriptor values calculated for the local interfaces around each side chain can also be displayed using '**Toggle labels and axes**' → '**Descriptor values'**. Pushing this toggle button will show the last local descriptor values selected by '**Coloring**'→ '**recolor-sc**', '**recolor-ab**' or '**recolor-sdi**' and '**Color inter/intra**'. (If none was chosen, it will display zeros)

You can also display the chain and residue labels, the coordinate axis, and the fibril axis using the '**Toggle labels**' menu button.
'**Reset view**' changes the view along the x, y, and z axes, respectively.

In '**Settings**' currently two options are available.

- The top left information box with clustering results can be shown or hidden (see section output for more details).
- The 'Settings' → '**Generate chains**' toggle button can be used to show the multiplicated structure which was used for eliminating end-of-protofilament effect which could cause error in calculating local interfaces in case of significantly warped structures. (In case of small warping, the structure is only elongated by a factor of three, however for large warping, the structure could be multiplied by a larger (odd) number. See [3] for more details.)

Top left box displayed in the graphical window shows:

- Structure ID (PDB code and number within the database),
- protein name,
- number of protofilaments and RMSD values of the fit for extension,
- for each protofilament: ID / Clustering results. (This will show 0 if no clustering is available)

2. Evaluation mode:

You can analyse local interfaces of a structure through steps 0-4 under 'ACWF Wizard: Evaluation' as follows. During this process, the ***Top left box*** displayed in the graphical window is used as a guide.

Step 0: Open an amyloid fibril structure.

Step 1: Press '**1: Assign protofilaments**', then click on each protofilaments once. The plugin will color it if, it was recognized as a protofilament based on its H-bonds. If a protofilament is recolored, the H-bond system was ambiguous. Press again the button, and click on a different part of the protofilament, away from close contacts with other protofilaments.

Step 2: press '**2: Setup fibrils**', the plugin makes a cleanup to the structure then generates additional strands. These are temporarily stored in new objects obj1, obj2 … For making images in the later steps, you might want to hide some of them by clicking on them in the menu.

Step 3: press '**3: Assign local interfaces**' to make the plugin find local interfaces for each sidechain. This may take a minute.

Step 4: press '**4: Sliding window**' to calculate (using CCP4 programs) the 3 descriptors for each local interface. Step 4 may take several minutes depending on the size of the structure.) During this process the structure is colored as follows: green: is the side chain for which the local interface is analysed; blue: its neighbouring residues contributing the local surface; green: residues forming the molecular surface of the opposing side of the interface. Calculated data for each allocated interface is displayed above the command line.

Resulting descriptor values are shown on screen, also these are automatically saved as described in the Output section. At this point you can also use different coloring to visualize descriptor values Sc, Ab and SDi, save images, toggle labels, as described in the Viewing options above.

Step 5: Press '**5: Rese**t' before opening and analyzing a new structure. This will clear program memory. Make sure to change the working directory inside PyMol or save output files manually, or they will be overwritten. Note, if one of the steps 1-4 fail, the usual reason is that the structure does not meet one of the criteria listed in the section about limitations. This case '**5: Reset**' should be clicked too before proceeding to the evaluation of the next structure in order to get proper results.

Output:

Database folders: for each structure of the database, data of analysis and images generated by the user can be found in a separate folder with the same name as the PDB structure. Evaluation mode also generates these output files. For each filament there is a separate set of output files, numbering starting from 0.

- '**monomer_X.pdb'** containing a single cleaned up protofilament.
- '**contacts_X.txt'** for each residue number its contacts are listed as residue number and protofilament number. If only itself is listed, that amino acid has no contacts. These contacts are used during calculation.
- '**contacts_info_X.txt'** same, but additional information is listed: the type of amino acid, the area of sidechain contacts, the area of backbone contacts.
- '**acwi-out6_X.txt'** results of sliding window method containing all interactions: residue number, amino acid type, Sc, area used for Sc calculation, Ab, SDi, list of IDs of contacting residues for surface A and B separately.
- '**acwi-out6i_X.txt'** similar output file but containing only intra contacts of a protofilament.

- ‘**acwi-out6_X-lengths.txt’** and ‘**acwi-out6i_X-lengths.txt’** accompanying the output files there is one contain the coordinates of start and end points of the line segments used for the calculation of SDi.

Known problems:

- AREAIMOL and SC may fail to calculate surface areas, if the protein includes elements other than C, N, O, S or unknown sidechains or simply too many atoms.
- Antiparallel fibril structures are incompatible with the plugin
- The extended fibril may deviate from the real fibril, if chains are in very different conformation.

References:

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

[2] Sulyok-Eiler, M. *et al.* Unravelling the Complexity of Amyloid Peptide Core Interfaces. *Journal of Chemical Information and Modeling*, **64** (22), 8628–8640 (2024)

[3] Coming soon

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

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

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