ZEAL

This manual describes how to use the command-line version of ZEAL in MATLAB, available for download from github link XXXX. For details on the method itself, see the publication [publication](https://academic.oup.com/bioinformatics):

F. Ljung and I. André, ZEAL: Protein structure alignment based on shape similarity, *Bioinformatics* XX (2020)

# Requirements

ZEAL depends on the following classes

* **PDB.m**  Handles reading/fetching/parsing of PDB files
* **molShape.m** Generates a molecular shape function (the molecular surface by default) from PDB object data
* **ZC.m** Handles computation of Zernike-Canterakis moments and shape reconstruction

## Toolboxes

ZEAL uses functions from these toolboxes which have to be installed (included in [MATLAB Runtime](https://www.mathworks.com/products/compiler/matlab-runtime.html) if ZEAL run as standalone program)

* gads\_toolbox G*lobal Optimization Toolbox*
* image\_toolbox *Image Processing Toolbox*
* optimization\_toolbox *Optimization Toolbox*
* statistics\_toolbox *Statistics and Machine Learning Toolbox*
* symbolic\_toolbox *Symbolic Math Toolbox*

# Abbreviations

* ZC Zernike-Canterakis
* MS Molecular surface
* SA/SAS Solvent-accessible surface
* vdw van der Waals
* PDB Protein Data Bank

# Input

ZEAL can be run in two modes depending on if one or two protein structures are given as input. Protein structrues can either be given as files ('filename.pdb' or 'filepath/filename.pdb' if not in same directory ) or as 4-letter PDB ID codes (eg. '5MOK'). If a 5-letter PDB ID code is given (eg. '5MOKA'), then the last letter is assumed to be the chain ID ('A') that should be selected for analysis. By default, all chains of the structure (first model if several exist) is used, excluding hydrogen atoms and any heteroatoms. If any alternate atom locations are defined, the 'A' state is selected. See the section [Changing default parameters](#MW_H_56E64AE2) to change what structure data should be used in the analysis.

**Single mode** If one structure is given, then ZEAL will compute the Zernike-Canterkis shape descriptor for the structure (defined in the object as the fixed structure)

% Single mode

shapeData\_pdbFile = ZEAL('5mok.pdb'); % -> reads pdb file

shapeData = ZEAL('5mokA'); % -> downloads the structure from PDB

**Align mode** Adding a second structure with parameter 'rot' will start the shape alignment search in ZEAL where the second structure (defined as rotating in the object) is rotated

% Align mode

% shapeAlignData\_pdbFile = ZEAL('5mok.pdb', 'rot', '2ho1.pdb'); % -> reads PDB files and and performs the shape alignment

shapeAlignData = ZEAL('5mokA', 'rot', '2ho1A'); % -> Downloads structures and performs the shape alignment

# Output

**Single mode** The shape descriptors and ZC moments are accessed from the property fixed.ZC.Descriptors

shapeData.fixed.ZC.Descriptors

or using the method getShapeDescripors

% data for fixed (default)

ZCDs\_fix = getShapeDescritors(shapeData);

% data for fixed

ZCDs\_fix = getShapeDescritors(shapeData, 'fixed');

% data for rotating

ZCDs\_rot = getShapeDescritors(shapeData, 'rotating');

The moments are accessed from the property fixed.ZC.Moments

shapeData.fixed.ZC.Moments

where the field Values contains the complex-valued moments, Indiceslist contains the n,l,m indices (col 1-3) for each moment with real part (col 4) and image part (col 5) separately.

The complex moments can also be accessed using the getMoments method

% data for fixed (default)

ZCmom\_fix = getMoments(shapeData);

% data for fixed

ZCmom\_fix = getMoments(shapeData, 'fixed');

% data for rotating

ZCmom\_rot = getMoments(shapeData, 'rotating');

**Align mode** The shape descriptors and ZC moments for the rotating structure can be access as above, but from the property rotating instead. The ZEAL score is accessed with

shapeAlignData.Score

and the Euclidean distance between the shape descriptors with

shapeAlignData.ZCDdistance

# Optional input

The parameters listed below are set by default, but can be changed (described in [Changing default settings](#MW_H_56E64AE2)).

settingsTable = getZEALSettingsTable()

# Changing default parameters

All parameters above can be changed by giving a comma-seperated argument list to ZEAL, or as a Matlab structure with fields equal to names of parameters.

### Example: Using argument list

ZEAL('5mok','fix\_chainID','B') % selects chain B for shape analysis

ZEAL('5mok','fix\_chainID','B', 'Shape', 'SAS', 'GridRes', 100) % + sets the solvent accessible surface as the shape function computed on a 100x100x100 grid

### Example: Using input structure

inputStruct.fix\_chainID = 'B';

inputStruct.Shape = 'SAS';

inputStruct.GridRes = 100;

ZEAL('5mok', inputStruct)

# Saving

Aligned structrues can be saved to PDB files using the save2pdb method. By default, both the fixed and rotating structure are exported to the current directory (see below on how to change). The names of the new pdb files have the format originalname\_ZEAL.pdb (this can't be changed). Also, HETATM records are omitted but this can be changed (se below).

save2pdb(shapeAlignData)

To export ***all records***, use

save2pdb(shapeAlignData, 'includeAll', true)

If the original pdb file contains multiple chains, and the alignment was done with respect to chain X of the fixed structure and chain Y of the rotating structure, the coordinates are transformed so that the centroid of the X chain and Y chain are placed at the origo, and coordinates of the rotating structure are rotated relative that center.

To ***include HETATM records***, use

save2pdb(shapeAlignData, 'includeHetatoms', true)

To ***save to specific directory***, use

save2pdb(shapeAlignData, 'folderPath', '/Users/yourUserName/Desktop')