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:

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 if ZEAL run as standalone program)

- `bioinformatics_toolbox` *Bioinformatics toolbox*
- `gads_toolbox` *Global 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 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
```

```
Running ZEAL in single mode
 Importing fixed structure: 5mok.pdb
 Computing shape function for fixed structure
 Computing ZC moments for fixed structure
```

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

```
Running ZEAL in single mode
 Importing fixed structure: 5mokA
 Computing shape function for fixed structure
 Computing ZC moments for fixed structure
```

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

```
Running ZEAL in Align mode
 Importing fixed structure: 5mokA
 Importing rotating structure: 2ho1A
 Computing shape function for fixed structure
 Computing shape function for rotating structure
 Computing ZC moments for fixed structure
 Computing ZC moments for rotating structure


 ////////////////////////////////////////////////////////////////////////////

 Searching for best shape superposition

 Stopping after 300 function evaluations

 ----------------------------------------------------------------------------
 Current best score      Euler (zyz)          iteration       time (s)
 ----------------------------------------------------------------------------
  0.31    0.00 0.00 0.00   0     0.2
  0.51    3.14 1.57 3.14   1     0.4
  0.58    4.82 2.03 2.56   20     3.5
  0.61    4.91 1.92 2.18   23     4.0
  0.62    4.63 2.09 1.78   27     4.7
  0.70    4.45 2.04 2.08   31     5.4
  0.74    4.15 1.92 1.89   34     5.9
```

```
  0.78    3.87 1.75 1.87   35     6.1
  0.79    4.03 1.71 2.05   43     7.4
  0.80    3.83 1.63 2.12   51     9.0
  0.81    3.77 1.62 2.12   109    21.9
 ------------------------------------------------------------------------

 Search completed after 59.9 s.

 Best score 0.81 found after 109 iterations (21.9 s)

 using Euler angles (zyz) [3.766 1.616 2.120]

 ////////////////////////////////////////////////////////////////////////
```

Output

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

```
shapeData.fixed.ZC.Descriptors
```

```
ans = 121×1
    0.0272
    0.0000
    0.0440
    0.0069
    0.0021
    0.0027
    0.0343
    0.0190
    0.0016
    0.0068
     ⋮
```

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
```

```
ans = struct with fields:
    IndicesList: [1771×5 double]
     CellValues: [21×21×41 double]
         Values: [1771×1 double]
```

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
```

```
ans = 0.8153
```

and the Euclidean distance between the shape descriptors with

```
shapeAlignData.ZCDdistance
```

```
ans = 0.0253
```

Optional input

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

```
settingsTable = getZEALSettingsTable()
```

`settingsTable = 20×5 table`

	parameter	type	default value	expected values	description
1	'Order'	'integer'	20	'>0'	'The maximum exp...
2	'ChiCoeffPath'	'char'	'[pwd '/chi_coeffic...	'folder path'	'Path to folder ...
3	'GridRes'	'integer'	64	'>0'	'The side length...
4	'Shape'	'char'	'MS'	"MS'/ 'SAS'/ 'vdw'/ ...	'The type of mol...
5	'ProbeRadius'	'double'	1.4000	'>=0'	'The radius of t...
6	'SmearFactor'	'double'	0.3000	'>0, <1'	'Fraction of gri...
7	'ShellThickness'	'integer'	2	'>0'	'Thickness of su...
8	'FunEvals'	'integer'	300	'>0'	'Number of ZEAL ...
9	'AlignLater'	'logical'	0	'true/false'	'If false then Z...

	parameter	type	default value	expected values	description
10	'fix_includeHe...	'logical'	0	'true/false'	'Flag to indicat...
11	'rot_includeHe...	'logical'	0	'true/false'	'"---" in rotati...
12	'fix_includeHa...	'logical'	0	'true/false'	'Flag to indicat...
13	'rot_includeHa...	'logical'	0	'true/false'	'"---" in rotati...
14	'fix_chainID'	'char'	'A'	"all' or 1 letter'	'The chain ID th...

⋮

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 surfac
```

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)
```

```
outputting PDB in file 5mokA_ZEAL.pdb ...
 done! closing file...
outputting PDB in file 2ho1A_ZEAL.pdb ...
 done! closing file...
```

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')
```