Package 'PdPDB'

January 20, 2025

Depends plyr, stats, utils, graphics, grDevices, dendextend, tseries,

Type Package

Title Pattern Discovery in PDB Structures of Metalloproteins

Version 2.0.1

Date 2017-11-09

Author Luca Belmonte <luca.belmonte@unimore.it>, Sheref S. Mansy
     <sheref.mansy@unitn.it>

Maintainer Luca Belmonte <luca.belmonte@unimore.it>

Description Looks for amino acid and/or nucleotide patterns and/or small
     ligands coordinated to a given prosthetic centre. Files have to be in the local
     file system and contain proper extension.

License MIT + file LICENSE

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-11-10 18:41:22 UTC

Contents

PdPDB   *Pattern Discovery in PDB Structures of Metalloproteins*

Description

Looks for amino acid and/or nucleotide patterns coordinated to a given prosthetic centre. It also
accounts for small molecule ligands. Patterns are aligned, clustered and translated to logo-like
sequences to infer coordination motifs.

1

Usage

```
PdPDB(path, metal, n, perc, interactive, dropsReplicate)
```

Arguments

path	A string containing the path to the PDB directory.
metal	A string containing the PDB chemical symbol of the target prosthetic centre; e.g. SF4 for [4Fe-4S] cluster, ZN for zinc. The PDB chemical symbol is case sensitive for macOS.
n	A numerical value that contains the number or residue in following/preceding n positions from the ligated amino acid or nucleotide; if n=1 PdPDB searches for x(L)x motif-like chains, if n=2 for xx(L)xx. (L)igand.
perc	A numerical value about the minimum percent of letters in a column otherwise residues are dropped.
interactive	A numerical value. 0 interactive, 1 automated (will not cut dendrogram), 2 user decided cut. In mode 1 and 2 ExPASy amino acid frequencies are used as reference.
dropsReplicate	A numerical value. 0 keeps replicated patterns, 1 drops replicated patterns entry by entry, 2 keeps only unique patterns.

Value

PdPDB generates a list of ".csv" and ".svg" files that will be stored in the same folder of the analyzed pdb/cif files (see "path"), its output is as follows:

frequency.csv	PDB-like patterns (i.e. with PDB chem Ids). "-" and "+" are used for residues out of the n inspecting window or from different monomers, respectively. Patterns come along with their frequency.
alignment.csv	Ligand-aligned patterns with dashes, plus signs and gaps ("*"). See 'frequency.csv'.
following_X_enrichment.csv	
	n files. Each file contains enrichment score, z-score and statistics at up to n following positions. X is the +position from ligated residue.
ligands_enrichment.csv	
	Enrichment scores and statistics for ligands.
notLigands_enrichment.csv	
	Enrichment statistics for the whole specimen but ligands.
preceeding_X_enrichment.csv	
	As for "following" but this is meant for residues preceeding ligands. See "following_X_enrichment.csv."
root_enrichment.csv	
	Overall enrichment score.
logo_Y.csv	Y files. Each file contains the logo and consensus sequence for a cluster. Y is the cluster number.
dendrogram.svg	The dendrogram along with the user deciced cutoff and clusters.

`following_X_proportions.svg`
> Plot of the enrichment score per each amino acid in following positions.

`ligands_proportions.svg`
> Plot of the enrichment score per each amino acid in ligated position.

`notLigands_proportions.svg`
> Plot of the enrichment score per each amino acid in non ligated position.

`preceeding_X_proportions.svg`
> Plot of the enrichment score per each amino acid in preceeding positions.

`root_proportions.svg`
> Plot of the root enrichment score.

`logo_Y.svg` Plot of the logo and consensus sequence of the Yth cluster. The complete aligned cluster is given as homonym '.csv' file. Sequences come along with percentages. If the dendrogram is not cut the root logo is given.

`following_X_standardized.svg`
> Plot of the z-score per each amino acid in following positions.

`ligands_standardized.svg`
> Plot of the z-score per each amino acid in ligated position.

`notLigands_standardized.svg`
> Plot of the z-score per each amino acid in non ligated position.

`preceeding_X_standardized.svg`
> Plot of the z-score per each amino acid in preceeding positions.

`root_standardized.svg`
> Plot of the root z-score.

`patterns.csv` PDB like extracted patterns along with the PDB ID and metal IDs. Useful for debbugging. Needed for restore.

`PdPDB.log` PdPDB log file. Useful for debbugging. Needed for restore.

Note

Files have to be in the local file system and contain the ".pdb" or ".cif" extension. Output files use brackets to highlight ligands and/or 'L' in heading line.

Author(s)

Luca Belmonte, Sheref S. Mansy

References

Belmonte L, Mansy SS Patterns of Ligands Coordinated to Metallocofactors Extracted from the Protein Data Bank, Journal of Chemical Information and Modeling (accepted)

Examples

```
################ Defining path to PDBs
path_to_PDB="inst/extdata/PDB" # this is where pdb/cif files are stored

################ Research Parameters
```

```
metal="SF4"  # searches for [4fe-4s] coordinating patterns
n=1  # searches for x(L)x patterns, (L) coordinates to SF4
perc=20  # drops residues with less than the 20% of frequency
interactive= 0 # interactive. User decided references and dendrogram cut
dropsReplicate=0 # do not remove replicated patterns

################ Launch PdPDB
PdPDB(path_to_PDB,metal,n, perc, interactive, dropsReplicate)
```

Index