NAME

DownloadPDBFiles.pl - Download PDB files for PDB ID(s)

SYNOPSIS

DownloadPDBFiles.pl PDBID(s) or PDBIDsTextFile

DownloadPDBFiles.pl [-c, --colmode *colnum | collabel*] [-d, --dataLocationURL *PDB URL*] [--DensityMap *yes | no*] [--DensityMapMode *XRayElectronDensity, CryoEMDensity, Auto*] [--DensityMapLocationURLCryoEM *Map URL*] [--DenistyMapLocationURLXRay *Map URL*] [--EDMapLocationSuffixes *CompositeMap,None,...*] [--EDMapTypes][-h, --help] [--indelim *comma | semicolon*] [-m, --mode <IDsOnCmdLine | IDsInFile>] [--PDBIDsCol *number | string*] [-p, --PDBFormat *PDB, CIF or Auto*] [-w, --WorkingDir dirname] PDBID(s) or PDBIDsTextFile

DESCRIPTION

Download PDB files corresponding to PDB IDs specified in a column in a CSV/TSV text file or on the command line as space delimited parameters.

It is also possible to download x-ray electron density and cryo-EM density maps for the specified PDB IDs.

OPTIONS

-c, --colmode *colnum | collabel*

Specify how columns are identified in a *TextFile* containing PDB IDs: using column number or column label. Possible values: *colnum or collabel*. Default value: *colnum*.

-d, --DataLocationURL *PDB URL*

Specify location of PDB URL where data files are available for download. Default value: *http://www.rcsb.org/pdb/files/*.

--DensityMap *yes | no*

Download x-ray electron density and cryo-EM density map file(s). Possible values: *Yes or No*. Default value: *no*.

--DensityMapMode *XRayElectronDensity, CryoEMDensity, Auto*

Specify type of density map files to download. Possible values: *XRayElectronDensity, CryoEMDensity, Auto*. Default value: *Auto*. The x-ray electron density and cryo-EM density map files are downloaded for *XRayElectronDensity* and *CryoEMDensity* values. The availability of both XRayElectronDensity and XRayElectronDensity is checked for *Auto* value by attempting to download x-ray map files followed by cryo-EM map files.

X-ray Electron Density (ED) map file(s) are downloaded in CCP4 and MTZ format. Three different types of ED map files may be downloaded using option --EDMapTypes: CompositeMap (2Fobs - Fcalc), DifferenceMap (Fobs - Fcalc), ReflectionMap. The format of ED data in first two file types is CCP4. The third file type contains ED data in MTZ format.

The names of the downloaded ED files are derived from input PDB IDs as shown below:

```
CompositeMap (2Fobs - Fcalc):  <PDBID>.ccp4
DifferenceMap (Fobs - Fcalc): <PDBID>_Diff.ccp4
ReflectionMap:  <PDBID>_Map.mtz
```

CryoEM density map file(s) are also downloaded in CCP4 format. The names of the cyroEM density map files is derived from EMDB ID in downloaded PDB or CIF file:

```
CryoEMFile:  emd_<EMDBID>.map.gz
Path: <CryoEMMapLocationURL>/EMD-<EMDBID>/map/emd_<EMDBID>.map.gz
```

--DensityMapLocationURLCryoEM *Map URL*

Specify location of cryoEM map URL where data files are available for download. Default value: *ftp://ftp.wwpdb.org/pub/emdb/structures/*.

The cryo-EM map files are also availabe at the following FTP server:

```
ftp://ftp.ebi.ac.uk/pub/databases/emdb/structures/
```

The cryoEM database ID is automatically retrieved from the downloaded PDB or CIF file. It is used to

generate the complete path name of the cryoEM map files:

```
ftp://ftp.wwpdb.org/pub/emdb/structures/EMD-<ID>/map/emd_<ID>.map.gz
```

In addition to map file, the following metadata file is automatically downloaded from FTP server:

```
ftp://ftp.wwpdb.org/pub/emdb/structures/EMD-<ID>/header/emd_<ID>.xml
```

--DenistyMapLocationURLXRay *Map URL*

Specify location of x-ray electron density map URL where data files are available for download. Default value: *http://www.ebi.ac.uk/pdbe/coordinates/files/*.

--EDMapLocationSuffixes *CompositeMap,None,...*

Specify file root suffixes for generating file names for x-ray electron density map files on a remote server. It is a pariwise comma delimited list of EDMapTypes and remote file suffixes. Default value: *CompositeMap, None, DifferenceMap, _diff, ReflectionMap, _map*.

The default names of the x-ray ED map files available on the server are shown below:

```
CompositeMap (2Fobs - Fcalc): <LowercasePDBID>.ccp4
DifferenceMap (Fobs - Fcalc): <LowercasePDBID>_diff.ccp4
ReflectionMap: <LowercasePDBID>_map.mtz
```

--EDMapTypes *CompositeMap,DifferenceMap,ReflectionMap,All*

Specify types of x-ray Electron Density (ED) map file(s) to download. It is either a comma delimited list of valid file types or All available file types. Possible values: *CompositeMap, DifferenceMap, ReflectionMap, All*. Default value: *CompositeMap,DifferenceMap*.

The CompositeMap (2Fobs - Fcalc) and DifferenceMap (Fobs - Fcalc) correspond to ED map data in CCP4 format. The ReflectionMap corresponds to ED map data in MTZ format.

-h, --help

Print this help message.

--indelim *comma | semicolon*

Input delimiter for CSV *TextFile* containing PDB IDs. Possible values: *comma or semicolon*. Default value: *comma*. For TSV files, this option is ignored and *tab* is used as a delimiter.

-m, --mode <IDsOnCmdLine | IDsInFile>

Indicate how PDB IDs are specified: PDB IDs are either present as space delimited command line parameters or in a specific column in a CSV/TSV text file. Possible values: *IDsOnCmdLine or IDsInFile*. Default: *IDsOnCmdLine*.

-p, --PDBIDsCol *number | string*

Column used to identify PDB ID(s) in a text file. Default value: First column containing text string PDB_ID or <PDBID>.

For *colnum* value of -c, --colmode option, input value is a column number. Example: *1*.

For *collabel* value of -c, --colmode option, input value is a column label. Example: *PDB_ID*.

This option is ignored during *IDsOnCmdLine* value of m, --mode option.

--PDBFormat *PDB, CIF or Auto*

Specify file format for downloading PDB files. Possible values: *PDB, CIF, auto*. Default value: *Auto*. The PDBID.pdb and PDBID.cif files are downloaded for *PDB* and *CIF* option values. The availability of PDB fies in both *PDB* and *CIF* format is checked for *Auto* option by attempting to download PDB.pdb file followed by PDBID.cif file.

The *PDB* format files are usually not available for structures determined using cryo-EM methodology.

-w, --WorkingDir *dirname*

Location of working directory. Default: current directory.

EXAMPLES

To retrieve a PDB file for PDB ID 2HYY and generate a local 2HYY.pdb file, type:

```
% DownloadPDBFiles.pl 2HYY
```

To retrieve a PDB file for PDB ID 2HYY along with electron density files and generate a local 2HYY.pdb and electron density map files 2HYY.ccp4 and 2HYY_diff.ccp4 corresponding to composit (2Fo - Fc) and difference maps (Fo - Fc), type:

```
% DownloadPDBFiles.pl --densityMap yes 2HYY
```

To retrieve PDB file for 5K12 in CIF format along with cryo-EM density file and generate a local 5K12.cif and density map file emd_8194.map.gz, type:

```
% DownloadPDBFiles.pl --densityMap yes --pdbFormat CIF 5K12
```

To retrieve PDB files for multiple PDB IDs 2HYY and 1KV2 and generate corresponding local PDB files, type:

```
% DownloadPDBFiles.pl 2HYY 1KV2
```

To retrieve PDB files for multiple PDB IDs 2HYY and 1KV2 and generate corresponding local PDB files along with appropriate x-ray electron density and cryo-EM density files, type:

```
% DownloadPDBFiles.pl --densityMap yes 2HYY 5K12
```

To download PDB files for PDB IDs present in column name PDB_ID or PDBID in SamplePDBIDs.csv file and generate correponding PDB files, type

```
% DownloadPDBFiles.pl -m IDsInFile SamplePDBIDs.csv
```

To download PDB files for PDB IDs present in a specific column name in SamplePDBIDs.csv file and generate correponding PDB files, type

```
% DownloadPDBFiles.pl -m IDsInFile -c collabel -p PDB_ID SamplePDBIDs.csv
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

ExtractFromPDBFiles.pl, InfoPDBFiles.pl, ModifyPDBFiles.pl

COPYRIGHT

Copyright (C) 2019 Manish Sud. All rights reserved.

This file is part of MayaChemTools.

MayaChemTools is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 3 of the License, or (at your option) any later version.