NAME

TopologicalPharmacophoreAtomTripletsFingerprints.pl - Generate topological pharmacophore atom triplets fingerprints for SD files

SYNOPSIS

TopologicalPharmacophoreAtomTripletsFingerprints.pl SDFile(s)...

TopologicalPharmacophoreAtomTripletsFingerprints.pl [--AromaticityModel *AromaticityModelType*] [--AtomTripletsSetSizeToUse *ArbitrarySize* | *FixedSize*] [-a, --AtomTypesToUse *"AtomType1, AtomType2..."*] [--AtomTypesWeight *"AtomType1, Weight1, AtomType2, Weight2..."*] [--CompoundID *DataFieldName or LabelPrefixString*] [--CompoundIDLabel *text*] [--CompoundIDMode] [--DataFields *"FieldLabel1, FieldLabel2,..."*] [-d, --DataFieldsMode *All* | *Common* | *Specify* | *CompoundID*] [--DistanceBinSize *number*] [-f, --Filter *Yes* | *No*] [--FingerprintsLabelMode *FingerprintsLabelOnly* | *FingerprintsLabelWithIDs*] [--FingerprintsLabel *text*] [-h, --help] [-k, --KeepLargestComponent *Yes* | *No*] [--MinDistance *number*] [--MaxDistance *number*] [--OutDelim *comma* | *tab* | *semicolon*] [--output *SD* | *FP* | *text* | *all*] [-o, --overwrite] [-q, --quote *Yes* | *No*] [-r, --root *RootName*] [-u, --UseTriangleInequality *Yes* | *No*] [-v, --VectorStringFormat *ValuesString, IDsAndValuesString* | *IDsAndValuesPairsString* | *ValuesAndIDsString* | *ValuesAndIDsPairsString*] [-w, --WorkingDir dirname] SDFile(s)...

DESCRIPTION

Generate topological pharmacophore atom triplets fingerprints [Ref 66, Ref 68-71] for *SDFile(s)* and create appropriate SD, FP or CSV/TSV text file(s) containing fingerprints vector strings corresponding to molecular fingerprints.

Multiple SDFile names are separated by spaces. The valid file extensions are *.sdf* and *.sd*. All other file names are ignored. All the SD files in a current directory can be specified either by **.sdf* or the current directory name.

Based on the values specified for --AtomTypesToUse, pharmacophore atom types are assigned to all non-hydrogen atoms in a molecule and a distance matrix is generated. Using --MinDistance, --MaxDistance, and --DistanceBinSize values, a binned distance matrix is generated with lower bound on the distance bin as the distance in distance matrix; the lower bound on the distance bin is also used as the distance between atom pairs for generation of atom triplet identifiers.

A pharmacophore atom triplets basis set is generated for all unique atom triplets constituting atom pairs binned distances between --MinDistance and --MaxDistance. The value of --UseTriangleInequality determines whether the triangle inequality test is applied during generation of atom triplets basis set. The lower distance bound, along with specified pharmacophore types, is used during generation of atom triplet IDs.

```
Let:

P = Valid pharmacophore atom type

Px = Pharmacophore atom x
Py = Pharmacophore atom y
Pz = Pharmacophore atom z

Dmin = Minimum distance corresponding to number of bonds between two atoms
Dmax = Maximum distance corresponding to number of bonds between two atoms
D = Distance corresponding to number of bonds between two atom

Bsize  = Distance bin size
Nbins = Number of distance bins

Dxy = Distance or lower bound of binned distance between Px and Py
Dxz = Distance or lower bound of binned distance between Px and Pz
Dyz = Distance or lower bound of binned distance between Py and Pz

Then:

PxDyz-PyDxz-PzDxy = Pharmacophore atom triplet IDs for atom types Px,
                    Py, and Pz

For example: H1-H1-H1, H2-HBA-H2 and so on

For default values of Dmin = 1 , Dmax = 10 and Bsize = 2:
```

```
the number of distance bins, Nbins = 5, are:

[1, 2] [3, 4] [5, 6] [7, 8] [9 10]

and atom triplet basis set size is 2692.

Atom triplet basis set size for various values of Dmin, Dmax and Bsize in
conjunction with usage of triangle inequality is:

Dmin    Dmax   Bsize   UseTriangleInequality   TripletBasisSetSize
1       10     2       No                      4960
1       10     2       Yes                     2692 [ Default ]
2       12     2       No                      8436
2       12     2       Yes                     4494
```

Using binned distance matrix and pharmacohore atom types, occurrence of unique pharmacohore atom triplets is counted.

The final pharmacophore atom triples count along with atom pair identifiers involving all non-hydrogen atoms constitute pharmacophore topological atom triplets fingerprints of the molecule.

For *ArbitrarySize* value of --AtomTripletsSetSizeToUse option, the fingerprint vector correspond to only those topological pharmacophore atom triplets which are present and have non-zero count. However, for *FixedSize* value of --AtomTripletsSetSizeToUse option, the fingerprint vector contains all possible valid topological pharmacophore atom triplets with both zero and non-zero count values.

Example of *SD* file containing topological pharmacophore atom triplets fingerprints string data:

```
... ...
... ...
$$$$
... ...
... ...
... ...
41 44  0  0  0  0  0  0  0  0999 V2000
 -3.3652    1.4499    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
... ...
2  3  1  0  0  0  0
... ...
M  END
>  <CmpdID>
Cmpd1

>  <TopologicalPharmacophoreAtomTripletsFingerprints>
FingerprintsVector;TopologicalPharmacophoreAtomTriplets:ArbitrarySize:
MinDistance1:MaxDistance10;696;NumericalValues;IDsAndValuesString;Ar1-
Ar1-Ar1 Ar1-Ar1-H1 Ar1-Ar1-HBA1 Ar1-Ar1-HBD1 Ar1-H1-H1 Ar1-H1-HBA1 Ar1
-H1-HBD1 Ar1-HBA1-HBD1 H1-H1-H1 H1-H1-HBA1 H1-H1-HBD1 H1-HBA1-HBA1 H1-
HBA1-HBD1 H1-HBA1-NI1 H1-HBD1-NI1 HBA1-HBA1-NI1 HBA1-HBD1-NI1 Ar1-...;
46 106 8 3 83 11 4 1 21 5 3 1 2 2 1 1 1 100 101 18 11 145 132 26 14 23
28 3 3 5 4 61 45 10 4 16 20 7 5 1 3 4 5 3 1 1 1 1 5 4 2 1 2 2 2 1 1 1
119 123 24 15 185 202 41 25 22 17 3 5 85 95 18 11 23 17 3 1 1 6 4 ...

$$$$
... ...
... ...
```

Example of *FP* file containing topological pharmacophore atom triplets fingerprints string data:

```
#
# Package = MayaChemTools 7.4
# Release Date = Oct 21, 2010
#
# TimeStamp = Fri Mar 11 15:38:58 2011
#
# FingerprintsStringType = FingerprintsVector
```

```
#
# Description = TopologicalPharmacophoreAtomTriplets:ArbitrarySize:M...
# VectorStringFormat = IDsAndValuesString
# VectorValuesType = NumericalValues
#
Cmpd1 696;Ar1-Ar1-Ar1 Ar1-Ar1-H1 Ar1-Ar1-HBA1 Ar1-Ar1-HBD1...;;46 106...
Cmpd2 251;H1-H1-H1 H1-H1-HBA1 H1-H1-HBD1 H1-H1-NI1...;4 1 3 1 1 2 2...
... ...
... ..
```

Example of CSV *Text* file containing topological pharmacophore atom triplets fingerprints string data:

```
"CompoundID","TopologicalPharmacophoreAtomTripletsFingerprints"
"Cmpd1","FingerprintsVector;TopologicalPharmacophoreAtomTriplets:Arbitr
arySize:MinDistance1:MaxDistance10;696;NumericalValues;IDsAndValuesStri
ng;Ar1-Ar1-Ar1 Ar1-Ar1-H1 Ar1-Ar1-HBA1 Ar1-Ar1-HBD1 Ar1-H1-H1 Ar1-H1-HB
A1 Ar1-H1-HBD1 Ar1-HBA1-HBD1 H1-H1-H1 H1-H1-HBA1 H1-H1-HBD1 H1-HBA1-HBA
1 H1-HBA1-HBD1 H1-HBA1-NI1 H1-HBD1-NI1 HBA1-HBA1-NI1 HBA1-HBD1-NI1 A...;
46 106 8 3 83 11 4 1 21 5 3 1 2 2 1 1 1 100 101 18 11 145 132 26 14 23
28 3 3 5 4 61 45 10 4 16 20 7 5 1 3 4 5 3 1 1 1 1 5 4 2 1 2 2 2 1 1 1
119 123 24 15 185 202 41 25 22 17 3 5 85 95 18 11 23 17 3 1 1 6 4 ...
... ...
... ...
```

The current release of MayaChemTools generates the following types of topological pharmacophore atom triplets fingerprints vector strings:

```
FingerprintsVector;TopologicalPharmacophoreAtomTriplets:ArbitrarySize:
MinDistance1:MaxDistance10;696;NumericalValues;IDsAndValuesString;Ar1-
Ar1-Ar1 Ar1-Ar1-H1 Ar1-Ar1-HBA1 Ar1-Ar1-HBD1 Ar1-H1-H1 Ar1-H1-HBA1 Ar1
-H1-HBD1 Ar1-HBA1-HBD1 H1-H1-H1 H1-H1-HBA1 H1-H1-HBD1 H1-HBA1-HBA1 H1-
HBA1-HBD1 H1-HBA1-NI1 H1-HBD1-NI1 HBA1-HBA1-NI1 HBA1-HBD1-NI1 Ar1-...;
46 106 8 3 83 11 4 1 21 5 3 1 2 2 1 1 1 100 101 18 11 145 132 26 14 23
28 3 3 5 4 61 45 10 4 16 20 7 5 1 3 4 5 3 1 1 1 1 5 4 2 1 2 2 2 1 1 1
119 123 24 15 185 202 41 25 22 17 3 5 85 95 18 11 23 17 3 1 1 6 4 ...

FingerprintsVector;TopologicalPharmacophoreAtomTriplets:FixedSize:MinD
istance1:MaxDistance10;2692;OrderedNumericalValues;ValuesString;46 106
8 3 0 0 83 11 4 0 0 0 1 0 0 0 0 0 0 0 0 21 5 3 0 0 1 2 2 0 0 1 0 0 0
0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 100 101 18 11 0 0 145 132 26
14 0 0 23 28 3 3 0 0 5 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 61 45 10 4 0
0 16 20 7 5 1 0 3 4 5 3 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 5 ...

FingerprintsVector;TopologicalPharmacophoreAtomTriplets:FixedSize:MinD
istance1:MaxDistance10;2692;OrderedNumericalValues;IDsAndValuesString;
Ar1-Ar1-Ar1 Ar1-Ar1-H1 Ar1-Ar1-HBA1 Ar1-Ar1-HBD1 Ar1-Ar1-NI1 Ar1-Ar1-P
I1 Ar1-H1-H1 Ar1-H1-HBA1 Ar1-H1-HBD1 Ar1-H1-NI1 Ar1-H1-PI1 Ar1-HBA1-HB
A1 Ar1-HBA1-HBD1 Ar1-HBA1-NI1 Ar1-HBA1-PI1 Ar1-HBD1-HBD1 Ar1-HBD1-...;
46 106 8 3 0 0 83 11 4 0 0 0 1 0 0 0 0 0 0 0 0 21 5 3 0 0 1 2 2 0 0 1
0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 100 101 18 11 0 0 145
132 26 14 0 0 23 28 3 3 0 0 5 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 61 ...
```

OPTIONS

--AromaticityModel *MDLAromaticityModel* | *TriposAromaticityModel* | *MMFFAromaticityModel* | *ChemAxonBasicAromaticityModel* | *ChemAxonGeneralAromaticityModel* | *DaylightAromaticityModel* | *MayaChemToolsAromaticityModel*

Specify aromaticity model to use during detection of aromaticity. Possible values in the current release are: *MDLAromaticityModel, TriposAromaticityModel, MMFFAromaticityModel, ChemAxonBasicAromaticityModel, ChemAxonGeneralAromaticityModel, DaylightAromaticityModel or MayaChemToolsAromaticityModel*. Default value: *MayaChemToolsAromaticityModel*.

The supported aromaticity model names along with model specific control parameters are defined in AromaticityModelsData.csv, which is distributed with the current release and is available under lib/data directory. Molecule.pm module retrieves data from this file during class instantiation and makes it available to method DetectAromaticity for detecting aromaticity corresponding to a specific model.

--AtomTripletsSetSizeToUse *ArbitrarySize* | *FixedSize*

Atom triplets set size to use during generation of topological pharmacophore atom triplets fingerprints.

Possible values: *ArbitrarySize* | *FixedSize*; Default value: *ArbitrarySize*.

For *ArbitrarySize* value of --AtomTripletsSetSizeToUse option, the fingerprint vector correspond to only those topological pharmacophore atom triplets which are present and have non-zero count. However, for *FixedSize* value of --AtomTripletsSetSizeToUse option, the fingerprint vector contains all possible valid topological pharmacophore atom triplets with both zero and non-zero count values.

-a, --AtomTypesToUse *"AtomType1,AtomType2,..."*

Pharmacophore atom types to use during generation of topological phramacophore atom triplets. It's a list of comma separated valid pharmacophore atom types.

Possible values for pharmacophore atom types are: *Ar, CA, H, HBA, HBD, Hal, NI, PI, RA*. Default value [Ref 71] : *HBD,HBA,PI,NI,H,Ar*.

The pharmacophore atom types abbreviations correspond to:

```
HBD: HydrogenBondDonor
HBA: HydrogenBondAcceptor
PI :  PositivelyIonizable
NI : NegativelyIonizable
Ar : Aromatic
Hal : Halogen
H : Hydrophobic
RA : RingAtom
CA : ChainAtom
```

AtomTypes::FunctionalClassAtomTypes module is used to assign pharmacophore atom types. It uses following definitions [Ref 60-61, Ref 65-66]:

```
HydrogenBondDonor: NH, NH2, OH
HydrogenBondAcceptor: N[!H], O
PositivelyIonizable: +, NH2
NegativelyIonizable: -, C(=O)OH, S(=O)OH, P(=O)OH
```

--CompoundID *DataFieldName or LabelPrefixString*

This value is --CompoundIDMode specific and indicates how compound ID is generated.

For *DataField* value of --CompoundIDMode option, it corresponds to datafield label name whose value is used as compound ID; otherwise, it's a prefix string used for generating compound IDs like LabelPrefixString<Number>. Default value, *Cmpd*, generates compound IDs which look like Cmpd<Number>.

Examples for *DataField* value of --CompoundIDMode:

```
MolID
ExtReg
```

Examples for *LabelPrefix* or *MolNameOrLabelPrefix* value of --CompoundIDMode:

```
Compound
```

The value specified above generates compound IDs which correspond to Compound<Number> instead of default value of Cmpd<Number>.

--CompoundIDLabel *text*

Specify compound ID column label for CSV/TSV text file(s) used during *CompoundID* value of --DataFieldsMode option. Default value: *CompoundID*.

--CompoundIDMode *DataField* | *MolName* | *LabelPrefix* | *MolNameOrLabelPrefix*

Specify how to generate compound IDs and write to FP or CSV/TSV text file(s) along with generated fingerprints for *FP* | *text* | *all* values of --output option: use a *SDFile(s)* datafield value; use molname line from *SDFile(s)*; generate a sequential ID with specific prefix; use combination of both MolName and LabelPrefix with usage of LabelPrefix values for empty molname lines.

Possible values: *DataField* | *MolName* | *LabelPrefix* | *MolNameOrLabelPrefix*. Default value: *LabelPrefix*.

For *MolNameAndLabelPrefix* value of --CompoundIDMode, molname line in *SDFile(s)* takes precedence over sequential compound IDs generated using *LabelPrefix* and only empty molname values are replaced with sequential compound IDs.

This is only used for *CompoundID* value of --DataFieldsMode option.

--DataFields *"FieldLabel1,FieldLabel2,..."*

Comma delimited list of *SDFiles(s)* data fields to extract and write to CSV/TSV text file(s) along with generated fingerprints for *text* | *all* values of --output option.

This is only used for *Specify* value of --DataFieldsMode option.

Examples:

```
Extreg
MolID,CompoundName
```

-d, --DataFieldsMode *All* | *Common* | *Specify* | *CompoundID*

Specify how data fields in *SDFile(s)* are transferred to output CSV/TSV text file(s) along with generated fingerprints for *text* | *all* values of --output option: transfer all SD data field; transfer SD data files common to all compounds; extract specified data fields; generate a compound ID using molname line, a compound prefix, or a combination of both. Possible values: *All* | *Common* | *specify* | *CompoundID*. Default value: *CompoundID*.

--DistanceBinSize *number*

Distance bin size used to bin distances between atom pairs in atom triplets. Default value: *2*. Valid values: positive integers.

For default --MinDistance and --MaxDistance values of 1 and 10 with --DistanceBinSize of 2 [Ref 70], the following 5 distance bins are generated:

```
[1, 2] [3, 4] [5, 6] [7, 8] [9 10]
```

The lower distance bound on the distance bin is uses to bin the distance between atom pairs in atom triplets. So in the previous example, atom pairs with distances 1 and 2 fall in first distance bin, atom pairs with distances 3 and 4 fall in second distance bin and so on.

In order to distribute distance bins of equal size, the last bin is allowed to go past --MaxDistance by up to distance bin size. For example, --MinDistance and --MaxDistance values of 2 and 10 with --DistanceBinSize of 2 generates the following 6 distance bins:

```
[2, 3] [4, 5] [6, 7] [8, 9] [10 11]
```

-f, --Filter *Yes* | *No*

Specify whether to check and filter compound data in SDFile(s). Possible values: *Yes or No*. Default value: *Yes*.

By default, compound data is checked before calculating fingerprints and compounds containing atom data corresponding to non-element symbols or no atom data are ignored.

--FingerprintsLabelMode *FingerprintsLabelOnly* | *FingerprintsLabelWithIDs*

Specify how fingerprints label is generated in conjunction with --FingerprintsLabel option value: use fingerprints label generated only by --FingerprintsLabel option value or append topological atom pair count value IDs to --FingerprintsLabel option value.

Possible values: *FingerprintsLabelOnly* | *FingerprintsLabelWithIDs*. Default value: *FingerprintsLabelOnly*.

Topological atom pairs IDs appended to --FingerprintsLabel value during *FingerprintsLabelWithIDs* values of --FingerprintsLabelMode correspond to atom pair count values in fingerprint vector string.

FingerprintsLabelWithIDs value of --FingerprintsLabelMode is ignored during *ArbitrarySize* value of --AtomTripletsSetSizeToUse option and topological atom triplets IDs not appended to the label.

--FingerprintsLabel *text*

SD data label or text file column label to use for fingerprints string in output SD or CSV/TSV text file(s) specified by --output. Default value: *TopologicalPharmacophoreAtomTripletsFingerprints*.

-h, --help

Print this help message.

-k, --KeepLargestComponent *Yes* | *No*

Generate fingerprints for only the largest component in molecule. Possible values: *Yes or No*. Default value: *Yes*.

For molecules containing multiple connected components, fingerprints can be generated in two different ways: use all connected components or just the largest connected component. By default, all atoms except for the largest connected component are deleted before generation of fingerprints.

--MinDistance *number*

Minimum bond distance between atom pairs corresponding to atom triplets for generating topological pharmacophore atom triplets. Default value: *1*. Valid values: positive integers and less than --MaxDistance.

--MaxDistance *number*

Maximum bond distance between atom pairs corresponding to atom triplets for generating topological pharmacophore atom triplets. Default value: *10*. Valid values: positive integers and greater than --MinDistance.

--OutDelim *comma* | *tab* | *semicolon*

Delimiter for output CSV/TSV text file(s). Possible values: *comma, tab, or semicolon* Default value: *comma*.

--output *SD* | *FP* | *text* | *all*

Type of output files to generate. Possible values: *SD, FP, text, or all*. Default value: *text*.

-o, --overwrite

Overwrite existing files.

-q, --quote *Yes* | *No*

Put quote around column values in output CSV/TSV text file(s). Possible values: *Yes or No*. Default value: *Yes*.

-r, --root *RootName*

New file name is generated using the root: <Root>.<Ext>. Default for new file names: <SDFileName><TopologicalPharmacophoreAtomTripletsFP>.<Ext>. The file type determines <Ext> value. The sdf, fpf, csv, and tsv <Ext> values are used for SD, FP, comma/semicolon, and tab delimited text files, respectively.This option is ignored for multiple input files.

-u, --UseTriangleInequality *Yes* | *No*

Specify whether to imply triangle distance inequality test to distances between atom pairs in atom triplets during generation of atom triplets basis set generation. Possible values: *Yes or No*. Default value: *Yes*.

Triangle distance inequality test implies that distance or binned distance between any two atom pairs in an atom triplet must be less than the sum of distances or binned distances between other two atoms pairs and greater than the difference of their distances.

```
For atom triplet PxDyz-PyDxz-PzDxy to satisfy triangle inequality:

Dyz > |Dxz - Dxy| and Dyz < Dxz + Dxy
Dxz > |Dyz - Dxy| and Dyz < Dyz + Dxy
Dxy > |Dyz - Dxz| and Dxy < Dyz + Dxz
```

-v, --VectorStringFormat *ValuesString, IDsAndValuesString* | *IDsAndValuesPairsString* | *ValuesAndIDsString* | *ValuesAndIDsPairsString*

Format of fingerprints vector string data in output SD, FP or CSV/TSV text file(s) specified by --output option. Possible values: *ValuesString, IDsAndValuesString* | *IDsAndValuesPairsString* | *ValuesAndIDsString* | *ValuesAndIDsPairsString*. Defaultvalue: *ValuesString*.

Default value during *FixedSize* value of --AtomTripletsSetSizeToUse option: *ValuesString*. Default value during *ArbitrarySize* value of --AtomTripletsSetSizeToUse option: *IDsAndValuesString*.

ValuesString option value is not allowed for *ArbitrarySize* value of --AtomTripletsSetSizeToUse option.

Examples:

```
FingerprintsVector;TopologicalPharmacophoreAtomTriplets:ArbitrarySize:
MinDistance1:MaxDistance10;696;NumericalValues;IDsAndValuesString;Ar1-
Ar1-Ar1 Ar1-Ar1-H1 Ar1-Ar1-HBA1 Ar1-Ar1-HBD1 Ar1-H1-H1 Ar1-H1-HBA1 Ar1
-H1-HBD1 Ar1-HBA1-HBD1 H1-H1-H1 H1-H1-HBA1 H1-H1-HBD1 H1-HBA1-HBA1 H1-
HBA1-HBD1 H1-HBA1-NI1 H1-HBD1-NI1 HBA1-HBA1-NI1 HBA1-HBD1-NI1 Ar1-...;
46 106 8 3 83 11 4 1 21 5 3 1 2 2 1 1 1 100 101 18 11 145 132 26 14 23
28 3 3 5 4 61 45 10 4 16 20 7 5 1 3 4 5 3 1 1 1 1 5 4 2 1 2 2 2 1 1 1
119 123 24 15 185 202 41 25 22 17 3 5 85 95 18 11 23 17 3 1 1 6 4 ...

FingerprintsVector;TopologicalPharmacophoreAtomTriplets:FixedSize:MinD
istance1:MaxDistance10;2692;OrderedNumericalValues;ValuesString;46 106
8 3 0 0 83 11 4 0 0 0 1 0 0 0 0 0 0 0 0 0 21 5 3 0 0 1 2 2 0 0 1 0 0 0
0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 100 101 18 11 0 0 145 132 26
14 0 0 23 28 3 3 0 0 5 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 61 45 10 4 0
```

```
0 16 20 7 5 1 0 3 4 5 3 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 5 ...

FingerprintsVector;TopologicalPharmacophoreAtomTriplets:FixedSize:MinD
istance1:MaxDistance10;2692;OrderedNumericalValues;ValuesAndIDsPairsSt
ring;46 Ar1-Ar1-Ar1 106 Ar1-Ar1-H1 8 Ar1-Ar1-HBA1 3 Ar1-Ar1-HBD1 0 Ar1
-Ar1-NI1 0 Ar1-Ar1-PI1 83 Ar1-H1-H1 11 Ar1-H1-HBA1 4 Ar1-H1-HBD1 0 Ar1
-H1-NI1 0 Ar1-H1-PI1 0 Ar1-HBA1-HBA1 1 Ar1-HBA1-HBD1 0 Ar1-HBA1-NI1 0
Ar1-HBA1-PI1 0 Ar1-HBD1-HBD1 0 Ar1-HBD1-NI1 0 Ar1-HBD1-PI1 0 Ar1-NI...
```

-w, --WorkingDir *DirName*

Location of working directory. Default value: current directory.

EXAMPLES

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs along with fingerprints vector strings data in ValuesString format, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl -r SampleTPATFP
  -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of fixed size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs along with fingerprints vector strings data in ValuesString format, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl
  --AtomTripletsSetSizeToUse FixedSize -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create SampleTPATFP.sdf, SampleTPATFP.fpf and SampleTPATFP.csv files with CSV file containing sequential compound IDs along with fingerprints vector strings data in ValuesString format, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl --output all
  -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs along with fingerprints vector strings data in ValuesString format and atom triplets IDs in the fingerprint data column label starting with Fingerprints, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl
  --FingerprintsLabelMode FingerprintsLabelWithIDs --FingerprintsLabel
  Fingerprints -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances not satisfying triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs along with fingerprints vector strings data in ValuesString format, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl
  --UseTriangleInequality No -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 6 distance bins spanning distances from 1 through 12 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs along with fingerprints vector strings data in ValuesString format, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl
  --UseTriangleInequality Yes --MinDistance 1 --MaxDistance 12
  --DistanceBinSIze 2 -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 6 distance bins spanning distances from 1 through 12 using "HBD,HBA,PI, NI, H, Ar" atoms with distances satisfying

triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs along with fingerprints vector strings data in ValuesString format, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl
  --AtomTypesToUse "HBD,HBA,PI,NI,H,Ar" --UseTriangleInequality Yes
  --MinDistance 1 --MaxDistance 12 --DistanceBinSIze 2
  --VectorStringFormat ValuesString -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs from molecule name line along with fingerprints vector strings data in ValuesString format, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl --DataFieldsMode
  CompoundID -CompoundIDMode MolName  -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs using specified data field along with fingerprints vector strings data in ValuesString format, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl --DataFieldsMode
  CompoundID -CompoundIDMode DataField --CompoundID Mol_ID
  -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing sequential compound IDs using combination of molecule name line and an explicit compound prefix along with fingerprints vector strings data, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl --DataFieldsMode
  CompoundID -CompoundIDMode MolnameOrLabelPrefix
  --CompoundID Cmpd --CompoundIDLabel MolID -r SampleSampleTPATFP
  -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing specific data fields columns along with fingerprints vector strings data, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl --DataFieldsMode
  Specify --DataFields Mol_ID -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create a SampleTPATFP.csv file containing common data fields columns along with fingerprints vector strings data, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl --DataFieldsMode
  Common -r SampleTPATFP -o Sample.sdf
```

To generate topological pharmacophore atom triplets fingerprints of arbitrary size corresponding to 5 distance bins spanning distances from 1 through 10 using default atoms with distances satisfying triangle inequality and create SampleTPATFP.sdf, SampleTPATFP.fpf and SampleTPATFP.csv files containing all data fields columns in CSV file along with fingerprints data, type:

```
% TopologicalPharmacophoreAtomTripletsFingerprints.pl --DataFieldsMode
  All  --output all -r SampleTPATFP -o Sample.sdf
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

InfoFingerprintsFiles.pl, SimilarityMatricesFingerprints.pl, AtomNeighborhoodsFingerprints.pl, ExtendedConnectivityFingerprints.pl, MACCSKeysFingerprints.pl, PathLengthFingerprints.pl, TopologicalAtomPairsFingerprints.pl, TopologicalAtomTorsionsFingerprints.pl,

COPYRIGHT TopologicalPharmacophoreAtomPairsFingerprints.pl