NAME

TorsionStrainEnergyAlerts

SYNOPSIS

import TorsionStrainEnergyAlerts

DESCRIPTION

TorsionStrainEnergyAlerts module provides the following functions:

TorsionStrainEnergyAlerts.GetTorsionLibraryFilePath, TorsionStrainEnergyAlerts.IdentifyTorsionLibraryAlertsForRotatableBonds, TorsionStrainEnergyAlerts.ListTorsionLibraryInfo, TorsionStrainEnergyAlerts.TorsionStrainEnergyAlerts

FUNCTIONS

TorsionStrainEnergyAlerts.GetTorsionLibraryFilePath

```
GetTorsionLibraryFilePath(self)
```

Get torsion strain library file path.

Arguments:

```
None
```

Returns:

```
FilePath (str): Torsion strain library path.
```

TorsionStrainEnergyAlerts.IdentifyTorsionLibraryAlertsForRotatableBonds

```
IdentifyTorsionLibraryAlertsForRotatableBonds(self, Mol)
```

Identify torsion strain library alerts for a molecule by matching rotatable bonds against SMARTS patterns specified for torsion rules in torsion energy library file.

Arguments:

```
Mol (object): RDKit molecule object.
```

Returns:

```
bool: True - Molecule contains strained torsions; False - Molecule
    contains no strained torsions or rotatable bonds.
dict or None: Torsion alerts information regarding matching of
    rotatable bonds to torsion strain library.
```

Example(s):

```
from TorsionAlerts.TorsionStrainEnergyAlerts import
    TorsionStrainEnergyAlerts

StrainEnergyAlerts = TorsionStrainEnergyAlerts()
AlertsStatus, AlertsInfo = StrainEnergyAlerts.
    IdentifyTorsionLibraryAlertsForRotatableBonds(RDKitMol)

RotBondsAlertsStatus = AlertsInfo["RotBondsAlertsStatus"]
TotalEnergy = AlertsInfo["TotalEnergy"]
TotalEnergyLowerBound = AlertsInfo["TotalEnergyLowerBound"]
TotalEnergyUpperBound = AlertsInfo["TotalEnergyUpperBound"]
AnglesNotObservedCount = AlertsInfo["AnglesNotObservedCount"]
MaxSingleEnergy = AlertsInfo["MaxSingleEnergy"])
MaxSingleEnergyAlertsCount = AlertsInfo[
    "MaxSingleEnergyAlertsCount"]

# List of rotatable bond IDs...
RotatableBondIDs = AlertsInfo["IDs"]
```

```
# Dictionaries containing information for rotatable bonds by using
# bond ID as key...
for ID in AlertsInfo["IDs"]:
    MatchStatus = AlertsInfo["MatchStatus"][ID]
    MaxSingleEnergyAlertStatus = AlertsInfo[
        "MaxSingleEnergyAlertStatus"][ID]
    AtomIndices = AlertsInfo["AtomIndices"][ID]
    TorsionAtomIndices = AlertsInfo["TorsionAtomIndices"][ID]
    TorsionAngle = AlertsInfo["TorsionAngle"][ID]
    HierarchyClassName = AlertsInfo["HierarchyClassName"][ID]
    HierarchySubClassName = AlertsInfo["HierarchySubClassName"][ID]
    TorsionRuleNodeID = AlertsInfo["TorsionRuleNodeID"][ID]
    TorsionRuleSMARTS = AlertsInfo["TorsionRuleSMARTS"][ID]
    EnergyMethod = AlertsInfo["EnergyMethod"][ID]
    AngleNotObserved = AlertsInfo["AngleNotObserved"][ID]
    Energy = AlertsInfo["Energy"][ID]
    EnergyLowerBound = AlertsInfo["EnergyLowerBound"][ID]
    EnergyUpperBound = AlertsInfo["EnergyUpperBound"][ID]
```

TorsionStrainEnergyAlerts.ListTorsionLibraryInfo

```
ListTorsionLibraryInfo(self)
```

List torsion strain library information.

Arguments:

```
None
```

Returns:

```
None
```

TorsionStrainEnergyAlerts.TorsionStrainEnergyAlerts

```
    TorsionStrainEnergyAlerts(self, AlertsMode = "TotalEnergy", TotalEnergyCutoff = 6.0,
MaxSingleEnergyCutoff = 1.8, RotBondsSMARTSMode = "SemiStrict", RotBondsSMARTSPattern =
None, TorsionLibraryFilePath = "auto", AlertTorsionsNotObserved = False)
```

Identify strained molecules based on torsion strain energy library [Ref 153] alerts by matching rotatable bonds against SMARTS patterns specified for torsion rules in a torsion energy library file. The molecules must have 3D coordinates. The default torsion strain energy library file, TorsionStrainEnergyLibrary.xml, is ia available in the directory containing this file.

The data in torsion strain energy library file is organized in a hierarchical manner. It consists of one generic class and six specific classes at the highest level. Each class contains multiple subclasses corresponding to named functional groups or substructure patterns. The subclasses consist of torsion rules sorted from specific to generic torsion patterns. The torsion rule, in turn, contains a list of peak values for torsion angles and two tolerance values. A pair of tolerance values define torsion bins around a torsion peak value.

A strain energy calculation method, 'exact' or 'approximate' [Ref 153], is associated with each torsion rule for calculating torsion strain energy. The 'exact' stain energy calculation relies on the energy bins available under the energy histogram consisting of 36 bins covering angles from -180 to 180. The width of each bin is 10 degree. The energy bins are are defined at the right end points. The first and the last energy bins correspond to -170 and 180 respectively. The torsion angle is mapped to a energy bin. An angle offset is calculated for the torsion angle from the the right end point angle of the bin. The strain energy is estimated for the angle offset based on the energy difference between the current and previous bins. The torsion strain energy, in terms of torsion energy units (TEUs), corresponds to the sum of bin strain energy and the angle offset strain energy.

```
Energy = BinEnergyDiff/10.0 * BinAngleOffset + BinEnergy[BinNum]

Where:

BinEnergyDiff = BinEnergy[BinNum] - BinEnergy[PreviousBinNum]
BinAngleOffset = TorsionAngle - BinAngleRightSide
```

The 'approximate' strain energy calculation relies on the angle difference between a torsion angle and the torsion peaks observed for the torsion rules in the torsion energy library. The torsion angle is matched to a torsion peak based on the value of torsion angle difference. It must be less than or equal to the value for the second tolerance 'tolerance2'. Otherwise, the torsion angle is not observed in the torsion energy library and a value of 'NA' is assigned for torsion energy along with the lower and upper bounds on energy at 95% confidence interval. The 'approximate' torsion energy (TEUs) for observed torsion angle is calculated using the following formula:

```
Energy = beta_1 * (AngleDiff ** 2) + beta_2 * (AngleDiff ** 4)
```

The coefficients 'beta_1' and 'beta_2' are available for the observed angles in the torsion strain energy library. The 'AngleDiff' is the difference between the torsion angle and the matched torsion peak.

For example:

```
<library>
    <hierarchyClass id1="G" id2="G" name="GG">
    ...
    </hierarchyClass>
    <hierarchyClass id1="C" id2="O" name="CO">
        <hierarchySubClass name="Ester bond I" smarts="O=[C:2][O:3]">
            <torsionRule method="exact" smarts=
                "[O:1]=[C:2]!@[O:3]~[CH0:4]">
                <angleList>
                    <angle score="56.52" tolerance1="20.00"
                    tolerance2="25.00" value="0.0"/>
                </angleList>
                <histogram>
                    <bin count="1"/>
                    ...
                </histogram>
                <histogram_shifted>
                    <bin count="0"/>
                    ...
                </histogram_shifted>
                <histogram_converted>
                    <bin energy="4.67... lower="2.14..." upper="Inf"/>
                    ...
                    <bin energy="1.86..." lower="1.58..." upper="2.40..."/>
                    ...
                   </histogram_converted>
            </torsionRule>
            <torsionRule method="approximate" smarts=
                "[cH0:1][c:2]([cH0])!@[O:3][p:4]">
                <angleList>
                <angle beta_1="0.002..." beta_2="-7.843...e-07"
                    score="27.14" theta_0="-90.0" tolerance1="30.00"
                    tolerance2="45.00" value="-90.0"/>
                ...
                </angleList>
                <histogram>
                    <bin count="0"/>
                     ...
                </histogram>
                <histogram_shifted>
                    <bin count="0"/>
                    ...
                </histogram_shifted>
            </torsionRule>
        ...
     ...
    </hierarchyClass>
     <hierarchyClass id1="N" id2="C" name="NC">
     ...
    </hierarchyClass>
    <hierarchyClass id1="S" id2="N" name="SN">
```

```
        ...
        </hierarchyClass>
        <hierarchyClass id1="C" id2="S" name="CS">
        ...
        </hierarchyClass>
        <hierarchyClass id1="C" id2="C" name="CC">
        ...
        </hierarchyClass>
        <hierarchyClass id1="S" id2="S" name="SS">
         ...
        </hierarchyClass>
    </library>
```

The rotatable bonds in a 3D molecule are identified using a default SMARTS pattern. A custom SMARTS pattern may be optionally specified to detect rotatable bonds. Each rotatable bond is matched to a torsion rule in the torsion strain energy library. The strain energy is calculated for each rotatable bond using the calculation method, 'exact' or 'approximate', associated with the matched torsion rule.

The total strain energy (TEUs) of a molecule corresponds to the sum of 'exact' and 'approximate' strain energies calculated for all matched rotatable bonds in the molecule. The total strain energy is set to 'NA' for molecules containing a 'approximate' energy estimate for a torsion angle not observed in the torsion energy library. In addition, the lower and upper bounds on energy at 95% confidence interval are set to 'NA'.

Arguments:

```
    AlertsMode (str): Torsion strain energy library alert types to use
        for issuing alerts. Possible values: TotalEnergy,
        MaxSingleEnergy, or TotalOrMaxSingleEnergy.
    TotalEnergyCutoff (float): Total strain strain energy (TEUs) cutoff.
    MaxSingleEnergyCutoff (float): Maximum single strain energy (TEUs)
        cutoff.
    RotBondsSMARTSMode (str): SMARTS pattern to use for identifying
        rotatable bonds in a molecule. Possible values: NonStrict,
        SemiStrict, Strict or Specify.
    RotBondsSMARTSPattern (str):SMARTS pattern for identifying rotatable
        bonds. This paramater is only valid for ’Specify’ value
        ’RotBondsSMARTSMode’.
    TorsionLibraryFilePath (str): A XML file name containing data for
        torsion starin energy library.
    AlertTorsionsNotObserved (bool): Issue alerts about molecules
        containing torsion angles not observed in torsion strain energy
        library.
```

Returns:

```
    object: An instantiated class object.
```

Example(s):

```
    StrainEnergyAlertsHandle = TorsionStrainEnergyAlerts()

    StrainEnergyAlertsHandle = TorsionStrainEnergyAlerts(AlertsMode = "TotalEnergy",
        TotalEnergyCutoff = 6.0, MaxSingleEnergyCutoff = 1.8, RotBondsSMARTSMode =
"SemiStrict",
        RotBondsSMARTSPattern = None, TorsionLibraryFilePath = "auto",
AlertTorsionsNotObserved = False)
```

The following sections provide additional details for the parameters.

AlertsMode: Torsion strain energy library alert types to use for issuing alerts about molecules containing rotatable bonds based on the calculated values for the total torsion strain energy of a molecule and the maximum single strain energy of a rotatable bond in a molecule.

Possible values: TotalEnergy, MaxSingleEnergy, or TotalOrMaxSingleEnergy

The strain energy cutoff values in terms of torsion energy units (TEUs) are used to filter molecules as shown below:

```
AlertsMode                AlertsEnergyCutoffs (TEUs)

TotalEnergy               >= TotalEnergyCutoff

MaxSingleEnergy           >= MaxSingleEnergyCutoff

TotalOrMaxSingleEnergy    >= TotalEnergyCutoff
                          or >= MaxSingleEnergyCutoff
```

TotalEnergyCutoff: Total strain strain energy (TEUs) cutoff [Ref 153] for issuing alerts based on total strain energy for all rotatable bonds in a molecule. This option is used during 'TotalEnergy' or 'TotalOrMaxSingleEnergy' values of 'AlertsMode' parameter.

The total strain energy must be greater than or equal to the specified cutoff value to identify a molecule containing strained torsions.

MaxSingleEnergyCutoff: Maximum single strain energy (TEUs) cutoff [Ref 153] for issuing alerts based on the maximum value of a single strain energy of a rotatable bond in a molecule. This option is used during 'MaxSingleEnergy' or 'TotalOrMaxSingleEnergy' values of 'AlertsMode' parameter.

The maximum single strain energy must be greater than or equal to the specified cutoff valueo to identify a molecule containing strained torsions.

RotBondsSMARTSMode: SMARTS pattern to use for identifying rotatable bonds in a molecule for matching against torsion rules in the torsion library. Possible values: NonStrict, SemiStrict, Strict or Specify. The rotatable bond SMARTS matches are filtered to ensure that each atom in the rotatable bond is attached to at least two heavy atoms.

The following SMARTS patterns are used to identify rotatable bonds for different modes:

```
NonStrict: [!$(*#*)&!D1]-&!@[!$(*#*)&!D1]

SemiStrict:
[!$(*#*)&!D1&!$(C(F)(F)F)&!$(C(Cl)(Cl)Cl)&!$(C(Br)(Br)Br)
&!$(C([CH3])([CH3])[CH3])]-!@[!$(*#*)&!D1&!$(C(F)(F)F)
&!$(C(Cl)(Cl)Cl)&!$(C(Br)(Br)Br)&!$(C([CH3])([CH3])[CH3])]

Strict:
[!$(*#*)&!D1&!$(C(F)(F)F)&!$(C(Cl)(Cl)Cl)&!$(C(Br)(Br)Br)
&!$(C([CH3])([CH3])[CH3])&!$([CD3](=[N,O,S])-!@[#7,O,S!D1])
&!$([#7,O,S!D1]-!@[CD3]=[N,O,S])&!$([CD3](=[N+])-!@[#7!D1])
&!$([#7!D1]-!@[CD3]=[N+])]-!@[!$(*#*)&!D1&!$(C(F)(F)F)
&!$(C(Cl)(Cl)Cl)&!$(C(Br)(Br)Br)&!$(C([CH3])([CH3])[CH3])]
```

The 'NonStrict' and 'Strict' SMARTS patterns are available in RDKit. The 'NonStrict' SMARTS pattern corresponds to original Daylight SMARTS specification for rotatable bonds. The 'SemiStrict' SMARTS pattern is derived from 'Strict' SMARTS pattern.

TorsionLibraryFilePath (str): Specify a XML file name containing data for torsion starin energy library hierarchy or use default file, TorsionEnergyLibrary.xml, available in the directory containing this file.

AlertTorsionsNotObserved: Issue alerts abpout molecules con ntaining torsion angles not observed in torsion strain energy library. It's not possible to calculate torsion strain energies for these torsions during 'approximate' match to a specified torsion in the library.

The 'approximate' strain energy calculation relies on the angle difference between a torsion angle and the torsion peaks observed for the torsion rules in the torsion energy library. The torsion angle is matched to a torsion peak based on the value of torsion angle difference. It must be less than or equal to the value for the second tolerance 'tolerance2'. Otherwise, the torsion angle is not observed in the torsion energy library and a value of 'NA' is assigned for torsion energy along with the lower and upper bounds on energy at 95% confidence interval.

AUTHOR

Manish Sud <msud@san.rr.com>

Collaborator

Pat Walters

COPYRIGHT

This module uses the torsion strain energy library developed by Gu, S.; Smith, M. S.; Yang, Y.; Irwin, J. J.; Shoichet, B. K. [Ref 153].

The torsion strain enegy library is based on the Torsion Library jointly developed by the University of Hamburg, Center for Bioinformatics, Hamburg, Germany and F. Hoffmann-La-Roche Ltd., Basel, Switzerland.

This file is part of MayaChemTools.