The new RDKit chirality code

A tag-team presentation from Greg and Ricardo@Schrödinger

Finding possible stereochemistry in molecules

T5 Informatics GmbH
greg.landrum@t5informatics.com
@dr_greg_landrum greglandrum

Status before v2020.09

- `AssignStereochemistry()` identifies stereogenic atoms and bonds and assigns "CIP" codes to them. This is called by default when constructing a molecule if sanitization is enabled.
  - Stereogenic atoms with assigned stereochemistry are assigned the property `_CIPCode` with an R/S label
  - Stereogenic atoms without assigned stereochemistry are assigned the property `_ChiralityPossible`
  - Stereogenic bonds with assigned stereochemistry have their BondStereo set to either `STEREOE` or `STEREOZ` and have StereoAtoms assigned
  - Assigned stereochemistry is removed from atoms/bonds that are not stereogenic
- Note that the "CIP" codes which are assigned are quite approximate. Ricardo will say more about this

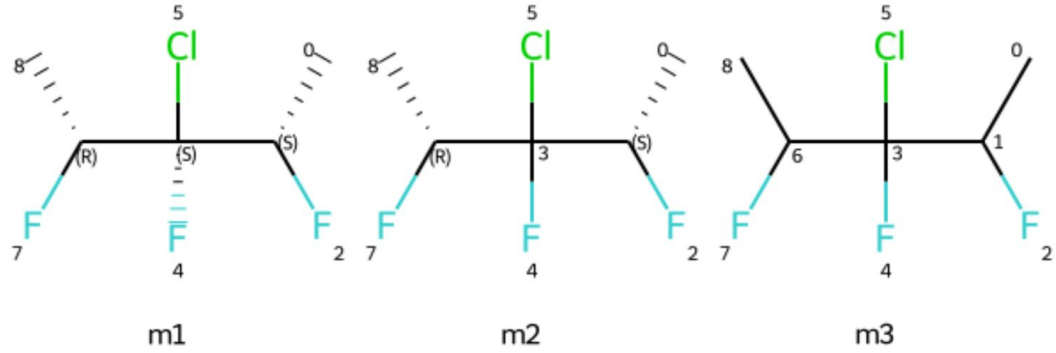
Example of results with previous RDKit versions

The potential chiral center on C3 in m2 is found, but it is not found in m3 since neither C1 nor C6 are specified.

In [103]:

```
IPythonConsole.drawOptions.addAtomIndices = True
IPythonConsole.drawOptions.addStereoAnnotation = True
m1 = Chem.MolFromSmiles('C[C@H](F)[C@](F)(Cl)[C@H](F)C')
m2 = Chem.MolFromSmiles('C[C@H](F)C(F)(Cl)[C@H](F)C')
m3 = Chem.MolFromSmiles('CC(F)C(F)(Cl)C(F)C')
Draw.MolsToGridImage((m1,m2,m3),legends=('m1','m2','m3'))
```

Out[103]:



In [90]:

```
# The old behavior, still the default for FindMolChiralCenters()
for m in (m1,m2,m3):
    print(Chem.FindMolChiralCenters(m,force=True,
                                    includeUnassigned=True))
```

```
[(1, 'S'), (3, 'S'), (6, 'R')]
[(1, 'S'), (3, '?'), (6, 'R')]
[(1, '?'), (6, '?')]
```

Changes in v2020.09

- A new algorithm is available with the function `FindPotentialStereo()`
- This identifies all stereogenic atoms and bonds, but does not assign CIP codes.
- The new code returns `StereoInfo` data structures which are more flexible and future proof than the current atom/bond centered labels:
  - `type: Atom_Tetrahedral, Bond_Double, Bond_Cumulene_Even, Bond_Atropisomer`
  - `specified: Unspecified, Specified, Unknown`
  - `centeredOn:` index of atom or bond
  - `descriptor: Tet_CW, Tet_CCW, Bond_Cis, Bond_Trans`
  - `controllingAtoms:` vector of the atoms the descriptor refers to

Example of results with v2020.09

The potential chiral center on C3 is found in m2 and m3

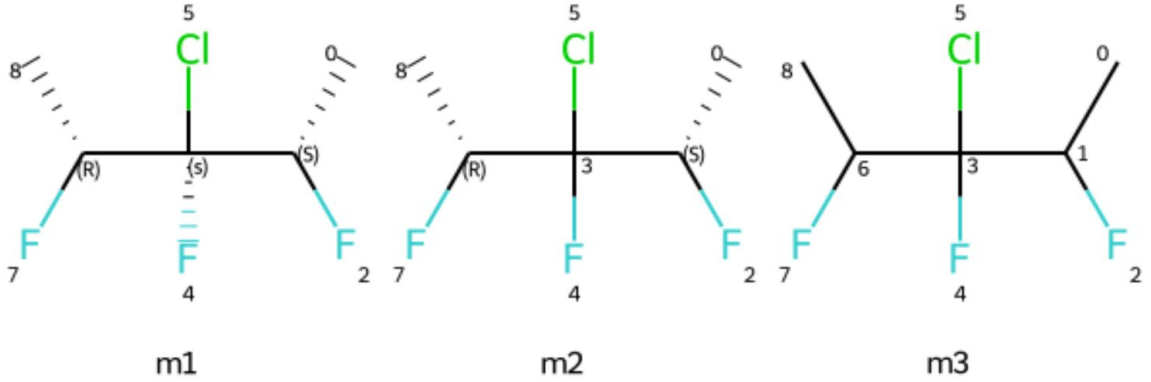
Additionally the correct CIP label - s - is assigned in m1

In [107]:

```
# Using the new code
for m in (m1,m2,m3):
    print(Chem.FindMolChiralCenters(m,force=True,
                                    includeUnassigned=True,
                                    useLegacyImplementation=False))
Draw.MolsToGridImage((m1,m2,m3),legends=('m1','m2','m3'))
```

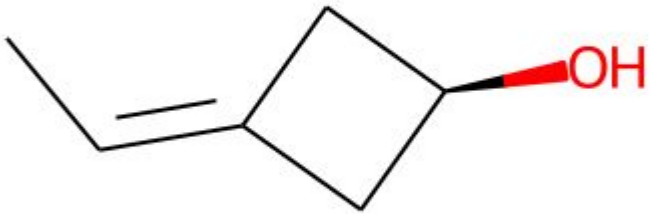
```
[(1, 'S'), (3, 's'), (6, 'R')]
[(1, 'S'), (3, '?'), (6, 'R')]
[(1, '?'), (3, '?'), (6, '?')]
```

Out[107]:



A situation we still don't handle correctly

Github #2984



The interaction of ring stereochemistry and double bond stereochemistry is not recognized/handled.

Ricardo is going to say more about this too