



Schrödinger

Why RegistrationHash?

Dan Nealschneider
Senior Staff Developer

<https://github.com/rdkit/rdkit/pull/5360>





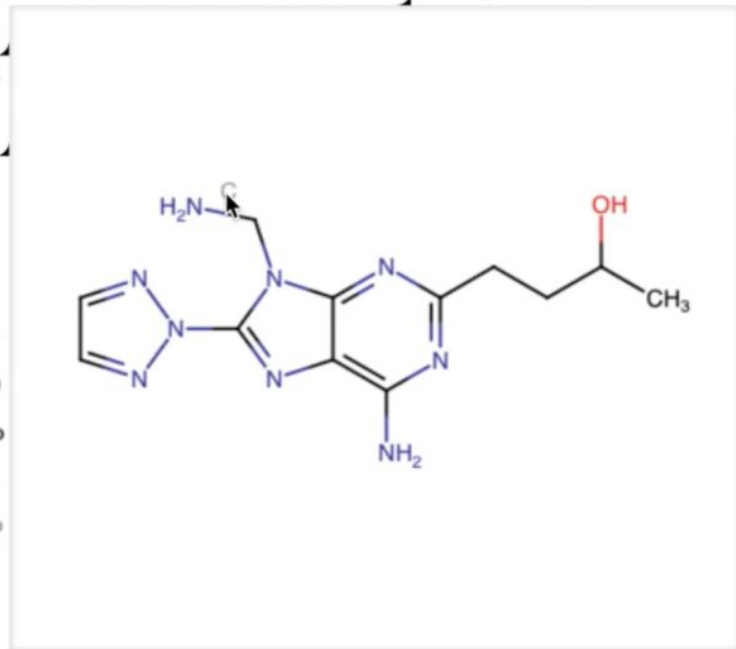
57

To preview properties as you sketch, add some Property columns to your Live Report.

ADVANCED

Open Live Report

A2A IC50

H
C
N
O
S
F
P
Cl
Br
I
*
A

Add Idea

Substructure Search

Similarity Search

Search Exact Match

☒ Preview Properties

Add Idea to Live Report

SEARCH BY ID

IMPORT FROM FILE

RUN ENUMERATION

Compound Structure
(Aligned)

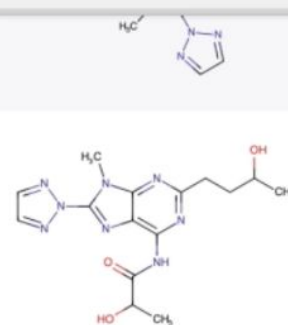
ID

Adenosine
Receptor
(A2A IC50)

A2A pIC50

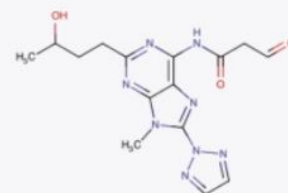
ADME
MPO

53



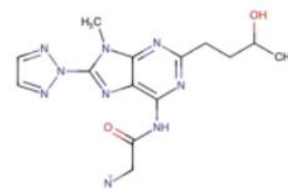
V103193

54



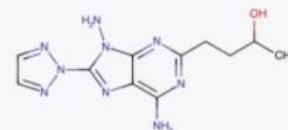
V103194

55



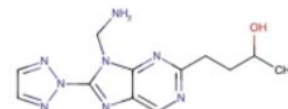
V103195

56



V105113

57



V105114

0.695

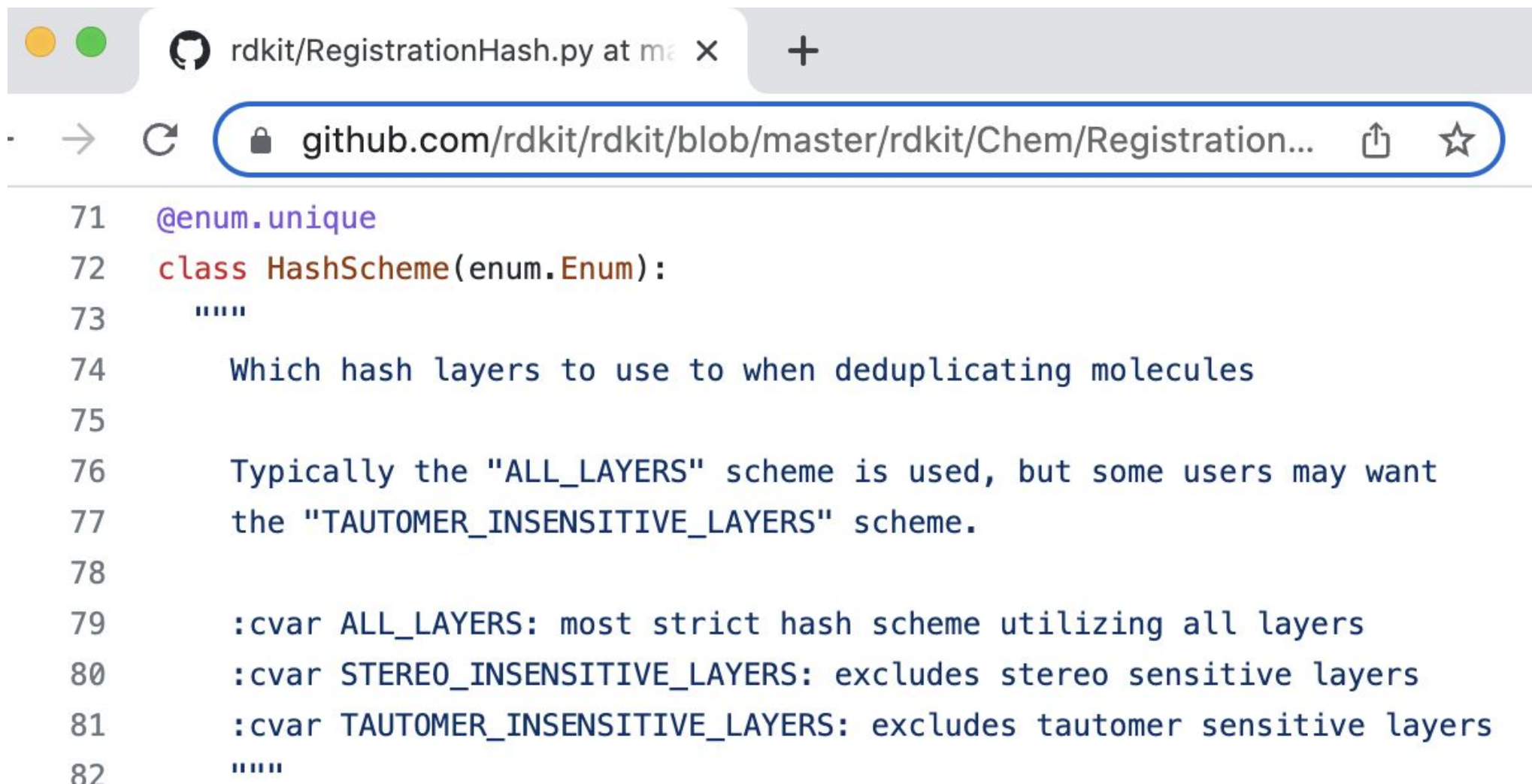
0.716

0.443

0.677



“Just chemically the same, man”



```
71 @enum.unique
72 class HashScheme(enum.Enum):
73     """
74     Which hash layers to use to when deduplicating molecules
75
76     Typically the "ALL_LAYERS" scheme is used, but some users may want
77     the "TAUTOMER_INSENSITIVE_LAYERS" scheme.
78
79     :cvar ALL_LAYERS: most strict hash scheme utilizing all layers
80     :cvar STEREO_INSENSITIVE_LAYERS: excludes stereo sensitive layers
81     :cvar TAUTOMER_INSENSITIVE_LAYERS: excludes tautomer sensitive layers
82     """
```

Why not SMILES?

Missing information

- *Enhanced stereochemistry*
- Polymer brackets
- Some annotations matter!
 - People use annotations for atropisomers

Extra information

- Map numbers
- Tautomeric specificity

CXSMILES

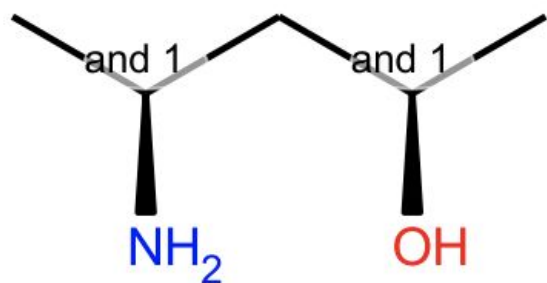
Missing information

- ? *Enhanced stereochemistry*
- ✓ Polymer brackets
- ✓ Some annotations matter!

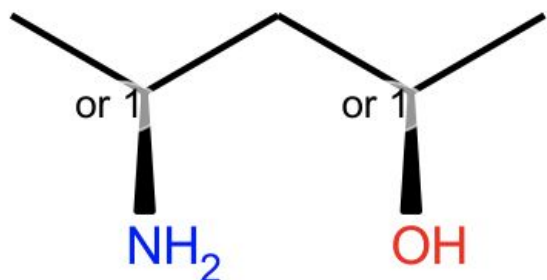
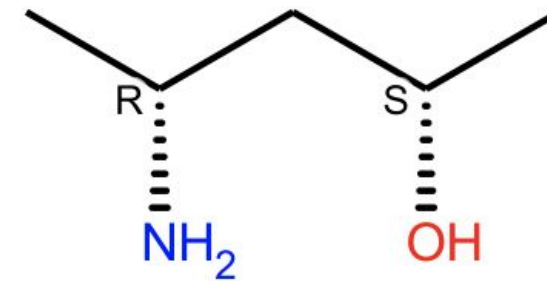
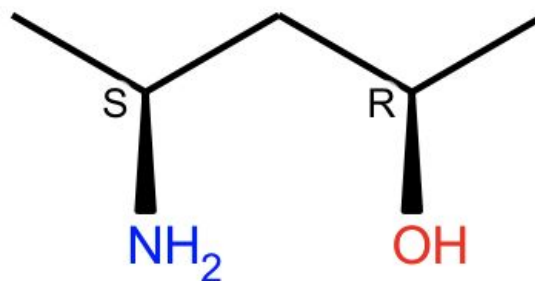
Extra information

- ✗ Map numbers
- ✗ Tautomeric specificity

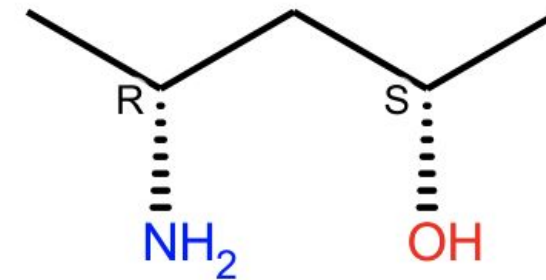
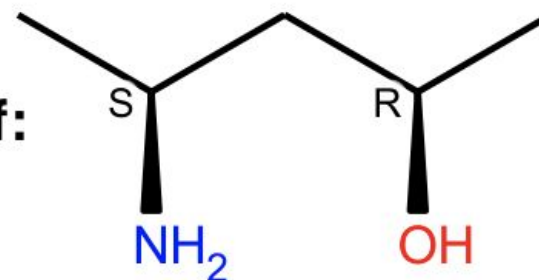
Enhanced stereochemistry refresher



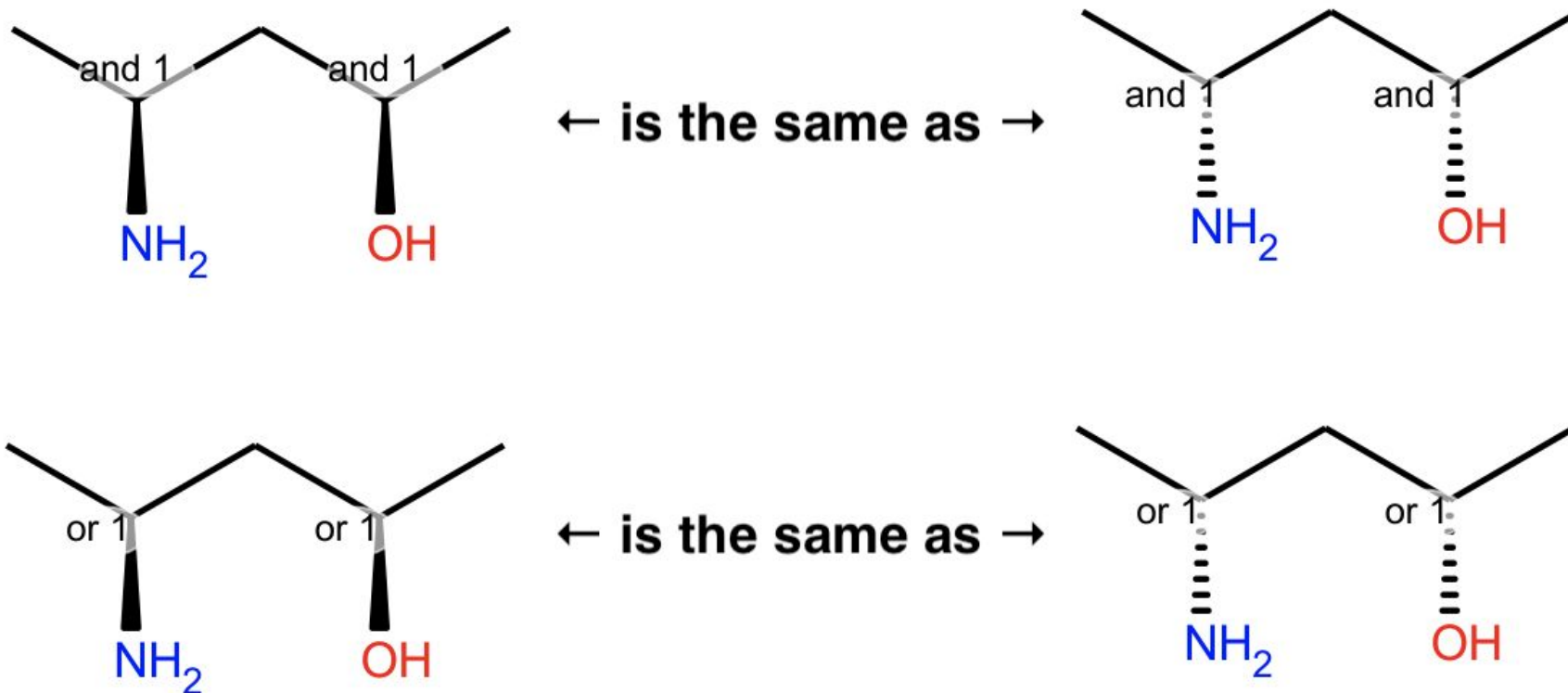
means both:



means either one of:



AND are equivalent. OR are equivalent



RegistrationHash

- Relies on CXSMILES support in RDKit
- Relies on molhash TautomerHash
- Validated in production in LiveDesign

Missing information

- ✓ *Enhanced stereochemistry*
- ✓ Polymer brackets
- ✓ Some annotations matter!

Extra information

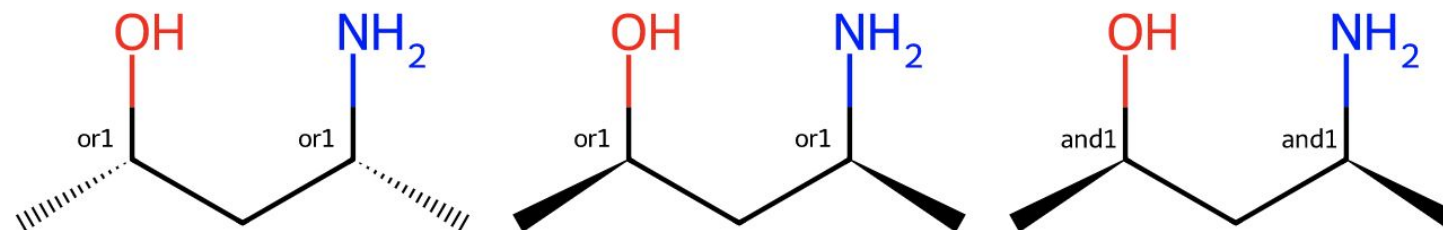
- ✓ Map numbers
- ✓ Tautomeric specificity


```
from rdkit.Chem.Draw import IPythonConsole
from rdkit.Chem import Draw
from rdkit.Chem import RegistrationHash
```

```
IPythonConsole.ipython_useSVG=True
IPythonConsole.drawOptions.addStereoAnnotation = True
```

```
In [2]: m1 = Chem.MolFromSmiles('C[C@H](O)C[C@H](N)C |o1:1,4|')
m2 = Chem.MolFromSmiles('C[C@@H](O)C[C@@H](N)C |o1:1,4|')
m3 = Chem.MolFromSmiles('C[C@@H](O)C[C@@H](N)C |&1:1,4|')
IPythonConsole.ShowMols([m1, m2, m3])
```

Out[2]:



```
In [3]: print(RegistrationHash.GetMolHash(RegistrationHash.GetMolLayers(m1)))
print(RegistrationHash.GetMolHash(RegistrationHash.GetMolLayers(m2)))
print(RegistrationHash.GetMolHash(RegistrationHash.GetMolLayers(m3)))
```

```
5812e64ea0c75ca980eab2cf633077684108627a
5812e64ea0c75ca980eab2cf633077684108627a
b86e4996c203e50a7280a824501c77f73e9e0f60
```

Acknowledgements

Chris Von Bargen
Greg Landrum
Hussein Faara
Rachel Walker
Ricardo Rodríguez-Schmidt

your name here?

<https://www.schrodinger.com/careers>: “Software Developer, Scientific Infrastructure”