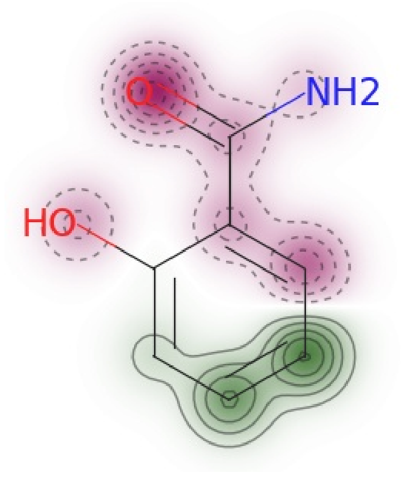
Calculating Atomic Contributions for Molecules Based on a Graph Convolutional QSAR Model

In an earlier tutorial we introduced the concept of model interpretability: understanding why a model produced the result it did. In this tutorial we will learn about atomic contributions, a useful tool for interpreting models that operate on molecules.

The idea is simple: remove a single atom from the molecule and see how the model's prediction changes. The "atomic contribution" for an atom is defined as the difference in activity between the whole molecule, and the fragment remaining after atom removal. It is a measure of how much that atom affects the prediction.

Contributions are also known as "attributions", "coloration", etc. in the literature. This is a model interpretation method [1], analogous to Similarity maps [2] in the QSAR domain, or occlusion methods in other fields (image classification, etc). Present implementation was used in [4].

Mariia Matveieva, Pavel Polishchuk. Institute of Molecular and Translational Medicine, Palacky University, Olomouc, Czech Republic.



Colab

This tutorial and the rest in this sequence can be done in Google colab. If you'd like to open this notebook in colab, you can use the following link.

Open in Colab

Setup

To run DeepChem within Colab, you'll need to run the following installation commands. This will take about 5 minutes to run to completion and install your environment. You can of course run this tutorial locally if you prefer. In that case, don't run these cells since they will download and install Anaconda on your local machine.

In [1]:
```
!curl -Lo conda_installer.py https://raw.githubusercontent.com/deepchem/deepchem/master/scripts/colab_install.py
import conda_installer
conda_installer.install()
!/root/miniconda/bin/conda info -e
```

```
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  3457  100  3457    0     0  24692      0 --:--:-- --:--:-- --:--:-- 24692
```

```
add /root/miniconda/lib/python3.7/site-packages to PYTHONPATH
python version: 3.7.12
fetching installer from https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86_64.sh
done
installing miniconda to /root/miniconda
done
installing openmm, pdbfixer
added conda-forge to channels
done
conda packages installation finished!
```

```
# conda environments:
#
base                  *  /root/miniconda
```

In [2]:
```
!pip install --pre deepchem
import deepchem
deepchem.__version__
```

```
Collecting deepchem
  Downloading deepchem-2.6.0.dev20211215231347-py3-none-any.whl (608 kB)

     |▌                               | 10 kB 25.3 MB/s eta 0:00:01
     |█                               | 20 kB 27.0 MB/s eta 0:00:01
     |█▌                              | 30 kB 30.7 MB/s eta 0:00:01
     |██                              | 40 kB 34.1 MB/s eta 0:00:01
     |██▋                             | 51 kB 37.7 MB/s eta 0:00:01
     |███▏                            | 61 kB 40.1 MB/s eta 0:00:01
     |███▋                            | 71 kB 36.7 MB/s eta 0:00:01
     |████▎                           | 81 kB 38.0 MB/s eta 0:00:01
     |████▊                           | 92 kB 39.3 MB/s eta 0:00:01
     |█████▍                          | 102 kB 34.8 MB/s eta 0:00:01
     |█████▉                          | 112 kB 34.8 MB/s eta 0:00:01
     |██████▍                         | 122 kB 34.8 MB/s eta 0:00:01
     |███████                         | 133 kB 34.8 MB/s eta 0:00:01
     |███████▌                        | 143 kB 34.8 MB/s eta 0:00:01
     |████████                        | 153 kB 34.8 MB/s eta 0:00:01
     |████████▌                       | 163 kB 34.8 MB/s eta 0:00:01
     |█████████▏                      | 174 kB 34.8 MB/s eta 0:00:01
     |█████████▋                      | 184 kB 34.8 MB/s eta 0:00:01
     |██████████▏                     | 194 kB 34.8 MB/s eta 0:00:01
     |██████████▊                     | 204 kB 34.8 MB/s eta 0:00:01
     |███████████▎                    | 215 kB 34.8 MB/s eta 0:00:01
     |███████████▉                    | 225 kB 34.8 MB/s eta 0:00:01
     |████████████▍                   | 235 kB 34.8 MB/s eta 0:00:01
     |████████████▉                   | 245 kB 34.8 MB/s eta 0:00:01
     |█████████████▍                  | 256 kB 34.8 MB/s eta 0:00:01
     |██████████████                  | 266 kB 34.8 MB/s eta 0:00:01
     |██████████████▌                 | 276 kB 34.8 MB/s eta 0:00:01
     |███████████████                 | 286 kB 34.8 MB/s eta 0:00:01
     |███████████████▌                | 296 kB 34.8 MB/s eta 0:00:01
     |████████████████▏               | 307 kB 34.8 MB/s eta 0:00:01
     |████████████████▋               | 317 kB 34.8 MB/s eta 0:00:01
     |█████████████████▏              | 327 kB 34.8 MB/s eta 0:00:01
     |█████████████████▊              | 337 kB 34.8 MB/s eta 0:00:01
     |██████████████████▎             | 348 kB 34.8 MB/s eta 0:00:01
     |██████████████████▉             | 358 kB 34.8 MB/s eta 0:00:01
     |███████████████████▍            | 368 kB 34.8 MB/s eta 0:00:01
     |███████████████████▉            | 378 kB 34.8 MB/s eta 0:00:01
     |████████████████████▌           | 389 kB 34.8 MB/s eta 0:00:01
     |█████████████████████           | 399 kB 34.8 MB/s eta 0:00:01
     |█████████████████████▌          | 409 kB 34.8 MB/s eta 0:00:01
     |██████████████████████          | 419 kB 34.8 MB/s eta 0:00:01
     |██████████████████████▋         | 430 kB 34.8 MB/s eta 0:00:01
     |███████████████████████▏        | 440 kB 34.8 MB/s eta 0:00:01
     |███████████████████████▋        | 450 kB 34.8 MB/s eta 0:00:01
     |████████████████████████▏       | 460 kB 34.8 MB/s eta 0:00:01
     |████████████████████████▊       | 471 kB 34.8 MB/s eta 0:00:01
     |█████████████████████████▎      | 481 kB 34.8 MB/s eta 0:00:01
     |█████████████████████████▉      | 491 kB 34.8 MB/s eta 0:00:01
     |██████████████████████████▍     | 501 kB 34.8 MB/s eta 0:00:01
     |██████████████████████████▉     | 512 kB 34.8 MB/s eta 0:00:01
     |███████████████████████████▌    | 522 kB 34.8 MB/s eta 0:00:01
     |████████████████████████████    | 532 kB 34.8 MB/s eta 0:00:01
     |████████████████████████████▌   | 542 kB 34.8 MB/s eta 0:00:01
     |█████████████████████████████   | 552 kB 34.8 MB/s eta 0:00:01
     |█████████████████████████████▋  | 563 kB 34.8 MB/s eta 0:00:01
     |██████████████████████████████▏ | 573 kB 34.8 MB/s eta 0:00:01
     |██████████████████████████████▋ | 583 kB 34.8 MB/s eta 0:00:01
     |███████████████████████████████▏| 593 kB 34.8 MB/s eta 0:00:01
     |███████████████████████████████▊| 604 kB 34.8 MB/s eta 0:00:01
     |████████████████████████████████| 608 kB 34.8 MB/s
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (from deepchem) (1.19.5)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from deepchem) (1.1.0)
Collecting rdkit-pypi
  Downloading rdkit_pypi-2021.9.3-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (20.6 MB)
     |████████████████████████████████| 20.6 MB 1.3 MB/s
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.7/dist-packages (from deepchem) (1.0.1)
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages (from deepchem) (1.4.1)
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from deepchem) (1.1.5)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas->deepchem) (2
018.9)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas->de
epchem) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from python-dateutil>=2.7.3->
pandas->deepchem) (1.15.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn
->deepchem) (3.0.0)
Installing collected packages: rdkit-pypi, deepchem
Successfully installed deepchem-2.6.0.dev20211215231347 rdkit-pypi-2021.9.3
```

```
Out[2]:  '2.6.0.dev'
```

A classification QSAR model for blood-brain barrier permeability

BBB permeability is the ability of compounds to enter the central nervous system. Here we use a dataset of relatively small compounds which are transported by diffusion without any carriers. The property is defined as log10(concentration in brain / concentration in blood). Compounds with a positive value (and 0) are labeled active, and others are labeled inactive. After modelling we will identify atoms favorable and unfavorable for diffusion.

First let's create the dataset. The molecules are stored in an SDF file.

In [8]:
```
import os
import pandas as pd
import deepchem as dc
import numpy as np
from rdkit import Chem
from rdkit.Chem import AllChem
from rdkit.Chem import Draw, PyMol, rdFMCS
from rdkit.Chem.Draw import IPythonConsole
from rdkit import rdBase
from deepchem import metrics
from IPython.display import Image, display
from rdkit.Chem.Draw import SimilarityMaps
import tensorflow as tf

current_dir = os.path.dirname(os.path.realpath('__file__'))
dc.utils.download_url(
    'https://raw.githubusercontent.com/deepchem/deepchem/master/examples/tutorials/assets/atomic_contributions_
    current_dir,
    'logBB.sdf'
)
DATASET_FILE =os.path.join(current_dir, 'logBB.sdf')
# Create RDKit mol objects, since we will need them later.
mols = [m for m in Chem.SDMolSupplier(DATASET_FILE) if m is not None ]
loader = dc.data.SDFLoader(tasks=["logBB_class"],
                           featurizer=dc.feat.ConvMolFeaturizer(),
                           sanitize=True)
dataset = loader.create_dataset(DATASET_FILE, shard_size=2000)
```

Now let's build and train a GraphConvModel.

In [10]:
```
np.random.seed(2020)
tf.random.set_seed(2020)
```

In [11]:
```
m = dc.models.GraphConvModel(1, mode="classification", batch_normalize=False, batch_size=100)
m.fit(dataset, nb_epoch=10)
```

```
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_14:0", shape=(331,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_po
ol_1/Reshape_13:0", shape=(331, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model/graph_pool_1/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a
large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_17:0", shape=(1646,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_p
ool_1/Reshape_16:0", shape=(1646, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_pool_1/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_20:0", shape=(1359,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_p
ool_1/Reshape_19:0", shape=(1359, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_pool_1/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_23:0", shape=(148,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_po
ol_1/Reshape_22:0", shape=(148, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model/graph_pool_1/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a
large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_11:0", shape=(331,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_co
nv_1/Reshape_10:0", shape=(331, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model/graph_conv_1/Cast:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a l
arge amount of memory.
  "shape. This may consume a large amount of memory." % value)
```

```
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_13:0", shape=(1646,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_c
onv_1/Reshape_12:0", shape=(1646, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_conv_1/Cast_1:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_15:0", shape=(1359,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_c
onv_1/Reshape_14:0", shape=(1359, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_conv_1/Cast_2:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_17:0", shape=(148,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_co
nv_1/Reshape_16:0", shape=(148, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model/graph_conv_1/Cast_3:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a
large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_19:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv
_1/Reshape_18:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_conv_1/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_21:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv
_1/Reshape_20:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_conv_1/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_23:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv
_1/Reshape_22:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_conv_1/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_25:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv
_1/Reshape_24:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_conv_1/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_27:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv
_1/Reshape_26:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_conv_1/Cast_8:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_29:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv
_1/Reshape_28:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_conv_1/Cast_9:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_14:0", shape=(331,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool
/Reshape_13:0", shape=(331, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_pool/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a large
amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_17:0", shape=(1646,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_poo
l/Reshape_16:0", shape=(1646, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_m
odel/graph_pool/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_20:0", shape=(1359,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_poo
l/Reshape_19:0", shape=(1359, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_m
odel/graph_pool/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
```

```
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_23:0", shape=(148,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool
/Reshape_22:0", shape=(148, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_pool/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a large
amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_14:0", shape=(334,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_po
ol_1/Reshape_13:0", shape=(334, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model/graph_pool_1/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a
large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_17:0", shape=(1838,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_p
ool_1/Reshape_16:0", shape=(1838, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_pool_1/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_20:0", shape=(1458,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_p
ool_1/Reshape_19:0", shape=(1458, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_pool_1/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_23:0", shape=(120,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_po
ol_1/Reshape_22:0", shape=(120, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model/graph_pool_1/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a
large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_11:0", shape=(334,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_co
nv_1/Reshape_10:0", shape=(334, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model/graph_conv_1/Cast:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a l
arge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_13:0", shape=(1838,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_c
onv_1/Reshape_12:0", shape=(1838, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_conv_1/Cast_1:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_15:0", shape=(1458,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_c
onv_1/Reshape_14:0", shape=(1458, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_conv_1/Cast_2:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_17:0", shape=(120,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_co
nv_1/Reshape_16:0", shape=(120, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model/graph_conv_1/Cast_3:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a
large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_14:0", shape=(334,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool
/Reshape_13:0", shape=(334, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_pool/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a large
amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_17:0", shape=(1838,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_poo
l/Reshape_16:0", shape=(1838, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_m
odel/graph_pool/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_20:0", shape=(1458,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_poo
l/Reshape_19:0", shape=(1458, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_m
odel/graph_pool/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
```

```
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_23:0", shape=(120,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool
/Reshape_22:0", shape=(120, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_mod
el/graph_pool/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a large
amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_14:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_p
ool_1/Reshape_13:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_pool_1/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_17:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_p
ool_1/Reshape_16:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_pool_1/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_20:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_p
ool_1/Reshape_19:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_pool_1/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_11:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_c
onv_1/Reshape_10:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_conv_1/Cast:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a
large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_13:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_c
onv_1/Reshape_12:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_conv_1/Cast_1:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_15:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_c
onv_1/Reshape_14:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_conv_1/Cast_2:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_14:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_poo
l/Reshape_13:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_m
odel/graph_pool/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_17:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_poo
l/Reshape_16:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_m
odel/graph_pool/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_20:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_poo
l/Reshape_19:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_m
odel/graph_pool/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool_1/
Reshape_23:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_p
ool_1/Reshape_22:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model/graph_pool_1/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_conv_1/
Reshape_17:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_c
onv_1/Reshape_16:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
```

```
as_model/graph_conv_1/Cast_3:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model/graph_pool/Re
shape_23:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model/graph_poo
l/Reshape_22:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras_m
odel/graph_pool/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a lar
ge amount of memory.
  "shape. This may consume a large amount of memory." % value)
```

Out[11]:
```
0.5348201115926107
```

Let's load a test set and see how well it works.

In [12]:
```python
current_dir = os.path.dirname(os.path.realpath('__file__'))
dc.utils.download_url(
    'https://raw.githubusercontent.com/deepchem/deepchem/master/examples/tutorials/assets/atomic_contributions_
    current_dir,
    'logBB_test_.sdf'
)
TEST_DATASET_FILE = os.path.join(current_dir, 'logBB_test_.sdf')
loader = dc.data.SDFLoader(tasks=["p_np"], sanitize=True,
                           featurizer=dc.feat.ConvMolFeaturizer())
test_dataset = loader.create_dataset(TEST_DATASET_FILE, shard_size=2000)
pred =  m.predict(test_dataset)
pred = np.argmax(np.squeeze(pred),axis=1)
ba = metrics.balanced_accuracy_score(y_true=test_dataset.y, y_pred=pred)
print(ba)
```

```
0.7444444444444445
```

The balanced accuracy is high enough. Now let's proceed to model interpretation and estimate the contributions of individual atoms to the prediction.

A fragment dataset

Now let's prepare a dataset of fragments based on the training set. (Any other unseen data set of interest can also be used). These fragments will be used to evaluate the contributions of individual atoms.

For each molecule we will generate a list of ConvMol objects. Specifying `per_atom_fragmentation=True` tells it to iterate over all heavy atoms and featurize a single-atom-depleted version of the molecule with each one removed.

In [13]:
```python
loader = dc.data.SDFLoader(tasks=[],# dont need task (moreover, passing the task can lead to inconsitencies in
                        featurizer=dc.feat.ConvMolFeaturizer(per_atom_fragmentation=True),
                        sanitize=True)
frag_dataset = loader.create_dataset(DATASET_FILE, shard_size=5000)
```

```
/usr/local/lib/python3.7/dist-packages/numpy/core/_asarray.py:83: VisibleDeprecationWarning: Creating an ndarray
from ragged nested sequences (which is a list-or-tuple of lists-or-tuples-or ndarrays with different lengths or
shapes) is deprecated. If you meant to do this, you must specify 'dtype=object' when creating the ndarray
  return array(a, dtype, copy=False, order=order)
/usr/local/lib/python3.7/dist-packages/deepchem/data/data_loader.py:885: VisibleDeprecationWarning: Creating an
ndarray from ragged nested sequences (which is a list-or-tuple of lists-or-tuples-or ndarrays with different len
gths or shapes) is deprecated. If you meant to do this, you must specify 'dtype=object' when creating the ndarra
y
  return np.array(features), valid_inds
```

The dataset still has the same number of samples as the original training set, but each sample is now represented as a list of ConvMol objects (one for each fragment) rather than a single ConvMol.

IMPORTANT: The order of fragments depends on the input format. If SDF, the fragment order is the same as the atom order in corresponding mol blocks. If SMILES (i.e. csv with molecules represented as SMILES), then the order is given by RDKit CanonicalRankAtoms

In [14]:
```python
print(frag_dataset.X.shape)
```

```
(298,)
```

We really want to treat each fragment as a separate sample. We can use a FlatteningTransformer to flatten the fragments lists.

In [15]:
```python
tr = dc.trans.FlatteningTransformer(frag_dataset)
frag_dataset = tr.transform(frag_dataset)
print(frag_dataset.X.shape)
```

```
(5111,)
```

Predicting atomic contributions to activity

Now we will predict the activity for molecules and for their fragments. Then, for each fragment, we'll find the activity difference: the change in activity when removing one atom.

Note: Here, in classification context, we use the probability output of the model as the activity. So the contribution is the probability difference, i.e. "how much a given atom increases/decreases the probability of the molecule being active."

In [16]:
```
# whole  molecules
pred = np.squeeze(m.predict(dataset))[:, 1] # probabilitiy of class 1
pred = pd.DataFrame(pred, index=dataset.ids, columns=["Molecule"])  # turn to dataframe for convinience

# fragments
pred_frags = np.squeeze(m.predict(frag_dataset))[:, 1]
pred_frags = pd.DataFrame(pred_frags, index=frag_dataset.ids, columns=["Fragment"])
```

We take the difference to find the atomic contributions.

In [17]:
```
# merge 2 dataframes by molecule names
df = pd.merge(pred_frags, pred, right_index=True, left_index=True)
# find contribs
df['Contrib'] = df["Molecule"] - df["Fragment"]
```

In [18]:
```
df
```

Out[18]:

	Fragment	Molecule	Contrib
C#CC1(O)CCC2C3C(C)CC4=C(CCC(=O)C4)C3CCC21C	0.756537	0.811550	0.055013
C#CC1(O)CCC2C3C(C)CC4=C(CCC(=O)C4)C3CCC21C	0.752759	0.811550	0.058791
C#CC1(O)CCC2C3C(C)CC4=C(CCC(=O)C4)C3CCC21C	0.747012	0.811550	0.064538
C#CC1(O)CCC2C3C(C)CC4=C(CCC(=O)C4)C3CCC21C	0.815878	0.811550	-0.004328
C#CC1(O)CCC2C3C(C)CC4=C(CCC(=O)C4)C3CCC21C	0.741805	0.811550	0.069745
...	...	...	...
c1cncc(C2CCCN2)c1	0.780473	0.813031	0.032559
c1cncc(C2CCCN2)c1	0.722649	0.813031	0.090383
c1cncc(C2CCCN2)c1	0.721607	0.813031	0.091425
c1cncc(C2CCCN2)c1	0.683299	0.813031	0.129732
c1cncc(C2CCCN2)c1	0.674451	0.813031	0.138581

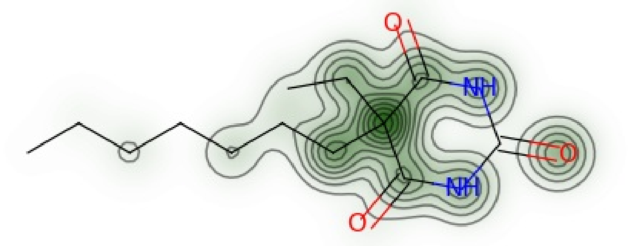
5111 rows × 3 columns

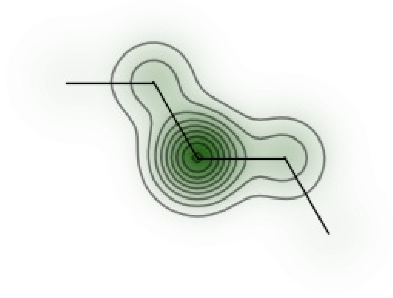
We can use the SimilarityMaps feature of RDKit to visualize the results. Each atom is colored by how it affects activity.

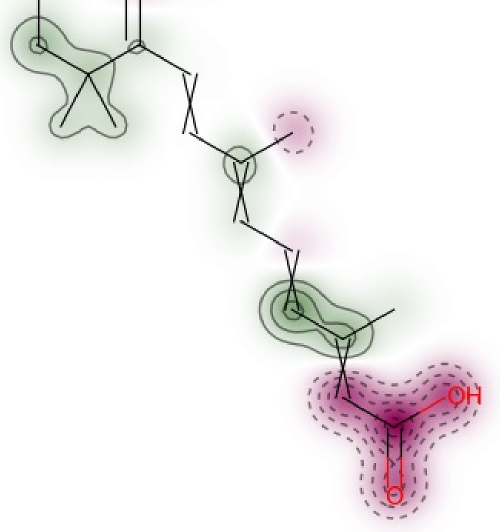
In [19]:
```
def vis_contribs(mols, df, smi_or_sdf = "sdf"):
    # input format of file, which was used to create dataset determines the order of atoms,
    # so we take it into account for correct mapping!
    maps = []
    for mol  in mols:
        wt = {}
        if smi_or_sdf == "smi":
            for n,atom in enumerate(Chem.rdmolfiles.CanonicalRankAtoms(mol)):
                wt[atom] = df.loc[mol.GetProp("_Name"),"Contrib"][n]
        if smi_or_sdf == "sdf":
            for n,atom in enumerate(range(mol.GetNumHeavyAtoms())):
                wt[atom] = df.loc[Chem.MolToSmiles(mol),"Contrib"][n]
        maps.append(SimilarityMaps.GetSimilarityMapFromWeights(mol,wt))
    return maps
```

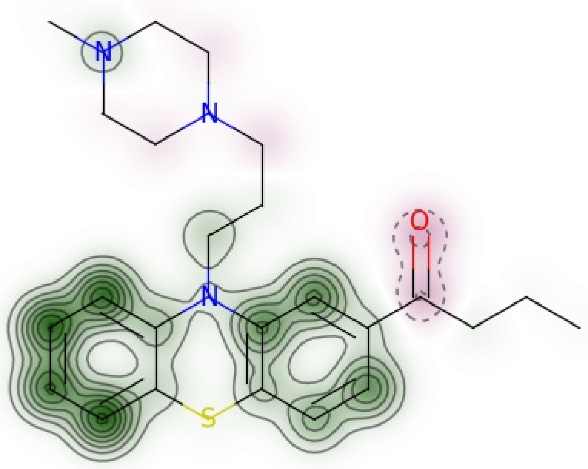
Let's look at some pictures:

In [20]:
```
np.random.seed(2000)
maps = vis_contribs(np.random.choice(np.array(mols),10), df)
```

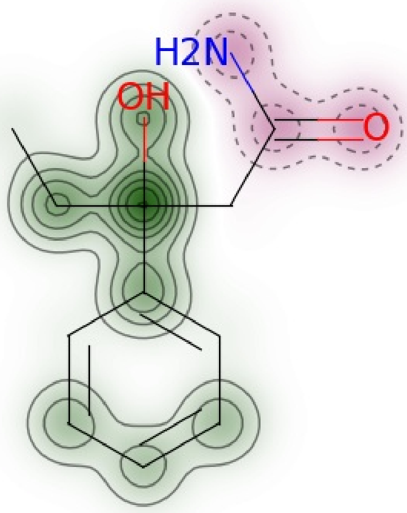

O
NH
O
NH
O



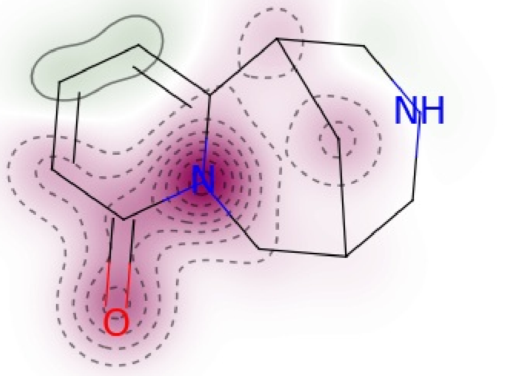

OH
O



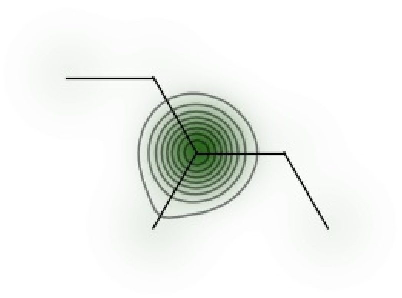
N
N
N
S
O

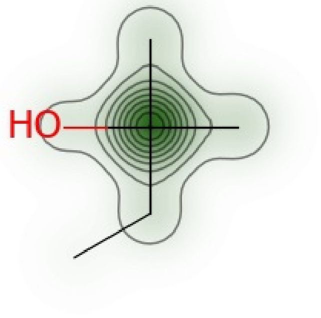


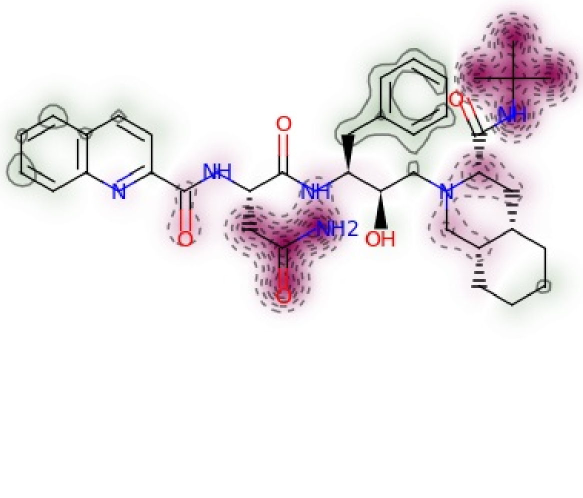
H2N
OH
O

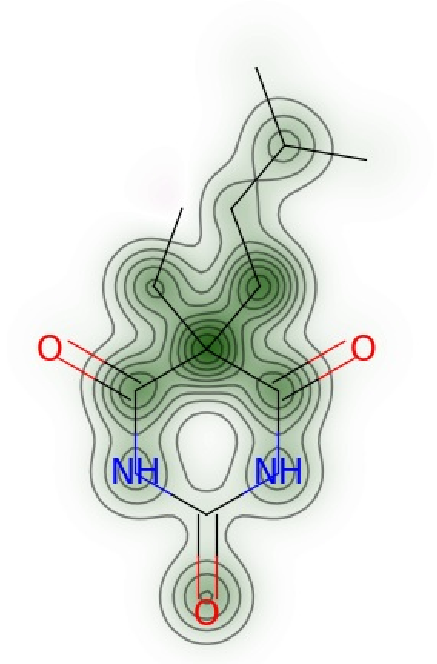


NH
N
O




HO


N
NH
O
O
NH2
O
NH
OH
N
O
NH



We can see that aromatics or aliphatics have a positive impact on blood-brain barrier permeability, while polar or charged heteroatoms have a negative influence. This is generally consistent with literature data.

A regression task

The example above used a classification model. The same techniques can also be used for regression models. Let's look at a regression task, aquatic toxicity (towards the water organism T. pyriformis).

Toxicity is defined as log10(IGC50) (concentration that inhibits colony growth by 50%). Toxicophores for T. pyriformis will be identified by atomic contributions.

All the above steps are the same: load data, featurize, build a model, create dataset of fragments, find contributions, and visualize them.

Note: this time as it is regression, contributions will be in activity units, not probability.

In [21]:
```
current_dir = os.path.dirname(os.path.realpath('__file__'))
dc.utils.download_url(
    'https://raw.githubusercontent.com/deepchem/deepchem/master/examples/tutorials/assets/atomic_contributions_
    current_dir,
    'Tetrahymena_pyriformis_Work_set_OCHEM.sdf'
)
DATASET_FILE =os.path.join(current_dir, 'Tetrahymena_pyriformis_Work_set_OCHEM.sdf')

# create RDKit mol objects, we will need them later
mols = [m for m in Chem.SDMolSupplier(DATASET_FILE) if m is not None ]
loader = dc.data.SDFLoader(tasks=["IGC50"],
                           featurizer=dc.feat.ConvMolFeaturizer(), sanitize=True)
dataset = loader.create_dataset(DATASET_FILE, shard_size=5000)
```

Create and train the model.

In [22]:
```
np.random.seed(2020)
tf.random.set_seed(2020)
m = dc.models.GraphConvModel(1, mode="regression", batch_normalize=False)
m.fit(dataset, nb_epoch=40)
```

```
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_14:0", shape=(291,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_3/Reshape_13:0", shape=(291, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_3/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_17:0", shape=(910,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_3/Reshape_16:0", shape=(910, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_3/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
```

```
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_20:0", shape=(663,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_3/Reshape_19:0", shape=(663, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_3/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_23:0", shape=(28,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph
_pool_3/Reshape_22:0", shape=(28, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model_1/graph_pool_3/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consu
me a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_11:0", shape=(291,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_conv_3/Reshape_10:0", shape=(291, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_conv_3/Cast:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consu
me a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_13:0", shape=(910,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_conv_3/Reshape_12:0", shape=(910, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_conv_3/Cast_1:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_15:0", shape=(663,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_conv_3/Reshape_14:0", shape=(663, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_conv_3/Cast_2:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_17:0", shape=(28,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph
_conv_3/Reshape_16:0", shape=(28, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model_1/graph_conv_3/Cast_3:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consu
me a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_19:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_
conv_3/Reshape_18:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model_1/graph_conv_3/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_21:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_
conv_3/Reshape_20:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model_1/graph_conv_3/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_23:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_
conv_3/Reshape_22:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model_1/graph_conv_3/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_25:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_
conv_3/Reshape_24:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model_1/graph_conv_3/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_27:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_
conv_3/Reshape_26:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model_1/graph_conv_3/Cast_8:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_29:0", shape=(0,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_
conv_3/Reshape_28:0", shape=(0, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_keras
_model_1/graph_conv_3/Cast_9:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume
a large amount of memory.
```

```
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_14:0", shape=(291,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_2/Reshape_13:0", shape=(291, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_2/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_17:0", shape=(910,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_2/Reshape_16:0", shape=(910, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_2/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_20:0", shape=(663,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_2/Reshape_19:0", shape=(663, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_2/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_23:0", shape=(28,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph
_pool_2/Reshape_22:0", shape=(28, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model_1/graph_pool_2/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consu
me a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_14:0", shape=(307,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_3/Reshape_13:0", shape=(307, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_3/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_17:0", shape=(944,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_3/Reshape_16:0", shape=(944, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_3/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_20:0", shape=(693,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_3/Reshape_19:0", shape=(693, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_3/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_23:0", shape=(16,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph
_pool_3/Reshape_22:0", shape=(16, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model_1/graph_pool_3/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consu
me a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_11:0", shape=(307,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_conv_3/Reshape_10:0", shape=(307, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_conv_3/Cast:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consu
me a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_13:0", shape=(944,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_conv_3/Reshape_12:0", shape=(944, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_conv_3/Cast_1:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_15:0", shape=(693,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_conv_3/Reshape_14:0", shape=(693, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_conv_3/Cast_2:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_17:0", shape=(16,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph
_conv_3/Reshape_16:0", shape=(16, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model_1/graph_conv_3/Cast_3:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consu
```

```
me a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_14:0", shape=(307,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_2/Reshape_13:0", shape=(307, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_2/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_17:0", shape=(944,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_2/Reshape_16:0", shape=(944, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_2/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_20:0", shape=(693,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/grap
h_pool_2/Reshape_19:0", shape=(693, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_k
eras_model_1/graph_pool_2/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_23:0", shape=(16,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph
_pool_2/Reshape_22:0", shape=(16, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv_ker
as_model_1/graph_pool_2/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consu
me a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_14:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_pool_3/Reshape_13:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_pool_3/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_17:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_pool_3/Reshape_16:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_pool_3/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_20:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_pool_3/Reshape_19:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_pool_3/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
3/Reshape_23:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_pool_3/Reshape_22:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_pool_3/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_11:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_conv_3/Reshape_10:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_conv_3/Cast:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may con
sume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_13:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_conv_3/Reshape_12:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_conv_3/Cast_1:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_15:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_conv_3/Reshape_14:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_conv_3/Cast_2:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_conv_
3/Reshape_17:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_conv_3/Reshape_16:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
```

```
_keras_model_1/graph_conv_3/Cast_3:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_14:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_pool_2/Reshape_13:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_pool_2/Cast_4:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_17:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_pool_2/Reshape_16:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_pool_2/Cast_5:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_20:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_pool_2/Reshape_19:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_pool_2/Cast_6:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
/usr/local/lib/python3.7/dist-packages/tensorflow/python/framework/indexed_slices.py:450: UserWarning: Convertin
g sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/private__graph_conv_keras_model_1/graph_pool_
2/Reshape_23:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/private__graph_conv_keras_model_1/gra
ph_pool_2/Reshape_22:0", shape=(None, 64), dtype=float32), dense_shape=Tensor("gradient_tape/private__graph_conv
_keras_model_1/graph_pool_2/Cast_7:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may c
onsume a large amount of memory.
  "shape. This may consume a large amount of memory." % value)
```

```
Out[22]: 0.12407124519348145
```

Load the test dataset and check the model's performance.

```
In [23]:
current_dir = os.path.dirname(os.path.realpath('__file__'))
dc.utils.download_url(
    'https://raw.githubusercontent.com/deepchem/deepchem/master/examples/tutorials/assets/atomic_contributions_
    current_dir,
    'Tetrahymena_pyriformis_Test_set_OCHEM.sdf'
)




TEST_DATASET_FILE = os.path.join(current_dir, 'Tetrahymena_pyriformis_Test_set_OCHEM.sdf')
loader = dc.data.SDFLoader(tasks=["IGC50"], sanitize= True,
                           featurizer=dc.feat.ConvMolFeaturizer())
test_dataset = loader.create_dataset(TEST_DATASET_FILE, shard_size=2000)
pred = m.predict(test_dataset)
mse = metrics.mean_squared_error(y_true=test_dataset.y, y_pred=pred)
r2 = metrics.r2_score(y_true=test_dataset.y, y_pred=pred)
print(mse)
print(r2)
```

```
0.2381780323921622
0.784334539071699
```

Load the training set again, but this time set `per_atom_fragmentation=True`.

```
In [24]:
loader = dc.data.SDFLoader(tasks=[], # dont need any task
                           sanitize=True,
                           featurizer=dc.feat.ConvMolFeaturizer(per_atom_fragmentation=True))
frag_dataset = loader.create_dataset(DATASET_FILE, shard_size=5000)
tr = dc.trans.FlatteningTransformer(frag_dataset) # flatten dataset and add ids to each fragment
frag_dataset = tr.transform(frag_dataset)
```

```
/usr/local/lib/python3.7/dist-packages/numpy/core/_asarray.py:83: VisibleDeprecationWarning: Creating an ndarray
from ragged nested sequences (which is a list-or-tuple of lists-or-tuples-or ndarrays with different lengths or
shapes) is deprecated. If you meant to do this, you must specify 'dtype=object' when creating the ndarray
  return array(a, dtype, copy=False, order=order)
/usr/local/lib/python3.7/dist-packages/deepchem/data/data_loader.py:885: VisibleDeprecationWarning: Creating an
ndarray from ragged nested sequences (which is a list-or-tuple of lists-or-tuples-or ndarrays with different len
gths or shapes) is deprecated. If you meant to do this, you must specify 'dtype=object' when creating the ndarra
y
  return np.array(features), valid_inds
```

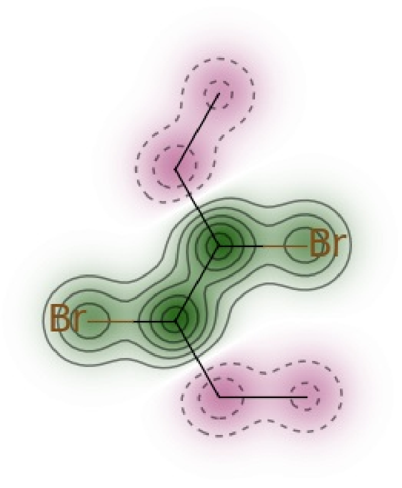
Compute the activity differences.

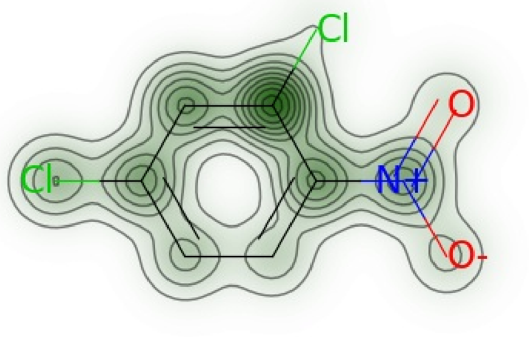
```
In [25]:
# whole molecules
pred = m.predict(dataset)
pred = pd.DataFrame(pred, index=dataset.ids, columns=["Molecule"])  # turn to dataframe for convenience
```

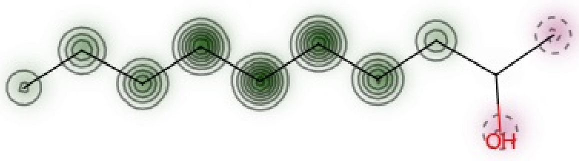
```
# fragments
pred_frags = m.predict(frag_dataset)
pred_frags = pd.DataFrame(pred_frags, index=frag_dataset.ids, columns=["Fragment"])  # turn to dataframe for co

# merge 2 dataframes by molecule names
df = pd.merge(pred_frags, pred, right_index=True, left_index=True)
# find contribs
df['Contrib'] = df["Molecule"] - df["Fragment"]
```

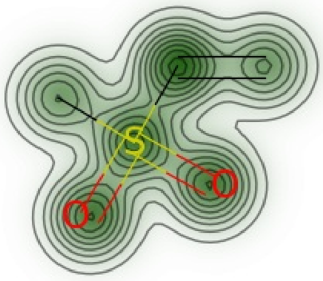
Lets take some molecules with moderate activity (not extremely active/inactive) and visualize the atomic contributions.

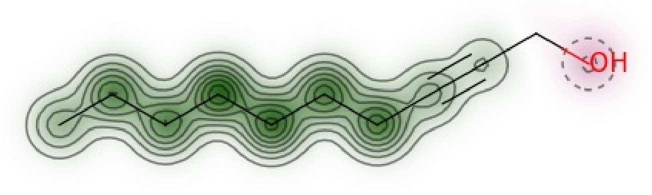
```
In [26]: maps = vis_contribs([mol for mol in mols if float(mol.GetProp("IGC50"))>3 and float(mol.GetProp("IGC50"))<4][:1
```

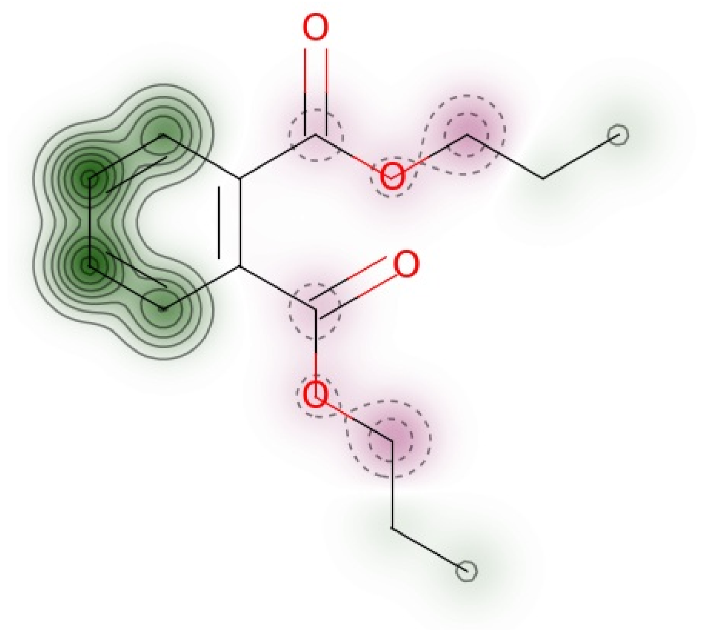


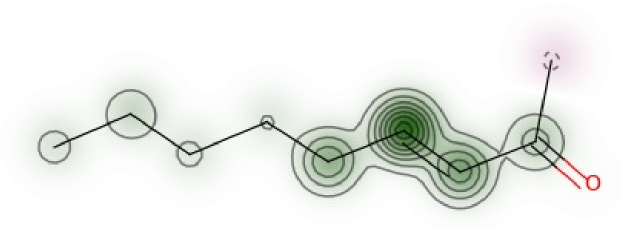


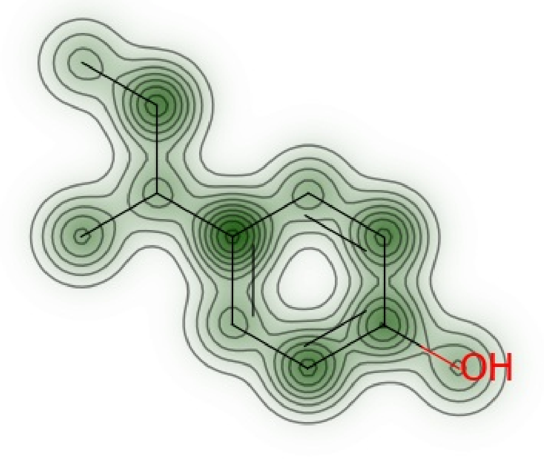

OH

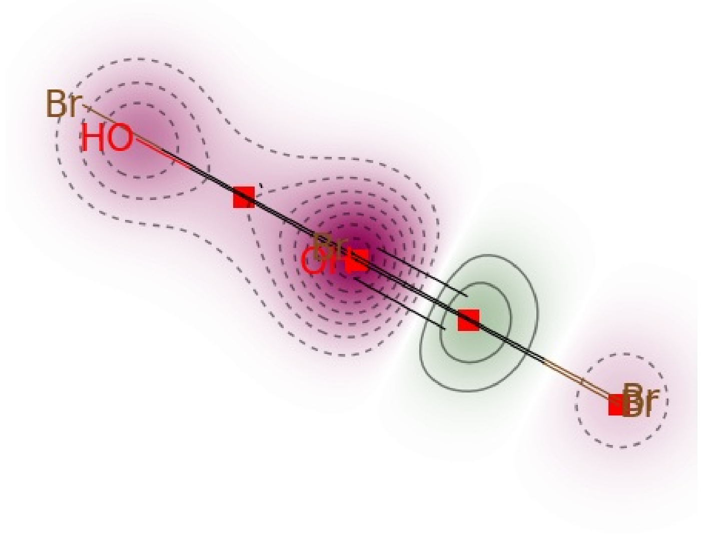

S
O
O

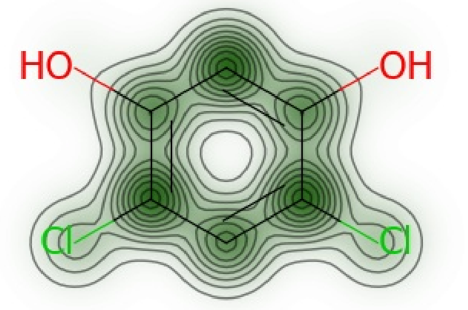

OH


O
O
O
O


O


OH





We can see that known toxicophores are in green, namely nitro-aromatics, halo-aromatics, long alkyl chains, and aldehyde; while carboxylic groups, alcohols, and aminos are detoxyfying, as is consistent with literature [3]

Appendix

In this tutorial we operated on SDF files. However, if we use CSV files with SMILES as input, the order of the atoms in the dataframe DOES NOT correspond to the original atom order. If we want to recover the original atom order for each molecule (to have it in our main dataframe), we need to use RDKit's Chem.rdmolfiles.CanonicalRankAtoms. Here are some utilities to do this.

We can add a column with atom ids (as in input molecules) and use the resulting dataframe for analysis with any other software, outside the "python-rdkit-deepchem" environment.

In [27]:
```python
def get_mapping(mols, mol_names):
    """perform mapping:
    atom number original <-> atom number(position)
    after ranking (both 1-based)"""
    # mols - RDKit mols
    # names  - any seq of strings
    # return list of nested lists: [[molecule, [atom , atom, ..], [...]]
    assert(len(mols)==len(mol_names))
    mapping = []
    for m,n in zip(mols, mol_names):
        atom_ids = [i+1 for i in list(Chem.rdmolfiles.CanonicalRankAtoms(m))]
        mapping.append([n, atom_ids])
    return mapping
```

In [28]:
```python
def append_atomid_col(df, mapping):
    # add column with CORRECT atom number(position)
    for i in mapping:
        df.loc[i[0],"AtomID"] = i[1]
    return df
```

Bibliography:

1. Polishchuk, P., O. Tinkov, T. Khristova, L. Ognichenko, A. Kosinskaya, A. Varnek & V. Kuz'min (2016) Structural and Physico-Chemical Interpretation (SPCI) of QSAR Models and Its Comparison with Matched Molecular Pair Analysis. Journal of Chemical Information and Modeling, 56, 1455-1469.

2. Riniker, S. & G. Landrum (2013) Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods. Journal of Cheminformatics, 5, 43.

3. M. Matveieva, M. T. D. Cronin, P. Polishchuk, Mol. Inf. 2019, 38, 1800084.

4. Matveieva, M., Polishchuk, P. Benchmarks for interpretation of QSAR models. J Cheminform 13, 41 (2021). https://doi.org/10.1186/s13321-021-00519-x

Congratulations! Time to join the Community!

Congratulations on completing this tutorial notebook! If you enjoyed working through the tutorial, and want to continue working with DeepChem, we encourage you to finish the rest of the tutorials in this series. You can also help the DeepChem community in the following ways:

Star DeepChem on GitHub

This helps build awareness of the DeepChem project and the tools for open source drug discovery that we're trying to build.

Join the DeepChem Discord

The DeepChem Discord hosts a number of scientists, developers, and enthusiasts interested in deep learning for the life sciences. Join the conversation!