Fingerprint Embedding using cold split for BindingDB_Kd

April 20, 2021

[1]:
```
import numpy as np
import random
```

Importing the BindingDB_Kd dataset

[2]:
```
from tdc.multi_pred import DTI
bindingDB_data = DTI(name = 'BindingDB_Kd')
```

```
Found local copy...
Loading...
Done!
```

[3]:
```
bindingDB_data
```

[3]:
```
<tdc.multi_pred.dataloader.DTI at 0x7ffdede3f2e0>
```

[4]:
```
# fixing random seed for reproducibility
random.seed(123)
np.random.seed(123)
```

Cold split method implementation on Drug column. Need to verify if column name attribute is required or not ?

[5]:
```
bindingDB_cold_split = bindingDB_data.get_split(method = 'cold_split',␣
↪column_name = 'Drug', seed = 42, frac = [0.7, 0.1, 0.2])
```

[6]:
```
train_set = bindingDB_cold_split['train']
```

[7]:
```
#train_set = train_set.head(10)
train_set
```

[7]:

	Drug_ID	Drug	\
0	4316.0	COc1ccc(CNS(=O)(=O)c2ccc(S(N)(=O)=O)s2)cc1	
1	4293.0	NS(=O)(=O)c1ccc(S(=O)(=O)NCc2cccs2)s1	
2	1612.0	COc1ccc(N2CC(O)c3cc(S(N)(=O)=O)sc3S2(=O)=O)cc1	
3	4369102.0	CCN[C@H]1CN(CCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21	
4	68844.0	CCN[C@H]1CN(CCCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21	
…	…	…	

```
34696  76311094.0           O=C(NCCCCS(=O)(=O)c1ccccc1)c1ccc2nccn2c1
34697  57406853.0  CCOc1cc(C(C)(C)C)ccc1C1=N[C@@](C)(c2ccc(Cl)cc2…
34698  58573469.0  CC(C)[C@@H](CS(=O)(=O)C(C)C)N1C(=O)[C@@](C)(CC…
34699    113557.0                        CCCCCCCOC1OC(CO)C(O)C(O)C1O
34700    113557.0                        CCCCCCCOC1OC(CO)C(O)C(O)C1O
```

	Target_ID	Target	Y
0	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…	0.490
1	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…	0.830
2	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…	0.160
3	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…	0.320
4	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…	0.130
…	…	…	…
34696	P43490	MNPAAEAEFNILLATDSYKVTHYKQYPPNTSKVYSYFECREKKTEN…	860.000
34697	NaN	MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKD…	2.900
34698	NaN	MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKD…	0.045
34699	P08191	MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLA…	17.000
34700	P08191	MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLA…	19.000

[34701 rows x 5 columns]

Apply fingerprint embedding from rdkit

[8]:
```
from rdkit.Chem.Fingerprints import FingerprintMols
```

```
RDKit WARNING: [02:46:00] Enabling RDKit 2019.09.3 jupyter extensions
```

Added a new column 'Drug_vector to store the fingerprint embedded values applied on Drug column'

[9]:
```
dataSource = train_set.drop(['Target_ID', 'Drug_ID', 'Y'], axis=1)
dataSource['Drug_vector'] = ''
dataSource
```

[9]:

```
                                                  Drug  \
0            COc1ccc(CNS(=O)(=O)c2ccc(S(N)(=O)=O)s2)cc1
1                 NS(=O)(=O)c1ccc(S(=O)(=O)NCc2cccs2)s1
2         COc1ccc(N2CC(O)c3cc(S(N)(=O)=O)sc3S2(=O)=O)cc1
3         CCN[C@H]1CN(CCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
4        CCN[C@H]1CN(CCCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
…                                                   …
34696          O=C(NCCCCS(=O)(=O)c1ccccc1)c1ccc2nccn2c1
34697  CCOc1cc(C(C)(C)C)ccc1C1=N[C@@](C)(c2ccc(Cl)cc2…
34698  CC(C)[C@@H](CS(=O)(=O)C(C)C)N1C(=O)[C@@](C)(CC…
34699                       CCCCCCCOC1OC(CO)C(O)C(O)C1O
34700                       CCCCCCCOC1OC(CO)C(O)C(O)C1O

                                                  Target Drug_vector
```

```
0      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
1      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
2      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
3      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
4      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
…                                                    …          …
34696  MNPAAEAEFNILLATDSYKVTHYKQYPPNTSKVYSYFECREKKTEN…
34697  MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKD…
34698  MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKD…
34699  MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLA…
34700  MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLA…

[34701 rows x 3 columns]
```

Reference taken from rdkit

[10]:
```
from rdkit import Chem
import tensorflow as tf
import torch

for ind, drug in enumerate(dataSource['Drug']):
    mol = Chem.MolFromSmiles(drug)
    fp = Chem.RDKFingerprint(mol)
    intmap = map(int, fp.ToBitString())
    dataSource['Drug_vector'][ind] = np.array(list(intmap))


dataSource
```

[10]:
```
                                                    Drug  \
0              COc1ccc(CNS(=O)(=O)c2ccc(S(N)(=O)=O)s2)cc1
1                   NS(=O)(=O)c1ccc(S(=O)(=O)NCc2cccs2)s1
2          COc1ccc(N2CC(O)c3cc(S(N)(=O)=O)sc3S2(=O)=O)cc1
3          CCN[C@H]1CN(CCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
4         CCN[C@H]1CN(CCCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
…                                                      …
34696            O=C(NCCCCS(=O)(=O)c1ccccc1)c1ccc2nccn2c1
34697  CCOc1cc(C(C)(C)C)ccc1C1=N[C@@](C)(c2ccc(Cl)cc2…
34698  CC(C)[C@@H](CS(=O)(=O)C(C)C)N1C(=O)[C@@](C)(CC…
34699                          CCCCCCCOC1OC(CO)C(O)C(O)C1O
34700                          CCCCCCCOC1OC(CO)C(O)C(O)C1O

                                                   Target  \
0      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
1      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
2      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
3      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
```

```
4      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
…                                                    …
34696  MNPAAEAEFNILLATDSYKVTHYKQYPPNTSKVYSYFECREKKTEN…
34697  MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKD…
34698  MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKD…
34699  MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLA…
34700  MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLA…

                                              Drug_vector
0      [0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, …
1      [0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, …
2      [0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, …
3      [0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, …
4      [0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, …
…                                                    …
34696  [1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, …
34697  [1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, …
34698  [0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, …
34699  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …
34700  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …

[34701 rows x 3 columns]
```

Using DataLoader to perform clustering on the Drug_vector column

[11]:
```
from torch.utils.data import DataLoader

train_loader = DataLoader(dataset=dataSource['Drug_vector'], shuffle=True,␣
↪batch_size=256)
```

[12]:
```
from sklearn.cluster import AgglomerativeClustering
clustering = AgglomerativeClustering(linkage='average', n_clusters=50)

clusters=[]
for i, batch in enumerate(train_loader):
    clustering.fit(batch)
    clusters.extend(clustering.labels_)
```

The cluster column to the datasource specifies which the cluster it belongs to. Added the cluster array to the column which was fetched by performing Agglomerative Clustering in above step

[13]:
```
dataSource['Clusters'] = clusters
```

[14]:
```
dataSource
```

[14]:
```
                                                  Drug  \
0              COc1ccc(CNS(=O)(=O)c2ccc(S(N)(=O)=O)s2)cc1
1                   NS(=O)(=O)c1ccc(S(=O)(=O)NCc2cccs2)s1
```

```
2          COc1ccc(N2CC(O)c3cc(S(N)(=O)=O)sc3S2(=O)=O)cc1
3          CCN[C@H]1CN(CCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
4        CCN[C@H]1CN(CCCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
…                                                      …
34696            O=C(NCCCCS(=O)(=O)c1ccccc1)c1ccc2nccn2c1
34697  CCOc1cc(C(C)(C)C)ccc1C1=N[C@@](C)(c2ccc(Cl)cc2…
34698  CC(C)[C@@H](CS(=O)(=O)C(C)C)N1C(=O)[C@@](C)(CC…
34699                          CCCCCCCCOC1OC(CO)C(O)C(O)C1O
34700                          CCCCCCCCOC1OC(CO)C(O)C(O)C1O

                                                    Target  \
0      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
1      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
2      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
3      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
4      MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
…                                                      …
34696  MNPAAEAEFNILLATDSYKVTHYKQYPPNTSKVYSYFECREKKTEN…
34697  MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKD…
34698  MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKD…
34699  MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLA…
34700  MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLA…

                                              Drug_vector  Clusters
0      [0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, …         4
1      [0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, …         2
2      [0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, …        24
3      [0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, …         0
4      [0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, …        16
…                                                      …       …
34696  [1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, …        12
34697  [1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, …        19
34698  [0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, …        46
34699  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …        10
34700  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …         0

[34701 rows x 4 columns]
```

Grouping the dataframe based on the clusters formed

[15]:
```
by_clusters = dataSource.groupby(['Clusters'])
```

[16]:
```
for cluster, data in by_clusters:
    print(f"First 2 entries for cluster: ", cluster)
    print("------------------------")
    print(data.head(2))
```

```
First 2 entries for cluster:  0
```

```
------------------------
                                                    Drug  \
3      CCN[C@H]1CN(CCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
45  Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…

                                                  Target  \
3   MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
45  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…

                                             Drug_vector  Clusters
3   [0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, …         0
45  [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …         0
First 2 entries for cluster:  1
------------------------
                                                    Drug  \
12          NS(=O)(=O)c1ccc(C(=O)NCc2ccc(F)c(F)c2F)cc1
105  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…

                                                  Target  \
12   SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPL…
105  MACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHF…

                                             Drug_vector  Clusters
12   [0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, …         1
105  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …         1
First 2 entries for cluster:  2
------------------------
                                                    Drug  \
1                NS(=O)(=O)c1ccc(S(=O)(=O)NCc2cccs2)s1
100  CNC(=S)N[C@@H]1[C@@H](O)[C@H](O[C@@H]2[C@@H](C…

                                                  Target  \
1    MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
100  MACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHF…

                                             Drug_vector  Clusters
1    [0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, …         2
100  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …         2
First 2 entries for cluster:  3
------------------------
                                                    Drug  \
131  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…
230  CC(C)(CN)c1nc(-c2ccncc2)c(-c2ccc3c(c2)CCC3N=O)…

                                                  Target  \
131  MMLSLNNLQNIIYNPVIPFVGTIPDQLDPGTLIVIRGHVPSDADRF…
230  MAALSGGGGGGAEPGQALFNGDMEPEAGAGAGAAASSAADPAIPEE…
```

```
                                                Drug_vector  Clusters
131  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …         3
230  [1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, …         3
First 2 entries for cluster:  4
------------------------
                                                    Drug  \
0          COc1ccc(CNS(=O)(=O)c2ccc(S(N)(=O)=O)s2)cc1
56  Cc1ccc(-n2nc(C(C)(C)C)cc2NC(=O)Nc2ccc(OCCN3CCO…

                                                  Target  \
0   MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
56  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…

                                               Drug_vector  Clusters
0   [0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, …         4
56  [1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, …         4
First 2 entries for cluster:  5
------------------------
                                      Drug  \
15                   NS(=O)(=O)c1ccccc1
28  Cc1cc(C)c2ccc(NC(=O)[C@H](C)N)nc2n1

                                                  Target  \
15  MASPDWGYDDKNGPEQWSKLYPIANGNNQSPVDIKTSETKHDTSLK…
28  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…

                                               Drug_vector  Clusters
15  [0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …         5
28  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, …         5
First 2 entries for cluster:  6
------------------------
                                                    Drug  \
27  CN[C@H](C)C(=O)N[C@H](C(=O)N1CCC[C@H]1C(=O)N[C…
58  Cc1ccc(-n2nc(C(C)(C)C)cc2NC(=O)Nc2ccc(OCCN3CCO…

                                                  Target  \
27  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
58  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…

                                               Drug_vector  Clusters
27  [0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, …         6
58  [1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, …         6
First 2 entries for cluster:  7
------------------------
                                                    Drug  \
33        C[C@H](N)C(=O)Nc1nc(-c2ccccc2)c(-c2ccccc2)s1
90  CC1(C)SSC(C)(C)[C@@H](NC(=O)[C@@H](N)Cc2ccc(O)…
```

```
                                                  Target  \
33  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
90  MEPAPSAGAELQPPLFANASDAYPSACPSAGANASGPPGARSASSL…

                                              Drug_vector  Clusters
33  [1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, …         7
90  [0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, …         7
First 2 entries for cluster:  8
------------------------
                                                    Drug  \
18  NS(=O)(=O)c1ccc(C(=O)NCCOCCOCCN(CC(=O)[O-])CC(…
53  Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…

                                                  Target  \
18  MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
53  MRGARGAWDFLCVLLLLLRVQTGSSQPSVSPGEPSPPSIHPGKSDL…

                                              Drug_vector  Clusters
18  [0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, …         8
53  [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …         8
First 2 entries for cluster:  9
------------------------
                                           Drug  \
7       NS(=O)(=O)c1ccc(C(=O)NCc2ccccc2)cc1
10  NS(=O)(=O)c1ccc(C(=O)NCc2cc(F)ccc2F)cc1

                                                  Target  \
7   SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPL…
10  SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPL…

                                              Drug_vector  Clusters
7   [0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, …         9
10  [0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, …         9
First 2 entries for cluster:  10
------------------------
                                                     Drug  \
47   Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…
117  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…

                                                   Target  \
47   HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
117  MADNFSLHDALSGSGNPNPQGWPGAWGNQPAGAGGYPGASYPGAYP…

                                               Drug_vector  Clusters
47   [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …         10
117  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …         10
First 2 entries for cluster:  11
------------------------
```

```
                                                    Drug  \
44  Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…
49  Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…

                                                  Target  \
44  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
49  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…

                                             Drug_vector  Clusters
44  [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …        11
49  [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …        11
First 2 entries for cluster:  12
------------------------
                                                    Drug  \
75   COc1cc2c(N3CCN(C(=O)Nc4ccc(OC(C)C)cc4)CC3)ncnc…
124  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…

                                                  Target  \
75   HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
124  MSNVPHKSSLPEGIRPGTVLRIRGLVPPNASRFHVNLLCGEEQGSD…

                                             Drug_vector  Clusters
75   [1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, …        12
124  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …        12
First 2 entries for cluster:  13
------------------------
                                                    Drug  \
70   CCN(CC)CCNC(=O)c1c(C)[nH]c(/C=C2\C(=O)Nc3ccc(F…
138  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…

                                                  Target  \
70   HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
138  MAFSGSQAPYLSPAVPFSGTIQGGLQDGLQITVNGTVLSSSGTRFA…

                                             Drug_vector  Clusters
70   [1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, …        13
138  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …        13
First 2 entries for cluster:  14
------------------------
                                                    Drug  \
8              NS(=O)(=O)c1ccc(C(=O)NCc2ccc(F)cc2)cc1
37  C[C@H](N)C(=O)Nc1nc(COc2cccc3ccccc23)c(Cc2cccc…

                                                  Target  \
8   SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPL…
37  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…

                                             Drug_vector  Clusters
```

```
8   [0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, …        14
37  [0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, …        14
First 2 entries for cluster:  15
------------------------
                                                      Drug  \
32        C[C@H](N)C(=O)Nc1nc(-c2ccccc2)c(C(C)(C)C)s1
76  COc1cc2c(N3CCN(C(=O)Nc4ccc(OC(C)C)cc4)CC3)ncnc…

                                                    Target  \
32  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
76  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…

                                               Drug_vector  Clusters
32  [0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, …        15
76  [1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, …        15
First 2 entries for cluster:  16
------------------------
                                                     Drug  \
4  CCN[C@H]1CN(CCCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
9          NS(=O)(=O)c1ccc(C(=O)NCc2ccc(F)cc2F)cc1

                                                     Target  \
4  MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
9  SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPL…

                                               Drug_vector  Clusters
4  [0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, …        16
9  [0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, …        16
First 2 entries for cluster:  17
------------------------
                                                      Drug  \
36      C[C@H](N)C(=O)Nc1nc(COc2ccccc2)c(Cc2ccccc2)s1
52  Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…

                                                    Target  \
36  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
52  MRGARGAWDFLCVLLLLLRVQTGSSQPSVSPGEPSPPSIHPGKSDL…

                                               Drug_vector  Clusters
36  [0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, …        17
52  [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …        17
First 2 entries for cluster:  18
------------------------
                                                       Drug  \
99   CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…
220                  O=C(Nc1ccncc1)C1(c2ccccc2)CCOCC1

                                                     Target  \
```

```
99   MACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHF…
220  MSAVALPRVSGGHDEHGHLEEFRTDPIGLMQRVRDECGDVGTFQLA…

                                                 Drug_vector  Clusters
99   [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …         18
220  [1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, …         18
First 2 entries for cluster:  19
------------------------

                                                    Drug  \
114  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…
202         O=C(O)COc1ccc(OCCNC[C@@H](O)COc2ccccc2)cc1

                                                  Target  \
114  MADNFSLHDALSGSGNPNPQGWPGAWGNQPAGAGGYPGASYPGAYP…
202  MAPWPHENSSLAPWPDLPTLAPNTANTSGLPGVPWEAALAGALLAL…

                                                 Drug_vector  Clusters
114  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …         19
202  [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …         19
First 2 entries for cluster:  20
------------------------

                                                    Drug  \
136  CNC(=S)N[C@@H]1[C@@H](O)[C@H](O[C@@H]2[C@@H](C…
267  NCCCNC(=O)CCCCCNC(=O)[C@H]1O[C@@H](n2cnc3c(N)n…

                                                  Target  \
136  MAFSGSQAPYLSPAVPFSGTIQGGLQDGLQITVNGTVLSSSGTRFA…
267  MGNAAAAKKGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQFDR…

                                                 Drug_vector  Clusters
136  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …         20
267  [1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, …         20
First 2 entries for cluster:  21
------------------------

                                                    Drug  \
19              NS(=O)(=O)Oc1ccc2c3c(c(=O)oc2c1)CCCCC3
22   CN[C@@H](C)C(=O)N[C@H](C(=O)N1CCC[C@H]1C(=O)NC…

                                                  Target  \
19   MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
22   MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…

                                                 Drug_vector  Clusters
19   [1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, …         21
22   [0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, …         21
First 2 entries for cluster:  22
------------------------

                                                    Drug  \
```

```
21  CN[C@@H](C)C(=O)N[C@H](C(=O)N1CCC[C@H]1C(=O)N[…
41              C[C@H](N)C(=O)Nc1nc2c(s1)CCc1ccccc1-2

                                                  Target  \
21  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
41  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…

                                             Drug_vector  Clusters
21  [0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, …        22
41  [0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, …        22
First 2 entries for cluster:  23
------------------------

                                                    Drug  \
43  Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…
48  Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…

                                                  Target  \
43  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
48  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…

                                             Drug_vector  Clusters
43  [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …        23
48  [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …        23
First 2 entries for cluster:  24
------------------------

                                                    Drug  \
2     COc1ccc(N2CC(O)c3cc(S(N)(=O)=O)sc3S2(=O)=O)cc1
50  Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…

                                                  Target  \
2   MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
50  HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…

                                             Drug_vector  Clusters
2   [0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, …        24
50  [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …        24
First 2 entries for cluster:  25
------------------------

                                                     Drug  \
129  CCCCNC(=S)N[C@@H]1[C@@H](O)[C@H](O[C@@H]2[C@@H…
407  CN1CCN(c2ccc3nc(-c4c(N)c5c(F)cccc5[nH]c4=O)[nH…

                                                   Target  \
129  MMLSLNNLQNIIYNPVIPFVGTIPDQLDPGTLIVIRGHVPSDADRF…
407  MALDVKSRAKRYEKLDFLGEGQFATVYKARDKNTNQIVAIKKIKLG…

                                              Drug_vector  Clusters
129  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …        25
```

```
407  [1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, …        25
First 2 entries for cluster:  26
------------------------
                                                    Drug  \
203           CCCC(=O)Nc1ccc(OCC(O)CNC(C)C)c(C(C)=O)c1
262  C[C@@H](NC(=O)CCCCCNC(=O)[C@H]1O[C@@H](n2cnc3c…

                                                  Target  \
203  MAPWPHENSSLAPWPDLPTLAPNTANTSGLPGVPWEAALAGALLAL…
262  MGNAAAAKKGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQFDR…

                                             Drug_vector  Clusters
203  [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …        26
262  [1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, …        26
First 2 entries for cluster:  27
------------------------
                                                    Drug  \
250  CCN1CCN(Cc2ccc(-c3cc4c(N[C@H](C)c5ccccc5)ncnc4…
381  CCOc1cc2ncc(C#N)c(Nc3ccc(F)c(Cl)c3)c2cc1NC(=O)…

                                                  Target  \
250  GEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIP…
381  MAWRCPRMGRVPLAWCLALCGWACMAPRGTQAEESPFVGNPGNITG…

                                             Drug_vector  Clusters
250  [1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, …        27
381  [1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, …        27
First 2 entries for cluster:  28
------------------------
                                                    Drug  \
67   CCN(CC)CCNC(=O)c1c(C)[nH]c(/C=C2\C(=O)Nc3ccc(F…
120  CNC(=S)N[C@@H]1[C@@H](O)[C@H](O[C@@H]2[C@@H](C…

                                                  Target  \
67   HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
120  MSNVPHKSSLPEGIRPGTVLRIRGLVPPNASRFHVNLLCGEEQGSD…

                                             Drug_vector  Clusters
67   [1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, …        28
120  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …        28
First 2 entries for cluster:  29
------------------------
                                                     Drug  \
6    CN[C@@H]1CN(c2cccc(OC)c2)S(=O)(=O)c2sc(S(N)(=O…
20   CN[C@@H](C)C(=O)N[C@H](C(=O)N1CCC[C@H]1C(=O)N[…

                                                   Target  \
6    MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
```

```
20  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…

                                          Drug_vector  Clusters
6   [0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, …        29
20  [0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, …        29
First 2 entries for cluster:  30
------------------------
                                             Drug  \
216                              c1ccc(-c2c[nH]cn2)cc1
264  NC(=O)CCCCCNC(=O)OC[C@H]1O[C@@H](n2cnc3c(N)ncn…

                                            Target  \
216  MSAVALPRVSGGHDEHGHLEEFRTDPIGLMQRVRDECGDVGTFQLA…
264  MGNAAAAKKGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQFDR…

                                          Drug_vector  Clusters
216  [0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, …        30
264  [1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, …        30
First 2 entries for cluster:  31
------------------------
                                             Drug  \
29             CCCCOc1ccc(-c2nnc(NC(=O)[C@H](C)N)s2)cc1
327  CNC(=O)[C@@H]1CCCN1C(=O)[C@H](NS(=O)(=O)Cc1ccc…

                                            Target  \
29   MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
327  MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRANT…

                                          Drug_vector  Clusters
29   [0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, …        31
327  [0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, …        31
First 2 entries for cluster:  32
------------------------
                                             Drug  \
14   NS(=O)(=O)c1ccc(C(=O)NCc2c(F)c(F)c(F)c(F)c2F)cc1
31                 Cc1sc(NC(=O)[C@H](C)N)nc1-c1ccccc1

                                            Target  \
14   SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPL…
31   MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…

                                          Drug_vector  Clusters
14   [0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, …        32
31   [0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, …        32
First 2 entries for cluster:  33
------------------------
                                              Drug  \
16   NS(=O)(=O)c1ccc(C(=O)NCCOCCOCCN(CC(=O)[O-])CC(…
```

```
26  C[C@@H](C(=O)N[C@H](C(=O)N1CCC[C@H]1C(=O)N[C@@…

                                                  Target  \
16  MASPDWGYDDKNGPEQWSKLYPIANGNNQSPVDIKTSETKHDTSLK…
26  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…

                                             Drug_vector  Clusters
16  [0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, …        33
26  [0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, …        33
First 2 entries for cluster:  34
------------------------

                                                    Drug  \
5          COc1cccc(N2CCc3cc(S(N)(=O)=O)sc3S2(=O)=O)c1
313  CCCNC(=O)[C@@H]1CCCN1C(=O)CC1(O)c2ccccc2-c2ccc…

                                                  Target  \
5    MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
313  MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRANT…

                                             Drug_vector  Clusters
5    [0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, …        34
313  [0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, …        34
First 2 entries for cluster:  35
------------------------

                                                    Drug  \
162               N#Cc1ccccc1OCC(O)CNCCNC(=O)Nc1ccccc1
257  NC(=O)CCCCCNC(=O)[C@H]1O[C@@H](n2cnc3c(N)ncnc3…

                                                  Target  \
162  MGAGVLVLGASEPGNLSSAAPLPDGAATAARLLVPASPPASLLPPA…
257  MGNAAAAKKGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQFDR…

                                             Drug_vector  Clusters
162  [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, …        35
257  [1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, …        35
First 2 entries for cluster:  36
------------------------

                                                    Drug  \
24  CN[C@@H](C)C(=O)N[C@H](C(=O)N1C[C@@H](Oc2ccccc…
95  CSCC[C@H](NC(=O)[C@H](Cc1ccc(OS(=O)(=O)O)cc1)N…

                                                  Target  \
24  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
95  MELLKLNRSVQGTGPGPGASLCRPGAPLLNSSSVGNLSCEPPRIRG…

                                             Drug_vector  Clusters
24  [0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, …        36
95  [1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, …        36
```

```
First 2 entries for cluster:  37
------------------------
                                                      Drug  \
229  CC(COc1nccc(-c2nc(-c3ccncc3)c(-c3ccc4c(c3)CCC4…
421  CCN1CCN(Cc2ccc(NC(=O)Nc3ccc(Oc4cc(NC)ncn4)cc3)…

                                                    Target  \
229  MAALSGGGGGGAEPGQALFNGDMEPEAGAGAGAAASSAADPAIPEE…
421  MAKQYDSVECPFCDEVSKYEKLAKIGQGTFGEVFKARHRKTGQKVA…

                                                Drug_vector  Clusters
229  [1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, …        37
421  [1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, …        37
First 2 entries for cluster:  38
------------------------
                                                      Drug  \
157        CC(=O)Nc1cc2c(cc1[N+](=O)[O-])-c1ccccc1C2=O
276  NC(=O)[C@@H](CCCN=C(N)N)NC(=O)[C@@H](CCCN=C(N)…

                                                    Target  \
157  MSGETTRLTEPQLRELAARGAAELDGATATDMLRWTDETFGDIGGA…
276  MGNAAAAKKGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQFDR…

                                                Drug_vector  Clusters
157  [0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, …        38
276  [1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, …        38
First 2 entries for cluster:  39
------------------------
                                                      Drug  \
42   Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc…
265  NCCCC[C@@H](NC(=O)CCCCCNC(=O)[C@H]1O[C@@H](n2c…

                                                    Target  \
42   HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
265  MGNAAAAKKGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQFDR…

                                                Drug_vector  Clusters
42   [1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, …        39
265  [1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, …        39
First 2 entries for cluster:  40
------------------------
                                                      Drug  \
17                                    NS(=O)(=O)c1ccccc1
103  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…

                                                    Target  \
17   MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKP…
103  MACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHF…
```

```
                                                Drug_vector  Clusters
17   [0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …        40
103  [1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, …        40
First 2 entries for cluster:  41
------------------------
                                                  Drug  \
40             C[C@H](N)C(=O)Nc1nc2c(ccc3ccccc32)s1
387  COc1cc2c(Nc3ccc(Br)cc3F)ncnc2cc1OCC1CCN(C)CC1

                                                 Target  \
40   MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
387  MAWRCPRMGRVPLAWCLALCGWACMAPRGTQAEESPFVGNPGNITG…

                                                Drug_vector  Clusters
40   [1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, …        41
387  [1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, …        41
First 2 entries for cluster:  42
------------------------
                                                   Drug  \
85   COc1cc2c(N3CCN(C(=O)Nc4ccc(OC(C)C)cc4)CC3)ncnc…
315            CNC(=O)[C@@H]1CCCN1C(=O)Cc1ccc2c(c1)OCO2

                                                 Target  \
85   HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
315  MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRANT…

                                                Drug_vector  Clusters
85   [1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, …        42
315  [0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, …        42
First 2 entries for cluster:  43
------------------------
                                                   Drug  \
13         NS(=O)(=O)c1ccc(C(=O)NCc2cc(F)c(F)c(F)c2)cc1
107  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…

                                                 Target  \
13   SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPL…
107  MACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHF…

                                                Drug_vector  Clusters
13   [0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, …        43
107  [1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, …        43
First 2 entries for cluster:  44
------------------------
                                                   Drug  \
34   C[C@H](N)C(=O)Nc1nc(-c2ccc(Cc3ccccc3)cc2)c(-c2…
35   C[C@H](N)C(=O)Nc1nc(-c2cccc(Cc3ccccc3)c2)c(Cc2…
```

```
                                                  Target  \
34  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
35  MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…

                                               Drug_vector  Clusters
34  [1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, …        44
35  [0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, …        44
First 2 entries for cluster:  45
------------------------
                                                 Drug  \
218                O=C1c2cc(O)ccc2-c2ccc(O)cc21
364  C#Cc1cccc(Nc2ncnc3cc(OCCOC)c(OCCOC)cc23)c1

                                                  Target  \
218  MSAVALPRVSGGHDEHGHLEEFRTDPIGLMQRVRDECGDVGTFQLA…
364  MSSPRAVVQLGKAQPAGEELATANQTAQQPSSPAMRRLTVDDFEIG…

                                               Drug_vector  Clusters
218  [0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …        45
364  [1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, …        45
First 2 entries for cluster:  46
------------------------
                                                   Drug  \
54   Cc1ccc(-n2nc(C(C)(C)C)cc2NC(=O)Nc2ccc(OCCN3CCO…
148                                 CCCCCCCCCC(=O)[O-]

                                                  Target  \
54   HHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKH…
148  MGLEKTVKEKLSFEGVGIHTGEYSKLIIHPEKEGTGIRFFKNGVYI…

                                               Drug_vector  Clusters
54   [1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, …        46
148  [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …        46
First 2 entries for cluster:  47
------------------------
                                                   Drug  \
39   C[C@H](N)C(=O)Nc1nc(COC2Cc3ccccc3C2)c(Cc2ccccc…
369  CN[C@H]1C[C@@H]2O[C@](C)([C@H]1OC)n1c3ccccc3c3…

                                                  Target  \
39   MSDAVSSDRNFPNSTNLPRNPSMADYEARIFTFGTWIYSVNKEQLA…
369  MSSPRAVVQLGKAQPAGEELATANQTAQQPSSPAMRRLTVDDFEIG…

                                               Drug_vector  Clusters
39   [0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, …        47
369  [1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …        47
First 2 entries for cluster:  48
```

```
------------------------
                                                    Drug  \
123  CO[C@@H]1O[C@H](CO)[C@@H](O[C@@H]2O[C@H](CO)[C…
350      COc1cc2c(Nc3ccc(Br)cc3F)ncnc2cc1OCC1CCN(C)CC1

                                                  Target  \
123  MSNVPHKSSLPEGIRPGTVLRIRGLVPPNASRFHVNLLCGEEQGSD…
350  MGAIGLLWLLPLLLSTAAVGSGMGTGQRAGSPAAGPPLQPREPLSY…

                                             Drug_vector  Clusters
123  [1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, …        48
350  [1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, …        48
First 2 entries for cluster:  49
------------------------
                                              Drug  \
180         N#Cc1ccccc1OCC(O)CNCCNC(=O)Nc1ccccc1
325  COCCNC(=O)[C@@H]1CCCN1C(=O)[C@H](N)C1CCCCC1

                                                  Target  \
180  MGQPGNGSAFLLAPNGSHAPDHDVTQERDEVWVVGMGIVMSLIVLA…
325  MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRANT…

                                             Drug_vector  Clusters
180  [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, …        49
325  [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …        49
```

[ ]: