In [147]:
```
import pandas as pd
```

In [144]:
```
pwd;
```

Out[144]:
```
'C:\\Users\\L03121898\\Desktop\\cadi22'
```

In [148]:
```
x=pd.read_csv("sample.csv")
```

In [153]:
```
x.head()
```

Out[153]:

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	Smiles
0	CHEMBL4450911	NaN	303.19	C15H18BNO3S	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	NaN	303.19	C15H18BNO3S	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	NaN	307.18	C14H18BNO4S	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C14H18BNO4S	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	Small molecule	311.40	C15H21NO4S	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...

In [154]:
```
x.tail()
```

Out[154]:

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	Smiles
38	CHEMBL350414	Small molecule	405.52	C24H23NO3S	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	C24H23NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...
40	CHEMBL409721	Small molecule	410.41	C19H27N2O6P	CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@...
41	CHEMBL271224	Small molecule	411.57	C24H29NO3S	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
42	CHEMBL271224	Large molecule	411.57	C24H29NO3S	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [155]:
```
x.shape
```

Out[155]:
```
(43, 5)
```

In [156]:
```
x.columns
```

Out[156]:
```
Index(['ChEMBL ID', 'Type', 'Molecular Weight', 'Molecular Formula', 'Smiles'], dtype='object')
```

In [151]:
```
x["Molecular Weight"].sum()
```

Out[151]: 15811.72

In [150]:
```
x["Molecular Weight"].min()
```

Out[150]: 303.19

In [149]:
```
x["Molecular Weight"].max()
```

Out[149]: 411.57

In [56]:
```
x["Molecular Weight"].mean()
```

Out[56]: 367.71441860465126

In [57]:
```
x["Molecular Weight"].median()
```

Out[57]: 371.5

In [59]:
```
x["Molecular Weight"].idxmax()
```

Out[59]: 41

In [79]:
```
x.iloc[41:42]
```

Out[79]:

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	Smiles
41	CHEMBL271224	Small molecule	411.57	C24H29NO3S	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [157]:
```
x2=x[["ChEMBL ID"]]
```

In [158]:
```
vaccine={"vaccine_name":["a","b","c","d"],"activity":["not active","active","
```

In [123]:
```
data={"a":1,"b":2,"c":3,"d":4,"e":5}
```

In [159]:
```
pd.Series(data)
```

Out[159]:
```
a    1
b    2
c    3
d    4
e    5
dtype: int64
```

In [160]:
```
y=pd.Series(vaccine)
```

In [161]:
```
y
```

Out[161]:
```
vaccine_name                    [a, b, c, d]
activity        [not active, active, 1, 2, 3]
dtype: object
```

In [162]:
```
y.to_csv("vaccine_data.csv")
```

In [163]:
```
x.head()
```

Out[163]:

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	Smiles
0	CHEMBL4450911	NaN	303.19	C15H18BNO3S	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	NaN	303.19	C15H18BNO3S	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	NaN	307.18	C14H18BNO4S	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C14H18BNO4S	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	Small molecule	311.40	C15H21NO4S	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...

In [63]:
```
x["ChEMBL ID"].value_counts()
```

Out[63]:
```
CHEMBL271224     2
CHEMBL411298     1
CHEMBL252417     1
CHEMBL404117     1
CHEMBL271223     1
CHEMBL252003     1
CHEMBL254282     1
CHEMBL251804     1
CHEMBL437595     1
CHEMBL4202767    1
CHEMBL257026     1
CHEMBL4444926    1
CHEMBL3235416    1
CHEMBL254493     1
CHEMBL254495     1
CHEMBL254703     1
CHEMBL258683     1
CHEMBL350414     1
CHEMBL163454     1
CHEMBL409721     1
CHEMBL252391     1
CHEMBL4450911    1
CHEMBL257270     1
CHEMBL401397     1
CHEMBL4439726    1
CHEMBL4460127    1
CHEMBL253224     1
CHEMBL405232     1
CHEMBL258333     1
CHEMBL257727     1
CHEMBL4451026    1
CHEMBL257726     1
CHEMBL404044     1
CHEMBL253428     1
CHEMBL400527     1
CHEMBL412123     1
CHEMBL409713     1
CHEMBL269997     1
CHEMBL257229     1
CHEMBL269996     1
CHEMBL271225     1
CHEMBL398545     1
Name: ChEMBL ID, dtype: int64
```

In [165]:
```
x["Type"].value_counts()
```

Out[165]:
```
Small molecule      36
Median molecules     1
Large molecule       1
Name: Type, dtype: int64
```

In [164]:
```
x[x["Type"].str.contains("Large", na=False)]
```

Out[164]:

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	Smiles
42	CHEMBL271224	Large molecule	411.57	C24H29NO3S	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [166]:
```
x[x["Type"].str.contains("Small", na=False)]
```

Out[166]:

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	
4	CHEMBL253224	Small molecule	311.40	C15H21NO4S	CC[C@@H](C)[C@H](S)(C
5	CHEMBL405232	Small molecule	313.33	C15H24NO4P	CC(C)C[C@@H](N)P(=(Cc
6	CHEMBL258333	Small molecule	315.39	C17H17NO3S	O=C(CS)N[C@@H](Cc1ccc(-c2cc
7	CHEMBL257727	Small molecule	329.42	C18H19NO3S	C[C@H](S)C(=O)N[Cc2cc
9	CHEMBL257726	Small molecule	343.45	C19H21NO3S	CC[C@H](S)C(=O)N[Cc
10	CHEMBL404044	Small molecule	343.45	C19H21NO3S	CC(C)(S)C(=O)N[Cc2cc
12	CHEMBL400527	Small molecule	345.46	C19H23NO3S	CC[C@@H](C)[C@H](S)(C
13	CHEMBL412123	Small molecule	347.35	C18H22NO4P	N[C@H](Cc1ccccc1)P(=(Cc
14	CHEMBL409713	Small molecule	355.33	C16H22NO6P	C[C[C@H]1CCCN1C(=O)OCc
15	CHEMBL269997	Small molecule	357.48	C20H23NO3S	CC(C)[C@H](S)C(=O)N[C
16	CHEMBL257229	Small molecule	369.49	C21H23NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)
17	CHEMBL269996	Small molecule	371.50	C21H25NO3S	CC[C@H](C)[C@H](S)C(=O)N[C
18	CHEMBL271225	Small molecule	371.50	C21H25NO3S	CC(C)C[C@H](S)C(=O)N[C
19	CHEMBL401397	Small molecule	371.50	C21H25NO3S	CC[C@@H](C)[C@H](S)C(=O)N[C
20	CHEMBL257270	Small molecule	371.50	C21H25NO3S	CCCC[C@H](S)C(=O)N[C
21	CHEMBL398545	Small molecule	371.50	C21H25NO3S	CC[C@@H](C)[C@H](S)C(=O)N[C@
22	CHEMBL252391	Small molecule	371.50	C21H25NO3S	CC[C@@H](C)[C@H](S)C(=O)N[C
23	CHEMBL252417	Small molecule	375.84	C17H14ClN3O3S	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3
24	CHEMBL404117	Small molecule	383.51	C22H25NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)
25	CHEMBL271223	Small molecule	385.53	C22H27NO3S	CC(C)(C)C[C@H](S)C(=O)N[C
26	CHEMBL252003	Small molecule	387.50	C21H25NO4S	CC[C@@H](C)[C@H](S)(C
27	CHEMBL254282	Small molecule	387.50	C21H25NO4S	CC[C@@H](C)[C@H](S)(C

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	
28	CHEMBL251804	Small molecule	387.50	C21H25NO4S	CC[C@@H](C)[C@H](S)(C
29	CHEMBL437595	Small molecule	391.49	C23H21NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)
30	CHEMBL4202767	Small molecule	391.85	C20H22ClNO5	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c(
31	CHEMBL411298	Small molecule	397.41	C19H28NO6P	CC(C)C[C[C@H]1CCCN1C(
32	CHEMBL257026	Small molecule	397.54	C23H27NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)
33	CHEMBL3235416	Small molecule	401.47	C20H27N5O4	(Cc1c[nH]cn1)NC(=O)CN(
34	CHEMBL254493	Small molecule	401.53	C22H27NO4S	CC[C@@H](C)[C@H](S)(Cc
35	CHEMBL254495	Small molecule	401.53	C22H27NO4S	CC[C@@H](C)[C@H](S)(C
36	CHEMBL254703	Small molecule	401.53	C22H27NO4S	CC[C@@H](C)[C@H](S)(C
37	CHEMBL258683	Small molecule	405.39	C20H24NO6P	C[C@@H](NC(=O)OCc1ccccc1)P(=
38	CHEMBL350414	Small molecule	405.52	C24H23NO3S	O=C(N[C@@H](Cc1ccc(-c2ccc
39	CHEMBL163454	Small molecule	405.52	C24H23NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc
40	CHEMBL409721	Small molecule	410.41	C19H27N2O6P	CC(=O)N[C@@H](C)C(=O)N1CCC
41	CHEMBL271224	Small molecule	411.57	C24H29NO3S	O=C(N[C@@H](Cc1ccc(-c2ccc

In [167]:
```
x[x["Type"].str.contains("Small|Median", na=False)]
```

Out[167]:

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	
4	CHEMBL253224	Small molecule	311.40	C15H21NO4S	CC[C@@H](C)[C@H](S (
5	CHEMBL405232	Small molecule	313.33	C15H24NO4P	CC(C)C[C@@H](N)P (C
6	CHEMBL258333	Small molecule	315.39	C17H17NO3S	O=C(CS)N[C@@H](Cc1ccc(-c2c
7	CHEMBL257727	Small molecule	329.42	C18H19NO3S	C[C@H](S)C(=O)N| c2c
9	CHEMBL257726	Small molecule	343.45	C19H21NO3S	CC[C@H](S)C(=O)N|
10	CHEMBL404044	Small molecule	343.45	C19H21NO3S	CC(C)(S)C(=O)N| c2c
11	CHEMBL253428	Median molecules	345.46	C19H23NO3S	CC[C@@H](C)[C@H](S (
12	CHEMBL400527	Small molecule	345.46	C19H23NO3S	CC[C@@H](C)[C@H](S (
13	CHEMBL412123	Small molecule	347.35	C18H22NO4P	N[C@H](Cc1ccccc1)P (C
14	CHEMBL409713	Small molecule	355.33	C16H22NO6P	C [C@H]1CCCN1C(=O)OC
15	CHEMBL269997	Small molecule	357.48	C20H23NO3S	CC(C)[C@H](S)C(=O)N|
16	CHEMBL257229	Small molecule	369.49	C21H23NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1
17	CHEMBL269996	Small molecule	371.50	C21H25NO3S	CC[C@H](C)[C@H](S)C(=O)N|
18	CHEMBL271225	Small molecule	371.50	C21H25NO3S	CC(C)C[C@H](S)C(=O)N|
19	CHEMBL401397	Small molecule	371.50	C21H25NO3S	CC[C@@H](C)[C@H](S)C(=O)N[C
20	CHEMBL257270	Small molecule	371.50	C21H25NO3S	CCCC[C@H](S)C(=O)N|
21	CHEMBL398545	Small molecule	371.50	C21H25NO3S	CC[C@@H](C)[C@H](S)C(=O)N[C@
22	CHEMBL252391	Small molecule	371.50	C21H25NO3S	CC[C@@H](C)[C@H](S)C(=O)N|
23	CHEMBL252417	Small molecule	375.84	C17H14ClN3O3S	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc
24	CHEMBL404117	Small molecule	383.51	C22H25NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1
25	CHEMBL271223	Small molecule	385.53	C22H27NO3S	CC(C)(C)C[C@H](S)C(=O)N|
26	CHEMBL252003	Small molecule	387.50	C21H25NO4S	CC[C@@H](C)[C@H](S (

	ChEMBL ID	Type	Molecular Weight	Molecular Formula	
27	CHEMBL254282	Small molecule	387.50	C21H25NO4S	CC[C@@H](C)[C@H](S ((
28	CHEMBL251804	Small molecule	387.50	C21H25NO4S	CC[C@@H](C)[C@H](S ((
29	CHEMBL437595	Small molecule	391.49	C23H21NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1
30	CHEMBL4202767	Small molecule	391.85	C20H22ClNO5	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c
31	CHEMBL411298	Small molecule	397.41	C19H28NO6P	CC(C)C [C@H]1CCCN1C
32	CHEMBL257026	Small molecule	397.54	C23H27NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1
33	CHEMBL3235416	Small molecule	401.47	C20H27N5O4	(Cc1c[nH]cn1)NC(=O)CN
34	CHEMBL254493	Small molecule	401.53	C22H27NO4S	CC[C@@H](C)[C@H](S (C
35	CHEMBL254495	Small molecule	401.53	C22H27NO4S	CC[C@@H](C)[C@H](S ((
36	CHEMBL254703	Small molecule	401.53	C22H27NO4S	CC[C@@H](C)[C@H](S ((
37	CHEMBL258683	Small molecule	405.39	C20H24NO6P	C[C@@H](NC(=O)OCc1ccccc1)P
38	CHEMBL350414	Small molecule	405.52	C24H23NO3S	O=C(N[C@@H](Cc1ccc(-c2cc
39	CHEMBL163454	Small molecule	405.52	C24H23NO3S	O=C(O)[C@H](Cc1ccc(-c2ccccc2)c
40	CHEMBL409721	Small molecule	410.41	C19H27N2O6P	CC(=O)N[C@@H](C)C(=O)N1CC
41	CHEMBL271224	Small molecule	411.57	C24H29NO3S	O=C(N[C@@H](Cc1ccc(-c2cc

In [168]:
```
del x["Molecular Formula"]
```

In [170]:
```
x.drop("Type",axis=1)
```

Out[170]:

	ChEMBL ID	Molecular Weight	Smiles
0	CHEMBL4450911	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...
5	CHEMBL405232	313.33	CC(C)C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
6	CHEMBL258333	315.39	O=C(CS)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
7	CHEMBL257727	329.42	C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
8	CHEMBL4451026	343.26	O=C(N[C@@H](Cc1ccccc1)B(O)O)C(CS)Cc1ccccc1
9	CHEMBL257726	343.45	CC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(...
10	CHEMBL404044	343.45	CC(C)(S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
11	CHEMBL253428	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc2ccccc2c...
12	CHEMBL400527	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc2ccccc1...
13	CHEMBL412123	347.35	N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
14	CHEMBL409713	355.33	C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O
15	CHEMBL269997	357.48	CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1...
16	CHEMBL257229	369.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
17	CHEMBL269996	371.50	CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc...
18	CHEMBL271225	371.50	CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc...
19	CHEMBL401397	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(-c2ccc...
20	CHEMBL257270	371.50	CCCC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)...
21	CHEMBL398545	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1-c1cc...
22	CHEMBL252391	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc...
23	CHEMBL252417	375.84	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3C)o2)C(=O)O)[...
24	CHEMBL404117	383.51	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
25	CHEMBL271223	385.53	CC(C)(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2...
26	CHEMBL252003	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(Oc2ccc...
27	CHEMBL254282	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(Oc2cccc...
28	CHEMBL251804	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1Oc1cc...
29	CHEMBL437595	391.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
30	CHEMBL4202767	391.85	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c(C)c2)cc1OC(C)C
31	CHEMBL411298	397.41	CC(C)C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc...
32	CHEMBL257026	397.54	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...

	ChEMBL ID	Molecular Weight	Smiles
33	CHEMBL3235416	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
36	CHEMBL254703	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...
40	CHEMBL409721	410.41	CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@...
41	CHEMBL271224	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
42	CHEMBL271224	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [ ]:
```
axis=0
axis=1
```

In [88]:
```
y=pd.read_csv("sample.csv")
y1=pd.read_csv("sample.csv")
```

In [89]:
```
combined=pd.concat([y,y1], axis=0)
```

In [90]:
```
combined.shape
```

Out[90]: (86, 5)

In [91]:
```
combined.to_csv("double.csv")
```

In [171]:
```
good= x.loc[x["Molecular Weight"]<=350]
```

In [172]:
```
good
```

Out[172]:

	ChEMBL ID	Type	Molecular Weight	Smiles
0	CHEMBL4450911	NaN	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	NaN	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	NaN	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	Small molecule	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...
5	CHEMBL405232	Small molecule	313.33	CC(C)C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
6	CHEMBL258333	Small molecule	315.39	O=C(CS)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
7	CHEMBL257727	Small molecule	329.42	C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
8	CHEMBL4451026	NaN	343.26	O=C(N[C@@H](Cc1ccccc1)B(O)O)C(CS)Cc1ccccc1
9	CHEMBL257726	Small molecule	343.45	CC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(...
10	CHEMBL404044	Small molecule	343.45	CC(C)(S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
11	CHEMBL253428	Median molecules	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc2ccccc2c...
12	CHEMBL400527	Small molecule	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc2ccccc1...
13	CHEMBL412123	Small molecule	347.35	N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O

In [97]:
```
bad= x.loc[x["Molecular Weight"]>350]
```

In [173]:
```
bad
```

Out[173]:

	ChEMBL ID	Type	Molecular Weight	Smiles
14	CHEMBL409713	Small molecule	355.33	C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O
15	CHEMBL269997	Small molecule	357.48	CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1...
16	CHEMBL257229	Small molecule	369.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
17	CHEMBL269996	Small molecule	371.50	CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc...
18	CHEMBL271225	Small molecule	371.50	CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc...
19	CHEMBL401397	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(-c2ccc...
20	CHEMBL257270	Small molecule	371.50	CCCC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)...
21	CHEMBL398545	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1-c1cc...
22	CHEMBL252391	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc...
23	CHEMBL252417	Small molecule	375.84	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3C)o2)C(=O)O)[...
24	CHEMBL404117	Small molecule	383.51	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
25	CHEMBL271223	Small molecule	385.53	CC(C)(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2...
26	CHEMBL252003	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(Oc2ccc...
27	CHEMBL254282	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(Oc2cccc...
28	CHEMBL251804	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1Oc1cc...
29	CHEMBL437595	Small molecule	391.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
30	CHEMBL4202767	Small molecule	391.85	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c(C)c2)cc1OC(C)C
31	CHEMBL411298	Small molecule	397.41	CC(C)C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc...
32	CHEMBL257026	Small molecule	397.54	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...

	ChEMBL ID	Type	Molecular Weight	Smiles
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...
40	CHEMBL409721	Small molecule	410.41	CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@...
41	CHEMBL271224	Small molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
42	CHEMBL271224	Large molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [138]:
```
good["label"]="good";
```

```
C:\Users\L03121898\AppData\Local\Schrodinger\PyMOL2\envs\Cadi22\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [174]:
```
good.head()
```

Out[174]:

	ChEMBL ID	Type	Molecular Weight	Smiles
0	CHEMBL4450911	NaN	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	NaN	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	NaN	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	Small molecule	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...

In [101]:
```
bad["label"]="bad"
```

```
C:\Users\L03121898\AppData\Local\Schrodinger\PyMOL2\envs\Cadi22\lib\site-pa
ckages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://p
andas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-vi
ew-versus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [102]:
```
bad.head()
```

Out[102]:

	ChEMBL ID	Type	Molecular Weight	Smiles	label
14	CHEMBL409713	Small molecule	355.33	C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O	bad
15	CHEMBL269997	Small molecule	357.48	CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1...	bad
16	CHEMBL257229	Small molecule	369.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...	bad
17	CHEMBL269996	Small molecule	371.50	CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc...	bad
18	CHEMBL271225	Small molecule	371.50	CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc...	bad

In [175]:
```
conbined=pd.concat([good,bad],axis=0)
```

In [106]:
```
conbined
```

Out[106]:

	ChEMBL ID	Type	Molecular Weight	Smiles
0	CHEMBL4450911	NaN	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	NaN	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	NaN	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	Small molecule	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...
5	CHEMBL405232	Small molecule	313.33	CC(C)C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
6	CHEMBL258333	Small molecule	315.39	O=C(CS)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
7	CHEMBL257727	Small molecule	329.42	C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
8	CHEMBL4451026	NaN	343.26	O=C(N[C@@H](Cc1ccccc1)B(O)O)C(CS)Cc1ccccc1
9	CHEMBL257726	Small molecule	343.45	CC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(...
10	CHEMBL404044	Small molecule	343.45	CC(C)(S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
11	CHEMBL253428	Median molecules	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc2ccccc2c...
12	CHEMBL400527	Small molecule	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc2ccccc1...
13	CHEMBL412123	Small molecule	347.35	N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
14	CHEMBL409713	Small molecule	355.33	C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O
15	CHEMBL269997	Small molecule	357.48	CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1...
16	CHEMBL257229	Small molecule	369.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
17	CHEMBL269996	Small molecule	371.50	CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc...
18	CHEMBL271225	Small molecule	371.50	CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc...
19	CHEMBL401397	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(-c2ccc...
20	CHEMBL257270	Small molecule	371.50	CCCC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)...
21	CHEMBL398545	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1-c1cc...
22	CHEMBL252391	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc...

	ChEMBL ID	Type	Molecular Weight	Smiles
23	CHEMBL252417	Small molecule	375.84	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3C)o2)C(=O)O)[...
24	CHEMBL404117	Small molecule	383.51	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
25	CHEMBL271223	Small molecule	385.53	CC(C)(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2...
26	CHEMBL252003	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(Oc2ccc...
27	CHEMBL254282	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(Oc2cccc...
28	CHEMBL251804	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1Oc1cc...
29	CHEMBL437595	Small molecule	391.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
30	CHEMBL4202767	Small molecule	391.85	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c(C)c2)cc1OC(C)C
31	CHEMBL411298	Small molecule	397.41	CC(C)C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc...
32	CHEMBL257026	Small molecule	397.54	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...
40	CHEMBL409721	Small molecule	410.41	CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@...
41	CHEMBL271224	Small molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
42	CHEMBL271224	Large molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [108]:
```
conbined["new_molwt"]=conbined["Molecular Weight"]+10
```

In [109]:
```
conbined.head()
```

Out[109]:

	ChEMBL ID	Type	Molecular Weight	Smiles	label	new_molwt
0	CHEMBL4450911	NaN	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1	good	313.19
1	CHEMBL4444926	NaN	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1	good	313.19
2	CHEMBL4439726	NaN	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O	good	317.18
3	CHEMBL4460127	NaN	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O	good	317.18
4	CHEMBL253224	Small molecule	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...	good	321.40

In [110]:
```
x
```

Out[110]:

	ChEMBL ID	Type	Molecular Weight	Smiles
0	CHEMBL4450911	NaN	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	NaN	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	NaN	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	Small molecule	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...
5	CHEMBL405232	Small molecule	313.33	CC(C)C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
6	CHEMBL258333	Small molecule	315.39	O=C(CS)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
7	CHEMBL257727	Small molecule	329.42	C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
8	CHEMBL4451026	NaN	343.26	O=C(N[C@@H](Cc1ccccc1)B(O)O)C(CS)Cc1ccccc1
9	CHEMBL257726	Small molecule	343.45	CC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(...
10	CHEMBL404044	Small molecule	343.45	CC(C)(S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
11	CHEMBL253428	Median molecules	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc2ccccc2c...
12	CHEMBL400527	Small molecule	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc2ccccc1...
13	CHEMBL412123	Small molecule	347.35	N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
14	CHEMBL409713	Small molecule	355.33	C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O
15	CHEMBL269997	Small molecule	357.48	CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1...
16	CHEMBL257229	Small molecule	369.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
17	CHEMBL269996	Small molecule	371.50	CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc...
18	CHEMBL271225	Small molecule	371.50	CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc...
19	CHEMBL401397	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(-c2ccc...
20	CHEMBL257270	Small molecule	371.50	CCCC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)...
21	CHEMBL398545	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1-c1cc...
22	CHEMBL252391	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc...

	ChEMBL ID	Type	Molecular Weight	Smiles
23	CHEMBL252417	Small molecule	375.84	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3C)o2)C(=O)O)[...
24	CHEMBL404117	Small molecule	383.51	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
25	CHEMBL271223	Small molecule	385.53	CC(C)(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2...
26	CHEMBL252003	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(Oc2ccc...
27	CHEMBL254282	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(Oc2cccc...
28	CHEMBL251804	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1Oc1cc...
29	CHEMBL437595	Small molecule	391.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
30	CHEMBL4202767	Small molecule	391.85	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c(C)c2)cc1OC(C)C
31	CHEMBL411298	Small molecule	397.41	CC(C)C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc...
32	CHEMBL257026	Small molecule	397.54	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...
40	CHEMBL409721	Small molecule	410.41	CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@...
41	CHEMBL271224	Small molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
42	CHEMBL271224	Large molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [113]:
```
x[(x["Molecular Weight"]>400) & (x["Molecular Weight"]<410)]
```

Out[113]:

	ChEMBL ID	Type	Molecular Weight	Smiles
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...

In [115]:
```
x3=x[(x["Molecular Weight"]>400) & (x["Molecular Weight"]<410)]
```

In [116]:
```
x3
```

Out[116]:

	ChEMBL ID	Type	Molecular Weight	Smiles
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...

In [117]:
```
x[x["Molecular Weight"].between(400,410)]
```

Out[117]:

	ChEMBL ID	Type	Molecular Weight	Smiles
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...

In [118]:
```
x.sort_values("Molecular Weight", ascending=False)
```

Out[118]:

	ChEMBL ID	Type	Molecular Weight	Smiles
42	CHEMBL271224	Large molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
41	CHEMBL271224	Small molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
40	CHEMBL409721	Small molecule	410.41	CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
32	CHEMBL257026	Small molecule	397.54	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
31	CHEMBL411298	Small molecule	397.41	CC(C)C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc...
30	CHEMBL4202767	Small molecule	391.85	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c(C)c2)cc1OC(C)C
29	CHEMBL437595	Small molecule	391.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
28	CHEMBL251804	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1Oc1cc...
27	CHEMBL254282	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(Oc2cccc...
26	CHEMBL252003	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(Oc2ccc...
25	CHEMBL271223	Small molecule	385.53	CC(C)(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2...
24	CHEMBL404117	Small molecule	383.51	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
23	CHEMBL252417	Small molecule	375.84	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3C)o2)C(=O)O)[...
22	CHEMBL252391	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc...
21	CHEMBL398545	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1-c1cc...

	ChEMBL ID	Type	Molecular Weight	Smiles
20	CHEMBL257270	Small molecule	371.50	CCCC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)...
19	CHEMBL401397	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(-c2ccc...
18	CHEMBL271225	Small molecule	371.50	CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc...
17	CHEMBL269996	Small molecule	371.50	CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc...
16	CHEMBL257229	Small molecule	369.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
15	CHEMBL269997	Small molecule	357.48	CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1...
14	CHEMBL409713	Small molecule	355.33	C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O
13	CHEMBL412123	Small molecule	347.35	N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
11	CHEMBL253428	Median molecules	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc2ccccc2c...
12	CHEMBL400527	Small molecule	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc2ccccc1...
10	CHEMBL404044	Small molecule	343.45	CC(C)(S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
9	CHEMBL257726	Small molecule	343.45	CC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(...
8	CHEMBL4451026	NaN	343.26	O=C(N[C@@H](Cc1ccccc1)B(O)O)C(CS)Cc1ccccc1
7	CHEMBL257727	Small molecule	329.42	C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
6	CHEMBL258333	Small molecule	315.39	O=C(CS)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
5	CHEMBL405232	Small molecule	313.33	CC(C)C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
4	CHEMBL253224	Small molecule	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...
2	CHEMBL4439726	NaN	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
1	CHEMBL4444926	NaN	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
0	CHEMBL4450911	NaN	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1

In [119]:
```
x["ChEMBL ID"].value_counts()
```

Out[119]:
```
CHEMBL271224     2
CHEMBL411298     1
CHEMBL252417     1
CHEMBL404117     1
CHEMBL271223     1
CHEMBL252003     1
CHEMBL254282     1
CHEMBL251804     1
CHEMBL437595     1
CHEMBL4202767    1
CHEMBL257026     1
CHEMBL4444926    1
CHEMBL3235416    1
CHEMBL254493     1
CHEMBL254495     1
CHEMBL254703     1
CHEMBL258683     1
CHEMBL350414     1
CHEMBL163454     1
CHEMBL409721     1
CHEMBL252391     1
CHEMBL4450911    1
CHEMBL257270     1
CHEMBL401397     1
CHEMBL4439726    1
CHEMBL4460127    1
CHEMBL253224     1
CHEMBL405232     1
CHEMBL258333     1
CHEMBL257727     1
CHEMBL4451026    1
CHEMBL257726     1
CHEMBL404044     1
CHEMBL253428     1
CHEMBL400527     1
CHEMBL412123     1
CHEMBL409713     1
CHEMBL269997     1
CHEMBL257229     1
CHEMBL269996     1
CHEMBL271225     1
CHEMBL398545     1
Name: ChEMBL ID, dtype: int64
```

In [120]:
```
x.drop_duplicates("ChEMBL ID")
```

Out[120]:

	ChEMBL ID	Type	Molecular Weight	Smiles
0	CHEMBL4450911	NaN	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	NaN	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	NaN	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	Small molecule	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...
5	CHEMBL405232	Small molecule	313.33	CC(C)C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
6	CHEMBL258333	Small molecule	315.39	O=C(CS)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
7	CHEMBL257727	Small molecule	329.42	C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
8	CHEMBL4451026	NaN	343.26	O=C(N[C@@H](Cc1ccccc1)B(O)O)C(CS)Cc1ccccc1
9	CHEMBL257726	Small molecule	343.45	CC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(...
10	CHEMBL404044	Small molecule	343.45	CC(C)(S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
11	CHEMBL253428	Median molecules	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc2ccccc2c...
12	CHEMBL400527	Small molecule	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc2ccccc1...
13	CHEMBL412123	Small molecule	347.35	N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
14	CHEMBL409713	Small molecule	355.33	C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O
15	CHEMBL269997	Small molecule	357.48	CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1...
16	CHEMBL257229	Small molecule	369.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
17	CHEMBL269996	Small molecule	371.50	CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc...
18	CHEMBL271225	Small molecule	371.50	CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc...
19	CHEMBL401397	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(-c2ccc...
20	CHEMBL257270	Small molecule	371.50	CCCC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)...
21	CHEMBL398545	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1-c1cc...
22	CHEMBL252391	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc...

	ChEMBL ID	Type	Molecular Weight	Smiles
23	CHEMBL252417	Small molecule	375.84	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3C)o2)C(=O)O)[...
24	CHEMBL404117	Small molecule	383.51	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
25	CHEMBL271223	Small molecule	385.53	CC(C)(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2...
26	CHEMBL252003	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(Oc2ccc...
27	CHEMBL254282	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(Oc2cccc...
28	CHEMBL251804	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1Oc1cc...
29	CHEMBL437595	Small molecule	391.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
30	CHEMBL4202767	Small molecule	391.85	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c(C)c2)cc1OC(C)C
31	CHEMBL411298	Small molecule	397.41	CC(C)C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc...
32	CHEMBL257026	Small molecule	397.54	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...
40	CHEMBL409721	Small molecule	410.41	CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@...
41	CHEMBL271224	Small molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [121]:
```
x.drop_duplicates("ChEMBL ID",keep="last")
```

Out[121]:

	ChEMBL ID	Type	Molecular Weight	Smiles
0	CHEMBL4450911	NaN	303.19	C[C@H](CS)C(=O)N[C@H](B(O)O)c1ccc2ccccc2c1
1	CHEMBL4444926	NaN	303.19	C[C@H](CS)C(=O)N[C@@H](B(O)O)c1ccc2ccccc2c1
2	CHEMBL4439726	NaN	307.18	C[C@H](CS)C(=O)N[C@@H](Cc1coc2ccccc12)B(O)O
3	CHEMBL4460127	NaN	307.18	C[C@H](CS)C(=O)N[C@H](Cc1coc2ccccc12)B(O)O
4	CHEMBL253224	Small molecule	311.40	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C...
5	CHEMBL405232	Small molecule	313.33	CC(C)C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
6	CHEMBL258333	Small molecule	315.39	O=C(CS)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
7	CHEMBL257727	Small molecule	329.42	C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
8	CHEMBL4451026	NaN	343.26	O=C(N[C@@H](Cc1ccccc1)B(O)O)C(CS)Cc1ccccc1
9	CHEMBL257726	Small molecule	343.45	CC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(...
10	CHEMBL404044	Small molecule	343.45	CC(C)(S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O
11	CHEMBL253428	Median molecules	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc2ccccc2c...
12	CHEMBL400527	Small molecule	345.46	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc2ccccc1...
13	CHEMBL412123	Small molecule	347.35	N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O
14	CHEMBL409713	Small molecule	355.33	C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O
15	CHEMBL269997	Small molecule	357.48	CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1...
16	CHEMBL257229	Small molecule	369.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
17	CHEMBL269996	Small molecule	371.50	CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc...
18	CHEMBL271225	Small molecule	371.50	CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc...
19	CHEMBL401397	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(-c2ccc...
20	CHEMBL257270	Small molecule	371.50	CCCC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)...
21	CHEMBL398545	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1-c1cc...
22	CHEMBL252391	Small molecule	371.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc...

	ChEMBL ID	Type	Molecular Weight	Smiles
23	CHEMBL252417	Small molecule	375.84	Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3C)o2)C(=O)O)[...
24	CHEMBL404117	Small molecule	383.51	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
25	CHEMBL271223	Small molecule	385.53	CC(C)(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2...
26	CHEMBL252003	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(Oc2ccc...
27	CHEMBL254282	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(Oc2cccc...
28	CHEMBL251804	Small molecule	387.50	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1Oc1cc...
29	CHEMBL437595	Small molecule	391.49	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
30	CHEMBL4202767	Small molecule	391.85	CCOc1c(Cl)cc(C(=O)Nc2ccc(C(=O)O)c(C)c2)cc1OC(C)C
31	CHEMBL411298	Small molecule	397.41	CC(C)C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc...
32	CHEMBL257026	Small molecule	397.54	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](...
33	CHEMBL3235416	Small molecule	401.47	CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1ccccc1)...
34	CHEMBL254493	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1OCc1c...
35	CHEMBL254495	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cc...
36	CHEMBL254703	Small molecule	401.53	CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc...
37	CHEMBL258683	Small molecule	405.39	C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1cc...
38	CHEMBL350414	Small molecule	405.52	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...
39	CHEMBL163454	Small molecule	405.52	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](S...
40	CHEMBL409721	Small molecule	410.41	CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@...
42	CHEMBL271224	Large molecule	411.57	O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H]...

In [122]:
```
x.describe()
```

Out[122]:

	Molecular Weight
count	43.000000
mean	367.714419
std	34.073065
min	303.190000
25%	344.455000
50%	371.500000
75%	397.475000
max	411.570000

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: