

## Merging filtered and fit data with CCL properties

In [1]:

```
import pandas as pd
import numpy as np

from scipy.optimize import curve_fit
import scipy.optimize as opt
from sklearn.metrics import r2_score

from tqdm import tqdm
import warnings
warnings.filterwarnings("ignore")

import os
_FOLDER = "database/"
_FOLDER_2 = "results/"

from functions.merging import *
```

In [2]:

```
cell_features = pd.read_csv(_FOLDER+"Cell_Line_Features_PANCAN_simple_MOBEM.tsv", sep="Wt")
drug_features = pd.read_csv(_FOLDER_2+"drug_features_with_pubchem_properties.csv")
```

## Merging with filtered and fitted data

In [7]:

```
df_list = ["fit_1234_02", "fit_auc_02", "filt_1234_02", "filt_auc_02"]
df_list_2 = [_FOLDER_2 + x + ".csv" for x in df_list]
```

In [8]:

```
df = pd.read_csv(df_list_2[0])
df.head(2)
```

Out[8]:

	CELL_LINE_NAME	COSMIC_ID	DRUG_ID	DRUGID_COSMICID	FOLD_DILUTION	MAX_CONC	fd_num_0	fd_num
0	SF539	905984	200	200_905984	2	1.024	0	0.111
1	TC-71	1240221	1149	1149_1240221	2	5.000	0	0.111

2 rows × 32 columns

In [12]:

```
i=0
df = pd.read_csv(df_list_2[i])
merged_df = merge_drug_cells(df, cell_features, drug_features,save_CCL_properties=True, _FOLDER_to_save=_FOLDER_2)
```

In [13]:

```
for i in range(len(df_list_2)):
    df = pd.read_csv(df_list_2[i])
    merged_df = merge_drug_cells(df, cell_features, drug_features)
    new_file_name = _FOLDER_2 + df_list[i] + "_merged.csv"
    merged_df.to_csv(new_file_name, index=False)
    print(new_file_name)
    print(f"Wt Drug Profiles+CCL {df.shape}, Merged Data: {merged_df.shape}")
    gr = merged_df.groupby("DRUG_ID").agg(count=("CELL_LINE_NAME", "count"))
    print("WtN unique drugs:", gr.shape[0])
    print("WtN of 10+ drug records:", gr[gr["count"]>10].shape[0])
```

```
results/fit_1234_02_merged.csv
    Drug Profiles+CCL (1527, 32), Merged Data: (1423, 1142)
    N unique drugs: 115
    N of 10+ drug records: 41
results/fit_auc_02_merged.csv
    Drug Profiles+CCL (3062, 33), Merged Data: (2812, 1143)
    N unique drugs: 146
    N of 10+ drug records: 56
results/filt_1234_02_merged.csv
    Drug Profiles+CCL (2108, 30), Merged Data: (1964, 1140)
    N unique drugs: 125
    N of 10+ drug records: 51
results/filt_auc_02_merged.csv
    Drug Profiles+CCL (14084, 31), Merged Data: (12772, 1141)
    N unique drugs: 193
    N of 10+ drug records: 122
```

In [ ]: