

1. Cel.

Celem laboratorium nr 9 było znalezienie najlepszych struktur lead z bazy ZINC, na podstawie score'ów uzyskanych z dockowania Autodock Vina oraz rescoreing oddt_cli w porównaniu do struktur inhibitorów AMCase z bazy ChEMBL.

2. Metodyka.

W tym celu wykorzystano regule 3 I przy użyciu paczki Lipinskiego przefiltrowano bazy ZINC, struktury te przekonwertowano w format smiles, a następnie wytworzono z nich fingerprinty morgana. Dzięki temu można porównywać fp Morgana inhibitorów z bazy ChEMBL z lead'ami z bazy ZINC. Przez porównanie (wykorzystanie słownika, więcej w notebook'u) można było wydobyc najbardziej podobne do inhibitorów AMCase'y, lead'y z bazy ZINC.

Następnie z tych przefiltrowanych lead'ów stworzono smiles'y, które przekonwertowane w format pdbqt zostały przepuszczone przez AutoDock Vina oraz rescoreing oddt_cli. Z takich danych, ponownie, można wyciągnąć najlepsze wyniki i ich struktury.

3. Wyniki.

Trzy najlepsze struktury (najlepsze score'y) wykazane przez AutoDock Vina to:

-ZINC75567398 (-9,1)

-ZINC70451801 (-9,1)

-ZINC00184014 (-8,4)

Z rescoreingu niestety udało się wyciągnąć jeden związek, który był wysoko oceniony przez wszystkie funkcje i jest to związek o ID ZINC75567398. Niestety nie można było wykorzystać funkcji PLEC, gdyż w czasie wyliczania tej funkcji użytkownikowi pojawiał się error.

```
/home/danjan/Documents/Projektowanie/Anaconda/anaconda3/envs/lab7/lib/python3.7/site-packages/oddt/virtualscreening.py:349: UserWarning: The chunksize (100) seems to be too large.
% self.chunksize)
/home/danjan/Documents/Projektowanie/Anaconda/anaconda3/envs/lab7/lib/python3.7/site-packages/oddt/virtualscreening.py:355: UserWarning: Falling back to sub-methods multithreading as the number of molecules is less than cores (1 < 4)
'%(%i < %i)' % (len(first_chunk), self.n_cpu))
```

Figure 1: Error. UserWarning: The chunksize (100) seems to be too large oraz UserWarning: Falling back to sub-methods multithreading as the number of molecules is less than cores (1 < 4)

4. Wnioski.

Okazuje się, iż najlepsza struktura wykazana przy użyciu AutoDock oraz przy użyciu rescoringu jest ZINC75567398. To oznacza, że na tej właśnie strukturze moglibyśmy opierać nasz development leku. Niestety jest to za mała ilość związków. By zdobyć większą bazę, kod wykonany przez użytkownika należałoby poprawić, a także powinno się znaleźć przyczynę wspomnianego wyżej błędu. Dzięki temu można by było stworzyć szerszą bazę, co dałoby szerszą możliwość rozwoju. Dodatkowo w czasie pracy na strukturach z bazy ZINC zauważono, iż w pliku roboczym (pass.smi) znajdują się powtórzenia ID związków lub nawet struktur smiles. Trzeba było to uwzględnić w czasie wykonywania opisanych metod i najlepiej poprawić, tak, żeby wyniki były jak najbardziej wiarygodne.

```
594 Cc1cc(C)n2cc(C[NH+]3CCc4cccc4C3)[NH+]c2n1 ZINC16035732
595 [NH3+]C1C(=O)N(Cc2ccc(Cl)cc2)c2cccc21 ZINC16082910
596 [NH3+]C1C(=O)N(Cc2ccc(Cl)cc2)c2cccc21| ZINC16082910
597 [NH3+]C1C(=O)N(Cc2cccc(Cl)c2)c2cccc21 ZINC16082912
598 [NH3+]C1C(=O)N(Cc2cccc(Cl)c2)c2cccc21 ZINC16082912
```

Figure 2: Powtórzenia w pliku pass.smi, id jak i struktur smiles.