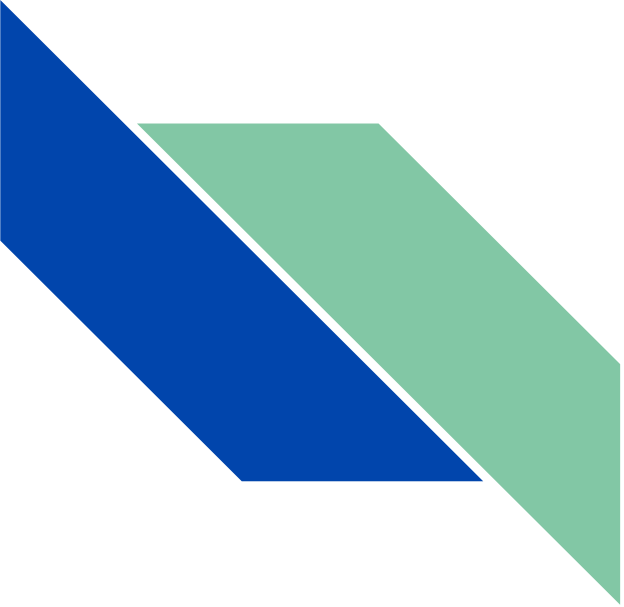


Reinforced Genetic Algorithm

Piotr Faron, Artur Górak, Kamil Harłacz, Yurii Titov

Plan prezentacji

1. Dane wejściowe
2. Dokowanie i test programów dokujących
3. Algorytm genetyczny
4. Wyniki

Dane wejściowe

Dwa pliki z danymi na temat:

- Białek, czyli naszych celów biologicznych - format .pdb zawierający dane strukturalne. Mamy dane dla 5 białek:
  - receptory serotoninowe 5-HT1A i 5-HT7
  - receptor dopaminowy D2
  - receptor histaminowy H1
  - receptor adrenergiczny beta2
- Pewnych związkach aktywnych do tych białek - format .sdf, czyli zawierający informacje na temat struktury cząsteczek, dane numeryczne oraz właściwości fizykochemiczne. Nas interesuje, szczególnie pole 11 zawierające wartości Ki, czyli poziom aktywności danego związku (im mniej tym związek bardziej aktywny)

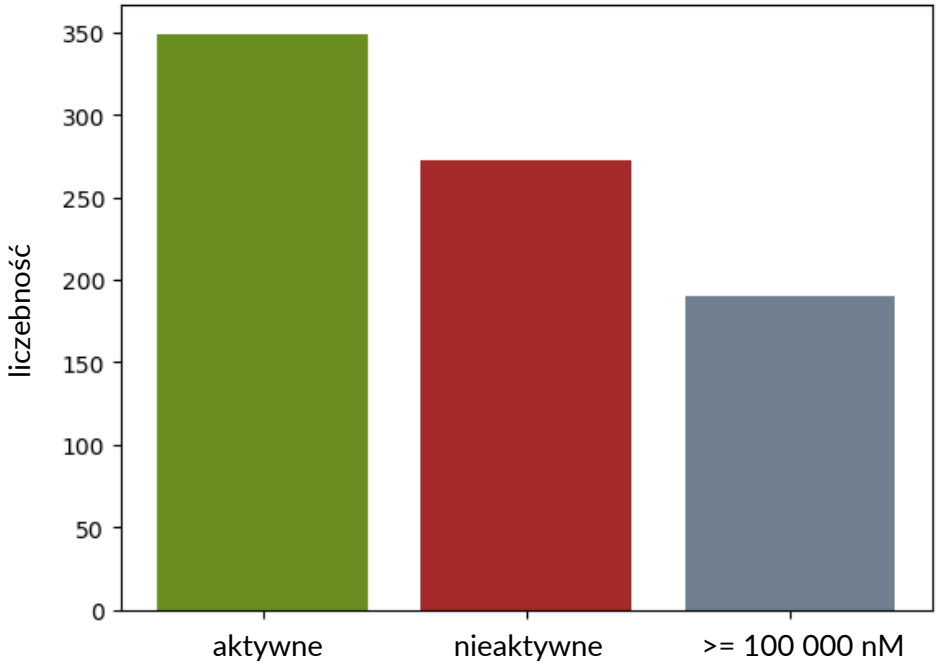
Obróbka danych wejściowych - zrozumienie Ki value

- Generalnie przyjmuje się, że Ki poniżej 1000 (jednostką są nM) determinuje aktywność,
- Ale w dalszych badaniach tak naprawdę bierze się pod uwagę związki o aktywności poniżej 100 nM.
- W badaniach laboratoryjnych nie wyznacza się z reguły dokładnej wartości Ki dla związków nieaktywnych, dlatego jeśli mamy wartości 100 000 to nie znaczy to że rzeczywiście taka wartość została wyznaczona tylko że w ten sposób zaznacza się nieaktywność związku (przez podanie bardzo dużej wartości Ki)

Związki aktywne i nieaktywne

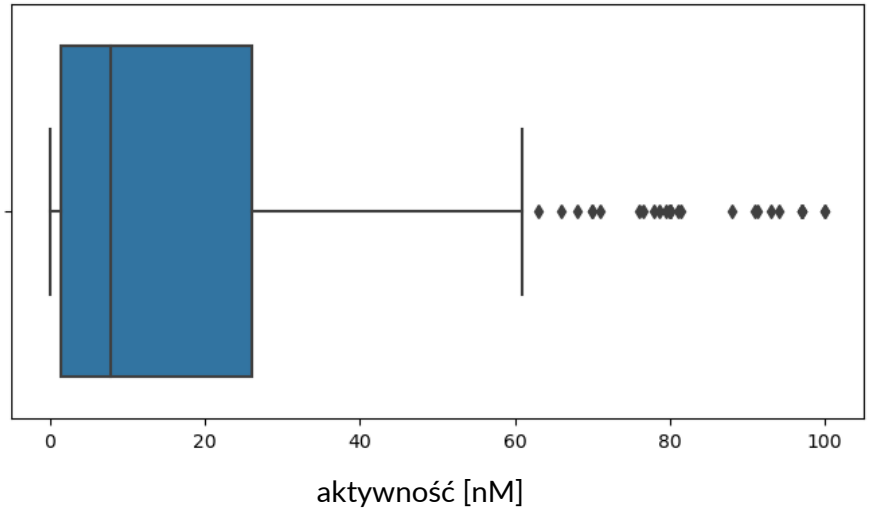
- Aktywne - o wartości Ki mniejszej od 100
- Nieaktywne - o wartości Ki od 100 do 100 000
- O niezdefiniowanym Ki (dla których, najprawdopodobniej, podana wartości Ki nie była dokładnie obliczona) - o wartości Ki większej od 100 000
- Do populacji początkowej trafiają tylko aktywne związki

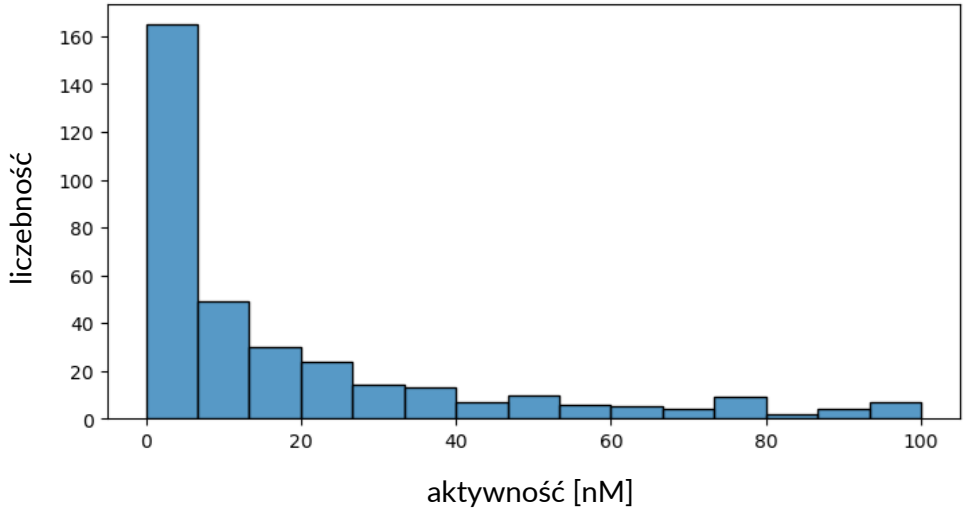
Związki aktywne i nieaktywne



Wyk. 1 *Porównanie liczebności związków aktywnych i nieaktywnych dla receptora beta2.*

Analiza rozkładu związków aktywnych



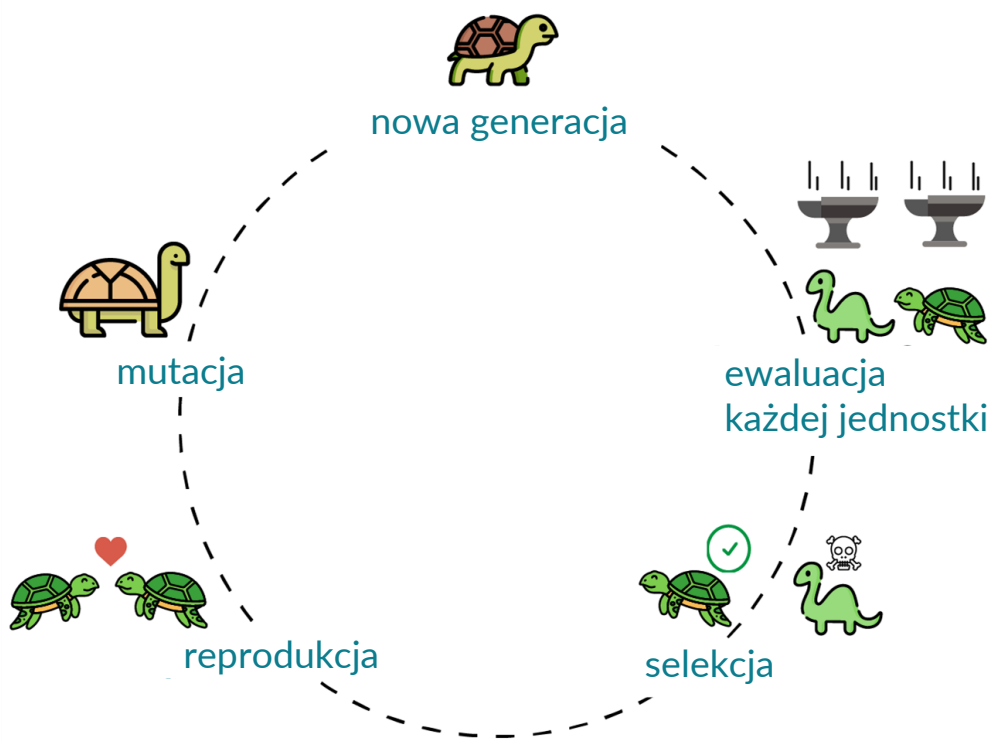


Wyk. 2 *Po lewej rozkład aktywnych związków na wykresie pudełkowym, po prawej liczebności poszczególnych klas.*

Wnioski

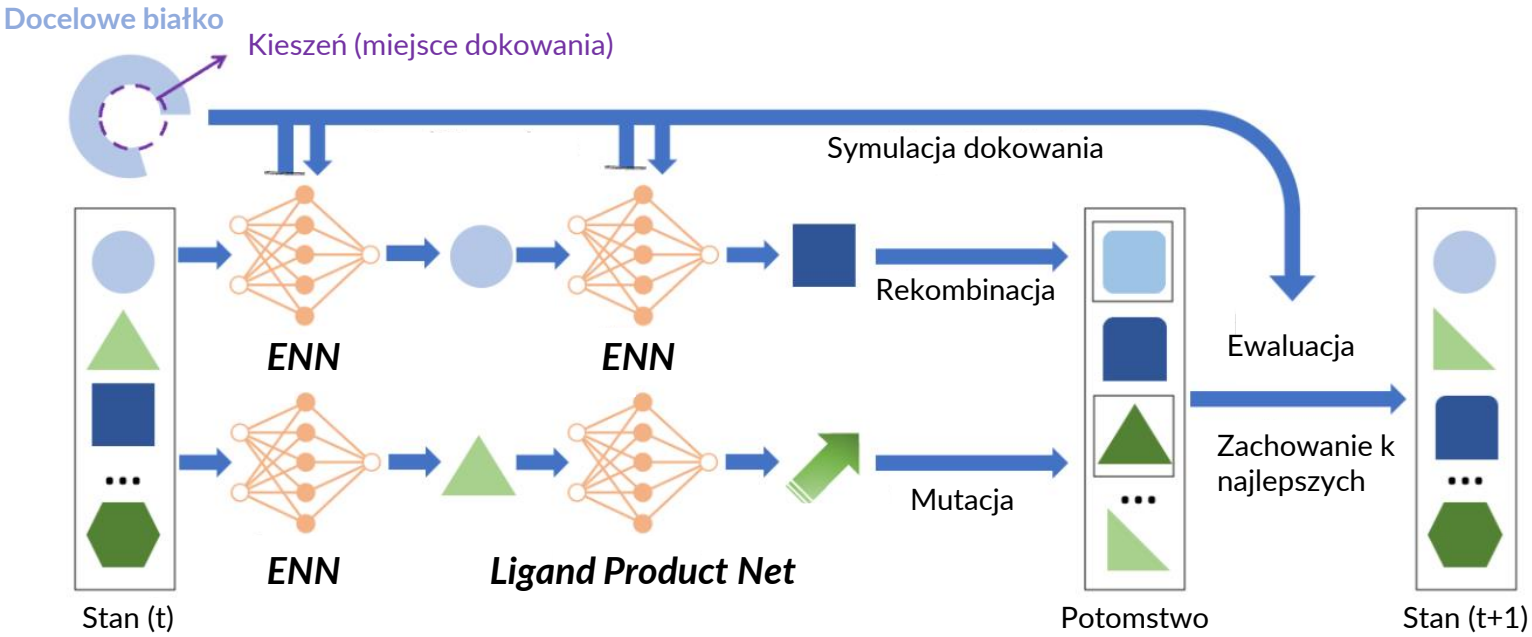
1. Zdecydowana większość wartości Ki w naszych danych znajduje się w relatywnej bliskości do zera. Jednak w przypadku każdego receptora mamy duże ilości wartości odstających.
2. Związki są rozdzielone na aktywne i nieaktywne prawie równolicznie
3. Około 15% związków znajdujących się w danych dla jednego receptora duplikuje się w danych dla innych 4 receptorów.
4. Informacja o jednym i tym samym związki może być zduplikowana w jednym i tym samym pliku. Ponadto podczas dane o aktywnościach niektórych zduplikowanych w obrębie jednego pliku związków różnią się.

Algorytm genetyczny



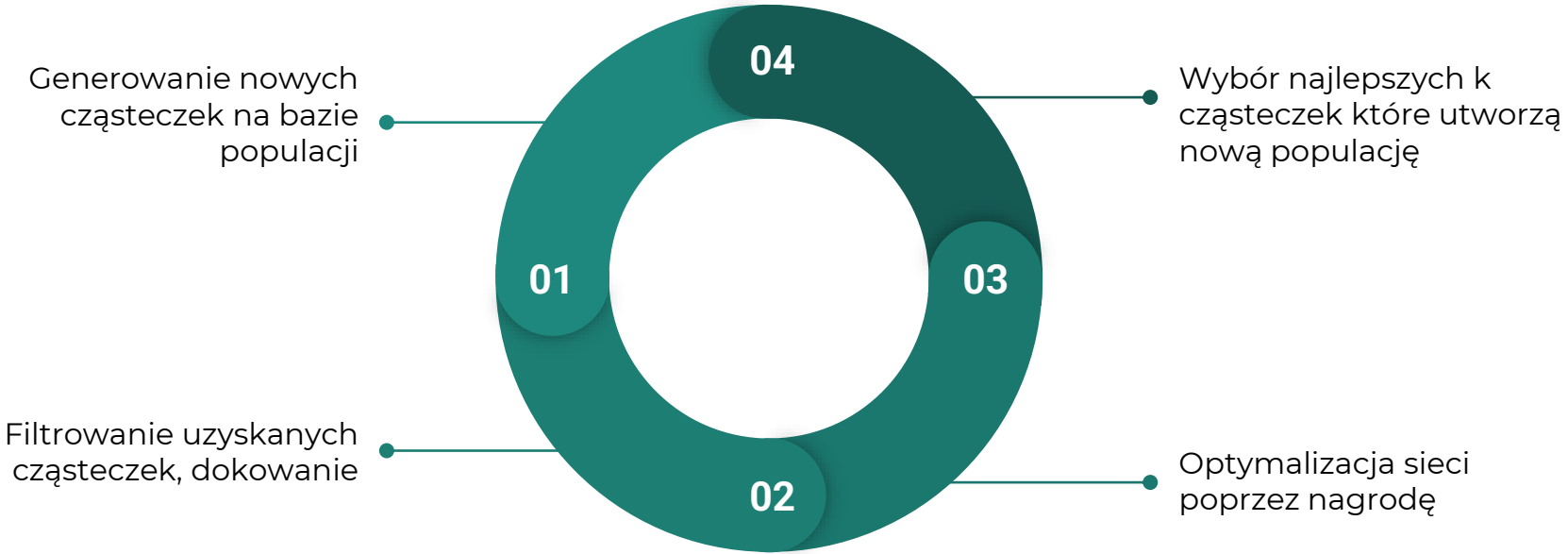
***Ryc. 1** Idea algorytmu genetycznego.*

Struktura sieci



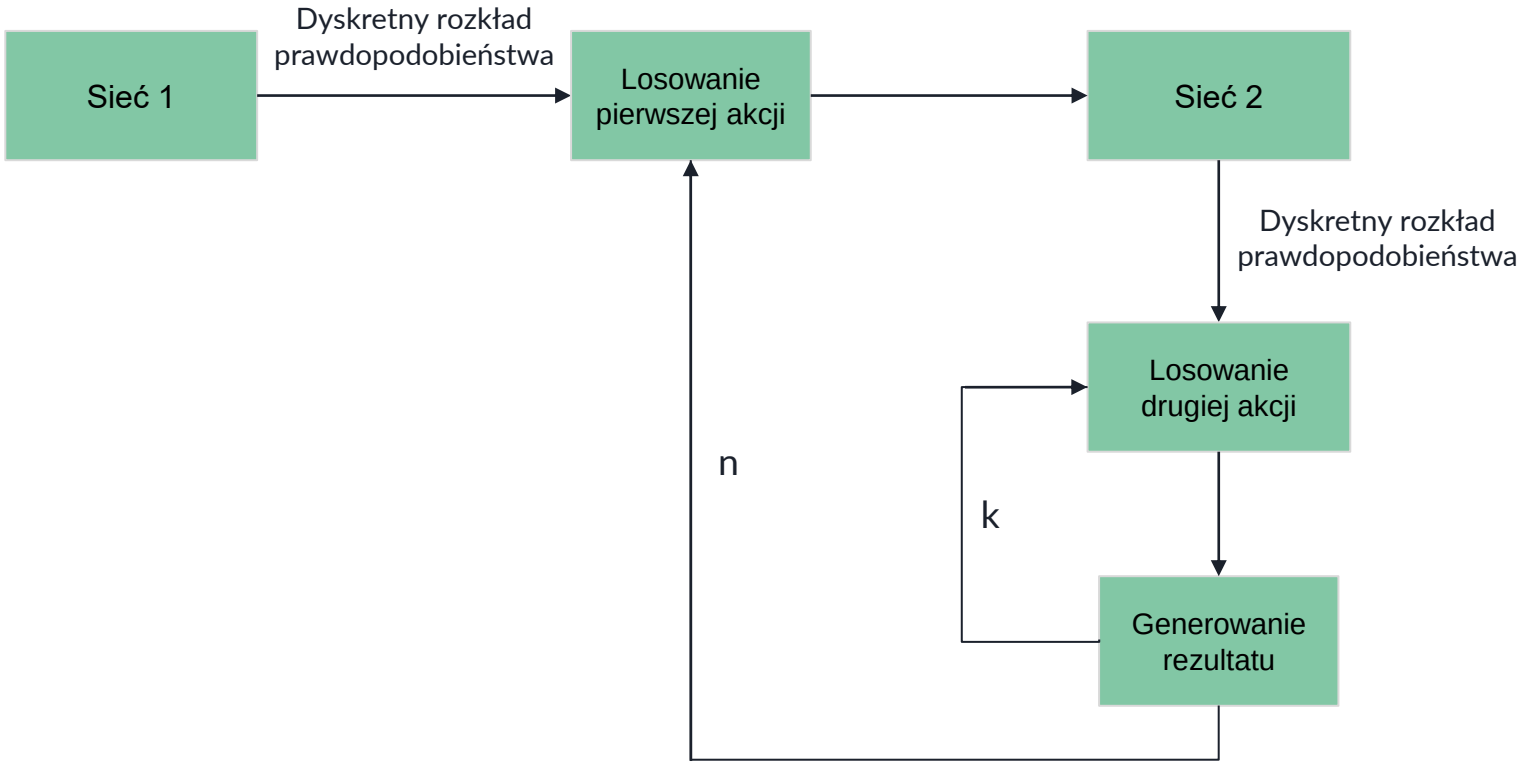
***Ryc. 2** Schemat Reinforced Genetic Algorithm.*

Działanie algorytmu



***Ryc. 3** Schemat przedstawiający jedną iterację algorytmu.*

Generacja związków



***Ryc. 4** Szczegółowy schemat generowania związków.*

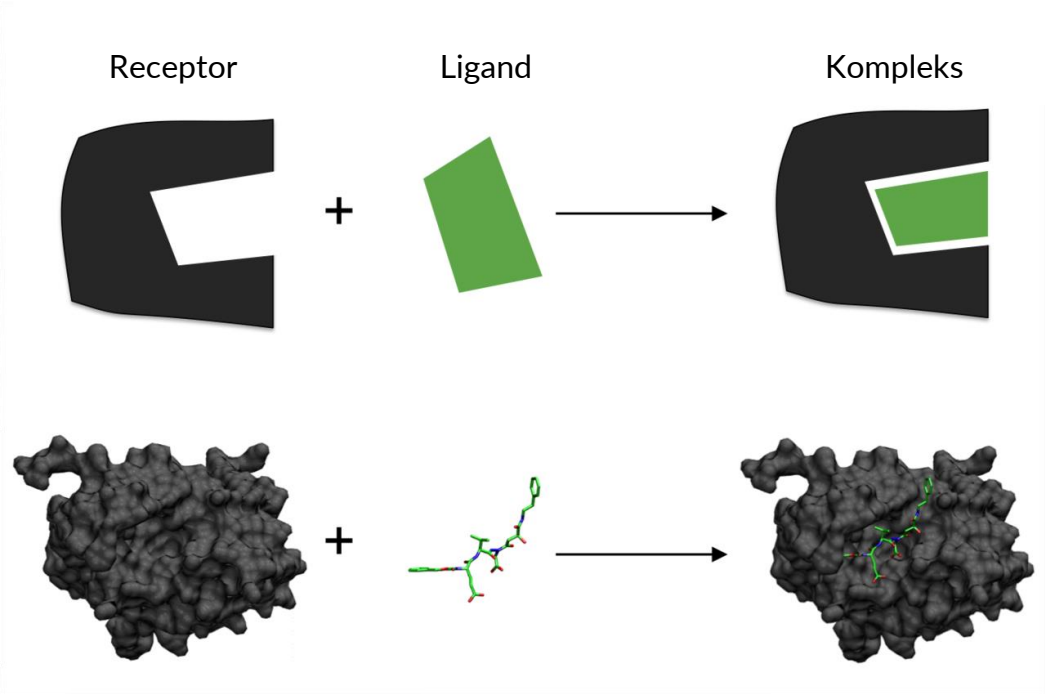
Policy Gradient

- Dla każdej trójki (pierwsza akcja, druga akcja, rezultat)
- Obliczamy sumę zlogarytmowanych prawdopodobieństw obydwu akcji
- Obliczamy nagrodę na podstawie docking score
- Mnożymy powyższe oraz wykonujemy krok optymalizacji

Hipotezy badawcze

- Zastosowany algorytm genetyczny poprawia funkcję oceny dokowania z populacji na populację.
- Algorytm dokujący jest zoptymalizowany tak, by zapewnić zadowalającą szybkość dokowania przy jednoczesnej dużej dokładności.
- Porównanie różnych metod mutacji związków (lub krzyżówek).

Dokowanie



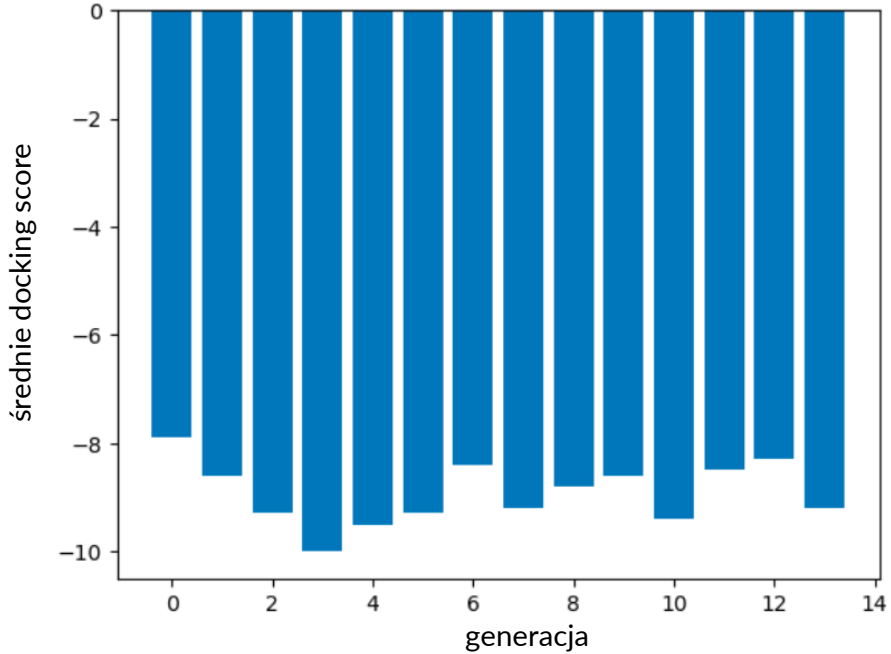
Ryc. 5 *Schemat procesu dokowania.*

Przetestowane narzędzia do dokowania

Tab. 1 *Czas dokowania [s] w zależności od narzędzi.*

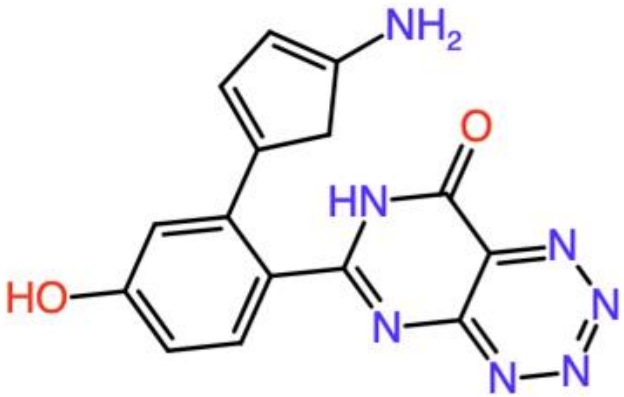
Nazwa narzędzia	**Czas dokowania w sekundach**
Vina	19.53
Smina	24.17
dockstring	26.2
docking_py	27.46
pyrx	∞
Dock6	∞
SwissDock	∞

Hipoteza: Wynik dokowania w kolejnych generacjach poprawia się



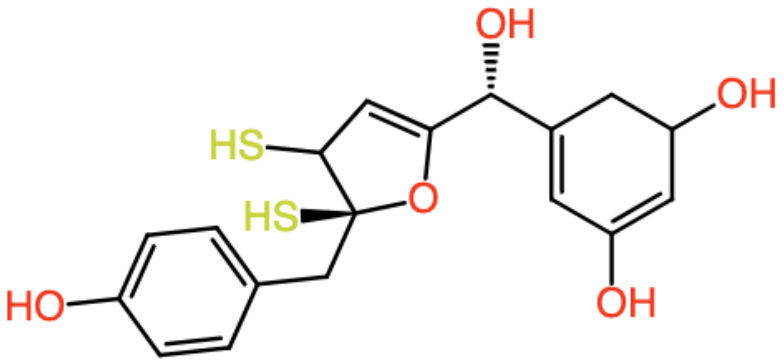
Wyk. 5 *Docking score w kolejnych generacjach dla beta2.*

Hipoteza: Mutacje dają porównywalnie dobre wyniki co rekombinacje



NC1=CC=C(c2cc(O)ccc2-c2nc3nnnnc3c(=O)[nH]2)C1

Ryc. 6 *SMILES z najlepszym wynikiem dokowania (-10) uzyskany wyłącznie przy pomocy rekombinacji.*



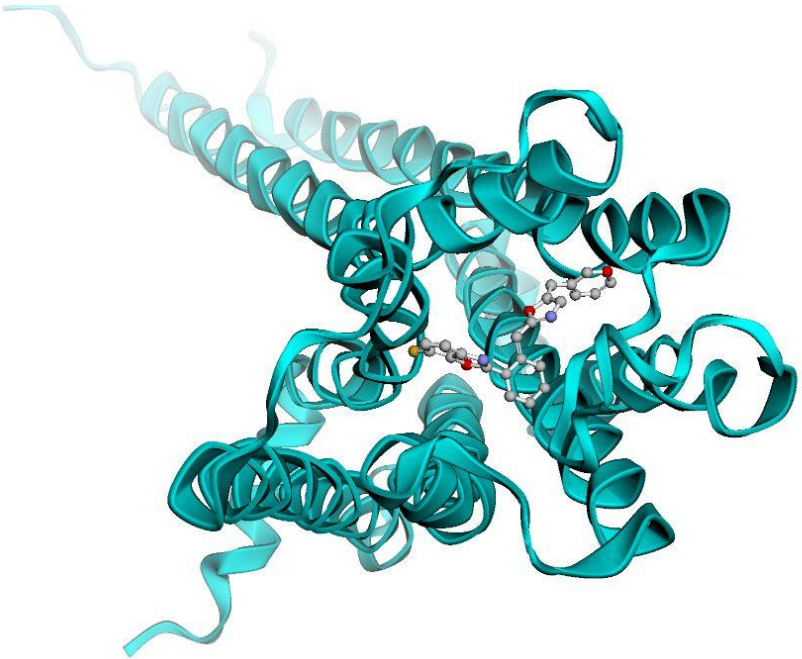
OC1=CC(O)CC([C@@H](O)C2=CC(S)[C@@](S)(Cc3ccc(O)cc3)O2)=C1

Ryc. 7 *SMILES z najlepszym wynikiem dokowania (-9.2) uzyskany wyłącznie przy pomocy mutacji.*

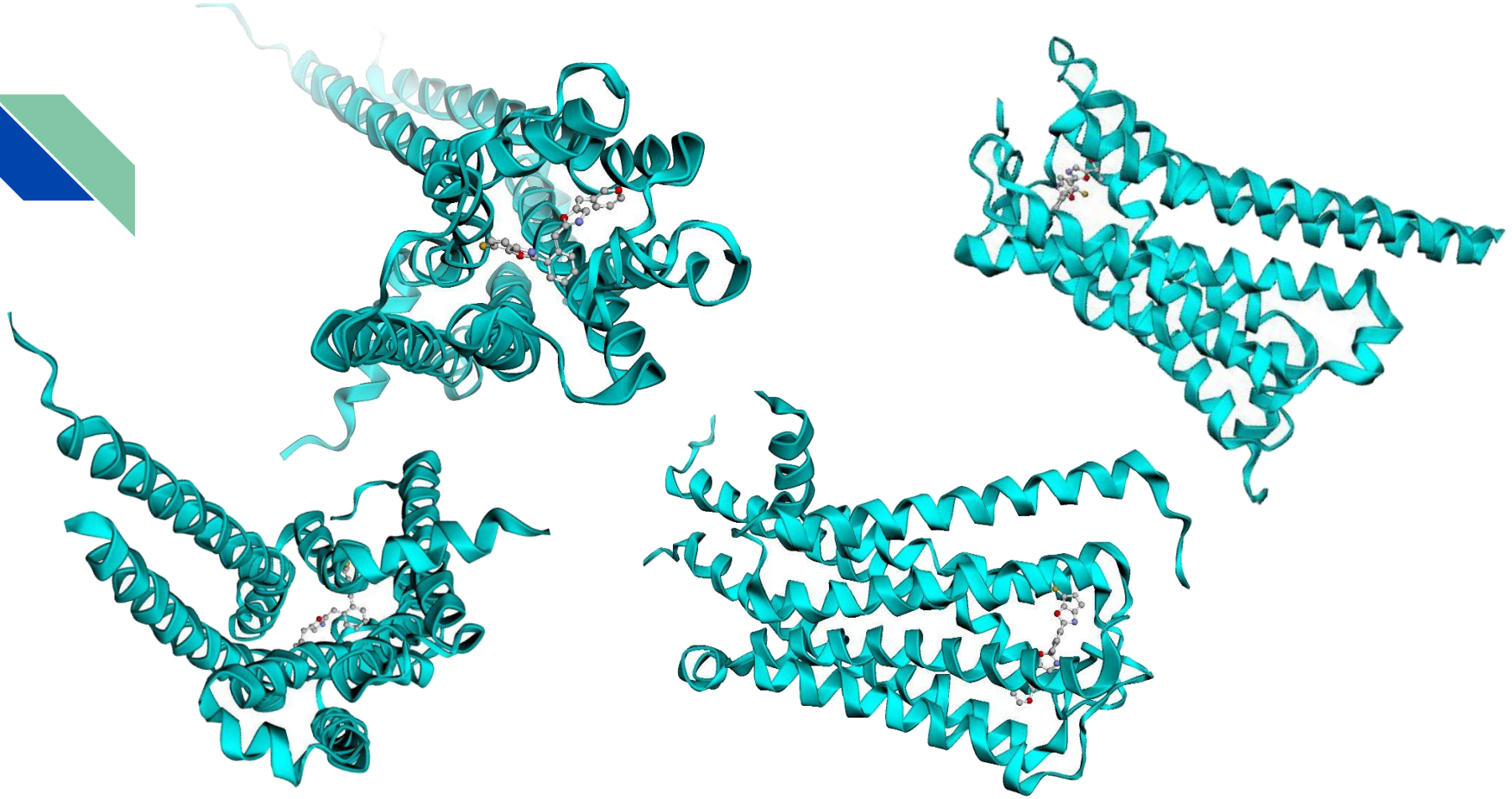
Wizualizacja

- Receptor: 5ht1a
- Docking score: -10.1
- SMILES:

  FC1=c2oc(C3CCCCC3=CC3=NCC(CC4=CC=COC4)O3)nc2=CC1



Ryc. 8 *Wizualizacja receptora 5ht1a oraz zadokowanego liganda.*

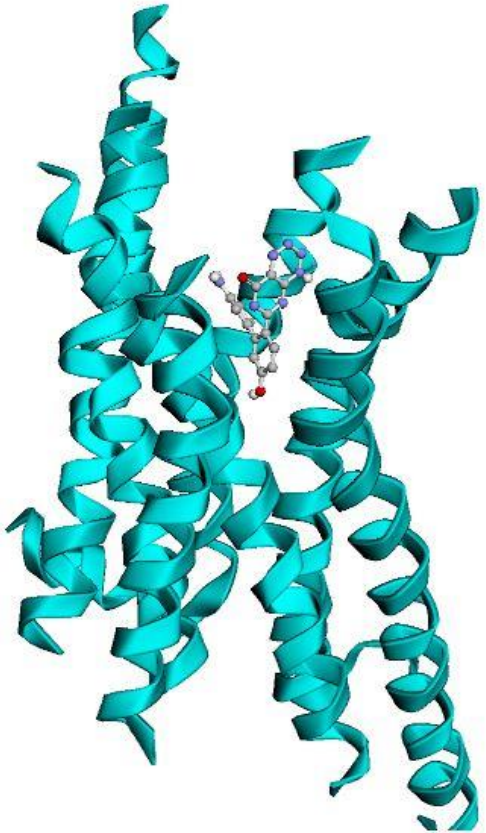


Ryc. 9 *Wizualizacja receptora 5ht1a oraz zadokowanego liganda z różnych perspektyw.*

Wizualizacja

- Receptor: beta2
- Docking score: -10.0
- SMILES:

  NC1=CC=C(c2cc(O)ccc2-c2nc3nnnnc3c(=O)[nH]2)C1

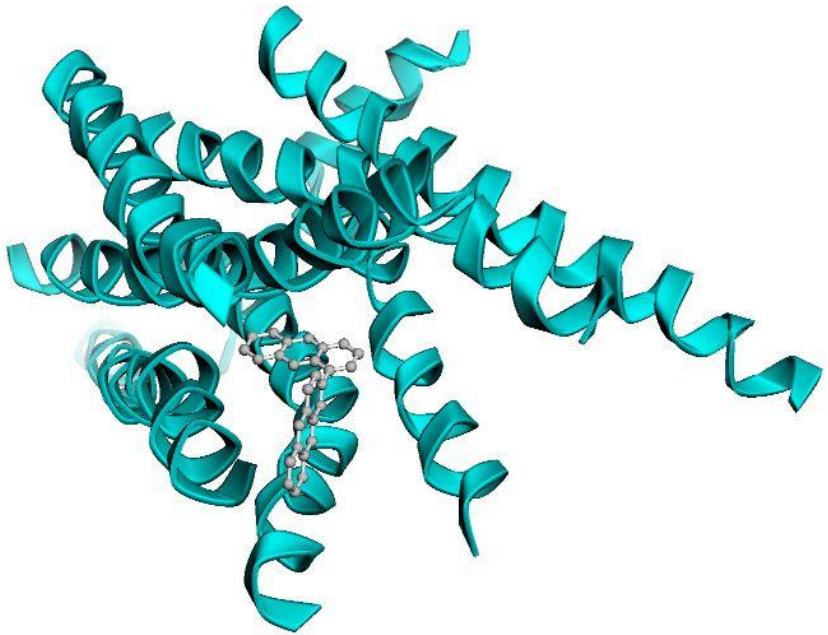


Ryc. 10 *Wizualizacja receptora beta2 oraz zadokowanego liganda.*

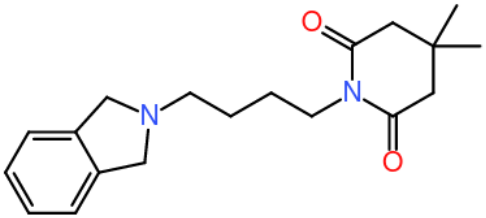
Wizualizacja

- Receptor: beta2
- Docking score: -9.2
- SMILES:

  C1=C(c2cccc3cc4c(cc23)CCCC4)Cc2cc3c(cc2C1)CCCC3

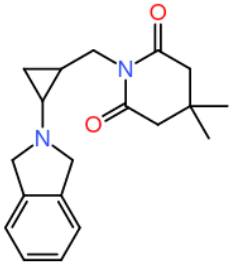


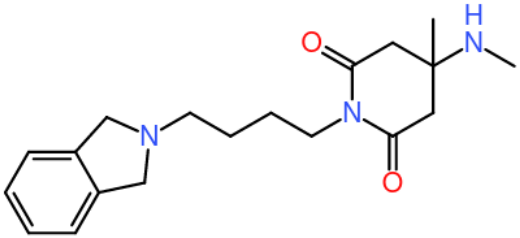
Ryc. 11 *Wizualizacja receptora d2 oraz zadokowanego liganda.*



Oryginał

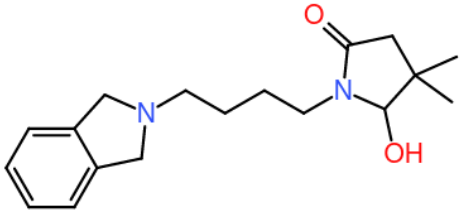
Add ring

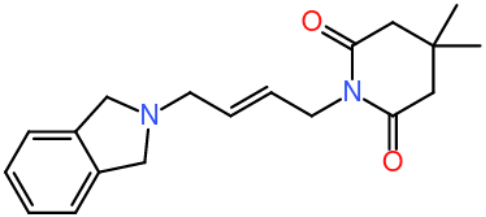




Insert atom

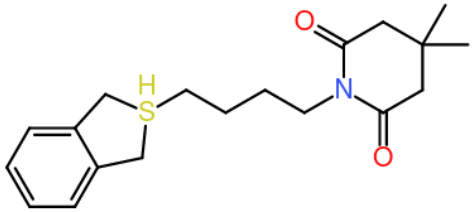
Delete atom

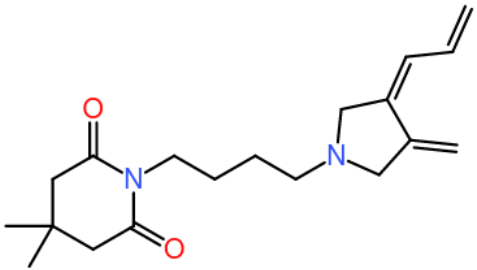




Change bond order

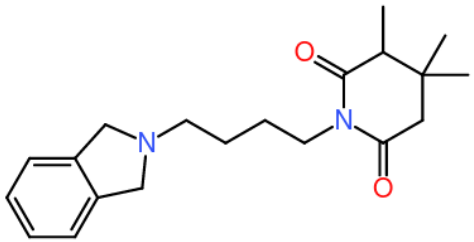
Change atom



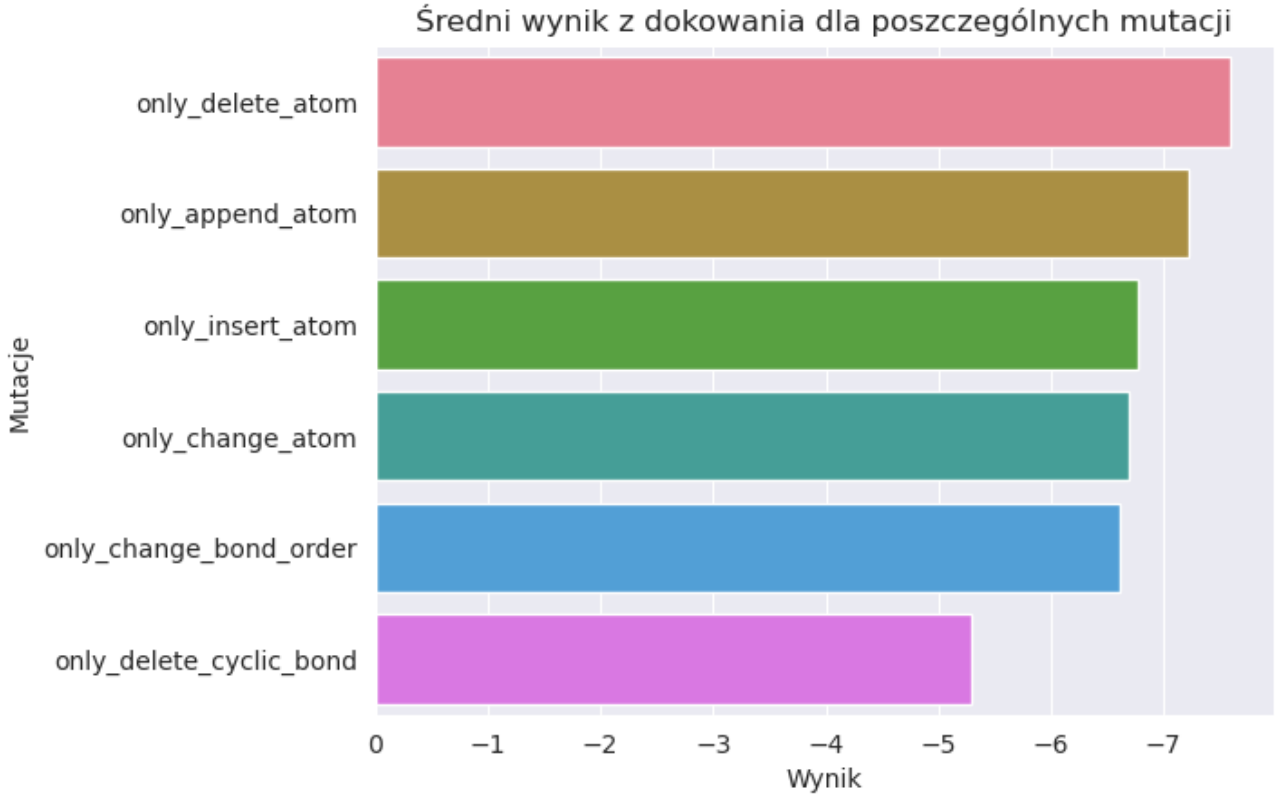


Delete cyclic bond

Append atom



Hipoteza: Porównanie różnych metod mutacji związków



Dziękujemy za uwagę