

Rozwój i walidacja funkcji scoringowych do modelowania oddziaływań z celami molekularnymi istotnymi dla radiofarmacji

Opiekunowie pracy:

Dr hab. Piotr Garnuszek, prof. NCBJ

Prof. dr hab. Jan Czesław Dobrowolski

Dr Piotr Lipiński

JOLANTA ZABORNIAK



Fundusze
Europejskie



Rzeczpospolita
Polska

Unia Europejska
Europejskie Fundusze
Strukturalne i Inwestycyjne



Cele pracy badawczej

- wykorzystanie metod obliczeniowych do szacowania energii oddziaływań kompleksów receptor – cząsteczka czynna
- rozwój i walidacja funkcji scoringowych do modelowania oddziaływań z celami molekularnymi istotnymi dla radiofarmacji
- cele istotne dla radiofarmacji wykorzystane w pracy badawczej:
 - CCK2R
 - GPCII
- opracowanie procedury obliczeniowej by odtwarzać dane eksperymentalne o wiązaniu

Cele molekularne analizowane w pracy

CCK2R

- receptor cholecystokininy podtypu 2
- receptor 7TM GPCR
- nadekspresja w raku rdzeniastym tarczycy
- kodowany przez gen CCKBR, który koduje receptor sprzężony z białkiem G dla gastryny i cholecystokininy (CCK)

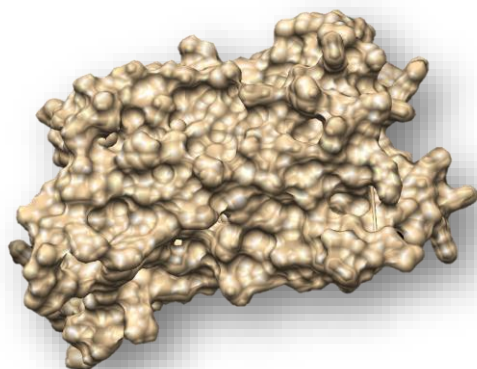
GCPII

- karboksypeptydaza glutaminianowa II (GCPII)
- specyficzny dla prostaty antygen błonowy (PSMA)
- glikoproteina błonowa klasy II
- ulega nadekspresji w raku prostaty

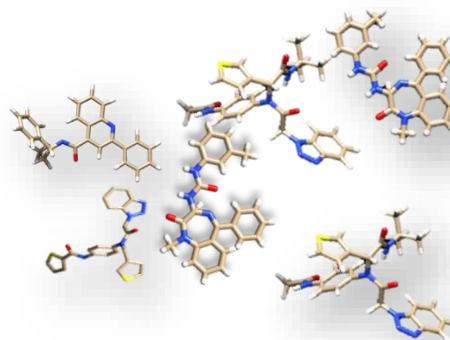


Dokowanie molekularne

receptor



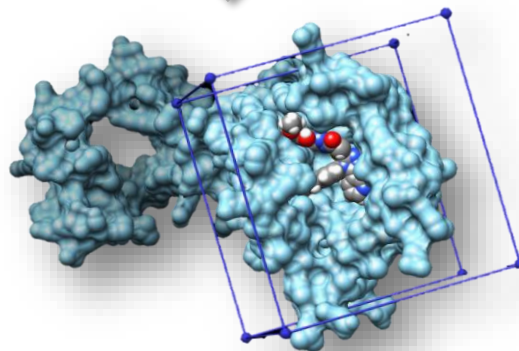
+



(potencjalne)
cząsteczki czynne



kompleks



SCORING



Fundusze Europejskie



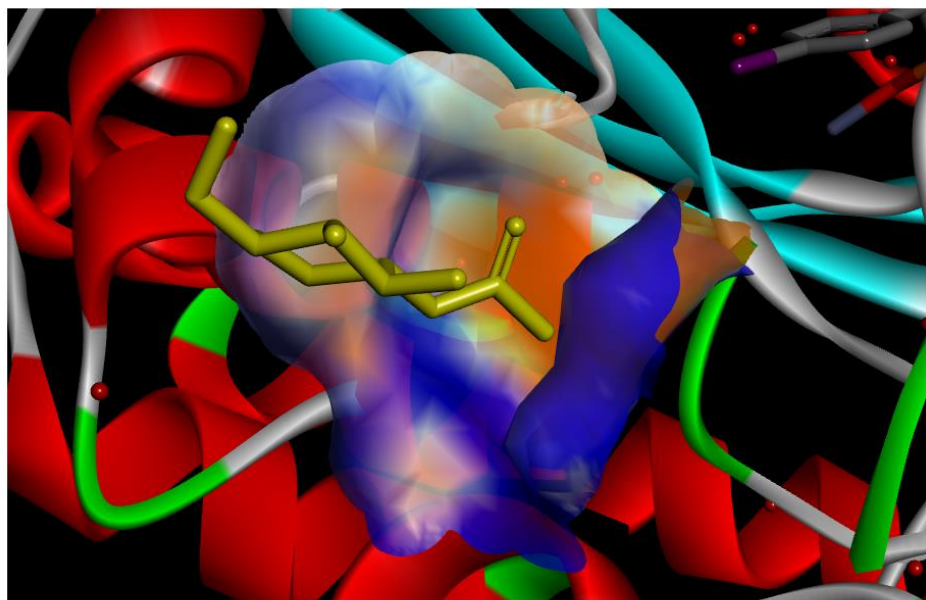
Rzeczpospolita
Polska

Unia Europejska
Europejskie Fundusze
Strukturalne i Inwestycyjne



Scoring

- *Scoring* jest krytycznym elementem dokowania
- Funkcje oceniające:
 - Oparte o pola siłowe
 - Empiryczne
 - Potencjały oparte na wiedzy (*knowledge-based*).



Scored poses

REMARK VINA RESULT MODEL 1: -8.184
REMARK VINA RESULT MODEL 2: -7.233
REMARK VINA RESULT MODEL 3: -7.178
REMARK VINA RESULT MODEL 4: -6.784

Dokowanie do GCPII

Struktury GCPII (używane do obliczeń)

2C6C	3BHX	3D7G
2C6G	3BI0	3D7H
2OR4	3BI1	3IWW
2PVV	3BXM	3SJF
2PVW	3D7D	3SJG
2XEJ	3D7F	3SJX

Zbiór ligandów

- 340 ligandów

**Dokowanie z domyślnymi
funkcjami oceniającymi
programu AutoDock VINA**

*Nazwy zaczerpnięte z nazewnictwa
obecnego w bazie RCSB PDB:

<https://www.rcsb.org/>



Fundusze
Europejskie



Rzeczpospolita
Polska

Unia Europejska
Europejskie Fundusze
Strukturalne i Inwestycyjne



Domyślna funkcja oceniająca Autodock Vina

$$E = \sum e_p(d)$$

$$d = r - R_I - R_j$$

$$e_p(d) = \begin{cases} w_1 \cdot Gauss_1(d) + \\ w_2 \cdot Gauss_2(d) + \\ w_3 \cdot Repulsion(d) + \\ w_4 \cdot Hydrophobic(d) + \\ w_5 \cdot HBond(d) \end{cases}$$

Scoring function weights and terms

Weight	Term
-0.0356	gauss ₁
-0.00516	gauss ₂
0.840	repulsion
-0.0351	hydrophobic
-0.587	hydrogen bonding
0.0585	N_{rot}

E – energia wiązania

d – odległość między „powierzchniami” atomów

r – odległość międzyatomowa

R_i i R_j promienie atomów w parze

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455-461.



Fundusze Europejskie



Rzeczpospolita
Polska

Unia Europejska
Europejskie Fundusze
Strukturalne i Inwestycyjne



Korelacja danych eksperymentalnych z domyślną funkcją scoringową

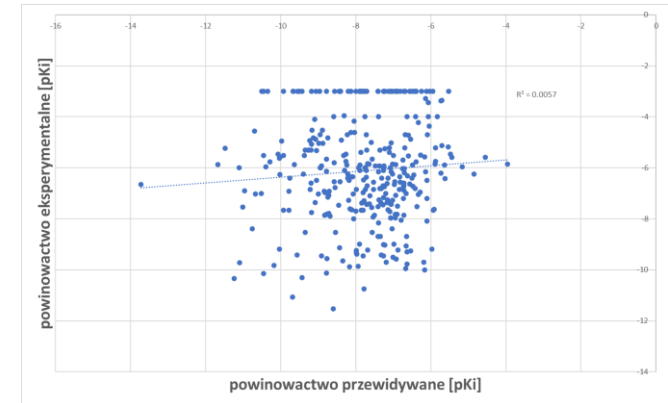
1. Domyślna funkcja scorująca – **brak mocy predykcyjnej**, $R^2 < 0.01$.

2. Próba stworzenia równań, w których wyrazami były zmienne uzyskane w wyniku dokowania do różnych struktur GCP11 (odpowiadające za różnego typu oddziaływania z tymi strukturami).

- a) Równania do 15 wyrazów.
- b) O różnej postaci (liniowe oraz nieliniowe).
- c) Wybór zmiennych za pomocą algorytmów genetycznych.

W sumie uzyskano 300 modeli.

3. W wyniku tej procedury udało się uzyskać poprawę korelacji, ale tylko do $R^2(\text{adj})$ około 0.40.



Deskryptory ze scoringu z dokowania

	Ilość wyrazów	R2 (adj)	
linear	6	0.2803	
	7	0.3032	
	8	0.3159	
	9	0.3243	
	10	0.3339	
	11	0.3409	
	12	0.3463	
	13	0.3519	
	14	0.3574	
	15	0.3624	
	simple quadratic	6	0.3105
		7	0.3186
		8	0.3314
		9	0.3438
		10	0.3544
11		0.3622	
12		0.3705	
13		0.3771	
14		0.3821	
	15	0.3983	
	6	0.3108	
	7	0.3295	
	8	0.3416	
	9	0.3494	

Modele oparte na deskryptorach 2D

4. Poszerzenie puli zmiennych o zmienne pochodzące z programu DRAGON.
5. 2000 deskryptorów 2D opisujących budowę cząsteczki (niezwiązane z dokowaniem).
6. Ponowny trening równań.
7. Znaczna poprawa jakości predykcji, R^2 (adj) nawet do 0.70.
8. Większość modeli bez wyrazów z dokowania!

Wszystkie deskryptory	Ilość wyrazów	r2 (adj)	
linear	6	0.5686	
	7	0.5933	
	8	0.6151	
	9	0.6443	
	10	0.6649	
	11	0.6597	
	12	0.6603	
	13	0.6798	
	14	0.7031	
	15	0.6894	
	simple quadratic	6	0.5839
		7	0.6026
		8	0.6170
		9	0.6227
		10	0.6404
11		0.6642	
12		0.6838	
13		0.6884	
14		0.6840	
15		0.7095	
simple cubic		6	0.5944
		7	0.6054
		8	0.6232
		9	0.6194
		10	0.6489
	11	0.6506	
	12	0.6660	

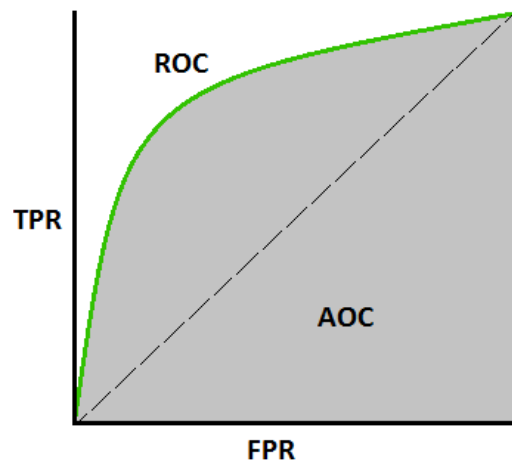
Przewidywanie aktywności

- Predykcja aktywności wg wybranych modeli
- Decoye – (Tldr's a Ligand Discovery Resource)
- Stosunek ilości:
 - true:decoys (16:544)
- Zdolność do prawidłowej klasyfikacji (wczesnego rozpoznania) za pomocą współczynnika EF

Analiza wyników

Krzywa AUC – ROC

Krzywa ROC – ilustruje sprawność binarnego klasyfikatora
Pole AUC



Czułość = $TP/(TP+FN)$

Swoistość = $TN/(TN+FP)$

FPR = 1 – swoistość

ROC – krzywa prawdopodobieństwa

AUC – pole powierzchni pod krzywą, stopień lub miara rozdzielności

EF (ang. *Enrichment Factor*) – współczynnik wzbogacenia

$$EF = \frac{H_{act}/H_{samp}}{N_{act}/N_{tot}}$$

Gdzie H_{act} to liczba prawdziwych ligandów w wybranej pierwszej części (np. 1%, 2%, 5% etc.) najlepszych wyników, H_{samp} to liczba uwzględnionych związków, N_{act} to liczba wszystkich prawdziwych ligandów w bibliotece, zaś N_{tot} to wielkość dokowanej biblioteki.

Wyniki

- Rzut monetą daje EF1% = 1.
- Klasyfikator w pełni skuteczny EF1% = 30.5.
- Porównanie między naszymi modelami EF1%. Najlepsze modele EF1% = 18.34.
- Nie ma związku między wynikami walidacji wewnętrznej a skutecznością w skryningu.
- Nie ma związku między skutecznością w skryningu a długością modeli.

1.0;18.34;PSMA-linear-8_2

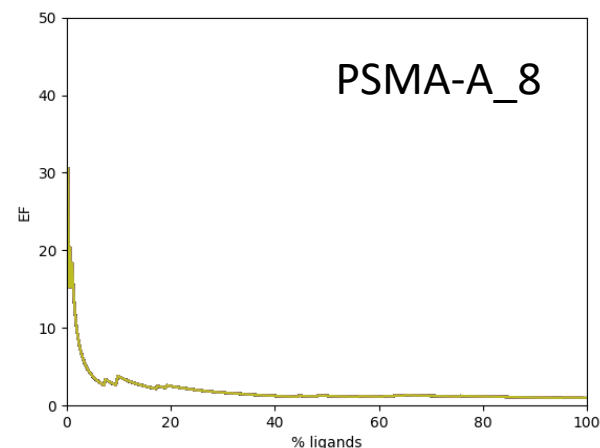
1.0;18.34;PSMA-linear-8_3

1.0;18.34;PSMA-linear-8_4

1.0;18.34;PSMA-linear-8_5

1.0;18.34;PSMA-linear-8_1

1.0;18.34;PSMA-A_8



Plany na przyszłość

- Próba uzyskania modeli przewidywania powinowactwa do GCPII z użyciem sieci neuronowych
- Przygotowanie publikacji do końca 2022 roku
- Staż krajowy
- Udział w szkole jesiennej projektu NOMATEN



Publikacje

- Lipiński P.F.J., Zaborniak J., Garnuszek P., Szurmak P., Virtual screening for small molecular non-covalent binders of the SARS-CoV-2 main protease, *Archives of Medical Science: AMS*, 2021, 17(3), 838.
- Garnuszek, P.; Karczmarczyk, U.; Maurin, M.; Sikora, A.; Zaborniak, J.; Pijarowska-Kruszyna, J.; Jaroń, A.; Wyczółkowska, M.; Wojdowska, W.; Pawlak, D.; Lipiński, P.F.J.; Mikołajczak, R. PSMA-D4 Radioligand for Targeted Therapy of Prostate Cancer: Synthesis, Characteristics and Preliminary Assessment of Biological Properties. *Int. J. Mol. Sci.* **2021**, 22, 2731.
- Piotr Garnuszek, Urszula Karczmarczyk, Michał Maurin, Arkadiusz Sikora, Jolanta Zaborniak, Justyna Pijarowska-Kruszyna, Antoni Jaroń, Monika Wyczółkowska, Wioletta Wojdowska, Dariusz Pawlak, Piotr FJ Lipiński, Renata Mikołajczak. PSMA-D4 Radioligand for Targeted Therapy of Prostate Cancer: Synthesis, Characteristics and Preliminary Assessment of Biological Properties. In: Letizia Giampietro, editor. *Prime Archives in Molecular Sciences: 2nd Edition*. Hyderabad, India: Vide Leaf. 2021.
 - [Publikacja artykułu jako rozdział w książce: ISBN: 978-81-953047-1-4.](#)



Dziękuję za uwagę



Fundusze
Europejskie



Rzeczpospolita
Polska

Unia Europejska
Europejskie Fundusze
Strukturalne i Inwestycyjne

