

PROTOKÓŁ DO AUTOMATYCZNEJ OCENY WYNIKÓW DOKOWANIA OPARTY O METODY UCZENIA MASZYNOWEGO

Sabina Smusz

Instytut Farmakologii PAN, Zakład Chemii Leków
Opiekun naukowy: Prof. dr hab. Andrzej J. Bojarski

Streszczenie: Dokowanie, czyli przewidywanie ułożenia liganda w miejscu wiążącym białka jest jedną z najbardziej popularnych procedur w zadaniach związanych z komputerowo wspomaganym projektowaniem leków [Breda i in. 2008]. Niezbędnym elementem do przeprowadzenia tego procesu jest znajomość struktury przestrzennej celu, która w przypadku braku danych krystalograficznych lub NMR jest przewidywana, najczęściej w procedurze modelowania homologicznego. Największe wyzwanie dokowania jest wciąż związane z odpowiednią analizą i interpretacją wyników. Można ją przeprowadzić poprzez ocenę wizualną, jednak jest to znacznie utrudnione w przypadku dużej liczby związków, lub przy wykorzystaniu różnych funkcji oceniających. Te, które są tworzone dla potrzeb wirtualnego przesiewania bibliotek związków chemicznych (ang. *virtual screening*) powinny umożliwiać przeprowadzenie rozróżnienia pomiędzy związkami wykazującymi wysokie prawdopodobieństwo posiadania rzeczywistej aktywności biologicznej, a tymi, dla których jest ono na tyle niskie, że mogą one zostać pominięte w dalszych pracach.

W przedstawionych badaniach opracowano protokół do automatycznej oceny wyników dokowania oparty o algorytmy uczenia maszynowego przetestowany na przykładzie receptorów serotoninowych 5-HT₆ i 5-HT₇. Wyniki dokowania przekształcane są do odpowiedniej reprezentacji – opisywane są za pomocą fingerprintu oddziaływań strukturalnych (ang. *Structural Interaction Fingerprint, SIFt* [Deng, Chuaqui, Singh 2004]) oraz deskryptorów Spectrophores [Bultinck i in. 2002]. Pierwszy sposób reprezentacji dostarcza informacji o oddziaływaniach pomiędzy zadokowanym ligandem, a każdym z aminokwasów danego receptora, z kolei deskryptory Spectrophores informują o konformacji liganda poprzez wartości własności atomowe, które są uzależnione od aktualnego ułożenia przestrzennego molekuly. Na tak przygotowanej reprezentacji przeprowadzono eksperymenty uczenia maszynowego przy wykorzystaniu 5 różnych algorytmów klasyfikacyjnych, które następnie poddano kilkietapowej analizie wyników. Uwzględniła ona jakość wykorzystanego modelu homologicznego, skuteczność działania metod uczenia maszynowego oraz różne konformacje zadokowanych związków w kompleksach ligand-białko.

Literatura:

Breda A., Basso L. A., Santos D. S., De Azevdo Jr W. F. 2008, Virtual Screening of Drugs: Score Functions, Docking and Drug Design. *Curr. Comput. Aided Drug Des* 4: 265-272.

Deng Z., Chuaqui C., Singh J. 2004, Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* 47: 337-344.

Bultinck P., Langenaeker W., Lahorte P., De Proft F., Geerlings P., Van Alsenoy C., Tollenaere J. P. 2002, The Electronegativity Equalization Method II: Applicability of Different Atomic Charge Schemes. *J. Phys. Chem.*, 106: 7895-7901.

Adres do korespondencji: smusz@if-pan.krakow.pl