

## Streszczenie rozprawy doktorskiej

Próba zrozumienia mechanizmów molekularnych zachodzących w żywych komórkach nie tylko ma na celu zaspokojenie ciekawości poznawczej naukowców, ale w znaczny sposób może się przyczynić do projektowania terapii, leków i metod zapobiegających procesom patologicznym. Kluczowym elementem zdobycia tej wiedzy jest pozyskanie informacji o strukturze przestrzennej białek uczestniczących w danym zjawisku. Niestety, ponieważ metody doświadczalne są drogie i czasochłonne, a wiele prób kończy się niepowodzeniem, pomocne okazują się metody komputerowe. Narzędzia modelujące struktury białek nie zawsze pozwalają jednak uzyskać właściwe rezultaty. Dlatego ważnym etapem projektowania struktur przestrzennych jest ocena jakości modeli. W tym celu można wykorzystać zgromadzone już dane o strukturach, np. sprawdzając, czy budowa przestrzenna danego modelu jest zbliżona do poznanych białek. Można również obliczać interakcje między poszczególnymi atomami i szacować, czy energetycznie taki model byłby stabilny w przyrodzie. Niekonwencjonalnym podejściem do problemu jest spojrzenie na proces modelowania białek od końca i zweryfikowanie, czy pełni on poprawnie swoją funkcję. Głównym celem niniejszej pracy było znalezienie optymalnych metod wyboru modeli białkowych, ze szczególnym uwzględnieniem białek transmembranowych.

Początkowa część rozprawy miała na celu zweryfikowanie pierwszej tezy, że „zastosowanie metod przepływowych w ocenie jakości modeli strukturalnych białek transmembranowych podnosi efektywność wyboru optymalnego modelu, charakteryzującego się także dobrą funkcjonalnością”. W tym celu poddano badaniom ponad 7,5 tys. modeli białek. Analizowano ich cechy strukturalne i charakterystyki prądowo-napięciowe, obliczone za pomocą metody przepływowej Poissona-Nersta-Plancka (PNP). Dodatkowo wykorzystano do oceny profile elektrostatyczne wewnątrz kanału. Do oceny strukturalnej używano miary pierwiastka średniokwadratowego odchylenia atomów (*Root Mean Square Deviation*, RMSD). W pracy Konopka i in., 2014, autorka niniejszej rozprawy wykazała, że zastosowanie dodatkowo oceny profili elektrostatycznych wewnątrz kanału (*Root Mean Square Error*, RMSE) wpływało na lepsze rozróżnienie modeli. RMSE szczególnie uwypuklał błędy strukturalne wewnątrz poru, które nie zawsze były widoczne w globalnej ocenie modelu, a są bardzo istotne przy transporcie jonów przez białka. W kolejnej pracy Dyrka i in., 2016, autorka rozprawy przeanalizowała zależność między strukturą, elektrostatyką i funkcją modeli. Zastosowanie parametrów funkcjonalnych znacznie ograniczyło zbiory modeli uznanych za dobre. Obliczone wartości parametrów konduktancji i selektywności modeli białek pozwoliły na wskazanie struktur poprawnych globalnie (z niskimi wartościami RMSD), z dobrze zachowaną elektrostatyką kanału (z niskimi wartościami RMSE). Z drugiej strony cechy funkcjonalne pozwalały na wyodrębnienie z grupy modeli o niskich RMSD, takich które były nieznacznie zniekształcone w stosunku do natywnego odpowiednika. Wychwycenie tych niewielkich zmian, które mają istotne znaczenie dla zachowania prawidłowej funkcji, jest kluczowym aspektem poprawy modelowanych białek. W ten sposób autorka rozprawy uzasadniła pierwszą tezę.

Druga teza zakładała, że „istnieją składowe funkcje energii całkowitej, które umożliwiają rozdzielanie modeli strukturalnych białek, uzyskanych z map kontaktów, na zbiory o orientacji prawidłowej i lustrzanej”. Modelowanie białek z uwzględnieniem ograniczeń w postaci par aminokwasów, których odległość w przestrzeni jest mniejsza niż zadana (z reguły 8 Å), jest w ostatnim czasie jedną ze skuteczniejszych metod predykcji struktur białkowych. Jednakże mapy kontaktów nie zawierają informacji o chiralności białka, dlatego modele mogą być zorientowane zgodnie z natywnym odpowiednikiem jak i tworzyć jego lustrzaną wersję. Powstałe modele lustrzane mogą jednak mieć znaczenie w przyrodzie i, w zależności od zewnętrznych czynników, być konkurencyjnymi formami danego białka. Aby umożliwić podział modeli na dwie grupy o odmiennej orientacji przestrzennej, najpierw przeanalizowano modele białek  $\alpha$ -helikalnych, ponieważ wizualna ocena ich skrętności jest bardziej intuicyjna niż  $\beta$ -struktur lub bardziej skomplikowanych kształtów. W pracy Kurczyńska i in., 2016 autorka rozprawy zaproponowała pierwszy raz systematyczną analizę ponad 1000 modeli białkowych. W pracy wykazano, że cechy strukturalne bez odniesienia do statystyki baz danych nie są wystarczające do rozdzielania modeli natywnie zorientowanych i lustrzanych. Pokazano również, że energia całkowita modeli prawidłowo zorientowanych i lustrzanych nie zawsze jest statystycznie różna dla obu typów modeli. Także liczba składowych funkcji energetycznej, które były statystycznie różne dla modeli lustrzanych i natywnie zorientowanych była zależna od jakości rekonstruowanych modeli. Ostatecznie najbardziej różnicujące okazały się składowe energii określające preferencje kątowe wynikające z wykresu Ramachandrana oraz prawdopodobieństwo przyjęcia przez dany aminokwas określonej konfiguracji kątów. W kolejnym etapie rozszerzono badania o domeny białkowe z różnych klas strukturalnych. Obliczenia prowadzono dla ponad 130 tys. modeli. W pracy Kurczyńska i Kotulska, 2018 autorka rozprawy wskazała, które składowe funkcje energetycznej są najbardziej różnicujące dla modeli natywnie zorientowanych i lustrzanych we wszystkich typach strukturalnych białek. Zaproponowała również zautomatyzowanie metody podziału obu typów modeli. Dzięki zastosowaniu klastrowania k-średnich i różnicujących składowych energetycznych wykazała słuszność drugiej tezy.