

### Streszczenie rozprawy doktorskiej

Badania wykonane przeze mnie w rozprawie doktorskiej pt. „Rozszerzenie pola siłowego UNRES o potencjały lokalne i dane z metod porównawczych w celu lepszego przewidywania struktur białek” składały się z czterech etapów:

- określenia wydajności pola siłowego UNRES w przewidywaniu struktur białek poprzez uczestnictwo w eksperymencie CASP10,
- wprowadzenia do pola siłowego UNRES lokalnych potencjałów korelujących położenia łańcuchów bocznych i łańcucha głównego otrzymanych na podstawie:
  - statystycznej analizie pętli białek znajdujących się w bazie PDB,
  - obliczeń fizykochemicznych wykonanych przy użyciu półempirycznej metody AM1,
- implementacji do pola siłowego UNRES więzów odpowiadających za uwzględnienie informacji o strukturze białek otrzymanych przy użyciu metod porównawczych (modelowania homologicznego).

Mój udział w grupie wFCPUNK w projekcie WeFold, polegał na przeprowadzeniu szeregu symulacji dynamiki molekularnej z wymianą replik (MREMD) w polu siłowym UNRES z wykorzystaniem informacji dostarczonych przez pozostałe grupy badawcze dotyczących przewidzianej struktury drugorzędowej i kontaktów między resztami aminokwasowymi. Wykonane przewidywania struktury białek, w szczególności T0740, którego struktura została najlepiej określona ze wszystkich metod biorących udział w eksperymencie CASP10, uwiaryściły zaletę uwzględnienia w symulacjach informacji otrzymanych przy użyciu innych metod bioinformatycznych.

W celu poprawy struktury przewidywanych białek postanowiliśmy wprowadzić do pola siłowego UNRES nowe lokalne potencjały korelujące położenia łańcuchów bocznych i głównego. Użyliśmy analizy statystycznej pętli i innych nieustrukturyzowanych fragmentów pochodzących z 4585 białek zdeponowanych w bazie PDB. Ze względu na niedostateczną różnorodność reszt aminokwasowych w badanych fragmentach użyliśmy uproszczonego alfabetu aminokwasowego, dzielącego 20 standardowych reszt aminokwasowych na 5 grup ze względu na ich właściwości fizykochemiczne. Otrzymane na tej podstawie potencjały średniej siły (PMF) zaimplementowaliśmy do pola siłowego UNRES używając szeregu Fouriera. Następnie przeprowadziłem optymalizację wag nowowprowadzonych potencjałów używając do tego celu zestawu ośmiu małych białek. Analiza przeprowadzonych symulacji MREMD przy wykorzystaniu



najlepszej wartości wagi, wynoszącej 0,57, wykazała znaczącą poprawę otrzymywanych struktur – średnia wartość RMSD spadła o 0,5 Å w stosunku do symulacji prowadzonych bez nowowprowadzonych potencjałów, a największa poprawę zaobserwowaliśmy w obrębie pętli i innych nieustrukturyzowanych elementów.

Postanowiliśmy również wyprowadzić takie same potencjały, jednak przy użyciu metod fizykochemicznych. Do tego celu użyliśmy wykonane poprzednio obliczenia blokowanych reszt aminokwasowych przy wykorzystaniu półempirycznej metody AM1 przez Kozłowską et al., uzupełnione o prolinę. Po otrzymaniu z nich PMF przeprowadziliśmy ich implementację w polu siłowym UNRES. Podczas optymalizacji wagi nowowprowadzonych potencjałów zaobserwowaliśmy, że nakładają się one częściowo z istniejącymi potencjałami torsyjnymi, dlatego obie te wagi musiałem dopasowywać łącznie. Jako treningowy zbiór użyłem osiem białek wykorzystywanych podczas testów potencjałów statystycznych. Spośród 12 przetestowanych zestawów wag, najlepszymi wynikami charakteryzował się zestaw, w którym nowowprowadzone potencjały użyte były z wagą 0,25 przy jednoczesnym obniżeniu wartości potencjałów torsyjnych o 0,50 – poprawa wartości RMSD dla białek treningowych wyniosła 0,41 Å. Wykonałem dodatkowe testy na 22 peptydach i białkach mieszczących się w zakresie 12-126 reszt aminokwasowych. Analiza otrzymanych wyników wykazała poprawę zarówno najniższej wartości RMSD z symulacji – o 0,33 Å (z 4,19 na 3,86 Å), jak i struktury reprezentującej jedną z pięciu grup, po wykonaniu klastrowania metodą minimum Warda – o 0,86 Å (z 7,8 na 6,94 Å).

Po wprowadzeniu nowych potencjałów zaimplementowałem do pola siłowego UNRES także więzy oparte o funkcję energii używaną w programie MODELLER, umożliwiające używanie w symulacjach informacji o strukturze lokalnej np. przewidzianej przy użyciu metod modelowania homologicznego. Aby określić ich skuteczność przeprowadziłem testy na pięciu białkach z eksperymentu CASP9, wymodelowanych przy użyciu programu MODELLER oraz czterech białek dwudomenowych z eksperymentu CASP10, które charakteryzowały się bardzo dobrym przewidzeniem poszczególnych domen, ale ich niedokładnym upakowaniem. Wykonane testy na białkach jednodomenowych wykazały, że uzyskiwane z symulacji w polu siłowym UNRES struktury przy użyciu więzów na modele homologiczne, charakteryzują się dokładnością zbliżoną do najlepszego z użytych modeli. Natomiast testy na białkach dwudomenowych, na które więzy nakładane były tylko na poszczególne domeny, uwidoczniły największą zaletę pola siłowego UNRES – zdolność do poprawnego upakowania elementów struktury drugo- i trzeciorzędowej. Trzy z czterech przetestowanych dwudomenowych białek charakteryzowały się nie tylko poprawionym ułożeniem domen w stosunku do użytych modeli, ale także większym podobieństwem do struktur eksperymentalnych niż struktury jakiejkolwiek metody biorącej udział w eksperymencie CASP10.