

Białka to duże lub bardzo duże cząsteczki, które mają kluczowe znaczenie dla procesów biochemicznych zachodzących w komórkach wszystkich organizmów i są niezbędnym elementem życia na poziomie molekularnym. O ile ustalenie budowy chemicznej molekuly białka, czyli kolejności aminokwasów występujących w jego łańcuchu (sekwencji aminokwasowej) obecnie wykonuje się rutynowo, wyznaczenie odpowiadającej mu struktury przestrzennej jest niezwykle trudnym zadaniem. Metody eksperymentalne są zbyt drogie i czasochłonne, aby mogły być zastosowane w przypadku wszystkich znanych białek. Liczba znanych sekwencji białkowych przewyższa obecnie o co najmniej dwa rzędy wielkości liczbę znanych struktur i na dodatek znacznie szybciej rośnie, co stymuluje rozwój metod obliczeniowych.

Cząsteczka białka jest stosunkowo długim, giętkim łańcuchem atomów połączonych wiązaniami. W stanie rozciągniętym istnieje możliwość dość swobodnej rotacji wokół tych wiązań. Łańcuch białka można zwinąć w przestrzeni do gęsto upakowanej konformacji na niezwykle wiele sposobów, z których tylko jeden jest tym właściwym, obserwowanym w żywych organizmach. Generowanie białkopodobnych, gęsto upakowanych konformacji, które spełniają reguły fizyki jest bardzo kosztowne obliczeniowo. Reguły te zaimplementowane są w postaci funkcji obliczającej energię układu, zazwyczaj jej złożoność obliczeniowa to $\mathcal{O}(N \ln N)$, co stanowi poważną przeszkodę w modelowaniu dużych białek (o dużej liczbie atomów N). Najważniejszym jednak problemem jest to, że na hiperpowierzchni energii białka znajduje się ogromna liczba bardzo stromych minimów, z których tylko jedno odpowiada stanowi natywnemu. Fizyczna natura funkcji energii zdefiniowanej dla wszystkich atomów białka zatem bardzo utrudnia jej minimalizację.

Często stosowanym rozwiązaniem jest wykorzystanie gruboziarnistej metody modelowania, w której grupę atomów zastępuje się pojedynczym centrum oddziaływań. Przykładem takiej metody jest zaproponowany ostatnio przez wnioskodawcę algorytm SURPASS, w którym jednej reszcie aminokwasowej odpowiada jeden pseudo-atom. Powoduje to znaczną redukcję (nieomalże dwudziestokrotną) liczby atomów N . Przede wszystkim jednak w modelu tym zaproponowano statystyczne pole siłowe, które redukuje bariery energetyczne i znacząco ułatwia przeszukiwanie przestrzeni stanów. Dzięki temu model SURPASS umożliwia dramatyczny wzrost wydajności obliczeniowej, uzyskane modele białek jednak istotnie odbiegają od rzeczywistych pełnoatomowych struktur. Kolejnym etapem obliczeń jest zatem odbudowa detali atomowych i udokładnianie modelu.

Celem niniejszego projektu jest opracowanie algorytmu do wieloskalowego modelowania struktur białek. Metoda ta wykorzystywała będzie model SURPASS i oprogramowanie ROSETTA, które połączone zostaną w jeden spójny protokół. W tym celu opracowanych zostanie szereg algorytmów usprawniających obie te metody a także umożliwiające ich wydajne połączenie. Proponowany rozwój algorytmu SURPASS umożliwi wysokowydajne modelowanie wielkich układów białkowych. Do modelu dodane zostaną nowe człony pola siłowego, co poprawi jego specyficzność. Niezmiernie ważnym elementem projektu jest wykorzystanie danych eksperymentalnych. Wykorzystane one zostaną do ograniczenia przeszukiwanego fragmentu przestrzeni stanów a także do wyboru najtrafniejszej konformacji końcowej.