

### Streszczenie

Stale nierozwiązanym problemem w biologii jest mechanizm w jaki białka przyjmują swoją specyficzną i zarazem funkcjonalną strukturę trzeciorzędową. Pomimo znajomości występujących w białkach oddziaływań oraz eksperymentalnych doniesień na temat przebiegu fałdowania, z powodu wyjątkowej złożoności struktury białek nie jest możliwe prześledzenie ani obliczeniowe odtworzenie całego procesu uzyskiwania przez białko struktury natywnej. Obecnie proponowane jest kilka modeli przebiegu fałdowania, które opierają się na różnych przesłankach eksperymentalnych i które kładą nacisk na odmienne czynniki fizykochemiczne dominujące w trakcie fałdowania. Odtworzenie procesu fałdowania metodą *in silico* stanowi jedno z podejść stosowanych w metodach przewidywania struktury białek na podstawie ich sekwencji aminokwasowej. Zarówno metody *ab initio* oraz metody korzystające z danych o znanych strukturach natywnych, w dalszym ciągu nie są wystarczające aby dostarczać trafnej informacji o strukturze. Zaprezentowany w niniejszej pracy model podejmuje próbę utworzenia metody przewidywania struktury białek, która zarówno wykorzystuje dane o dostępnych strukturach natywnych ale też opiera się na znanych mechanizmach fałdowania. Model ten dodatkowo wykorzystywany jest do wyjaśniania zjawisk związanych ze stabilizacją struktury białek. Główne założenie modelu stanowi wieloetapowość procesu fałdowania, polegająca na występowaniu obserwowanych eksperymentalnie stanów pośrednich. Z tego powodu wyróżniono dwa główne etapy strukturę Early Stage oraz strukturę Late Stage.

Etap Early Stage wykorzystuje informację o lokalnych preferencjach konformacji łańcucha głównego, które jak pokazano prowadzą do formowania się struktur drugorzędowych na pierwszych etapach fałdowania. Na tym etapie ważne jest zebranie informacji o relacji sekwencji do struktury w formie pozwalającej na określenie jak najtrafniejszej lokalnej konformacji. W celu uproszczenia definicji struktury lokalnej wprowadzono siedmioliterową klasyfikację konformacji

łańcucha głównego. Klasyfikacja ta opiera się na ograniczonej podprzestrzeni konformacyjnej w formie eliptycznej ścieżki na mapie Ramachandrana oraz podziale tej mapy na siedem stref definiujących siedem kodów strukturalnych. Na podstawie kodów strukturalnych tworzone są biblioteki motywów strukturalnych, których prawdopodobieństwo wystąpienia w znanych strukturach wykorzystywane jest do wyznaczania najbardziej prawdopodobnej struktury wczesnego pośrednika, będącej punktem startowym do kolejnych etapów wyznaczania struktury natywnej. W pracy przedstawiono dwie metody przewidywania tej struktury, z których jedna opiera się na tabeli kontyngencji prawdopodobieństwa dla tetrapeptydów natomiast druga, bardziej efektywna, na słownikach statystycznych. Zbudowano również narzędzie do wizualizacji relacji struktury do sekwencji w formie tabeli kontyngencji.

W kolejnym etapie – Late Stage - przyjmuje się dominujący wpływ zjawiska hydrofobowego na rozmieszczenie reszt aminokwasowych w białkach ze względu na ich właściwości. Idealny przypadek opisuje model Fuzzy Oil Drop (FOD) zakładający teoretyczny rozkład hydrofobowości w molekułę białka w formie trójwymiarowej funkcji Gaussa. Rozkład ten stanowi referencję, względem której modyfikowana jest struktura białka w trakcie symulacji procesu fałdowania. Poprzez wyznaczenie odległości pomiędzy obserwowanym rozkładem dla znanej struktury białka z rozkładem według FOD poprzez obliczenie entropii relatywnej, możliwa jest ilościowa ocena stopnia uformowania w molekułę jądra hydrofobowego. Procedura ta pozwala na analizę czynników stabilizujących natywne struktury białek. Przy pomocy modelu FOD pokazano znaczenie nieustrukturalnionych fragmentów łańcucha w procesie fałdowania oraz w tworzeniu kompleksów przez białka wiążące DNA. Przeanalizowano również współdziałanie efektu hydrofobowego oraz wiązań dwusiarczkowych jako czynników stabilizujących funkcjonalną strukturę białek.