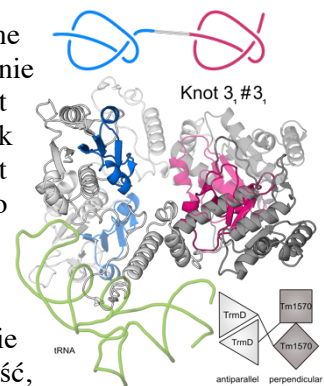


## Klasyfikacja topologiczna białek z mostkami cysteinowymi oraz wpływ zapętlenia na zwijanie białek

Niniejszy projekt poświęcony jest kilku niezależnym kierunkom badawczym: nietrywialnej topologii i zapętleniom, poprawności zwijania białek, sztucznej inteligencji oraz analizie danych („big-data”). W ostatnim czasie, dzięki opartemu o algorytmy sztucznej inteligencji programowi AlphaFold, uzyskane zostały przełomowe wyniki dotyczące przewidywania 3-wymiarowej struktury białek, których jakość jest porównywalna z wynikami otrzymywanymi eksperymentalnie. Bazy danych AlphaFold (stworzona przez Google) oraz ESMFold (Meta/Facebook) zawierają w sumie 800 milionów struktur przewidzianych metodami sztucznej inteligencji, których w ramach tego projektu użyjemy do znalezienia związków zapętlenia z funkcją biologiczną, określenia prawdopodobieństwa poprawnego zwijania białek, oraz charakterystyki innych własności białek.

Białka są podstawowymi budulcami żywych organizmów oraz pełnią ważne biologiczne funkcje. Białka to długie łańcuchy złożone z kilkuset lub nawet kilku tysięcy aminokwasów (istnieje takich 21 podstawowych typów aminokwasów). W odpowiednich warunkach, w procesie zwanych zwijaniem, takie łańcuchy przyjmują pewne szczególne 3-wymiarowe kształty określone jako stany natywne; przyjęcie takich konfiguracji jest konieczne do tego, by białka pełniły określone funkcje biologiczne. W związku z tym, sekwencja aminokwasów tworzących białko musi w jakiś sposób determinować jego 3-wymiarowy natywny kształt oraz funkcję. Jeśli białko nie zwinie się poprawnie do struktury natywnej, to nie jest ono w stanie pełnić odpowiedniej funkcji.

Jako że białka są długimi łańcuchami, to można podejrzewać, że mogą one tworzyć różnego rodzaju zapętlenia, np. takie jak węzły, które powstają spontanicznie na sznurku lub kablu. Przez wiele lat uważano, iż struktura białek jest zbyt skomplikowana, by mogły się zwinąć do zapętlonej struktury. Okazało się to jednak nieprawdą – zapętlone białka zostały znalezione, a kierowniczką tego projektu jest jedną z najbardziej znanych w świecie osób, które doprowadziły do rozwoju tego kierunku badawczego. Ostatnio, w oparciu o 200 milionów struktur przewidzianych przez AlphaFold, odkryła ona i sklasyfikowała 700.000 potencjalnie zawężonych białek. Ponadto, odkryła ona pierwsze podwójnie zapętlone białko, które pokazane jest na rysunku obok. Do kluczowych obecnie pytań należy to, jak istnienie takiej topologii wpływa na proces zwijania, stabilność, oraz funkcję białka.



Kierowniczką projektu odkryła także inne rodzaje zapętlonych struktur w białkach: lassa, sploty, theta-krzywe, etc. Najczęściej występujące z nich to lassa, których struktura składa się z dwóch części: pętli stworzonej przez główny łańcuch białka (i domkniętej przez (nie)kowalencyjny kontakt między dwoma aminokwasami), oraz jednego z końców białka przecinającego taką pętlę. Lassa są znane od ok. 10 lat, jednakże z powodu niewystarczającej ilości danych wciąż nie znamy odpowiedzi na pytania dotyczące ich podstawowych własności: jakie są ich funkcje w białkach, oraz jak bardzo złożone geometrycznie mogą być lassa. Dotychczas także nie zbadano szczegółowo, jak istnienie lass lub węzłów wpływa na prawdopodobieństwo poprawnego zwinienia białka.

Niepoprawne zwinienie białek uznawane jest za główną przyczynę chorób Alzheimera, Parkinsona, Huntingtona, i innych. Pewne badania wskazują na to, że za chorobę Parkinsona odpowiedzialne są także zawężone białka. Niezależnie, nasze wstępne eksperymentalne wyniki, dotyczące białek z proteomu E. Coli, wskazują na istnienie korelacji pomiędzy zapętleniem a prawdopodobieństwem złego zwinienia białka. W oparciu o analizę wszystkich białek z PDB znaleźliśmy, iż 61% białek (E. Coli), które unikają degradacji jest splątanych (większość z nich tworzy lassa). W ramach tego projektu, w oparciu o 850 milionów struktur zdeponowanych w bazach danych AlphaFold oraz ESMFold, planujemy rozszerzyć tę analizę.

Między innymi, w ramach tego projektu zamierzamy zidentyfikować i sklasyfikować wszystkie typy lass, splotów, theta krzywych oraz innych motywów przewidzianych przez AlphaFold i ESMFold. Wyniki takie powinny poszerzyć naszą wiedzę dotyczącą złożoności struktury białek, oraz korelacji pomiędzy ich splątaniem i strukturą. Nasze wstępne wyniki oparte o dane z bazy AlphaFold wskazują, że lassa występują częściej w białkach alosterycznych oraz w enzymach. Musi to być także wynikiem procesów ewolucyjnych. Na podstawie bazy AlphaFold zidentyfikowaliśmy 700.000 potencjalnie zawężonych białek, których zapętlenia i ich związek z mechanizmem zwijania także będziemy badać w ramach tego projektu. W ramach projektu przeprowadzimy też szerokie symulacje poprawności zwijania, które będziemy mogli porównać z eksperymentem. Jako że kierowniczką projektu oraz prowadzona przez nią grupa badawcza ma duże doświadczenie w badaniu zapętlonych białek, i m.in. ostatnio dokonała ona eksperymentalnej analizy podwójnie zapętlonego białka TrmD-TM1570, to jesteśmy przekonani, że wszystkie ambitne cele projektu w takim gronie zostaną osiągnięte.