

Oddziaływania białko-białko to ważny proces, niezbędny dla wielu funkcji komórki – takich jak reakcja na bodźce zewnętrzne. Niestety, bardzo ciężko jest je badać. Istnieją metody eksperymentalne zdolne do sprawdzenia czy dane białka oddziałują ze sobą, jednak ich zastosowanie wymaga zazwyczaj znalezienia uprzednio potencjalnych kandydatów. Moim celem jest wypełnienie tej luki korzystając z największego zasobu danych białkowych – sekwencyjnych baz danych.

W ostatnich latach pojawiła się nowa metoda do badania struktury i oddziaływań białkowych *in silico* - „Direct Coupling Analysis” (DCA). Jest to metoda statystyczna, która dla zbioru spokrewnionych sekwencji wyznacza model korelacji pomiędzy parami odpowiadających sobie aminokwasów w białkach. Pozwala to na oszacowanie tendencji do wspólnej mutacji, z reguły powodowanej wspólną presją ewolucyjną. Dotyczy ona zazwyczaj oddziałujących ze sobą aminokwasów - zatem korelacja w DCA sugeruje kontakt w strukturze białka. Co więcej, model koewolucji uzyskiwany poprzez DCA może być użyty do sprawdzenia jak dobrze dana sekwencja pasuje do analizowanego zbioru. Mając dwie grupy sekwencji reprezentujących oddziałujące ze sobą rodziny białek, możemy z kolei dopasować konkretnych partnerów, poprzez założenie, że poprawne dopasowanie to takie, które wykazuje najwyższą koewolucję. Wreszcie, DCA umożliwia oszacowanie całkowitej koewolucji dwóch zbiorów potencjalnych partnerów – co pomoże w ukierunkowaniu, często drogich, metod eksperymentalnych.

Celem mojego projektu jest rozszerzenie możliwości DCA poprzez wprowadzenie, dotychczas pominiętych, informacji biologicznych. Obecnie DCA traktuje aminokwasy jako niezależne i wzajemnie niepodobne obiekty, co nie ma odbicia w rzeczywistości. Wprowadzenie nowych danych do DCA nie jest zadaniem łatwym, ale jego wykonanie otworzy drogę do nowych zastosowań tej metody, np. do obecnie problematycznych szybko ewoluujących sekwencji. Moje wykształcenie bioinformatyczne, razem z dużym doświadczeniem moich opiekuna (Joanna Sułkowska, pionier badania białek przy użyciu DCA) i współpracownika (Martin Weigt, jeden z autorów tej metody) gwarantują powodzenie tego projektu. Dodatkowo, zmiana danych wejściowych pozwoli na redukcję rozmiaru macierzy używanych w obliczeniach. Wiąże się to z kolejnym celem tego projektu – poprawą czasu obliczeniowego, obecnie głównego ograniczenia metody. Poprawa wydajności pozwoli na zastosowanie DCA do poszukiwania potencjalnych oddziaływań używając całych białkowych baz danych.

Stworzoną przeze mnie nową wersję DCA zastosuję również do badań bardziej szczegółowych - a konkretnie koewolucji interfejsów białkowych. Skupię się na dwóch klasach białek, które łączy nietrywialna topologia - obie posiadają zawężony fragment w swojej strukturze. Takie białka, pomimo zachowanej ewolucyjnie struktury, cechuje wysoka zmienność sekwencyjna. To utrudnia skorzystanie z dotychczasowej wersji DCA – wprowadzone przeze mnie fizykochemiczne modyfikacje zniosą tę trudność. Dodatkowo, oba typy białek, które chcę zbadać można zaklasyfikować jako białka fuzyjne (początkowo oddzielne molekuly, które wyniku ewolucji zostały złączone w jeden długi łańcuch białkowy). Przeanalizuję interfejs istniejący pomiędzy tymi, kiedyś rozdzielnymi, domenami.

W szczególności, zbadam wewnętrzny, położony między potwórczeniami w strukturze łańcucha, interfejs dwukierunkowych transbłonowych transporterów. Te właśnie powtórzenia sugerują, że białka te powstały w wyniku fuzji homodimeru. W niektórych przypadkach taka fuzja spowodowała powstanie węzła, jednak nie jest to reguła. Planuję porównać przebieg ewolucji tych dwóch foldów, aby wyjaśnić te rozbieżności.

Dodatkowo, niedawno znaleźliśmy grupę prawdopodobnych fuzyjnych enzymów – złożonych z dwóch, zwykle występujących w postaci homodimeru, metylotransferaz. Ponadto, obie takie podjednostki występując pojedynczo są zawężone. Chcę je przyrównać do ich niezłączonych odpowiedników i przewidzieć w jaki sposób i poprzez jakie oddziaływania takie białka mogą funkcjonować. Także na podstawie interfejsu chcę wyciągnąć wnioski dotyczące struktury przestrzennej takiego białka fuzyjnego. Znalezienie podwójnie zawężonego białka, z jakim prawdopodobnie mamy tu do czynienia, spowoduje przełom z dziedziny związania białek i zrozumienia ich topologii.