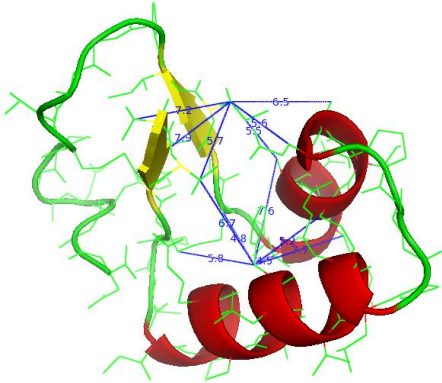


Znajomość struktury trójwymiarowej białek umożliwia poznanie ich funkcji i wykorzystanie np. w procesie projektowania leków. Niestety, w związku z tym, że metody eksperymentalne są skomplikowane oraz czasochłonne, obecnie mniej niż 20% białek o znanej sekwencji, posiada znaną strukturę. Dlatego poszukuje się innych metod na jej określenie. Są nimi np. komputerowe metody przewidujące tzw. miejsca kontaktowe, czyli punkty w strukturze białka leżące blisko siebie w przestrzeni. Dostarczają one istotnych informacji o sposobie związania się białka, które mogą posłużyć do odtworzenia jego struktury. (Rys. 1). Jeżeli miejsca kontaktowe byłyby wiarygodnie przewidziane, mogłyby być wykorzystane jako dodatkowe dane dla metod eksperymentalnych. To by z kolei doprowadziło do zmniejszenia ich kosztów i wymaganego czasu.



**Rysunek 1.** Przykładowe miejsca kontaktowe (niebieskie linie) w białku krabiny. Liczby oznaczają odległość wyrażoną w Ångstrmach.

Głównie dzięki metodom eksperymentalnym możemy poznać struktury białek i na ogół te metody dostarczają wiarygodnych danych. Zdarzały się jednak przypadki fałszerstw lub po prostu błędnego wyznaczenia struktury białka, które to włączano do zbioru białek o znanej strukturze. Jednym z nich jest fotoaktywne białko żółte, którego strukturę opublikowano w 1989 roku. Dopiero po 6 latach odkryto, że jest ona błędna i zaktualizowano dane o niej. Gdyby informacja o miejscach kontaktowych była dostępna, mogłaby być wykorzystana w pracy z metodami eksperymentalnymi, zapobiegając błędowi.

Dzisiejsze metody przewidywania miejsc kontaktowych wykorzystują tzw. mutacje skorelowane w dopasowaniu wielosekwencyjnym białka. Dopasowanie wielosekwencyjne to zbiór sekwencji aminokwasów białek związanych ewolucyjnie z analizowanym białkiem. Teoria mutacji skorelowanych mówi, że jeżeli dwie

...TVC	<b>C</b>	ASIVARSNFN	<b>I</b>	CRLPG...
...TVC	<b>C</b>	PSIVCFRNFN	<b>I</b>	LRVPA...
...TTC	<b>P</b>	PGIVAKSPFN	<b>V</b>	CRLPG...
...TTC	<b>P</b>	PSIVAARNFN	<b>V</b>	LRVPA...
...TSC	<b>S</b>	PGIVCHSPFN	<b>L</b>	CRLPG...
...TAC	<b>S</b>	PSIVAHSNFN	<b>L</b>	CRVPG...
...TTC	<b>C</b>	PSIVHSFNFN	<b>A</b>	LRLPA...

correlated

pozycje w sekwencji białka są istotne dla jego struktury lub funkcji, to gdy na jednej z tych pozycji dochodzi do mutacji aminokwasu, wówczas na drugiej pozycji dochodzi do mutacji skorelowanej, w celu zachowania odpowiednich interakcji (żółte kolumny z rysunku 2). W związku z tym, że miejsca kontaktowe są punktami istotnymi dla struktury białka, to ich obecność również może być określana za pomocą mutacji skorelowanych.

**Rysunek 2.** Mutacje skorelowane pomiędzy dwiema pozycjami w sekwencjach dopasowania wielosekwencyjnego.

Obecnie najlepsza metoda wykorzystująca mutacje skorelowane, gplmDCA [1], działa w oparciu o algorytm analizy sprzężeń bezpośrednich (z ang. Direct Coupling Analysis, DCA). Niestety, skuteczność przewidywania kontaktów przez gplmDCA wynosi jedynie 40% dla 100 przewidzianych miejsc kontaktowych. To nie wystarcza do rekonstrukcji struktury białka. Żadne z wcześniejszych badań związanych z DCA nie skupiało się na analizie błędnych wyników tej metody aby zrozumieć przyczyny ich występowania. Wiadomo, że mutacje skorelowane są bardziej związane z funkcją, niż strukturą białka, co prowadzi do błędów. Poza tym, skuteczność działania DCA jest zależna od cech analizowanego białka. Obecnie niemożliwe jest określenie wiarygodności wyników DCA. Dlatego nawet gdy metoda prawidłowo przewiduje kontakty w strukturze to ich wykorzystanie wiąże się z dużym ryzykiem popełnienia błędu. Jeżeli błędne wyniki zostałyby wyeliminowane, podniosłoby to jakość i wiarygodność przewidywanych miejsc kontaktowych.

W dotychczasowych badaniach zauważyliśmy, że nieprawidłowo przewidywane przez DCA miejsca kontaktowe mogą być skorelowane z fizykochemicznymi cechami tworzących je aminokwasów. Dlatego analizujemy częstotliwość występowania typów aminokwasów i ich położenie w strukturze białka wśród prawidłowo i błędnie przewidzianych miejsc kontaktowych. Zaproponowane przez nas miary zróżnicowania wartości DCA wśród przewidzianych dla danego białka miejsc kontaktowych umożliwiają określenie wiarygodności wyników DCA dla tego białka. Zamierzamy zbadać ten temat z uwzględnieniem klas strukturalnych białek. Kopro motorem przygotowywanej rozprawy doktorskiej jest prof. G. Vriend z Centrum Informatyki Molekularnej i Biomolekularnej w Nijmegen (Holandia), który pokazał, że informacja ukryta w dopasowaniu wielosekwencyjnym, może wskazać aminokwasy istotne dla funkcji białka. Ta wiedza wraz z zaproponowaną przez niego miarą korelacji funkcyjnej pomoże w opracowaniu przez nas procedury filtracji wyników DCA zależnych od funkcji białka, a nie struktury. Ostatnim etapem naszych badań będzie zaprojektowanie filtrów opartych na wynikach powyższych badań, które wyeliminują błędne wyniki gplmDCA i umożliwią ocenę zasadności stosowania DCA w analizie danego białka. Wyniki naszych badań przybliżą wykorzystanie metod komputerowych w procesie odtwarzania struktury białka.

[1] C. Feinauer i inni, *PLoS Comput Biol.*, **2014**, 10(10), e1003847.