

Zaawansowane metody rozpoznawania typów aminokwasów w przypisywaniu sygnałów NMR białek natywnie nieustrukturyzowanych

Diana Gągała

Kierownik: **dr Anna Zawadzka-Kazimierczuk**

Białka natywnie nieustrukturyzowane (ang. Intrinsically Disordered Proteins, IDPs) stanowią odrębną klasę białek, charakteryzującą się brakiem stabilnej struktury trzeciorzędowej w stanie natywnym. IDPs odgrywają zasadniczą rolę w procesach związanych z wiązaniem cząsteczek, biorą udział w mechanizmach regulacji, kontroli czy przekazywania sygnałów. Jednakże dynamiczna struktura IDPs może powodować ich błędne związanie i w konsekwencji agregację. Zmienione strukturalnie bądź funkcjonalnie IDPs przyczyniają się do rozwoju licznych jednostek chorobowych. Spektroskopia NMR umożliwia badanie struktur IDPs, ich dynamiki oraz oddziaływań, jednak największe wyzwanie stanowi przypisanie sygnałów jąder poszczególnych reszt aminokwasowych. Dynamiczna struktura IDPs powoduje uśrednianie się wartości przesunięć chemicznych, ze względu na nieustannie zmieniające się otoczenie chemiczne, co w konsekwencji prowadzi do silnego stłoczenia sygnałów i utrudnia interpretację wyników. Remedium stanowi wykorzystanie wysoko wymiarowych eksperymentów NMR (3D, 4D, 5D etc.) i unikalny proces ich analizy metodą SMFT (ang. Sparse Multidimensional Fourier Transform), który umożliwia obserwację dwuwymiarowych przekrojów widm wysoko wymiarowych z perspektywy widma niższej wymiarowości (tzw. bazowego), mającego wspólne wymiary z widmem wysoko wymiarowym. Sam proces przypisywania sygnałów można zautomatyzować przy pomocy programu TSAR (ang. Tool for SMFT-based Assignment of Resonances), który w oparciu o dwuwymiarowe przekroje przypisuje sygnały poszczególnych jąder reszt aminokwasowych w sekwencji białka. W celu ulepszenia procesu przypisania, przystosowano program TSAR do wykorzystywania informacji z programu Potenci (uwzględniającego pH, temperaturę i sekwencję białka) oraz zaimplementowano procedurę LDA (statystyczną metodę określającą prawdopodobieństwa, że wybrana reszta jest aminokwasem danego typu). Dzięki wprowadzeniu dodatkowych danych do programu TSAR, uzyskano poprawę przypisania sygnałów, jak i zmniejszenie liczby błędnie przypisanych i nieprzypisanych reszt. Najlepsze wyniki otrzymano dla wersji TSAR skombinowanej z algorytmem LDA: liczba poprawnych przypisań wzrosła o blisko 5 p.p., błędnych zmalała o ponad 0,6 p.p., a nieprzypisanych – zmalała o ponad 4 p.p.

Literatura:

- [1] Dziedzic-Letka A., Ożyhar A., „Białka inherentnie nieuporządkowane”, *Postępy w Biochemii*, 2012, 58 (1), 100-109.
- [2] Kazimierczuk K., Zawadzka-Kazimierczuk A., Koźmiński W., „Narrow peaks and high dimensionalities: Exploiting the advantages of random sampling”, *Journal of Magnetic Resonance*, 2009, 197, 219-228.
- [3] Zawadzka-Kazimierczuk A., Koźmiński W., Billeter M., „TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra”, *Journal of Biomolecular NMR*, 2012, 54(1), 81-95.
- [4] Nielsen J.T., Mulder F.A.A., „POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins”, *Journal of Biomolecular NMR*, 2018, 70, 141-165.
- [5] Romero J.A., Putko P., Urbańczyk M., Kazimierczuk K., Zawadzka-Kazimierczuk A., „Linear discriminant analysis reveals hidden patterns in NMR chemical shifts of intrinsically disordered proteins” *PLoS Comput Biol*, 2022, 18(10): e1010258.