

Białka są kluczowymi cząsteczkami biologicznym, które znajdują się we wszystkich organizmach żywych. Składają się one z 20 prostych związków chemicznych – aminokwasów. Sekwencja, tj. kolejność aminokwasów w łańcuchu białka determinuje jego strukturę przestrzenną, a ta, z kolei, decyduje o jego funkcji. Doświadczalne wyznaczenie struktury przestrzennej białka, jest procesem czasochłonnym, nie zawsze kończącym się sukcesem. W związku z tym tworzone są narzędzia bioinformatyczne, które na podstawie sekwencji są w stanie przewidywać strukturę białka. Skalę trudności tego problemu ilustruje fakt, że pomimo 60-letniej historii badań żadne z dostępnych metod nie są w stanie przewidzieć prawidłowej struktury przestrzennej dla wielu sekwencji białkowych.

Często stosowanym modelem w badaniach nad zależnością pomiędzy sekwencją a strukturą białek są tak zwane domeny splecionych helis typu  $\alpha$  – białka o regularnej strukturze, zaangażowane w wiele istotnych procesów takich jak transdukcja sygnałów biologicznych czy transport komórkowy. Struktury domen splecionych helis typu  $\alpha$  mogą zostać opisane poprzez równania parametryczne. Dzięki parametryzacji dowolna struktura przestrzenna domeny tego typu może zostać jednoznacznie opisana za pomocą zaledwie kilku parametrów liczbowych. Łatwość ilościowego opisu oznacza reprezentację struktur domen splecionych helis typu  $\alpha$  w sposób przystępny dla współczesnych metod obliczeniowych takich jak np. uczenie maszynowe.

Celem niniejszego projektu jest charakterystyka zależności pomiędzy sekwencją a strukturą przestrzenną domen splecionych helis typu  $\alpha$ . W tym celu wykorzystane zostaną nowoczesne algorytmy uczenia maszynowego oraz symulacji molekularnych. W ciągu ostatnich lat algorytmy tego typu zrewolucjonizowały nie tylko wiele aspektów życia codziennego, ale także miały ogromny wpływ na prowadzenie badań naukowych, między innymi w dziedzinie nauk biologicznych. Szerokim echem, także w doniesieniach medialnych, odbiła się również zaskakująca skuteczność algorytmu AlphaFold, przewidującego strukturę białek. Program ten, stworzony przez powiązaną z firmą Google grupę DeepMind, zdeklasował algorytmy innych grup badawczych, które są związane przez długi czas z tym polem badań. W proponowanym projekcie algorytmy uczenia maszynowego zostaną wykorzystane do stworzenia narzędzia które będzie w stanie przewidywać, z wykorzystaniem równań parametrycznych, strukturę domen splecionych helis typu  $\alpha$  na podstawie sekwencji.

Ponadto, efektem prowadzonych badań będzie systematyczny opis reguł i zależności sekwencyjno-strukturalnych w domenach splecionych helis typu  $\alpha$ . Część doświadczalna projektu pozwoli zweryfikować wykryte reguły i powiązania oraz opisać nowe białka o tej architekturze, niezaobserwowane dotąd w naturze.