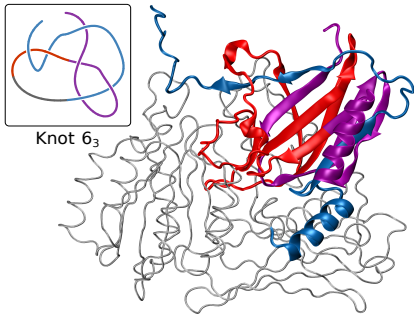


Biologiczny kod zawężenia - identyfikacja wzoru zawężenia w biomolekułach na podstawie metod AI

Badania prowadzone w ramach niniejszego projektu dotyczą trzech z pozoru odległych od siebie zagadnień: białek, węzłów, oraz sztucznej inteligencji. Białka są podstawowym budulcem organizmów żywych oraz pełnią w nich wiele istotnych funkcji. Podstawowymi elementami budowy białek jest 21 aminokwasów. Każde białko jest łańcuchem złożonym z kilkudziesięciu, kilkuset, lub nawet kilku tysięcy aminokwasów. Łańcuch taki w odpowiednich warunkach przyjmuje konkretny trójwymiarowy kształt, zwany strukturą natywną białka; przyjęcie takiego kształtu jest konieczne do tego, by dane białko mogło pełnić swoją biologiczną funkcję. W związku z tym, sekwencja aminokwasów tworzących białko (którą można uznać za długie słowo złożone z 21 rodzajów liter), musi w jakiś sposób determinować trójwymiarową strukturę natywną białka, a także jego funkcję. Zrozumienie tych związków jest jednym z celów biologii strukturalnej; jak się okazuje, jest to również jedno z największych wyzwań współczesnej nauki.



Rysunek 1 Białko z nowym typem węzła zidentyfikowane przez nas na podstawie danych z bazy AlphaFold.

Jako że białka są długimi łańcuchami, można podejrzewać, że mogą się na nich tworzyć węzły – podobnie jak na sznurku, kablu telefonicznym, itp. Przez wiele lat sądzono, iż z uwagi na ich skomplikowaną strukturę, węzły nie mogą istnieć w stanie natywnym białka. Okazuje się jednak, że nie jest to prawdą – węzły takie zostały znalezione, a jednym z najbardziej uznanych na świecie badaczy takich struktur jest kierowniczką niniejszego projektu. Należy podkreślić, iż informacja o tym, czy dany łańcuch jest zawężony, jest bardziej podstawowa niż konkretny trójwymiarowy kształt struktury natywnej – dlatego też badania zawężonych białek są tak ważne. Natomiast, niezależnie od wielu sukcesów w ramach tego kierunku badań, wiele zagadnień dotyczących zawężonych białek – w szczególności funkcje jakie pełnią w białkach węzły – wciąż wymaga wyjaśnienia. Warto też zaznaczyć, iż w oczywisty sposób badania takie łączą się z matematyczną teorią węzłów, której celem jest opisanie własności i klasyfikacja węzłów jako obiektów matematycznych.

Istotnym ograniczeniem w realizacji wyżej opisanych badań jest ilość dostępnych danych. Wyznaczenie trójwymiarowej struktury białka eksperymentalnie jest kosztowne i czasochłonne. Z dotychczasowych badań kierowniczki projektu wynika między innymi, iż białka zawężone stanowią kilka procent wszystkich wyznaczonych struktur białkowych – z jednej strony jest to duża liczba, ale też okazuje się ona zbyt mała, by odpowiedzieć na najważniejsze pytania o rolę białek z węzłami. Natomiast w ostatnim roku ilość danych ważnych dla biologii strukturalnej istotnie się zwiększyła, w wyniku przełomowych wyników otrzymanych przy zastosowaniu metod sztucznej inteligencji: stworzony przez Google program AlphaFold, oparty o algorytmy sztucznej inteligencji, przewidział strukturę ponad 400000 białek z 20 różnych genomów, prześcigając zakresem informacji bazę danych doświadczalnych. Z kolei wykonana już przez nasz zespół wstępna analiza takich danych dla organizmu ludzkiego wykazała, iż AlphaFold przewiduje istnienie węzłów, w tym także nieznanymi wcześniej rodzajów. W ramach niniejszego projektu zamierzamy rozszerzyć analizę zawężonych białek o wyniki otrzymane przez AlphaFold, a także stworzyć własne algorytmy sztucznej inteligencji do analizy takich białek. Mamy nadzieję, że podejście takie doprowadzi do odkryć analogicznej rangi, jak rezultaty podejścia opartego na programie AlphaFold w stosunku do tradycyjnych metod przewidywania struktury białek.

Ścisłej rzecz biorąc, w ramach tego projektu planujemy odpowiedzieć m.in. na takie fundamentalne pytania: jakie typy węzłów mogą istnieć w białkach, jaki kod aminokwasów może być odpowiedzialny za zawężenie białka (czyli też jak AlphaFold przewidział zawężoną strukturę nie opierając się o przykłady innych białek z węzłami), czy i jak można zaprojektować wzór aminokwasów w celu uzyskania „sztucznego” zawężonego białka. W tym celu wykonamy wszechstronny, topologiczny i biologiczny przegląd danych udostępnionych przez AlphaFold, a także wykorzystamy metody sztucznej inteligencji w celu wyznaczenia kodu zawężenia, przewidzenia struktury, oraz zaprojektowania sztucznych zawężonych struktur białek. Ponadto wykażemy, że metody sztucznej inteligencji mogą być stosowane do analizy topologicznej różnych typów zawężeń w białkach (tzw. lass, theta-krzywych, węzłów cysteinowych), co zostanie poparte rygorystyczną analizą matematyczną.

Niniejszy projekt będzie realizowany przez międzynarodowy, interdyscyplinarny zespół, złożony z ekspertów z zakresu metod sztucznej inteligencji i uczenia maszynowego w układach biologicznych (zespół czeski); matematyków, których głównym zadaniem będzie zastosowanie teorii węzłów m.in. do modelowania otwartych łańcuchów (zespół słoweński); oraz ekspertów zajmujących się badaniem białek zapętłonych (zespół polski). Należy podkreślić, iż metody sztucznej inteligencji są obecnie bardzo aktywnie rozwijane i stosowane w wielu dziedzinach. Można zatem mieć nadzieję, iż w niedalekiej przyszłości opracowane w projekcie metody znajdą zastosowania w analizie innych zagadnień biologicznych (np. dotyczących chromosomów lub RNA), pomogą w projektowaniu leków (jedno z białek wchodzące w skład SARS-CoV 2 posiada topologię slipknota) czy stabilnych materiałów opartych na nietrywialnej topologii polimeru.