

Sophia Bałdysz  
**A new identification method  
of phage cell wall-associated lytic proteins**

Bakteriofagowe białka lityczne (lizyny) są jednym z kluczowych czynników w procesie infekcji wirusowej. Ich rola jest dwójaka - umożliwiają wstrzyknięcie fagowego materiału genetycznego do komórki gospodarza poprzez lokalne nadtrawienie peptydoglikanu, a także rozkładają ścianę gospodarza uwalniając wiriony potomne pod koniec cyklu replikacyjnego. Choć białka te mają wspólną funkcję biologiczną to są kodowane przez wiele różnych rodzin genowych, wykazują się różnorodnością budowy i aktywności enzymatycznych. Różnorodność ta utrudnia identyfikację nowych kandydatów należących do tej grupy. Stanowi to problem, ponieważ białka te znajdują zastosowanie w wielu obszarach gospodarki i medycyny. W dobie rosnącej liczby zakażeń spowodowanych bakteriami antybiotykoopornymi wykorzystanie białek o właściwościach bakteriobójczych może być obiecującą terapią, a preparaty lecznicze zawierające inne białka o podobnych właściwościach są na etapie badań klinicznych. Co więcej, produkty zawierające enzymy lityczne już są wykorzystywane w kosmetyce. Zatem zapotrzebowanie na nowe lizyny do zastosowań medycznych i przemysłowych jest duże, ale ich znalezienie nie jest łatwym zadaniem. Tradycyjne metody poszukiwania nowych białek wymagają izolacji fagów ze środowiska, sekwencjonowania ich genomów i przeprowadzanie laboratoryjnych testów aktywności enzymatycznej. Tak zaprojektowany cykl badawczy jest żmudny, kosztochłonny i nie zawsze przynosi satysfakcjonujące efekty. Dlatego też kilka grup badawczych opracowało narzędzia bioinformatyczne do identyfikacji enzymów litycznych. Niestety obecnie wiele z tych programów nie jest dostępnych, a te które są, nie były uaktualnione od czasu opublikowania. Dodatkowo, większość modeli wykorzystywanych w tych narzędziach było trenowanych i testowanych na małych, nie zrównoważonych, powtarzalnych i/lub niewłaściwie opisanych zestawach danych. Ogranicza to wiarygodność oferowanych przez nie przewidywań a zatem ich użyteczność w procesie izolacji nowych lizyn.

Celem tej pracy doktorskiej było opracowanie nowej metody identyfikacji bakteriofagowych białek litycznych związanych ze ścianą komórkową z wykorzystaniem technik uczenia maszynowego. Konstrukcja takiego narzędzia wymagała dogłębnej analizy właściwości fizykochemicznych białek litycznych, zakonserwowanych domen związanych z lizynami i rodzin białkowych, które je grupują.

Aby skonstruować możliwie najdoskonalszy klasyfikator białek litycznych przetestowano zestaw modeli uczenia maszynowego i reprezentacji danych sekwencji. Najlepsza metoda opierała się na stosunkowo prostej sztucznej sieci neuronowej i połączeniu dwóch metod reprezentacji składu aminokwasów z wektorem właściwości fizykochemicznych całej sekwencji białkowej. Porównanie wytrenowanego modelu i obecnie dostępnych narzędzi do identyfikacji lizyn wykazało, że zaproponowane rozwiązanie przewyższa istniejące pod względem kilku kluczowych miar wiarygodności przewidywań, w tym miary F1 oraz czułości.

W trakcie realizacji projektu został także zebrany zestaw domen i rodzin związanych z lizynami. Wyniki przeprowadzonych eksperymentów wskazują, że wybrane domeny i rodziny umożliwiają skuteczniejszą identyfikację białek litycznych w porównaniu z wcześniej zebranymi zestawami modeli. Zastosowanie przygotowanego zestawu umożliwiło identyfikację białek litycznych z danych metagenomicznych pochodzących ze środowiska i od pacjentów. Dlatego też, zestaw ten może stanowić uzupełnienie opracowanego klasyfikatora uczenia maszynowego. Po ocenie *in silico* zarówno przewidywania klasyfikatora, jak i trafienia wybranych modeli zostały zweryfikowane *in vitro*. W tym celu wyprodukowano przewidziane białka lityczne i oceniono ich aktywność enzymatyczną za pomocą zymografii. W ten sposób udało się wyprodukować, oczyścić i potwierdzić aktywność lizyny związanej z wirionem, rozkładającej peptydoglikan *Proteus mirabilis*. Według mojej najlepszej wiedzy jest to pierwsze białko lityczne związane z wirionem zidentyfikowane dotychczas u bakteriofaga infekującego *Proteus mirabilis*, który jest istotnym patogenem oportunistycznym. Ponadto, wybrany zestaw modeli umożliwił identyfikację nowego enzymu w próbkach pochodzących od pacjentów. Enzym ten zabija bakterie z rodzaju *Rothia*. Ponadto, opracowane metody pozwoliły wyizolować z próbek metagenomicznych pochodzących od zwierząt lizyny, rozkładające ściany bakterii z rodzaju *Enterococcus*.

Podsumowując, efektem tej pracy jest opracowanie skutecznego algorytmu uczenia maszynowego, który umożliwi szybkie wyszukanie nowych bakteriofagowych białek litycznych. Białka te mogą znaleźć zastosowanie w medycynie i przemyśle, a w przyszłości mogą przyczynić się do rozwoju bioinformatyki genomów.