

## PROJEKT BAKTERIOFAG

Bakteriofagi **KP36**, **KP34**, **KP32**, sklasyfikowane do dwóch różnych rodzin (*Siphoviridae* i *Podoviridae*), są naturalnym wrogiem bakterii *Klebsiella pneumoniae* (pateczka zapalenia płuc). Nieznany jest mechanizm ich oddziaływania z polisacharydową otoczką szczepów bakteryjnych (numer serotypu otoczki K63 - dla fagów KP36 i KP34), który umożliwia bakteriofagowi jej uszkodzenie i przedostanie się do wnętrza bakterii. Rozpoznanie tego mechanizmu poszerzyłoby wiedzę o dodatkowe metody wykorzystania fagów i ich enzymów do zwalczania zakażeń bakteryjnych.

Za oddziaływanie z otoczką bakterii odpowiadają najprawdopodobniej białka zawarte w ogonkach bakteriofagów. Jednym z nich jest produkt genu **gp50** (GenBank, acc no YP\_009226010.1). Białko kodowane przez gen *gp50* zostało częściowo scharakteryzowane *in vitro*. Jednakże nieznana jest jego struktura ani sposób oddziaływania z molekułami otoczki bakteryjnej. Wstępne testy biologiczne wskazują, że białko które jest produktem tego genu wchodzi w interakcje z polisacharydami bakteryjnymi powodując ich depolimeryzację.

Należy przeprowadzić analizę bioinformatyczną, która przybliży wyjaśnienie zjawiska, uwzględniając poniższe etapy i stosując metody wprowadzone na wcześniejszych zajęciach [**lab x**].

1. Znaleźć 2 artykuły przeglądowe na temat bakteriofagów i ich enzymów/ badanych genów [**lab1**]
2. Znaleźć sekwencję białka, które jest produktem genu *gp50* badanego bakteriofaga KP36 (NCBI, EBI-EMBL, inne). Znaleźć w odpowiednich serwisach/bazach dodatkowe istotne informacje na temat tego genu i jego produktu [**lab2, lab3**].
3. Za pomocą odpowiednich narzędzi wyznaczyć zbiór białek homologicznych. Sprawdzić co to są za białka, z jakich organizmów, jakie pełnią funkcje. Rozszerzyć analizę wyznaczając białka wtórnie homologiczne (homologiczne do homologów pierwotnych). Spośród elementów tego zbioru wyznaczyć podzbiór bliskich homologów (biorąc pod uwagę wartość parametru *e*) [**lab5**]
4. Wykonać analizę wzbogacania zbioru genów z wykorzystaniem ontologii genów GO dla badanego genu i jego najlepszych homologów [**lab4**].
5. Za pomocą narzędzi do jakościowego (typu dot-plot) i ilościowego dopasowania par (PSA) wyznaczyć dopasowanie parami dla kilku (2-3) najwyższych homologów [**lab5, lab6**].
6. Za pomocą narzędzi do dopasowania wielosekwencyjnego MSA wyznaczyć logo rodziny białek i znaleźć obszary o wysokiej konserwacji. Przeanalizować wynik dla rozmaitych poziomów filtracji podobieństwa i kilku znacznie różniących się macierzy substytucji, wybrać najbardziej obiecujące MSA do dalszych analiz. Dla obszaru o najsilniejszej konserwacji przeprowadzić bardziej szczegółową analizę potencjalnych homologów [**lab7**].
7. Wyznaczyć drzewo filogenetyczne białek ze zbioru. Przeprowadzić analizę przynależności najlepszego zbioru białek do rodzin białkowych lub motywów strukturalnych (np. Pfam, PRINTS, Prosite, CATH-Gene3D, inne) [**lab8**].
8. Znaleźć homologi o znanej strukturze białka, zobrazować je za pomocą Pymola, przeanalizować potencjalne podobieństwa i różnice z poszukiwanym białkiem [**lab9**].
9. Zamodelować strukturę drugorzędową oraz trzeciorzędową poszukiwanego białka [**lab10**].
10. \* Zamodelować potencjalne oddziaływania uzyskanej struktury z polisacharydami otoczki bakteryjnej za pomocą narzędzi do dokowania.