

**„Analizy genomiczne jako podstawy szacowania potencjału metabolicznego
mikroorganizmów morskich”
mgr Michał Grabski**

Cyjanobakterie (sinice) to wszechobecna grupa mikroorganizmów zamieszkująca różnorodne środowiska. Od wód morskich, słonawych i słodkich, po siedliska lądowe, organizmy te rozwijają się w licznych biotach, wszystko dzięki ich ogromnym zdolnościom adaptacyjnym. Wiązanie azotu atmosferycznego i zdolność do przeprowadzania fotosyntezy w zbiornikach wodnych to tylko kilka z wielu cech umożliwiających tym tlenowym fotoautotrofom rozwój w warunkach niskiego poziomu składników odżywczych. Odporność sinic prowadzi do zjawisk bardzo szybkiego ich namnażania się, co powoduje tworzenie tzw. zakwitów, zmieniających skład mikroorganizmów w obrębie zajmowanego basenu. Stymulowana eutrofizacją i globalnym ociepleniem dominacja tych Gram-ujemnych prokariotów w wodach powierzchniowych jest częściowo zależna od produkcji metabolitów wtórnych. Związki te, pośrednicząc w interakcjach biotycznych i abiotycznych, nie tylko powodują zaburzenia w składzie mikroorganizmów, ale także powodują szkodliwe skutki zarówno dla organizmów wyższych, jak i niższych [Dittmann i in., 2012; Paerl i Otten 2013; Wang i in., 2021]. Właściwości te wzbudzają zainteresowanie naukowców, którzy poszukują niewykorzystanych dotąd zasobów substancji czynnych ukrytych w komórkach sinic.

Szerokie spektrum właściwości, w tym aktywność przeciwnowotworowa, przeciwdrobnoustrojowa lub przeciwwirusowa, koreluje ze strukturalną różnorodnością związków pochodzących z sinic. Metabolity wtórne zawierają wiele różnych grup cząsteczek, spośród których najbardziej znaczące są peptydy syntetyzowane bez udziału rybosomów [Calteau i in., 2014]. Te związki są wytwarzane przez modułowe enzymy, które tworzą struktury produktów zależne od zawartych w enzymach domen, dodając poszczególne aminokwasy do powstającej cząsteczki. Syntetazy peptydów nierybosomalnych (NRPS) i syntetazy poliketydów (PKS) katalizują biosyntezę cząsteczek z aminokwasów lub prekursorów acylo-CoA, wytwarzając odpowiednio peptydy i poliketydy [Walsh., 2004]. NRPS tworzą raczej produkty o małych cząsteczkach, ze względu na specyficzne procesy syntezy, w których każdy moduł, zbudowany z kilku domen, aktywuje i wiąże jeden konkretny aminokwas z powstającym peptydem. Minimalny moduł, który aktywuje i przenosi substrat do powstającego produktu peptydowego, składa się z domen adenylacji (A) i tiolacji (PCP). Wydłużenie łańcucha peptydu następuje poprzez tworzenie wiązań między dwoma tioestrami acylowymi, katalizowane przez domenę kondensacji (C). Sekwencyjne działania wyżej wymienionych domen podczas tworzenia produktu peptydowego kończą się na C-końcu

enzymu modułowego, gdzie domena tioesterazy (TE) uwalnia łańcuch albo przez hydrolizę, albo cyklizację [Marahiel i in., 1997]. Peptydy o wyższych masach cząsteczkowych są kodowane w genomie i podlegają syntezie rybosomalnej i modyfikacjom potranslacyjnym (RiPP). Geny prekursorowe kodujące tę klasę peptydów składają się z regionów determinujących N-końcową sekwencję liderową, która wiąże się z enzymami potranslacyjnie modyfikującymi peptydy w celu zmiany peptydu rdzeniowego na końcu C, który po przetworzeniu staje się dojrzałym związkiem [Oman i van der Donk., 2010].

Strukturalna złożoność poliketydów, związana z różnorodnością funkcjonalną, różni się w zależności od typu cząsteczki tworzącej syntetazę, klasyfikowanej ze względu na organizację strukturalną. Podobnie jak w enzymach NRPS, domeny PKS mogą być zorganizowane w sposób sekwencyjny, przesuując produkty między liniowym układem domen (typ I) lub mogą występować samodzielnie z minimalnym funkcjonalnym układem domen, składającym się z ketosyntetazy (KS), acylotransferazy (AT) i białka nośnikowego acylu (ACP) (typ II i III). Jednak podczas gdy typ III katalizuje iteracyjną kondensację w obrębie jednego enzymu, typ II działa podobnie jak syntetaza typu I, przesuując powstający produkt między modułami, które znajdują się w kilku enzymach. Ze względu na tę samą liniową organizację, domeny PKS mogą tworzyć hybrydowe enzymy z domenami NRPS, zwiększając różnorodność strukturalną tworzonych związków, poprzez włączenie zarówno jednostek acylowych, jak i aminoacylowych [Miyanaga i in., 2018].

W latach 2012-2022 rozwiązano struktury około 520 związków chemicznych pochodzących z sinic, co pokazuje, że sinice są obfitym źródłem biologicznie aktywnych produktów naturalnych [He i in., 2024]. Podejścia do eksploracji genomu okazały się jednak zawodne, ujawniając istnienie metabolitów wtórnych, podczas gdy identyfikowano geny nieulegające ekspresji lub obserwowano niewystarczającą wydajność produkcji danego związku chemicznego z powodu powolnego wzrostu sinic [Kehr i in., 2011; Calteau i in., 2014]. Sekwencjonowanie genomów może jednak pozwolić ustalić, czy badane czynniki są produktami rozpadu metabolizmu komórkowego, czy też docelowo wytwarzanymi związkami o fizjologicznej roli w organizmie.

Celem niniejszej rozprawy doktorskiej było określenie i analiza kompletnych sekwencji DNA genomów wybranych szczepów sinic, ze szczególnym uwzględnieniem genów kodujących enzymy zaangażowane w produkcję biologicznie aktywnych metabolitów. Przetestowano również nowatorskie podejście do wykorzystania analizy genomu w celu przewidywania syntezy związków przeciwwirusowych i przeciwnowotworowych.

Kompletny genom szczepu *Nostoc edaphicum* CCNP1411 został zsekwencjonowany i złożony *de novo*, ujawniając kolisty chromosom o 7 733 505 parach zasad (bp) i pięć kolistych plazmidów, co dało całkowity rozmiar genomu 8 316 316 bp [Artykuł nr 1]. Genom przeszukano pod kątem klastrów genów syntazy peptydów nierybosomalnych (NRPS), ujawniając cztery regiony w obrębie chromosomu, kodujące białka przypominające architekturę domeny NRPS. Sekwencje kodujące znalezione w badanym genomie wykazały podobieństwa do znanych klastrów genów biosyntezy peptydów, przypuszczalnie odpowiedzialnych za syntezę nostocyklopeptydów, cyjanopeptolin i anabaenopeptyn. Ilości nostocyklopeptydów zostały określone w surowym ekstrakcie *N. edaphicum* CCNP1411, co również przyczyniło się do wyjaśnienia struktury klastra genów nostocyklopeptydów, ujawniając dziewięć otwartych ramek odczytu (ORF), ułożonych podobnie do znanego systemu nostocyklopeptydów NRPS rozwiązanego przez Beckera i in. (2004). Dwa geny, *ncpA* i *ncpB*, kodują białka z powtarzalnymi modułami, które katalizują syntezę peptydów, obejmując domeny kondensacji (C), adenylacji (A) i białka nośnikowego peptydylu (PCP). Gen *ncpA* (11 334 pb) koduje trzy moduły, podczas gdy mający 14 157 bp gen *ncpB* koduje cztery moduły, gdzie domena C-końcowa obejmuje aktywność oksydoreduktazy, uwalniając peptyd z syntetazy. Specyficzność substratową domen adenylacji przewidywano przy użyciu kodu nadającego specyficzność, zaproponowanego przez Challis i Townsend (2000) oraz Stachelhaus i in. (1999), identyfikując substraty aminokwasowe, których sekwencja okazała się zgodna ze strukturami wykrytymi przez analizy LC-MS/MS. Analiza genomu *N. edaphicum* CCNP1411 podkreśliła potencjał tego szczepu do wytwarzania różnych peptydów nierybosomalnych.

Region genomu *N. edaphicum* CCNP1411 przypuszczalnie zawierający geny związane z produkcją anabaenopeptyn został poddany dalszym badaniom [Artykuł nr 2]. Główna struktura tego klastra genów znajduje się między pozycjami 2 265 881 bp i 2 288 626 bp na chromosomie, obejmując cztery geny: *aptA*, *aptB*, *aptC* i *aptD*. Stwierdziłem, że geny te kodują białka zawierające domeny adenylacji, zaangażowane w aktywację aminokwasów do syntezy peptydów. Gen *aptA* koduje białko z dwoma modułami, z dodatkową domeną epimeryzacji (E) na swoim C-końcu, odpowiedzialną za zmianę stereochemii adenylowanego aminokwasu. Gen *aptB* koduje białko z jednym modułem, podczas gdy *aptC* koduje dwa moduły z domeną metylotransferazy (M) znajdującą się między domenami A i PCP drugiego, kodowanego przez *aptC* fragmentu. Gen *aptD*, oprócz kodowania standardowych domen białkowych C, A i PCP, koduje również domenę tioesterazy na C-końcu białka, wymaganą do uwolnienia powstającego peptydu. Oprócz podobieństw w proponowanych aminokwasach aktywowanych przez domeny

adenylacji produktów wyżej wymienionych genów, konfiguracja D i metylacja w drugiej i piątej pozycji anabaenopeptyny były zgodne z wynikami analiz uzyskanymi zarówno przy użyciu metod LC-MS/MS, jak i genomicznych.

Podczas poszukiwań nowych leków pochodzących z sinic, *Pseudanabaena galeata* CCNP1313 wykazała silną aktywność ekstraktów i frakcji przeciwko komórkom nowotworowym i wirusom. Jednak nie określono żadnych substancji czynnych [Cegłowska M i in., 2022]. Dlatego w tej pracy zastosowano podejście „oddolne” w celu zbadania potencjału tego szczepu [Artykuł nr 3]. Określono sekwencję genomu *Pseudanabaena galeata* CCNP1313. Składa się ona z sześciu replikonów, w tym kolistego chromosomu (4 928 719 bp) i pięciu kolistych megaplazmidów, o całkowitej wielkości genomu 5 842 326 bp. Chociaż wcześniejsze badania wiązały aktywność biologiczną *P. galeata* CCNP1313 z peptydami, nie znaleziono rejonów w genomie kodujących w pełni funkcjonalne systemy NRPS, które podtrzymywałyby produkcję peptydów o sugerowanych długościach. Dwa regiony w chromosomie, kodujące domeny adenylacji, znajdujące się w dwóch oddzielnych *loci*, prawdopodobnie nie odpowiadają za produkcję peptydów, ponieważ domniemane reakcje kondensacji przeprowadzane przez syntetazy wykorzystują w syntezie produktu inne cząsteczki oprócz aminokwasów. Nietypowa reszta (seryna zamiast ewolucyjnie konserwowanego kwasu asparaginowego) w kieszeni wiążącej, kodowana przez gen pierwszego modułu syntazy peptydów nierybosomalnych (pozycja chromosomu 842 738 – 848 680 bp), może wiązać kwas hydroksylowy/karboksylowy zamiast aminokwasu, zatem domniemany produkt może nie zostać określony w ten sposób. Drugi klaster genów (pozycja chromosomu 1 530 477 – 1 548 363 bp) koduje mieszany hybrydowy system PKS/NRPS, prawdopodobnie produkujący hybrydy poliketydowo-aminokwasowe.

Odkryto również domniemany gen syntetazy lantipeptydowej, zlokalizowany w obrębie plazmidu (pPg_03 pozycja 9308 pb–12 523 bp), wraz z genem peptydu prekursorowego i genami powiązаныmi z mechanizmem wydzielniczym, co sugeruje potencjalną produkcję rybosomalnie syntetyzowanych i potranslacyjnie modyfikowanych peptydów (RiPP). Jednak brak reszt cysteiny w peptydzie prekursorowym wskazywał, że może on nie tworzyć lantioniny. Potencjalnie identyfikuje to nową klasę RiPP, znanych jako cyjanotyny. Wszystkie szlaki syntezy metabolitów wtórnych, odkryte na podstawie analizy genomu *P. galeata* CCNP1313, obejmowały białka, które wcześniej nie były adnotowane.

Podsumowując, analizy genomiczne szczepów sinic, przedstawione w tej rozprawie doktorskiej, wskazały na charakterystyczne cechy genów kodujących enzymy zaangażowane w syntezę związków biologicznie czynnych. Podejście do eksploracji genomu, opracowane w

tej pracy, wskazało nowy sposób poszukiwania związków o aktywnościach antyproliferacyjnych i przeciwwirusowych.

LITERATURA:

Becker JE, Moore RE, Moore BS. Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster: molecular basis for imine macrocyclization. *Gene*. 2004 Jan 21;325:35-42. doi: 10.1016/j.gene.2003.09.034. PMID: 14697508.

Cegłowska M, Szubert K, Grygier B, Lenart M, Plewka J, Milewska A, Lis K, Szczepański A, Chykunova Y, Barreto-Duran E, Pyrc K, Kosakowska A, Mazur-Marzec H. *Pseudanabaena galeata* CCNP1313-Biological Activity and Peptides Production. *Toxins (Basel)*. 2022 May 6;14(5):330. doi: 10.3390/toxins14050330. PMID: 35622577; PMCID: PMC9146944.

Challis GL, Ravel J, Townsend CA. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol*. 2000 Mar;7(3):211-24. doi: 10.1016/s1074-5521(00)00091-0. PMID: 10712928.

He, Y., Chen, Y., Tao, H. *et al.* Secondary metabolites from cyanobacteria: source, chemistry, bioactivities, biosynthesis and total synthesis. *Phytochem Rev* (2024). <https://doi.org/10.1007/s11101-024-09960-w>

Kehr JC, Gatte Picchi D, Dittmann E. Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes. *Beilstein J Org Chem*. 2011;7:1622-35. doi: 10.3762/bjoc.7.191. Epub 2011 Dec 5. PMID: 22238540; PMCID: PMC3252866.

Marahiel MA, Stachelhaus T, Mootz HD. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev*. 1997 Nov 10;97(7):2651-2674. doi: 10.1021/cr960029e. PMID: 11851476.

Miyanaga A, Kudo F, Eguchi T. Protein-protein interactions in polyketide synthase-nonribosomal peptide synthetase hybrid assembly lines. *Nat Prod Rep.* 2018 Nov 14;35(11):1185-1209. doi: 10.1039/c8np00022k. PMID: 30074030.

Oman TJ, van der Donk WA. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol.* 2010 Jan;6(1):9-18. doi: 10.1038/nchembio.286. PMID: 20016494; PMCID: PMC3799897.

Paerl HW, Otten TG. Harmful cyanobacterial blooms: causes, consequences, and controls. *Microb Ecol.* 2013 May;65(4):995-1010. doi: 10.1007/s00248-012-0159-y. Epub 2013 Jan 13. PMID: 23314096.

Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol.* 1999 Aug;6(8):493-505. doi: 10.1016/S1074-5521(99)80082-9. PMID: 10421756.

Walsh CT. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science.* 2004 Mar 19;303(5665):1805-10. doi: 10.1126/science.1094318. PMID: 15031493.

Artykuł nr 1

Fidor A., **Grabski M.**, Gawor J., Gromadka R., Węgrzyn G., Mazur-Marzec H. (2020). *Nostoc edaphicum* CCNP1411 from the Baltic Sea - A new producer of nostocyclopeptides. *Marine Drugs*, 18: 442.

Artykuł nr 2

Konkel R., **Grabski M.**, Cegłowska M., Wieczerek E., Węgrzyn G., Mazur-Marzec, H. (2022). Anabaenopeptins from *Nostoc edaphicum* CCNP1411. *International Journal of Environmental Research and Public Health*, 19: 12346.

Artykuł nr 3

Grabski M., Gawor J., Cegłowska M., Gromadka R., Mazur-Marzec H., Węgrzyn, G. (2024). Genome mining of *Pseudanabaena galeata* CCNP1313 indicates a new scope in the search for antiproliferative and antiviral agents. *Microorganisms*, 12: 1628.