“Genomic analyzes as the basis for estimating the metabolic potential of marine microorganisms”

Michał Grabski M.Sc.

Cyanobacteria is an ubiquitous group of microorganisms inhabiting variety of environments. From marine, brackish and fresh waters to terrestrial habitats, this phylum thrives in numerous biomes, all due to their vast adaptation abilities. Fixation of atmospheric nitrogen and ability to carry out photosynthesis in bodies of water are just few of many attributes enabling these oxygenic photoautotrophs to thrive under low nutrient conditions. Cyanobacterial resilience leads to unrestrained outbursts of their development, forming so-called blooms, shifting microbial community composition within a occupied basin. Fueled by eutrophication and global warming, intensity of domination over surface layer of these gram-negative prokaryotes is partially dependent on secondary metabolites production. These molecules, by mediating biotic and abiotic interactions, not only cause disturbances in the microbial community, but also cause harmful effects on both higher and lower organisms [Dittmann et al., 2012; Paerl and Otten 2013; Wang et al., 2021]. Those properties arouse the scientific interest regarding marine untapped resources of active agents concealed within cyanobacteria cells.

Various spectra of properties, including antitumor, antimicrobial or antiviral activities, correlate to the structural variety of cyanobacteria-derived compounds. Secondary metabolites harbor many divergent groups of molecules, among which the most prominent are peptides synthesized without involvement of ribosomes [Calteau et al., 2014]. Those compounds are produced by modular enzymes where structures of products depend on contained domains, adding components to nascent molecule. Non-ribosomal peptide synthetases (NRPSs) and polyketide synthetases (PKSs) conduct biosynthesis from either amino-acids or acyl-CoA precursors, producing peptides and polyketides, respectively [Walsh., 2004]. NRPSs assemble rather small-molecule products, due to the assembly line where each module, build from several domains, activates and binds one particular amino-acid to the nascent peptide. A minimal module that activates and carries substrate to the nascent peptide product consist of adenylation (A) and thiolation (PCP) domains. Chain elongation of the peptide occurs via formation of bonds between two acyl thioesters, catalyzed by the condensation domain (C). Sequential actions of the above mentioned domains on the peptide product finalized at the C-terminus of the modular enzyme where the thioesterase (TE) domain releases the chain either by hydrolysis or cyclization [Marahiel et al., 1997]. Structures of peptides of higher molecular masses are encoded within the genome and are liable to ribosomal synthesis and post-translational

modifications (RiPPs). Precursor genes encoding this class of peptides consist of regions determining the N-terminal leader sequence which binds to post-translational modification enzymes in order to modify the core peptide at the C-terminus, which becomes a mature compound after processing [Oman and van der Donk., 2010].

The structural complexity of poliketides, associated with a functional diversity, varies based on the type of the synthetase building molecule, classified due to the structural organization. Like in NRPSs enzymes, PKSs domains can be organized in a consecutive manner, shifting products between a linear arrangement of domains (type I), or can be found standalone with minimal functional arrangement of domains, consisting of ketosynthetase (KS), acyltransferase (AT) and acyl carrier protein (ACP) (type II and III). However, while type III catalyzes iterative condensation within one enzyme, type II functions alike a type I synthetase, shifting the nascent product in between modules, thus found in several enzymes. Due to the same assembly linear organization, PKS domains can make up hybrid enzymes with NRPSs domains, adding to structural variety of produced compounds by incorporating both acyl and aminoacyl units [Miyanaga et al., 2018].

Resolved structures of some 520 natural products derived from cyanobacteria were reported between 2012 and 2022, showing cyanobacteria as a prolific source of biologically active natural products [He et al., 2024]. Genome mining approaches came clutch revealing multitude of secondary metabolites, when encountering silent genes or insufficient yield of production of a chemical compound due to slow growth of cyanobacteria [Kehr et al., 2011; Calteau et al., 2014]. Gene sequencing can establish whether studied agents are breakdown products of cellular metabolism or deliberately produced compounds with physiological role in the source organism.

The aim of this PhD thesis was to determine and analyze complete DNA sequences of selected strains of cyanobacteria, with special attention to genes encoding enzymes involved in production of biologically active metabolites. The novel approach of the use of genome mining to predict the synthesis of antiviral and anticancer compounds has also been tested.

The complete genome of *Nostoc edaphicum* CCNP1411 has been sequenced and assembled *de novo* revealing a circular chromosome of 7,733,505 base pairs (bp) and five circular plasmids, resulting in a total genome size of 8,316,316 bp [**Article no. 1**]. The genome had been searched for non-ribosomal peptide synthetase (NRPS) gene clusters, revealing four spans resembling NRPS domain architecture within chromosome. Coding sequences found within studied genome expressed similarities to known peptide biosynthetic gene clusters putatively responsible for the synthesis of nostocyclopeptides, cyanopeptolins and

anabaenopeptins. Amounts of nostocyclopeptides were traced in crude extract of *N. edaphicum* CCNP1411, thus the structure of the nostocyclopeptide gene cluster was further elucidated, revealing nine open reading frames (ORFs), arranged similarly to known nostocyclopeptides NRPS system resolved by Becker et al. (2004). Two genes, *ncpA* and *ncpB*, encode proteins with repetitive modules that catalyze the synthesis of peptides, involving condensation (C), adenylation (A), and peptidyl carrier protein (PCP) domains. The *ncpA* gene (11,334 bp) encodes three modules, while the 14,157 bp-long *ncpB* encodes four modules, where the C-terminal domain encompass the oxidoreductase activity, releasing the peptide from the synthetase. Substrate specificity of the adenylation domains was predicted using a specificity-conferring code, proposed by Challis and Townsend (2000) and Stachelhaus et al. (1999), identifying amino acids substrates which sequence were found comaptible with structures detected by LC-MS/MS analyzes. The genomic resolution of *N. edaphicum* CCNP1411 highlighted the potential of this strain to produce various non-ribosomal peptides.

The suggested region of the *N. edaphicum* CCNP1411 genome, putatively containing genes for production of anabaenopeptins, was further studied [**Article no. 2**]. The core structure of this gene cluster is located between positions 2,265,881 bp and 2,288,626 bp on the chromosome, encompassing four genes: *aptA*, *aptB*, *aptC* and *aptD*. These genes were found to encode proteins containing adenylation domains, involved in the activation of amino-acids for peptide synthesis. The *aptA* gene encodes a protein with two modules, with an additional epimerisation (E) domain at its C-terminus responsible for changing stereochemistry of adenylated amino acid. The *aptB* gene encodes a protein with one module, while *aptC* encodes two modules with a methyltransferase (M) domain found between A and PCP domains of the second, *aptC*-encoded fragment. The *aptD* gene, besides coding for standard C, A and PCP protein domains, encodes also a thioesterase domain at the C-terminus of the protein, required for releasing the nascent peptide. Besides similarities in proposed amino acids activated by adenylation domains of products derived from above mentioned genes, D-configuration and methylation in second and fifth positions of the anabaenopeptin was compatible with results obtained using both LC-MS/MS and genomic methods.

In search for novel drugs derived from cyanobacteria, *Pseudanabaena galeata* CCNP1313 had shown a potent activity of extracts and fractions against cancer cells and viruses. However no active agents had been determined [Cegłowska M et al., 2022], thus a "bottom-up" approach was employed to explore the potency of this strain [**Article no. 3**]. The genome sequence of *Pseudanabaena galeata* CCNP1313 has been determined. It consists of

six replicons, including a circular chromosome (4,928,719 bp) and five circular megaplasmids, with a total genome size of 5,842,326 bp. Although previous studies linked the biological activity of *P. galeata* CCNP1313 to peptides, no fully functional NRPS systems were found to sustain production of peptides with suggested lengths. Two regions in the chromosome, encoding adenylation domains found within two separate spans, are unlikely to be responsible to produce peptides, as putative condensation reactions conducted by synthetases engage other molecules besides amino acids in the product synthesis. Unusual residue (serine instead of an evolutionary conserved aspartic acid) within the binding pocket, encoded by the gene for the first module of the non-ribosomal peptide synthase (chromosome position 842,738 – 848,680 bp), may bind hydroxy/carboxy acid instead of amino acid, thus the product may not be elucidated. The second gene cluster (chromosome position 1,530,477 – 1,548,363 bp), is a mixed PKS/NRPS hybrid system, possibly producing polyketide-amino acid hybrids.

A putative lanthipeptide synthetase gene within plasmid spans (pPg_03 position 9308 bp–12,523 bp) was also discovered, along with a gene for a precursor peptide, and associated secretion machinery, suggesting the potential production of ribosomally synthesized and post-translationally modified peptides (RiPPs). However, the lack of cysteine residues in the precursor peptide indicated that it may not form lanthionine. This potentially identifies a novel class of RiPPs, known as cyanotins. All secondary metabolite synthesis routes, discovered on the basis of the analysis of the *P. galeata* CCNP1313 genome, involved proteins which were not annotated before.

In summary, genomic analyses of cyanobacterial strains, presented in this PhD thesis, indicated characteristic features of genes encoding enzymes involved in synthesis of biologically active compounds. The genome mining approach, invented in this work, indicated a new scope in the search for antiproliferative and antiviral agents.

REFERENCES:

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, Pyrć 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.

ARTICLES INCLUDED IN THE PhD THESIS

Article no. 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.

Article no. 2

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

Article no. 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.