Project closing report

Comprehensive investigation of bacterial lipopeptides

Funding ID: OTKA-K 128659

Prepared by András Szekeres (University of Szeged) and Balázs Leitgeb (HUN-REN Biological Research Centre, Szeged)

During our project, about one hundred *Bacillus* strains were selected for our studies from the Szeged Microbiological collection as well as approximately 300 strains were isolated from various origin including soil, mushroom compost, corn-, pea-, tomato-, paprika-, carrot- and sweet potato rhizosphere. During the isolation procedure the colonies were selected based on the morphological characters and were completely purified. The *Bacillus* strains were identified firstly with the MIDI Sherlock Microbial Identification System (Microbial ID Inc., Newark, NJ, USA) based on the gas chromatographic profile of the extracted total cellular fatty acids. After the complete method setup, the identifications of the isolates were carried out in five replicates using the RTSBA6 library in connection of the Instant FAME Method. Samples with a similarity index (SI) ≥ 0.7 were considered as an acceptable identification. The isolated strains are belonged to the *B. agaradhaerens*, *B. alcalophilus*, *B. amyloliquefaciens*, *B. atrophaeus*, *B. cereus*, *B. licheniformis*, *B. megaterium*, *B. pumilus* (4), *B. simplex*, *B. subtilis* and *B. velezensis* species. For the isolates, whose the identity were questionable the molecular biological identifications were also applied using multi-gene approach. These molecular biological tools involved the 16S RNA, *gyrA*, *rpoB* and *recQ* genes, because it is difficult to identify the *Bacillus* species only via phylogenetic analysis of the generally applied16S rRNA gene sequence.

During the taxonomical examinations of the isolated *Bacillus* strains, the problems with the identification of *B. velezensis* and *B. amyloliquefaciens* isolates were occurred. These species are closely-related members of the "operational group *B. amyloliquefaciens*", which considered as a taxonomical unit above species level within the *Bacillus* genus. Separating these two taxa from one another have proven to be difficult to implement and could not push easily into the line of routine analyses, thus the goal of our examinations was to determine, whether the whole FAME profiling could be used to distinguish among these two species using both type strains and environmental isolates. Initially, the classification was confirmed by the partial sequences of the *gyrA* and *rpoB* genes and the classified *B. velezensis* and *B. amyloliquefaciens* field isolates and the type strains were considered as samples to develop

the identification method based on FAME profiles. The content of FAs was revealed, and the features were constructed from the analysis of sixteen *B. velezensis* strains (n = 3) and two *B. amyloliquefaciens* strains (n = 25). The 15:0 iso, 15:0 anteiso, 16:0, 17:0 iso and 17:0 anteiso have been primary FA components in both taxa. Especially, the FA 16:0, 17:0 iso and 17:0 have drawn a distinction between *B. velezensis* and *B. amyloliquefaciens*. The 2-D Plot built from PC 1 and PC 2 of Principal Component Analysis showed a separation of two taxa in n-dimensional space. The SI value, generated from calculations of distance in multi-dimensional space, illustrated the relation between analysing FA profiles and the mean FAs of library's database as its match. In our case, all identified samples exhibited the high matches with SI > 0.5 and well-SI-separations (> 0.1) confirming that the method is reliable with high confidence.

Based on our results so far, the FA-based identification used as a biomarker can be considered as a credible differentiating factor for *Bacillus* species that works even within closely related groups. Additionally, the available library in the MIS have contained altogether 40 Bacillus species, but numerous species were not included in it. During our work, based on the FA profiles of the molecularly identified isolates we constructed a new extended library, which contains the new *Bacillus* entries [Microorganisms 2022, 10(2), 418., Molecules 2023, 28(3), 1172].

Examinations of the produced surfactins from the ferment broth extracts of both culture collection elements and the environmental isolates were also carried out. An improved HPLC gradient elution method was applied capable to separate all lipopeptide components in our samples and to examine the separated, previously hidden, more non-polar fractions in order to identify the various surfactin molecules, with different amino acid sequences and aliphatic acid chain lengths to the full extent. When the structures of these compounds with greater masses were examined by MS2 analyses, isoforms of surfactin molecules containing aspartic acid 4-methyl ester (AME) in their fifth peptide position were recognized, also involving the newly described types ([Lxx4,AME5] and [AME5,Val7]). After the study of the MS2 spectra validated our suggestions regarding the amino acid sequence, we calculated and compared the ratios of the produced surfactin isoforms. The results of these semi-quantitative comparisons proved the presence of these variants to be remarkable amount. In the case of the highest m/z values (at m/z 1100 and 1114), only these new isoforms possessing AME5 could be detected and the formation of C18-[AME5] had a yield close to 100% [Molecules 2018, 23(9), 2224.].

The screening of the strains was also carried out regarding their surfactin production involving the recently known surfactin variants using an optimized triple quadrupole based multiple reaction monitoring method differentiating 158 surfactin variants from ferment broths. The yield of the total surfactins ranged in a wide concentration level, from 0.1 μg/ml to 65 μg/ml and the surfactin profiles were remarkable different for the producer strains. In the case of *Bacillus* strains isolated from vegetable rhizospheres, about half part of the isolates produced surfactins in average concentrations of 0.5 – 6.6 g/l. As a result of our qualitative measurements, a total number of 29 surfactin molecules were identified with 157 detected instances with different retention times suggesting numerous variations of branching within their fatty acid chains apart from alterations in their chain lengths and amino acid sequences. Comparing the relative amounts of the different surfactin variants and homologues, our data showed that the [Sur], [AME5] and [Val7] isoforms were the most dominant in all cases, while regarding the occurrence of surfactins with different fatty acid chain lengths, the C14 – C16 molecules had the largest area ratios [Molecules 2023, 28(3), 1172].

Since the newest members of these lipopeptides were described by us recently, there was no information that is available on their characteristic features, e.g., the dependence of their production from different cultivation parameters. Therefore, the effects of both the different carbon sources and various metal ions on the surfactin production of a selected strain were also investigated. The cultivation parameters of *B. subtilis* SZMC 6179J strain were modified by changing the carbon source of the culture medium from glucose to cellobiose, ethanol, starch, maltose, mannitol, fructose, sucrose, glycerol, or xylose, as well as by supplementing the original culture medium with manganese, copper, or nickel ions, to examine the differences regarding the production of surfactin molecules, and to compare their relative amounts to the ones in the sample from the original glucose medium. The detected surfactins were examined by their relative ratios of different variants, grouped by both their peptide sequence and the various chain lengths of the homologues. It was found that the use of carbon sources other than glucose had effects on the ratios of the produced surfactins, showing selectivity in certain cases, while the application of metal ions had a larger impact and enabled the discovery of surfactin variants possessing AME5 in their fifth amino acid position, which was the most dominant group in the extracted fermenting broth from the metal ion-supplemented media. Furthermore, these metal ions promoted the production of compounds having longer fatty acid chains: two thirds of the detected molecules were C16, C17, or C18 homologues [Molecules 2018, 23(10), 2675].

Within the topic of the purification as well as biological and chemical activity of the surfactin variants the best producer strains were selected for the fermentation and purification studies. For this purpose, a multi-step purification and separation process was developed to isolate surfactins from other contaminants found in the crude extract of the ferment broth and also to separate the different variants and homologues of this lipopeptide family. The method incorporates normal phase flash chromatography for pre-purifying the crude extract and two consecutive reverse phase high performance liquid chromatographic techniques; one with a preparative column to further cleanse the sample from contaminants, which was followed by a semi-preparative RP-HPLC for the isolation of the various surfactin molecules. The measurement of the relative amounts of lipopeptides in the crude extract and in each fraction of every step was carried out by HPLC-HESI-MS examinations, as well as the identification of the different surfactin variants detected in the fractions of the semi-preparative HPLC separation. For the *Bacillus subtilis* GBB64 strain, the measured weight of the crude extract was 492.25 mg and peak area ratio of surfactins in it was 21.35%. After the preparative flash chromatographic separation, fractions 4 – 7 were merged together and kept for further purification. The weight of the dry matter found in these fractions was 143.71 mg and the relative amount of surfactins was observed to be 30.44%. Fractions 9 – 13 of the preparative HPLC purification step were collected together, their dry matter weighting 30.18 mg altogether, while the integrated peak area of surfactins were 85.39% that of the total area of peaks detected in this sample, meaning that more than 98% of impurities detected in the crude extract were removed by these separation steps, however, about 75% of surfactins were also lost during the process. Examination of the fractions of the semi-preparative HPLC isolation technique showed that 9 different surfactin variants were isolated and identified in total, out of which three molecules were found to be completely purified, and three others were detected in relative amounts of more than 95% in some fractions. Furthermore, for the determination of the structural elements, the GC-MS and a HPLC-UV method were developed using standard compounds for the analysis of fatty acid- and peptide parts of surfactins, respectively. For the measurements of the fatty acids, methylation reaction was applied for the derivatization, while the amino acids were derivatised with o-phthaldialdehyde and 9-fluorenylmethyloxycarbonyl chloride in the first stage of our work, which was later supplemented with the derivatizations with Marfey's reagent.

In the in vitro antagonistic assays, the antagonisms of *Bacillus* isolates, were examined against phytopathogenic bacteria (*Pseudomonas syringae* SZMC 16160, *Erwinia amylovora* SZMC 21402, *Erwinia carotovora* SZMC 6190, *Xanthomonas campestris* SZMC 6182,

Agrobacteria tumefaciens SZMC 14554) from Szeged Microbiological Collection. The clearance zones due to bacterial growth inhibitions showed potentially antagonistic activities of *Bacillus* isolates. As observed, numerous isolates exhibited biocontrol properties on phytopathogenic bacteria with inhibition zones ranging from 1.0 to 16.67 mm. *B. velezensis* strains were considered to have substantial potential of biocontrol antagonizing against all testing pathogens. Remarkably, *E. amylovora* and *X. campestris* were significantly antagonized by a wide range of isolates. In addition, isolates slightly antagonized *A. tumefaciens*, *P. syringae* and *E. carotovora* with lesser activities [Molecules 2023, 28(3), 1172].

Altogether, in our papers and presentations the complete theoretical background and practical characteristics of the mass spectrometric measurement of surfactins were described deeply, which serve a useful guide to researchers for the comprehensive surfactin analysis. Applying the knowledge acquired during the mass spectrometric analysis and preparative purification of surfactins the methods were used also for the examination of other secondary metabolites [Toxins 2023, 15(2), 134; Toxins 2023, 15(3), 178].

For various lipopeptides, parameterization was performed by means of quantum chemical calculations, in order to calculate accurately the partial atomic charges. In the course of parameterization procedure, the restrained electrostatic potential method was used, to obtain the necessary atomic charges. Applying the calculated partial atomic charges, as well as assigning the missing force field parameters, preliminary theoretical calculations were carried out on lipopeptides, to assess the suitability of atomic charges and other parameters for the molecular modelling simulations. Based on the results derived from the preliminary calculations, it could be concluded that the applied parameterization procedure, as well as the calculated partial atomic charges proved to be suitable to study the three-dimensional structure and dynamic behaviour of lipopeptides by theoretical methods.

Various theoretical calculations were performed for different lipopeptides, such as surfactins, iturins, fengycins, fusaricidins, lichenysins and plipastatins. To investigate the structural properties of these lipopeptides, simulated annealing, simulated annealing-molecular dynamics and molecular dynamics (MD) calculations were carried out. Based on these simulations, a comprehensive conformational analysis and structural characterization were performed. The Φ-Ψ conformational spaces of amino acids were explored in detail, in the course of which Ramachandran and three-dimensional Ramachandran plots were constructed. Applying the latter type of Ramachandran plots, the conformational distributions were compared to one another, and conformational similarity indices were calculated, with

regard to the lipopeptides with various length of fatty acyl chains, as well as to the different isoforms. These results revealed that nor the length of fatty acyl chains, neither the type of isoforms produced significant effects on the conformational distributions of amino acids. To characterize the χ conformational spaces, the preferred rotamer states were determined in the case of the side-chains of amino acids. For the lipopeptides, the presence of different secondary structural elements (*i.e.* regular and inverse γ-turns, various types of β-turns), were examined along the entire sequences of peptides. The results led to the observations that types I, I', II, II', III and III' β-turns could be identified in certain tetrapeptide units of lipopeptides, as well as regular and inverse γ-turns could be found in certain tripeptide segments of molecules. Furthermore, the occurrence of stabilizing intramolecular H-bonds was investigated. Based on these results, it could be concluded that the $i \leftarrow i+3$ and $i \leftarrow i+2$ H-bonds, stabilizing the above-mentioned turn structures, appeared in the appropriate tetra- and tripeptide units of lipopeptides, respectively. Moreover, other intramolecular H-bonds, formed between the backbone atoms, as well as between the backbone and side-chain atoms, were observed, which contributed also to the structural stability of molecules. For the lipopeptides, cluster analyses were carried out, based on the backbone heavy atoms, and the conformationally related subfamilies, as well as their representative structures were identified.

In order to investigate the dynamic behaviour of above-mentioned lipopeptides, MD simulations were performed, as follows: *(1)* MD calculations started from different initial conformations; *(2)* MD calculations started from same conformational state using random initial velocities. On the basis of MD trajectories, the alterations of backbone and side-chain conformations, as well as the rearrangements of intramolecular H-bonds were examined as a function of time. The conformational interconversions between the characteristic structures were studied, and the structural stability of different conformational states was investigated. Applying the MD trajectories, RMSD values were calculated to further characterize the dynamic behaviour of lipopeptides, and additionally, RMSF values were calculated to study the fluctuations of the heavy atoms of backbone.

Further MD simulations were performed for the surfactins, iturins, fengycins, fusaricidins and lichenysins. These MD calculations were carried out applying explicit solvents, such as water, methanol, chloroform and dimethyl-sulfoxide, in order to study the effects of the various types of solvents on the characteristic structural features of lipopeptides. The results indicated that the different environments produced effects on the conformational properties of above-mentioned lipopeptides. These effects could be observed for the

secondary structural elements, rotamer states and intramolecular H-bonds, as well as for the RMSD and RMSF values.

Applying MD calculations, for the surfactins and iturins, the aggregation processes were examined, in order to characterize the association of these lipopeptides. In the course of these studies, it was investigated how these lipopeptides aggregate with each other, including dimer and other multimer aggregates.

Structure-activity relationship studies were also carried out for lipopeptides, in order to identify the possible relationships between the structural features and the bioactivity of molecules. The determination of these relationships was very difficult and time-consuming task, thus so far we have not succeeded in identifying structure-activity relationships in detail for lipopeptides.

The micelle-bound conformations were investigated for surfactins, iturins and fengycins. To characterize the micelle-bound conformations of lipopeptides, MD calculations were carried out on peptide-micelle systems. Based on the MD trajectories, the interactions between the lipopeptides and micelles were characterized by different structural features, *i.e.*, secondary structures, as well as intra- and intermolecular interplays. For the micelle-bound conformations of lipopeptides, the occurring secondary structural elements (*i.e.* γ-turns, β-turns and multiple turn structures), as well as other non-conventional structural motifs were determined, and the preferred rotamer states of the side-chains of amino acids were identified. Additionally, the presence of intramolecular H-bonds stabilizing the micelle-bound conformations of lipopeptides was investigated, as well as the appearance of intermolecular interactions formed between the lipopeptides and lipid molecules were studied. Based on the results, the micelle-bound conformations, as well as the interactions between the lipopeptides and micelles were characterized comprehensively for surfactins, iturins and fengycins.

To examine the interactions of surfactins, iturins and fengycins with membranes, as well as to investigate the mode of action of these molecules, MD calculations were performed on peptide-membrane ensembles. Based on the MD trajectories, the conformational changes and interconversions of molecules upon binding to the membranes were examined by monitoring of various structural features, as well as the different orientations of lipopeptides at the surface of membranes were studied. Moreover, the structural stability was investigated, not only for the lipopeptides, but also for the lipopeptide-membrane ensembles.

During our project, the results were presented altogether in 18 conference abstracts and in 10 journal articles as well as the research activities were also involved

into the education of BSc, MSc and PhD students resulting 6 diploma theses and 2 PhD dissertation.

Szeged, 29 April 2024