Content is avaliable at SCOPUS

Meat Technology — Special Issue 66/3

www.meatcon.rs ▪ www.journalmeattechnology.com

UDK: 663/664:579.864

ID: 179278089

https://doi.org/10.18485/meattech.2025.66.3.22

Original Scientific Paper

Bioinformatic discovery of novel collagen-binding aggregation factors in lactic acid bacteria

Emilija Đukanović[1], Luka Dragačević[2], Milan Kojić[2] and Darya Tsibulskaya[2]*

[1] *University of Belgrade. Faculty of Technology and Metallurgy, Department of Biochemical Engineering and Biotechnology, Karnegijeva 4, 11120 Belgrade, Serbia*

[2] *Institute of Virology, Vaccines and Sera, Torlak, Vojvode Stepe 458, 11221 Belgrade, Serbia*

ARTICLE INFO

Keywords:
Lactic acid bacteria
Aggregation Factors
Bacterial adhesins

ABSTRACT

Lactic acid bacteria (LABs) are a unique group of microorganisms found in the diet of nearly all humans and form an integral part of the healthy human microbiome. Some LAB strains exhibit a remarkable ability for autoaggregation, mediated by Snowflake Forming Collagen Binding Aggregation Factors (SFCBAFs)—a fascinating group of proteins described so far only in LABs. To date, only five SFCBAFs have been characterized in detail: AggL from *Lactococcus lactis*, AggE from *Enterococcus faecium*, AggLb from *Lacticaseibacillus paracasei*, AggLr from *Lactococcus raffinolactis*, and AggA from *Tetragenococcus halophilus*. In this study, we present bioinformatically predicted novel SFCBAF candidates and demonstrate their widespread distribution among LAB species. Furthermore, we provide evidence that such proteins may not be exclusive to LABs, as homologous sequences were also identified in phylogenetically distant bacteria such as *Staphylococcus aureus*, *Oceanobacillus* spp., *Bacillus* spp., and others, expanding our understanding of this unique protein family.

1. Introduction

Lactic acid bacteria (LAB) represent a unique group of microorganisms that are found in foods, such as meat products (*Carneiro et al.*, 2024), consumed by nearly every human on the planet and are also present in the microbiome of healthy individuals. In nature, LAB are not limited to food products—they can also be found in soil, on plants, and on or within both terrestrial and marine organisms. Thus, this group of bacteria is truly ubiquitous and is represented across nearly all ecological niches (*Akpoghelie et al*., 2025).

One of the remarkable properties of certain LAB strains is autoaggregation, ability to adhere among themselves or to other strains, forming visible aggregates *(Trunk, S. Khalil, & C. Leo*, 2018). There is substantial evidence suggesting that aggregation is closely linked to biofilm formation, adhesion, colonization, and host physiological functions *(Burgain et al., 2014; Du et al*., 2022; *Kragh et al*., 2016; *Miljkovic et al*., 2015).

Among wide range of different aggregation factors, there is a distinct group that is referred to as Snowflake Forming Collagen Binding Aggregation Factors (SFCBAFs). These are large proteins that contain adhesion domains in the first part of the protein, followed by an immunoglobulin-like domain and then repeats domains. They are exported from the cell to the surface and subsequently

***Corresponding author:** Darya Tsibulskaya, dtsibulskaya@torlak.rs

Paper received Jun 1st 2025. Paper accepted Jun 7th 2025.
The paper was presented at the 63rd International Meat Industry Conference "Food for Thought: Innovations in Food and Nutrition" – Zlatibor, October 05th-08th 2025.
Published by Institute of Meat Hygiene and Technology – Belgrade, Serbia.
This is an open access article CC BY licence (http://creativecommons.org/licences/by/4.0)

covalently anchored to the cell wall through a well-characterized mechanism involving the LPXTG cell wall anchor domain (*Siegel, Reardon, & Ton-That*, 2017). While the aggregation mechanism is still under investigation, it is believed that the adhesive properties are conferred by collagen-binding and immunoglobulin-like domains, whereas the repeat regions serve to spatially project the protein away from the cell surface (*Miljkovic et al*., 2016). At present, only five representatives of this group have been well characterized: AggL from *Lactococcus lactis*, AggE from *Enterococcus faecium*, AggLb from *Lacticaseibacillus paracasei*, AggLr from *Lactococcus raffinolactis*, and AggA from *Tetragenococcus halophilus (Endo et al., 2023; Kojic et al., 2011; Miljkovic et al., 2018, 2015; Veljović et al., 2017)*. In this study, we predict the presence of novel proteins from this group and perform their analysis.

2. Materials and methods

2.1. Selection of representatives for phylogenetic tree construction

To retrieve protein sequences for analysis, we used the local alignment tool BLASTP 2.16.0. Searches were conducted against NCBI databases using amino acid sequences of known SFCBAFs as queries. (*Sayers et al*., 2022). The presence of signal peptides in the retrieved proteins was predicted using SignalP 6.0 (*Teufel et al.*, 2022). Duplicate sequences were removed prior to further analysis.

2.2. Multiple sequence alignment and phylogenetic tree construction

Multiple sequence alignment of the protein sequences was performed using Clustal Omega, which employs hidden Markov models (HMMs) for improved accuracy (*Madeira et al*., 2024). The resulting phylogenetic tree was saved in Newick format (*Junier & Zdobnov*, 2010) and used for downstream clustering analysis.

2.3. Clustering of the phylogenetic tree

To assess diversity and select representatives from distinct evolutionary lineages, we calculated pairwise evolutionary distances between all taxa based on the phylogenetic tree generated by Clustal Omega. Using the resulting distance matrix, we applied KMeans clustering (Steinley, 2006), specifying ten clusters. From each cluster, we selected the most distant taxon from the cluster centre, i.e., the taxon showing the greatest divergence from the cluster's average profile. This approach allowed us to compile a set of phylogenetically diverse representatives for further analysis.

2.4. Domain structure analysis

To analyse the domain architecture of the proteins corresponding to the selected taxa, we used the online tool InterPro (*Paysan-Lafosse et al.*, 2023).

3. Results and discussion

3.1. Phylogenetic analysis of potential SFCBAF representatives

In order to assess the distribution of aggregation factors, we performed BLASTP 2.16.0 searches using sequences of previously described SFCBAF-type factors (AggL, AggE, AggLb, AggLr, and AggA) against the NCBI protein database (Sayers et al., 2022). All predicted protein sequences were subsequently screened for the presence of signal peptides required for protein cell export using SignalP 6.0, and sequences lacking a signal peptide were excluded. The remaining candidates were then analysed for the presence of LPxTG anchor motifs, taking into account known motif variants as described previously (*Malik et al.*, 2023) and our internally obtained data. As a result, we assembled a dataset of 246 proteins, including the previously described ones, each of which had less than 100% sequence identity to any other in the set. A phylogenetic tree was constructed based on Clustal Omega alignment (results shown in Supplementary). During our research we identified novel SFCBAF-like factors in non-LAB bacteria, including *Mycoplasma sp. P36-A1*, *Oceanobacillus* spp., *Virgibacillus* spp., *Gracilibacillus* spp., *Jeotgalicoccus halotolerans*, *Aliicoccus persicus*, *Pseudogracilibacillus* spp., *Ornithinibacillus* sp., *Bacillus* spp., *Amphibacillus* sp., *Atopostipes* spp., *Irregularibacter muris*, *Eubacterium nodatum*, *Corticicoccus populi*, *Coprococcus* spp., *Staphylococcus agnetis*, *Staphylococcus aureus*, *Mammaliicoccus stepanovicii*, and *Lentibacillus sp. JNUCC-1*. Interestingly, aside from *Mycoplasma sp. P36-A1*, all other bacteria were Gram-positive, which may indirectly suggest a conserved mechanism of cell surface anchoring for these proteins.

3.2. Domain structure analysis of potential SFCBAF representatives

To understand the diversity of newly predicted SFCBAF representatives, we clustered the phylogenetic tree into 10 groups based on pairwise distances between its leaves and selected one most "central" representative from each cluster. In addition, we included the longest (5511 amino acids) and the shortest (1063 amino acids) SFCBAF representatives. We also re-verified the presence of leader and anchor sequences. The list of the selected representatives is provided in Table 1.

Table 1. Selected representatives of SFCBAF

ID	Number of amino acids	Leader peptide	LPxTG anchor motifs
*Oenococcus oeni*_ WP_032822309.1	1875	MNKWFRRLSAVLMALVISLQYVAAGA	LPNTG
*Lacticaseibacillus paracasei*_WP_211757376.1	1963	MNKRKLGQQLYSLVLIILLFLQSSAQVVGA	MPNTG
*Lactococcus raffinolactis*_ WP_412921610.1	1667	MKKLWQFFALIILLMQSVSPGIVVA	LPQTS
*Lactococcus lactis*_ WP_058224032.1	1793	MEVFLKKHMRKIVSIFFILLQFIQPITAIA	LPKTG
*Streptococcus pacificus*_ WP_199575997.1	1063	MSKSIKLFFKSTLIAIFAIIGMIIASANTVEA	LPKTG
*Enterococcus faecium*_ WP_002335623.1	1918	MNRKKIDQFEGLKRVMPLIMLIVLLLQSVFSPVISVA	LPTKE
*Oceanobacillus sojae*_ WP_260053073.1	2523	MKKRAKRVSIFMIFLLVMQSFATGFAPIQTAYA	LPGTA
*Irregularibacter muris*_ WP_257533483.1	1700	MNRCKGKSRYNIALIVVLIMLFQMIIPSTLAVA	LPKTG
Candidatus *Enterococcus huntleyi*_WP_161843597.1	5511	MERLRKVLTIIGLLGILANIMPVNAFA	LPKTG
*Mammaliicoccus stepanovicii*_ WP_095088808.1	1458	MGKKLVVFSVFMLLLNLFSPFVNGEKVFA	LPQTG
Lactococcus lactis BGKP1_ AggL	1768	MEKKSRYATKFYVVLMMLSLVSQLFMPFLQVAA	LPATG
Enterococcus faecium BGGO9-28_AggE	1636	MENKSRYATKFYVVLMMLSLVSQLFVPVLQVAA	LPATG
Lacticaseibacillus paracasei BGNJ1-64_AggLb	2997	MNKKKIGQQIYSLVLIFLLFLQSSAQVVGA	MPNTG
Lactococcus raffinolactis BGTRK10-1_AggLr	1774	MKKLSKSSIFILMAVIILLQYVSPILA	LPKTS
*Tetragenococcus halophilus*_AggA	2298	MTFNHVKKVAMVFMLVVLVFQSFVSPLSAVA	LPKTG
*Enterococcus italicus*_ WP_251870029.1	2822	MENKSRYATKFYVVLMMLSLVSQLFVPVLQVEA	LPATG
*Tetragenococcus halophilus*_ WP_317915830.1	3346	MTFNHVKKVAMVFMLVVLVFQSFVSPLSAVA	LPKTG

We aligned the selected representatives once again relative to each other (Figure 1, A) and performed domain structure analysis using InterPro (Figure 1, B). All potential representatives showed a similar structural organisation to previously described SFCBAF-type proteins: following the leader peptide, all contained adhesion domains (InterPro ID: IPR008966) with collagen-binding regions. The bacterial adhesion domain consists of a β-sandwich formed by 9 β-strands arranged in 2 layers with a Greek-key topology, representing a subclass of the immunoglobulin-like fold. Such domains are commonly found in surface-associated proteins and are essential for bacterial adhesion to host cell surfaces (*Symersky et al*., 1997). The number of adhesion domains varied from 1 (in the longest protein) to 7 (in several representatives). These domains are followed by an immunoglobulin (Ig)-like domain (InterPro ID: IPR013783) characterized by a prealbumin-like fold. Ig-like domains are among the most widely distributed protein modules found in diverse organisms. They frequently mediate interactions, often through β-sheet pairing with other Ig-like domains. Ig-like folds are present not only in immunoglobulins but also in a variety of other proteins, including T-cell antigen receptors, cell adhesion molecules, MHC class I and II antigens, muscle proteins such as titin, and others*(Bork et al., 1994; Halaby et al.*, 1999; *Potapov et al.*, 2004; *Teichmann & Chothia*, 2000). We hypothesize that adhesion domains, together with Ig-like domain, are crucial for aggregation. After the Ig-like domain, two organizational variants are observed: in one group, the Ig-like domain is repeated 6 to 19 times (excluding the first domain); in the other group, the repeats consist of a Cna protein B-type domain (InterPro ID: IPR008454). This domain has been well-characterized in the Cna protein of *Staphylococcus aureus*. Most probably if forms an extended stalk structure that positions the ligand-binding domain away from the bacterial surface. Cna is a collagen-binding MSCRAMM (Microbial Surface Component Recognizing Adhesive Matrix Molecules) and is necessary for *S. aureus* cells to bind to cartilage (*Foster & Höök*, 1998; *Shimoji et al*., 2003). The number of such repeats among the selected representatives ranges from 3 to 50. Finally, only one SFCBAF representative, *Enterococcus faecium* WP_002335623.1, displays a hybrid structure: after the initial Ig-like domain, there is another Ig-like domain followed by six Cna protein B-type domains. At the C-terminus of each protein, an LPxTG anchor motif is present, although in some cases, it deviates from the canonical form. For example, in *Lactobacillus paracasei* WP_211757376.1 and *L. paracasei* BGNJ1-64 AggLb, it is represented as MPNTG, while in *E. faecium* WP_002335623.1, it appears as LPTKE.

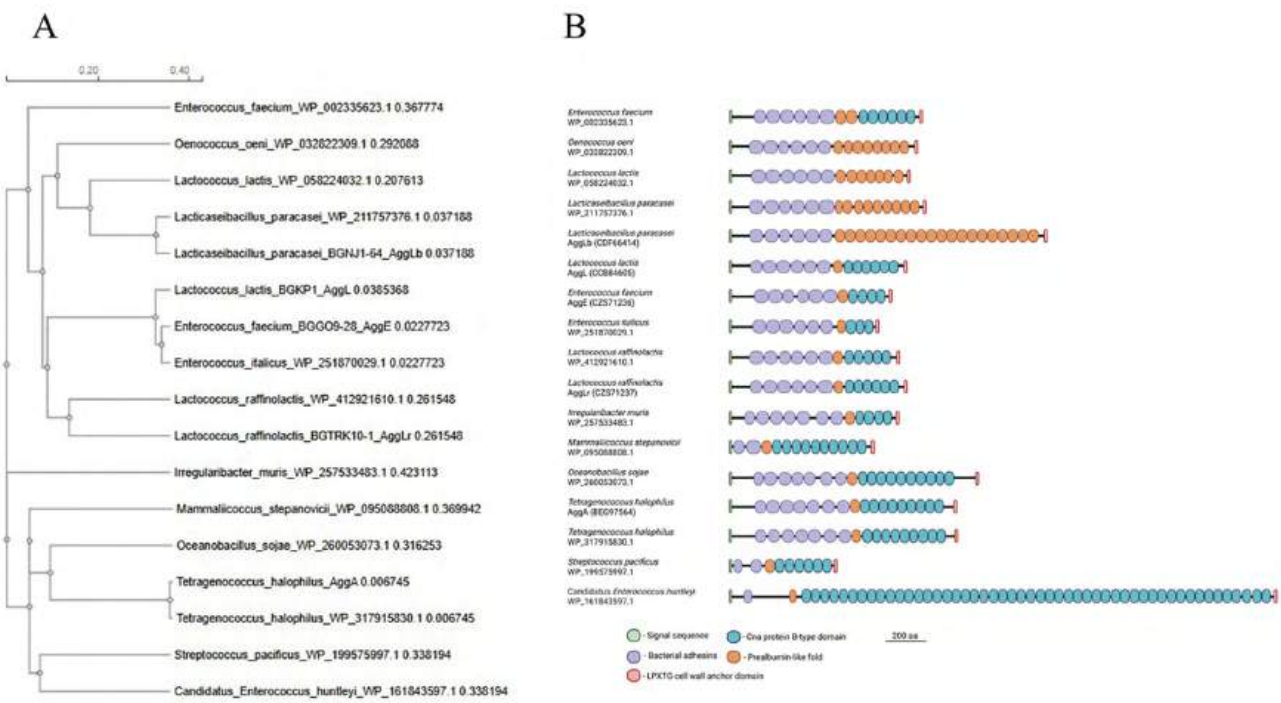

Figure 1. Analysis of SFCBAF representatives. **A** – Phylogenetic tree constructed based on Clustal Omega alignment of the selected SFCBAF representatives. **B** – Domain organisation of SFCBAF representatives.

4. Conclusion

Lactic acid bacteria represent a unique group of microorganisms closely associated with the food industry, making them of particular interest for research. The aggregation of lactic acid bacteria gives them a distinctive phenotype and new properties, including potential adhesion to the human gastrointestinal tract and the displacement of other pathogenic microorganisms. In this study, we demonstrated the prevalence of SFCBAF-type aggregation factors. Investigating such bacteria may help improve the properties of LAB used in the food industry, including in the meat industry.

Disclosure statement: No potential conflict of interest was reported by the authors.

Funding: This research was funded by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, Contract number 451-03-66/2024-03/200177.

Acknowledgements: The authors would like to thank the Ministry of Science, Technological Development and Innovation of the Republic of Serbia.

References

Akpoghelie, P. O., Edo, G. I., Ali, A. B. M., Yousif, E., Zainulabdeen, K., Owheruo, J. O., … Alamiery, A. A. (2025). Lactic acid bacteria: Nature, characterization, mode of action, products and applications. *Process Biochemistry*, 152, 1–28. https://doi.org/10.1016/J.PROCBIO.2025.02.010

Bork, P., Holm, L., & Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *Journal of Molecular Biology*, 242(4), 309–320. https://doi.org/10.1006/JMBI.1994.1582

Burgain, J., Scher, J., Francius, G., Borges, F., Corgneau, M., Revol-Junelles, A. M., … Gaiani, C. (2014). Lactic acid bacteria in dairy food: Surface characterization and interactions with food matrix components. *Advances in Colloid and Interface Science*, 213, 21–35. https://doi.org/10.1016/J.CIS.2014.09.005

Carneiro, K. O., Campos, G. Z., Scafuro Lima, J. M., Rocha, R. da S., Vaz-Velho, M., & Todorov, S. D. (2024). The Role of Lactic Acid Bacteria in Meat Products, Not Just as Starter Cultures. *Foods,* 13(9), 3170. https://doi.org/10.3390/FOODS13193170

Du, Y., Li, H., Shao, J., Wu, T., Xu, W. L., Hu, X., & Chen, J. (2022). Adhesion and Colonization of the Probiotic Lactobacillus plantarum HC-2 in the Intestine of Litopenaeus Vannamei Are Associated With Bacterial Surface Proteins. *Frontiers in Microbiology*, 13. https://doi.org/10.3389/FMICB.2022.878874/PDF

Endo, R., Hotta, S., Wakinaka, T., Mogi, Y., & Watanabe, J. (2023). Identification of an operon and its regulator required for autoaggregation in Tetragenococcus halophilus. *Applied and Environmental Microbiology*, 89(12). https://doi.org/10.1128/AEM.01458-23

Foster, T. J., & Höök, M. (1998). Surface protein adhesins of Staphylococcus aureus. *Trends in Microbiology*, 6(12), 484–488. https://doi.org/10.1016/S0966-842X(98)01400-0

Halaby, D. M., Poupon, A., & Mornon, J. P. (1999). The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Engineering*, 12(7), 563–571. https://doi.org/10.1093/PROTEIN/12.7.563

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics*, 26(13), 1669. https://doi.org/10.1093/BIOINFORMATICS/BTQ243

Kojic, M., Jovcic, B., Strahinic, I., Begovic, J., Lozo, J., Veljovic, K., & Topisirovic, L. (2011). Cloning and expression of a novel lactococcal aggregation factor from Lactococcus lactis subsp. lactis BGKP1. *BMC Microbiology*, 11. https://doi.org/10.1186/1471-2180-11-265

Kragh, K. N., Hutchison, J. B., Melaugh, G., Rodesney, C., Roberts, A. E. L., Irie, Y., … Bjarnsholt, T. (2016). Role of multicellular aggregates in biofilm formation. *MBio*, 7(2). https://doi.org/10.1128/MBIO.00237-16

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. N., … Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research*, 52(W1), W521–W525. https://doi.org/10.1093/NAR/GKAE241

Malik, A., Shoombuatong, W., Kim, C. B., & Manavalan, B. (2023). GPApred: The first computational predictor for identifying proteins with LPXTG-like motif using sequence-based optimal features. *International Journal of Biological Macromolecules*, 229, 529–538. https://doi.org/10.1016/J.IJBIOMAC.2022.12.315

Miljkovic, M., Bertani, I., Fira, D., Jovcic, B., Novovic, K., Venturi, V., & Kojic, M. (2016). Shortening of the Lactobacillus paracasei subsp. paracasei BGNJ1-64 AggLb Protein Switches Its Activity from Auto-aggregation to Biofilm Formation. *Frontiers in Microbiology*, *7*(SEP). https://doi.org/10.3389/FMICB.2016.01422

Miljkovic, M., Marinkovic, P., Novovic, K., Jovcic, B., Terzic-Vidojevic, A., & Kojic, M. (2018). AggLr, a novel aggregation factor in Lactococcus raffinolactis BGTRK10-1: its role in surface adhesion. *Biofouling*, 34(6), 685–698. https://doi.org/10.1080/08927014.2018.1481956

Miljkovic, M., Strahinic, I., Tolinacki, M., Zivkovic, M., Kojic, S., Golic, N., & Kojic, M. (2015). AggLb Is the Largest Cell-Aggregation Factor from Lactobacillus paracasei Subsp. paracasei BGNJ1-64, Functions in Collagen Adhesion,

and Pathogen Exclusion In Vitro. *PloS One*, 10(5). https://doi.org/10.1371/JOURNAL.PONE.0126387

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, 51(D1), D418–D427. https://doi.org/10.1093/NAR/GKAC993

Potapov, V., Sobolev, V., Edelman, M., Kister, A., & Gelfand, I. (2004). Protein--protein recognition: juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins. *Journal of Molecular Biology*, 342(2), 665–679. https://doi.org/10.1016/J.JMB.2004.06.072

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 50(D1), D20–D26. https://doi.org/10.1093/NAR/GKAB1112

Shimoji, Y., Ogawa, Y., Osaki, M., Kabeya, H., Maruyama, S., Mikami, T., & Sekizaki, T. (2003). Adhesive surface proteins of Erysipelothrix rhusiopathiae bind to polystyrene, fibronectin, and type I and IV collagens. *Journal of Bacteriology*, 185(9), 2739–2748. https://doi.org/10.1128/JB.185.9.2739-2748.2003

Siegel, S. D., Reardon, M. E., & Ton-That, H. (2017). Anchoring of LPXTG-like proteins to the gram-positive cell wall envelope. *Current Topics in Microbiology and Immunology*, 404, 159–175. https://doi.org/10.1007/82_2016_8,

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), 1–34. https://doi.org/10.1348/000711005X48266

Symersky, J., Patti, J. M., Carson, M., House-Pompeo, K., Teale, M., Moore, D., … Narayana, S. V. L. (1997). Structure of the collagen-binding domain from a Staphylococcus aureus adhesin. *Nature Structural Biology*, 4(10), 833–838. https://doi.org/10.1038/NSB1097-833,

Teichmann, S. A., & Chothia, C. (2000). Immunoglobulin superfamily proteins in Caenorhabditis elegans. *Journal of Molecular Biology*, 296(5), 1367–1383. https://doi.org/10.1006/JMBI.1999.3497

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., … Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, 40(7), 1023–1025. https://doi.org/10.1038/S41587-021-01156-3

Trunk, T., S. Khalil, H., & C. Leo, J. (2018). Bacterial autoaggregation. *AIMS Microbiology*, 4(1), 140–164. https://doi.org/10.3934/MICROBIOL.2018.1.140

Veljović, K., Popović, N., Miljković, M., Tolinački, M., Terzić-Vidojević, A., & Kojić, M. (2017). Novel Aggregation Promoting Factor AggE Contributes to the Probiotic Properties of Enterococcus faecium BGGO9-28. *Frontiers in Microbiology*, 8(SEP). https://doi.org/10.3389/FMICB.2017.01843

Authors info

Emilija Đukanović, https://orcid.org/0009-0004-6198-2500
Luka Dragačević, https://orcid.org/0000-0003-1662-0493
Milan Kojić, https://orcid.org/0000-0001-5645-750X
Darya Tsibulskaya, https://orcid.org/0000-0001-9145-7097