DOI: 10.25205/978-5-4437-1691-6-8

COMPARATIVE ANALYSIS OF AMINO ACID SEQUENCES OF S8 PEPTIDASE FROM *BACILLUS* SPP.

G. D. Milovidov, T. S. Tikhomirova, V. A. Nemashkalov

Institute for Biological Instrumentation RAS, Pushchino

✉ Knorih@gmail.com

Abstract

Serine proteases are widely used in various fields, including the food industry and medicine. Modern enzymes have high activity and broad substrate specificity, the characteristics of which are influenced by both external conditions and intramolecular features. This study examines the amino acid sequences of proteases belonging to the S8 family from *Bacillus* spp. to evaluate the intramolecular determinants affecting their enzymatic properties.

Members of the genus *Bacillus* are producers of a large number of enzymes with applications across various human activities [1–3]. Among these enzymes, a notable group includes serine proteases, also known as subtilisins. These hydrolases exhibit broad substrate specificity and are used in the production of functional foods, dietary supplements, cosmetics, and household chemicals. Furthermore, the fibrinolytic activity of microbial serine proteases, such as nattokinase, An6 fibrinase (BAF1), CFR 15 protease, and bacillokinase II, has demonstrated potential for thrombolytic therapy of cardiovascular diseases by targeting fibrin degradation [3].

Most serine proteases consist of a peptidase domain belonging to the S8 family and a signal peptide I9 [4, 5]. The I9 domain, consisting of at least 100 amino acids, acts as an intramolecular chaperone, facilitating the secretion of mature protease. This domain also exhibits transient inhibitory activity, which diminishes as it is cleaved by the associated protease. The presence of the I9 domain is crucial for maintaining high protease activity. In contrast, the peptidase domain S8, composed of 250–280 amino acids, is responsible for the primary proteolytic activity by a highly conserved active center, containing Asp, His, and Ser residues.

To assess the conservation level of S8 peptidases among *Bacillus* spp., a comparative analysis of amino acid sequences was conducted. A pairwise alignment of 2045 sequences of S8 peptidases from *Bacillus licheniformis*, *Bacillus subtilis*, and *Lederbergia lenta* (obtained from the NCBI Protein database https://www.ncbi.nlm.nih.gov/protein/) was performed using ClustalOmega software. Percent identity was then calculated using Python v.3.6.12. The results of this analysis revealed seven distinct clusters among the available sequences (Fig. 1a).

As can be seen (see Fig. 1a), the sequences of S8 peptidases for these strains are not conserved, which may affect the degree of their activity. However, the average amino acid composition for all S8 peptidases used in the analysis

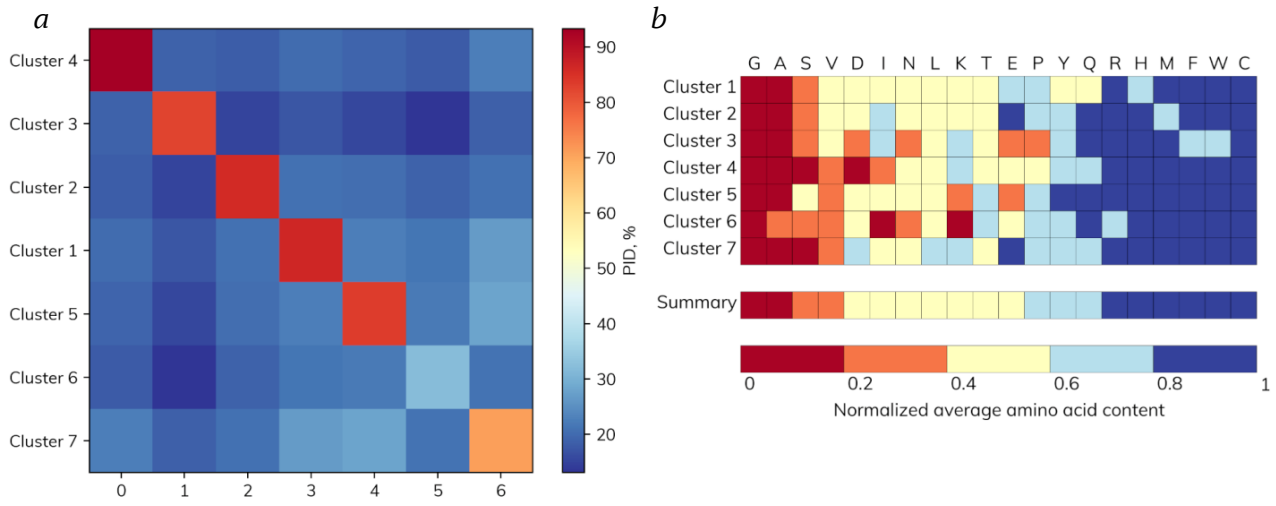

Fig. 1. Comparative analysis of 2045 amino acid sequences of S8 peptidases from *Bacillus licheniformis*, *Bacillus subtilis*, and *Lederbergia lenta* strains: *a* — symmetric matrix of average percent identity (PID) values; *b* — average amino acid composition. Normalization was carried out using the min-max method

was similar (Fig. 1b). S8 peptidases from *Bacillus* species are characterized by a high abundance of Gly and Ala residues, while Arg, His, Met, Phe, Trp, and Cys residues are virtually absent. Using the AlphaFold2 tool, tertiary structures were predicted based on consensus sequences for these seven distinct clusters of S8 peptidases (Fig. 2).

The overlay of the predicted tertiary structures for the seven S8 peptidase clusters reveals a highly conserved spatial organization of the enzyme core and active site, with the exception of disordered regions. Despite significant sequence variability, this indicates a high degree of conservation in both amino acid composition (see Fig. 2b) and intramolecular structure among serine proteases from *Bacillus* species. This structural consistency likely plays a critical role in defining the proteolytic activity of these enzymes.

The obtained results will be used both for further in-depth bioinformatics analysis and for identifying important sequence features responsible for the increased activity of proteases from *Bacillus* spp.

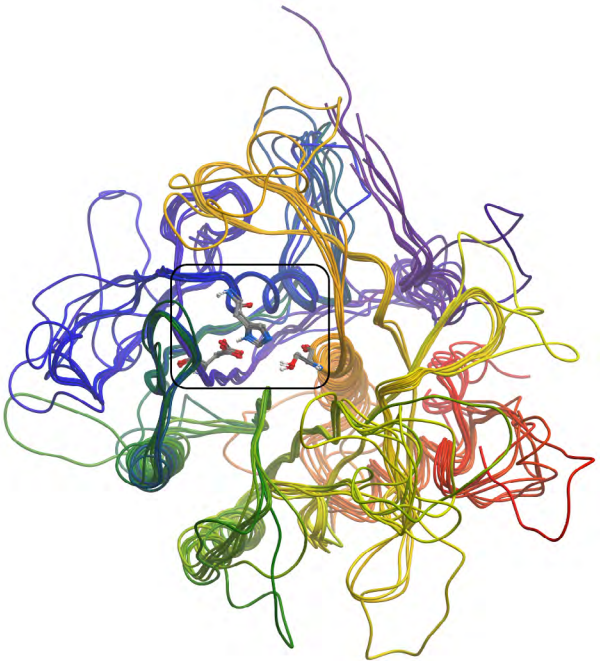

Fig. 2. Comparison of predicted tertiary structures of consensus S8 peptidase sequences from seven clusters

References

1. Gu J., Qiu Q., Yu Y. et al. Bacterial transformation of lignin: key enzymes and high-value products // Biotechnol. Biofuels Bioprod. 2024. Vol. 17. P. 2.

2. Sheng Y., Yang J., Wang C. et al. Microbial nattokinase: from synthesis to potential application // Food Funct. The Royal Society of Chemistry. 2023. Vol. 14, No. 6. P. 2568–2585.

3. Singh R., Gautam P., Sharma C., Osmolovskiy A. Fibrin and fibrinolytic enzyme cascade in thrombosis: Unravelling the role // Life. 2023. Vol. 13, No. 11. P. 2196.

4. Kojima S., Minagawa T., Miura K. The propeptide of subtilisin BPN' as a temporary inhibitor and effect of an amino acid replacement on its inhibitory activity // FEBS Lett. 1997. Vol. 411, No. 1. P. 128–132.

5. Li Y., Hu Z., Jordan F., Inouye M. Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding. Refolding and inhibitory abilities of propeptide mutants // J. Biol. Chem. 1995. Vol. 270, No. 42. P. 25127–25132.