



UvA-DARE (Digital Academic Repository)

Bacillus cereus spore and cell proteome dynamics

Gao, X.

Publication date
2022

[Link to publication](#)

Citation for published version (APA):

Gao, X. (2022). *Bacillus cereus spore and cell proteome dynamics*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 6

Discussion

Bacterial spores are resistant structures capable of survival under unfavorable conditions. In this thesis, *B. cereus* spores have been studied using proteomics. The specificity of the inner membrane protein composition is revealed in Chapter 2 in comparison with the protein composition of the vegetative cell membrane. The spore revival is studied by monitoring protein dynamics during spore germination and outgrowth in Chapter 3. We demonstrate that the spore protein composition can be affected by intracellular perturbations of protein concentrations in Chapter 4. Further, we explored the application of a new sample preparation method on *B. cereus* spores and vegetative cells, trying to improve the depth and width of protein identification and quantification in Chapter 5. However, there are still remaining challenges in spore research. In this chapter, challenges are discussed regarding the quantification method deployed and spore germination heterogeneity observed. In addition, emerging technologies or potential tools that can be used in spore research are also included.

6.1. Proteomics data acquisition approaches

Proteomics data obtained in this thesis is measured using label-free quantification, which has proven to be simple, low-cost and highly reproducible. However, compared with metabolic labeling methods, the label-free quantification is less accurate. The metabolic labeling methods that are frequently used include ^{15}N labeling and SILAC, which have been tried on *B. cereus* spores (unpublished data). Usually, *B. cereus* spores are generated in CDGS medium. To get ^{15}N labeled spores, vegetative cells need to finish sporulation in a medium where all nitrogen sources are replaced with ^{15}N labeled. CDGS medium contains several amino acids and it would be too expensive to replace all these amino acids with ^{15}N labeled ones. We therefore designed a new strategy where we tried to harvest fully ^{15}N labeled spores as follows: a single colony of *Saccharomyces cerevisiae* is cultured in a medium where the only nitrogen source is $^{15}\text{NH}_4\text{Cl}$ and transferred to fresh medium for several times until the yeast cells are fully labeled. Then the yeast cells are lysed and freeze dried to obtain fully ^{15}N labeled yeast extract which is used in a yeast-extract-based sporulation medium to generate fully ^{15}N labeled *B. cereus* spores. Although eventually we did obtain fully ^{15}N labeled *B. cereus* spores, the yield of ^{15}N yeast extract was too low to support further experimentation. The SILAC approach was also tried. It has been reported that high labeling efficiency in gram-positive *Staphylococcus aureus* and gram-negative *E. coli* can be achieved by adding high concentrations of isotope-labeled lysine and arginine together with

naturally occurring proline [1]. However, this strategy could not facilitate full incorporation of amino acids into *B. cereus* spores. In the future, chemical labeling methods, like iTRAQ and TMT, can be tried to facilitate *B. cereus* proteomics studies.

6.2. Proteomics data-analysis tools

Even though mass spectrometry is powerful and MS-based proteomics has been developing at an ever-increasing pace, the analysis of the data derived from the extreme complexity of the proteome is still challenging. In order to move this type of analysis forward more and more use is made of the capabilities of artificial intelligence (AI). Deep learning, a part of machine learning and AI, is changing the world. Different from classical statistics, deep learning based on artificial neural multi-layer networks can automatically learn complex features and patterns of the data without making assumptions about the underlying distributions and models, such as linearity and normality. Deep learning can achieve unprecedented accuracy by computational power and learning from large datasets. Deep learning could be a turnaround for the analysis of complex and multi-dimensional proteomics data. The architectures and developed DL-based software have been reviewed in detail [5,6]. Deep learning can be applied in the proteomics workflow to enhance data quality and in proteomics data analysis. The most well-known network architectures utilized in MS-based proteomics can generally be classified into deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), which differ in their number of neurons in each layer, and number of layers and types of connections between layers (Figure 6.1). Deep learning has demonstrated superior performance in retention time (RT) prediction, MS/MS spectrum prediction, de novo peptide sequencing, post-translational modification (PTM) prediction, major histocompatibility complex (MHC)-peptide binding prediction, and protein structure prediction [19]. Deep learning has been demonstrated to be more powerful than machine learning in fragmentation spectra prediction and retention time prediction improving peptide identification, which can also be used to build in silico spectral libraries for Data Independent Acquisition (DIA) data analysis. More accurate MS/MS spectra prediction can improve protein identification in Data Dependent Acquisition (DDA) data analysis by search engine DeepMass:Prism [7] or by post-processing tools such as Percolator [8].

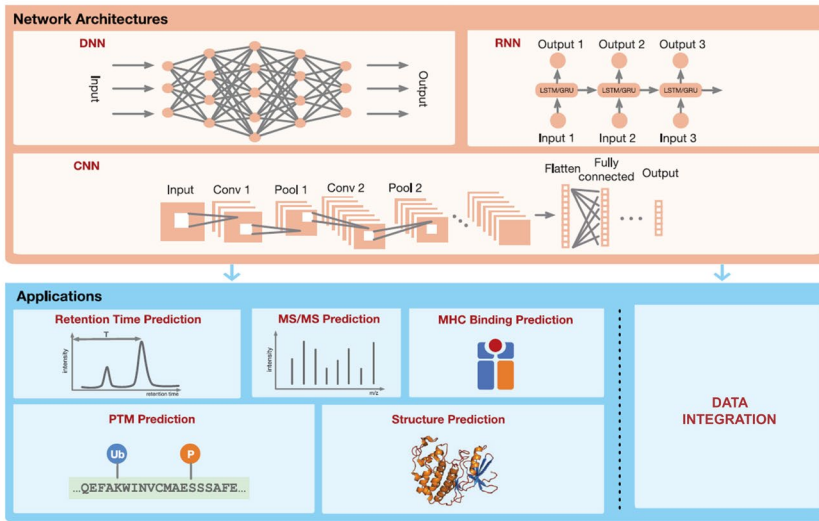


Figure 6.1. Overview of the network architectures of deep learning and its applications in proteomics. The most well-known network architectures utilized in MS-based proteomics can generally be classified into deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs). Deep learning has been mainly applied on retention time (RT) prediction, MS/MS spectrum prediction, de novo peptide sequencing, post-translational modification (PTM) prediction, major histocompatibility complex (MHC)-peptide binding prediction, protein structure prediction and multi-omics data integration. Adapted and modified from [5].

6.3. Spore germination heterogeneity and data-acquisition

In Chapter 3, protein dynamics during spore germination and outgrowth are monitored to explore the mechanism of germination. There are two points that can be optimized in the future to get a more comprehensive view of spore germination. One is to deal with the heterogeneity of spore germination where single-cell study is needed. Studies on single cell level are usually launched by flow cytometry and fluorescence microscopy. It has been reported that flow cytometry combined with the fluorescent reporter dye SYTO9, which only stains double-stranded DNA or RNA, can be applied to germinated *B. cereus* spores to evaluate germination responses at single spore levels [9]. Other research reported a similar strategy used on *B. cereus* spores to recognize phenotypes at different stages during germination and outgrowth under sorbic acid stress [10]. Antibody based immunofluorescence confocal microscopy has also been used to study single cell behaviour. Absolute protein quantification methods, such as single-cell Western Blot, can break the sensitivity limitation of assays using microfluidics [2], by which reactions can be run at nanoliter scale. The most

advanced device is the single-cell barcode chip (SCBC) [3]. There is a demand for multiplexed protein measurement and quantification which means simultaneously measuring multiple proteins in the same cell. Generally, the measurement of multiplexed proteins in a single cell is antibody-based. Mass cytometry is a broadly used multiplexed single-cell protein measurement method, which is similar to flow cytometry but analyzed by mass spectrometry rather than optically [4]. This method has been already commercialized under the name CyTOF (Fluidigm, CA, USA) and has been applied to many biological problems. But the limitation is the availability of desired antibodies, and target accessibility including creating new antibodies or the selection of antibodies which prevents unbiased results. The best way for bypassing an antibody is MS-based proteomics.

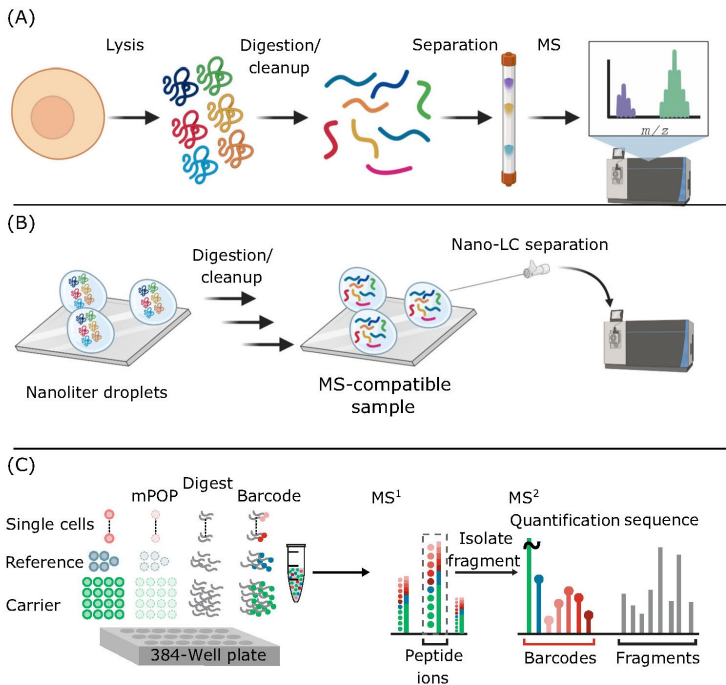


Figure 6.2. MS-based Single-Cell proteomics methods. (A) MS-based proteomics analysis steps. (B) The nanoPOTS (nanodroplet processing in one pot for trace samples) platforms miniaturize MS-ready sample preparation into 200-nl droplets and integrate all steps onto a microwell plate. (C) The SCoPE-MS (single-cell ProtEomics by mass spectrometry) platforms incorporate tandem mass tags (TMTs) and a carrier proteome to improve throughput and enable the quantification of peptides from individual cells. Abbreviation: nano-LC, nano-liquid chromatography; mPOP, minimal Proteomic sample preparation; MS¹, the first MS analysis; MS², the second MS analysis. Adapted from [2].

The bottleneck for MS-based single-cell proteomics is the miniscule amount of proteins in a single cell. When using traditional proteomics methods

sufficient materials which can come from biological systems or large cells, like human oocytes, are needed for analysis. The limitation of MS-based single-cell proteomics is the sensitivity in assay and detection. To perform single-cell proteomics, modifications are needed in sample preparation, LC peptide separation and MS analysis data collection. Currently there is not a state-of-the-art method that is suitable for all universal research aims, as each single-cell proteomics method is dependent on its particular purpose. Two platforms (Figure 6.2) are promising for single-cell MS-based proteomics: the nanodroplet processing in one pot for trace samples (nanoPOTS) platform [15] and the single-cell ProtEomics by mass spectrometry (SCoPE-MS) platform which is the first MS-based single-cell proteomics analysis of common laboratory cell lines [16,17].

6.4. Further challenges

To get a comprehensive understanding of spore germination, research on only the protein level is far from sufficient. Thus the phosphoproteome can also be analyzed, since phosphorylation is the most common PTM in signal transduction and widespread in germinating *B. subtilis* spores [15]. The first deep-learning framework MusiteDeep for predicting general and kinase-specific phosphorylation sites has been reported to achieve an improvement in the prediction of phosphorylation sites [18]. Metabolomics is another aspect to reveal germination and outgrowth mechanisms as phenotypes are intrinsically linked to the metabolites involved. Untargeted global metabolomics analysis can decipher spore germination and outgrowth and reveal information regarding metabolism. There are questions about the protein interactions in germinosomes, between coat proteins, between germinants and germination receptors and so on. Over the past few decades, although several different methods have been developed to analyze protein interactions, they are only focused on a small proteome subset with a relatively low throughput. The strategies developed for high-throughput detection of stable interactions and protein assemblies have been reviewed [13]. To date, five main MS-based technologies are widely used: yeast two-hybrid screens (Y2H screens) and related complementation assays, affinity purification mass spectrometry (AP-MS), proximity labelling approaches exemplified by BioID112 and APEX113, cross-linking mass spectrometry and protein co-fractionation coupled to mass spectrometry (CoFrac-MS).

Omics data sets, including genomics, transcriptomics, proteomics and metabolomics etc., allow researchers to report a greater level of data that can aid the understanding of the functional requirements for spores to survive and revive. Multi-omics integration analysis could provide

information on spore germination from several regulatory levels of this crucial developmental biology process. Deep learning can be applied to the integration of such multi-omics data with approaches used in for instance cancer research as an example [11]. Finally new software is constantly being developed and deployed. R scripts became available during the last five years for multi-omics data analysis under the name of mixOmics [12]. Clearly there is more to follow and such analysis tools will be instrumental together with appropriate experimental design to make sure that the current data deluge, also in the area of spore biology, continues to go beyond the gathering of heaps of data and will more and more lead to true new knowledge of the developmental biological processes involved.

Reference

1. Han, J., Yi, S., Zhao, X., Zheng, Y., Yang, D., Du, G., ... & Sun, X. Improved SILAC method for double labeling of bacterial proteome. *Journal of proteomics*, 2019, 194, 89-98.
2. Vistain, L. F., & Tay, S. Single-cell proteomics. *Trends in Biochemical Sciences*, 2021.
3. Ma, C. et al. A clinical microchip for evaluation of single immune cells reveals high functional heterogeneity in phenotypically similar T cells. *Nat. Med.* 17, 2011, 738–743
4. Bendall, S.C. et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 332, 2011, 687–696
5. Wen, B., Zeng, W. F., Liao, Y., Shi, Z., Savage, S. R., Jiang, W., & Zhang, B. (2020). Deep learning in proteomics. *Proteomics*, 20(21-22), 1900335
6. Meyer JG. Deep learning neural network tools for proteomics. *Cell Reports Methods*. 2021 May 17:100003.
7. Tiwary S, Levy R, Gutenbrunner P, Soto FS, Palaniappan KK, Deming L, Berndl M, Brant A, Cimermanic P, Cox J. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nature methods*. 2019 Jun;16(6):519-25.
8. Gessulat S, Schmidt T, Zolg DP, Samaras P, Schnatbaum K, Zerweck J, Knaute T, Rechenberger J, Delanghe B, Huhmer A, Reimer U. ProSIT: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature methods*. 2019 Jun;16(6):509-18.
9. Warda AK, Tempelaars MH, Abee T, Nierop Groot MN. Recovery of heat treated *Bacillus cereus* spores is affected by matrix composition and factors with putative functions in damage repair. *Frontiers in microbiology*. 2016 Jul 18;7:1096.
10. Van Melis CC, Groot MN, Tempelaars MH, Moezelaar R, Abee T. Characterization of germination and outgrowth of sorbic acid-stressed *Bacillus cereus* ATCC 14579 spores: phenotype and transcriptome analysis. *Food microbiology*. 2011 Apr 1;28(2):275-83.
11. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clinical Cancer Research*. 2018 Mar 15;24(6):1248-59.
12. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology*. 2017 Nov 3;13(11):e1005752.
13. Bludau I, Aebersold R. Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nature Reviews Molecular Cell Biology*. 2020 Jun;21(6):327-40.
14. Rosenberg A, Soufi B, Ravikumar V, Soares NC, Krug K, Smith Y, Macek B, Ben-Yehuda S. Phosphoproteome dynamics mediate revival of bacterial spores. *BMC biology*. 2015 Dec;13(1):1-9.
15. Zhu Y, Piehowski PD, Zhao R, Chen J, Shen Y, Moore RJ, Shukla AK, Petyuk VA, Campbell-Thompson M, Mathews CE, Smith RD. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nature communications*. 2018 Feb 28;9(1):1-0.
16. Budnik B, Levy E, Harmange G, Slavov N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome biology*. 2018 Dec;19(1):1-2.
17. Kelly RT. Single-cell proteomics: progress and prospects. *Molecular & Cellular Proteomics*. 2020 Nov 1;19(11):1739-48.
18. Wang D, Zeng S, Xu C, Qiu W, Liang Y, Joshi T, Xu D. MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction.

- Bioinformatics. 2017 Dec 15;33(24):3909-16.
19. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, 596(7873): 583-589.