高等学校化学学报
CHEMICAL JOURNAL OF CHINESE UNIVERSITIES
研究论文

细菌的蛋白质组成分析

——深入比较三种质谱方法

蒋　龑[1]，陈妍琳[2]，宋高瑜[3]，陈炎炎[4]，白　晶[1]，朱莹娣[3]，李　娟[3]
（1. 天津大学医学部生命科学学院，天津 30072；
2. 国科大杭州高等研究院分子医学院，杭州 310024；
3. 中国科学院杭州医学研究所，杭州 310022；4. 浙江工业大学信息学院，杭州 310023）

摘要　细菌蛋白质组成分析对于了解其生物学、生理学及其与环境的相互作用至关重要. 质谱是用于蛋白质分析的最有力工具之一，在蛋白质的分子量确定、表达水平测量及结构修饰分析等研究中不可或缺. 本文比较了蛋白质指纹质谱、自上而下蛋白质组学和自下而上蛋白质组学等3种广泛使用的质谱方法在细菌蛋白成分分析中的表现. 结果表明，自下而上蛋白质组学提供了最高的蛋白覆盖率，同时也在不同菌种之间显示出最大的蛋白质图谱重合度. 相比之下，蛋白质指纹质谱显示出最高的检测再现性以及菌种区分或鉴定的高效性. 自上而下蛋白组学检测到的细菌蛋白质数量明显少于自下而上蛋白质组学，但其可以与蛋白质指纹质谱兼容(二者均偏向于检测高丰度、高稳定和高亲水性的核糖体蛋白质)，在菌种鉴定的同时为蛋白标志物的发现提供重要手段. 本文对基于质谱的蛋白质组成分析方法进行了比较，为特定分析目标的方法选择提供了指导意见. 这将对于细菌感染诊断、抗生素耐受性分析和抗生素作用靶点发现等多个领域的研究具有重要价值.

关键词　细菌；蛋白质组成分析；质谱
中图分类号　O657.63　　**文献标志码**　A　　doi：10.7503/cjcu20240345

Bacterial Protein Profiling

——Comparison of Three Mass Spectrometry Methodologies

JIANG Yan[1], CHEN Yanlin[2], SONG Gaoyu[3], CHEN Yanyan[4], BAI Jing[1], ZHU Yingdi[3*], LI Juan[3*]
(*1. School of Life Sciences, Faculty of Medicine, Tianjin University, Tianjin* 300072, *China*;
2. School of Molecular Medicine, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, *China*;
3. Hangzhou Institute of Medicine(HIM), Chinese Academy of Sciences, Hangzhou 310022, *China*;
4. School of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, *China*)

Abstract　Profiling the protein composition of bacteria is essential for understanding their biology, physiology and

收稿日期：2024-07-08. 网络首发日期：2024-07-28.
联系人简介：朱莹娣，女，博士，研究员，主要从事质谱分析方法在疾病检测中的应用研究. E-mail: zhuyingdi@him.cas.cn
李　娟，女，博士，研究员，主要从事生物成像与蛋白-蛋白相互作用方面的研究. E-mail: lijuan@him.cas.cn
基金项目：国家重点研发计划重点专项(批准号：2020YFA090900)、国家自然科学基金(批准号：22174161, 21991080, 22304179)、2022年度浙江省引进培育领军型创新创业团队(批准号：2022R01006)和浙江省尖兵领雁研发攻关计划项目(批准号：2024SDYXS0003)资助.
Supported by the National Key Research and Development Program of China(No.2020YFA090900), the National Natural Science Foundation of China(Nos. 22174161, 21991080, 22304179), the Zhejiang Leading Innovation and Entrepreneurship Team, China(No. 2022R01006) and the Pioneer Research and Development Program of Zhejiang Province, China(No.2024SDYXS0003).

interaction with environment. Mass spectrometry has become a pivotal tool for protein analysis, facilitating the examination of expression levels, molecular masses and structural modifications. In this study, we compared the performance of three widely-used mass spectrometry methods, *i.e.*, matrix-assisted laser desorption/ionization (MALDI) protein fingerprinting, top-down proteomics and bottom-up proteomics, in the profiling of bacterial protein composition. It was revealed that bottom-up proteomics provided the highest protein coverage and exhibited the greatest protein profile overlap between bacterial species. In contrast, MALDI protein fingerprinting demonstrated superior detection reproducibility and effectiveness in distinguishing between bacterial species. Although top-down proteomics identified fewer proteins than bottom-up approach, it complemented MALDI fingerprinting in the discovery of bacterial protein markers, both favoring abundant, stable, and hydrophilic bacterial ribosomal proteins. This study represents the most systematic and comprehensive comparison of mass spectrometry-based protein profiling methodologies to date. It provides valuable guidelines for the selection of appropriate profiling strategies for specific analytical purposes. This will facilitate studies across various fields, including infection diagnosis, antimicrobial resistance detection and pharmaceutical target discovery.

Keywords Bacteria; Protein profiling; Mass spectrometry

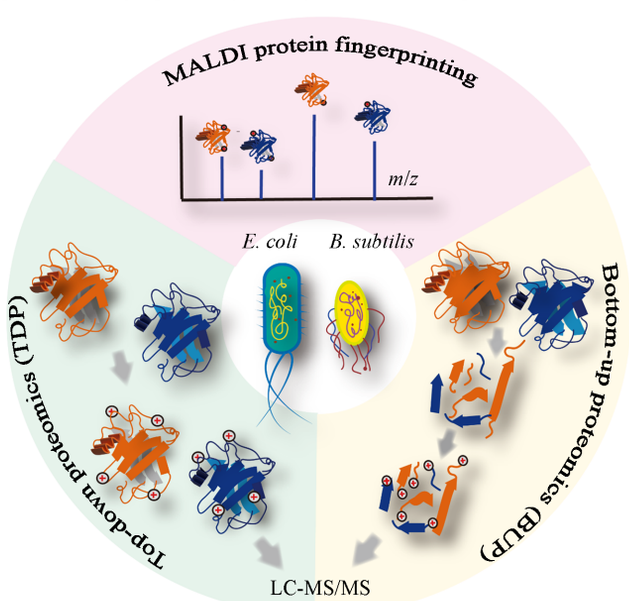
1 Introduction

Proteins are the key player in cellular processes, functioning as enzymes, structural components, signaling entities, and molecular transporters. Profiling the protein composition of bacteria is a crucial undertaking in microbiology and biotechnology. It provides substantial insights into bacterial biology, physiology and interactions with their environment. In the aspect of cell biology, protein profiling is essential for elucidating the molecular foundations of bacterial physiology and metabolism. By identifying proteins associated with metabolic pathways, researchers can map the biochemical networks that bacteria utilize for survival and growth under diverse conditions[1]. Such knowledge is invaluable for developing interventions to control pathogenic bacteria or to optimize beneficial strains for industrial applications. Profiling of bacterial proteins is also imperative for identifying virulence factors and mechanisms of antibiotic resistance[2]. Clarifying the proteins that confer antibiotic resistance enables the development of targeted strategy to inhibit the resistance or novel drugs that can circumvent the resistance mechanisms, thereby addressing the critical issue of drug-resistance bacterial strains. Analyzing bacterial protein composition moreover enhances the understanding of microbial ecology, symbiosis, and the role of bacteria in biogeochemical cycles. The information can be leveraged to optimize bacterial strains for various biotechnological applications, including fermentation, bioremediation, biofuel production and bioplastic synthesis[3]. Comprehensive insights into bacterial proteomes not only advance scientific knowledge but also facilitate practical applications, ultimately contributing to innovations in health care, industry and environmental management.

The most widely used techniques for protein composition profiling include capillary electrophoresis, liquid chromatography, gel electrophoresis and mass spectrometry. Capillary electrophoresis and liquid chromatography are mainly employed for protein separation, based on their size and charge or their interactions with the stationary phase, respectively. They are frequently combined with different detection methods such as ultraviolet-visible absorbance spectroscopy or mass spectrometry for precise quantification and/or identification of proteins[4,5]. Gel electrophoresis, particularly sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), are classic methods for separating and visualizing proteins in complex mixtures. SDS-PAGE separates proteins based on their molecular weight (or mass), while 2D-PAGE works based on both the protein mass and isoelectric point. They are also frequently followed by mass spectrometry to achieve detailed protein identification and

quantification[6,7]. Mass spectrometry itself measures proteins according to their mass-to-charge ratios. It is currently the most powerful tool available for profiling the protein composition of biological samples, owing to its unparalleled capabilities in protein identification, quantification and analysis of post-translation modifications. As a technique capable of label-free detection of multiple molecules simultaneously with high sensitivity and accuracy, mass spectrometry has been extensively employed in various chemical and biological analyses, despite its high cost. Among the diverse methodologies in mass spectrometry, matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS), top-down proteomics utilizing liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) and bottom-up proteomics utilizing also LC-MS/MS are three predominant tools for protein detection and profiling. MALDI-TOF MS is primarily used for rapid protein characterization and identification. It involves the ionization of proteins under laser irradiation, facilitated by laser energy-absorbing matrix molecules, followed by the separation and measurement of ions according to their mass-to-charge ratios by time-of-flight mass spectrometry. It is widely adopted in microbial identification due to its ability to generate characteristic mass spectra, or "fingerprints", in a rapid and high-throughput manner[8]. Top-down proteomics involves the direct analysis of intact proteins without prior digestion into peptides. Using LC-MS/MS, this approach provides detailed information on amino acid sequences and molecular masses of the protein analytes[9]. It is particularly advantageous for characterizing protein isoforms and detecting molecular modifications that might be lost or misinterpreted in peptide-based methods. Bottom-up proteomics, on the other hand, involves the enzymatic digestion of proteins into smaller peptides, which are than separated by liquid chromatography and subsequently analyzed by tandem mass spectrometry. This peptide-based method enables large-scale protein identification, quantification, and detailed mapping of the protein compositions in biological samples[10].

In this study, we conducted a comprehensive comparison of the three mass spectrometry methodologies for the profiling of bacterial protein components. *Escherichia coli* (*E. coli*, strain MG1655) and *Bacillus subtilis* (*B. subtilis*, strain 168) were analyzed as the representative of gram-negative bacteria and gram-positive bacteria, respectively. Both species have been widely adopted as model organisms for laboratory studies, due to their biosafety and ease of cultivation. The protein profiles of the two species were depicted by using the three methodologies, *i. e.*, MALDI protein fingerprinting using a Bruker rapifleX MALDI-TOF tissueTyper, top-down proteomic analysis using Q Exactive HF quadrupole-Orbitrap mass spectrometer, and bottom-up proteomics using an Orbitrap Exploris 480 mass spectrometer (Scheme 1). Fresh *E. coli* and *B. subtilis* cells after overnight culture under their optimal growth condition were collected for the analysis. The whole intact cells were directly subjected to MALDI fingerprinting. Top-down proteomics was conducted with proteins extracted from the bacterial cells using the widely used ethanol-acetonitrile-formic acid extraction protocol. Bottom-up proteomics was conducted with the peptides digested from the proteins by the



Scheme 1 Profiling of bacterial protein composition using three mass spectrometry methodologies

protease trypsin. This study provides critical insights into the strengths and limitations of each method, representing the most systematic and comprehensive comparison of mass spectrometry-based protein profiling methodologies to date. It offers valuable guidelines for selecting appropriate approaches for specific analytical purposes in the microbial field.

2 Experimental

2.1 Materials and Measurements

Sinapinic acid(SA)(matrix substance for MALDI-MS, ≥99.0%), ethanol(≥99.8%), formic acid(ACS reagent, ≥96.0%) and Luria-Bertani microbial growth medium were bought from Sigma-Aldrich(St. Gallen, Switzerland). Acetonitrile(HPLC grade) was purchased from Aventor Performance Materials(Center Valley, PA, USA). Trifluoroacetic acid(99.0%, extra pure) was obtained from Acros Organics(New Jersey, USA). Deionized water(18.2 MΩ·cm) was purified by an alpha Q Millipore system(Zug, Switzerland), and used in all aqueous solutions.

The bacterial morphological photos were taken by using JEOL JSM-IT800 Schottky field emission scanning electron microscope(SEM). MALDI-TOF MS measurements were performed on a Bruker rapifleX MALDI-TOF tissueTyper. Top-down proteomics were performed using Thermo Fisher UPLC Ultimate 3000-Q Exactive HF Hybrid Quadrupole Orbitrap Mass Spectrometer. Bottom-up proteomics were performed using Thermo Scientific EASY-nLC 1200-Orbitrap Exploris 480 mass spectrometer.

2.2 Bacterial Culture

E. coli and *B. subtilis* were gifts from Zhejiang Cancer Hospital, Zhejiang, China. Each strain was grown as a pre-culture in 2 mL of Luria-Bertani at 35—37 ℃ for 6 h with continuous shaking at 200 r/min. Thereafter, 100 μL of the pre-culture was added into 3 mL of Luria-Bertani and incubated 12—16 h at 35—37 ℃ with continuous shaking. Concentrations of bacterial cells in their fresh cultures were determined by measuring the optical density at 600 nm($OD_{600\,nm}$), with 1.0 $OD_{600\,nm}$ corresponding to *ca.* 8×10^8 cells/mL.

2.3 MALDI-TOF MS Measurement

Bacterial cells were separated from the culture media by centrifugation at 10000 r/min for 3 min, and washed twice with deionized water. The cellular pellet was resuspended in water with the concentration of *ca.* 5×10^8 cells/mL, and 1 μL of the cell suspension was deposited on a MALDI target plate, and dried at room temperature. 1 μL of SA matrix(15 mg/mL dissolved in the mixture of acetonitrile/water/trifluoroacetic acid(volume ratio, 50∶49.9∶0.1) was added to cover the dried cells, and dried at room temperature. MALDI-TOF MS measurements were performed on a Bruker rapifleX MALDI-TOF tissueTyper under linear positive mode in the mass range of *m/z* 2000—30000. Instrumental parameters were set as laser intensity 70%, laser attenuator with 30% offset and 40% range, 5000 laser shots accumulation for each spot, 20.0 Hz laser frequency, 10× detector gain, suppress up to 1000, 350 ns pulsed ion extraction. The mass spectra were pre-processed by moving average smoothing, statistical non-linear iterative peak(SNIP) baseline correction, and total ion current(TIC) intensity normalization. The intra- and inter-sample peak matching was performed using the Forward approach, with 1500 ppm(parts-per-million) tolerance and average(AVG) reference type. The peak picking was performed through MALDIquant, with signal-to-noise ratio(*S/N*) of 3, half window size of 60 and minimal peak intensity of 5×10^{-5}.

2.4 Top-Down Proteomic Analysis

Proteins were extracted from bacterial cell according to an ethanol/formic acid/acetonitrile extraction protocol, which is described in the Bruker MALDI Biotyper® 3.0 user manual(year 2011). Briefly, cells

were harvested from 1 mL of fresh culture by centrifugation at 10000 r/min for 3 min and washed twice with deionized water. The cell pellet was suspended in 300 μL of water, followed by adding 900 μL of ethanol and mixing thoroughly. The mixture of water and ethanol was completely removed by centrifugation at 10000 r/min for twice (2 min for each time) and air-drying for *ca.* 30 min. The obtained cell pellet was resuspended in 50 μL of formic acid/water (volume ratio, 70∶30). The mixture was vortexed thoroughly and let stand for 5 min, followed by adding 50 μL of acetonitrile. The final mixture was vortexed thoroughly for another 5 min. Thus, intracellular proteins were extracted into the solvent. After centrifugation at 10000 r/min for 2 min, the supernatant containing the extracted proteins was harvested and subjected to top-down proteomic analysis using LC-MS/MS (Thermo Fisher UPLC Ultimate 3000-Q Exactive HF Hybrid Quadrupole Orbitrap Mass Spectrometer).

2.5 Bottom-Up Proteomic Analysis

Proteins were extracted from bacteria using the same method as described above in Section 2.4. After methanol-chloroform precipitation, the proteins were digested in trypsin protease to generate short peptides. After peptide quantification using a colorimetric assay kit (Pierce™, Thermo Fisher), an identical quantity of peptides was analyzed for each sample using LC-MS/MS (Orbitrap Exploris 480 mass spectrometer) under the data-dependent acquisition mode. Proteins were identified through the observation of unique peptides by using MaxQuant proteomic data searching software with the searching result displayed using Scaffold tool. The relative quantity of each protein was determined by label-free quantification.

3 Results and Discussion

3.1 Characterization of Bacterial Cells

E. coli can divide every 20—30 min[11] and can reach an overnight cell density of more than 10^9 cells per milliliter when growing in an optimal environment, *i. e.*, growing in a rich liquid broth medium Luria-Bertani at 37 ℃ with aeration in this study. The cells were typically rod-shaped, with a diameter of 0.5—1 μm and a length of 1—3 μm [Fig. 1(A) and (B)]. The growth rate and cell shape of *B. subtilis* were similar to that of *E. coli* under the optimal growth condition, *i. e.*, growing in Luria-Bertani at 35 ℃ with aeration. Compared to *E. coli*, the cells of *B. subtilis* were slightly longer, around 2—4 μm [Fig. 1(C) and (D)]. Interestingly, the cell surface of *B. subtilis* appeared much smoother than *E. coli* [Fig. 1(D) *vs.* (B)], likely attributed to their difference in the cell wall composition and the surface appendages. Gram-positive bacteria like *B. subtilis* possess a thick peptidoglycan layer on the cell surface, conferring a uniform and rigid cell structure. Gram-negative bacteria like *E. coli* have a much thinner peptidoglycan layer located between an inner cytoplasmic membrane and an outer membrane containing lipopolysaccharides, and the cells also typically possess numerous surface structures like pili, flagella and fimbriae[12]. The presence of the additional outer membrane and surface structures contribute to their rougher appearance. Protein profiles of the two species were collected by using the three mass spectrometry approaches, and the resulting profiles were subsequently

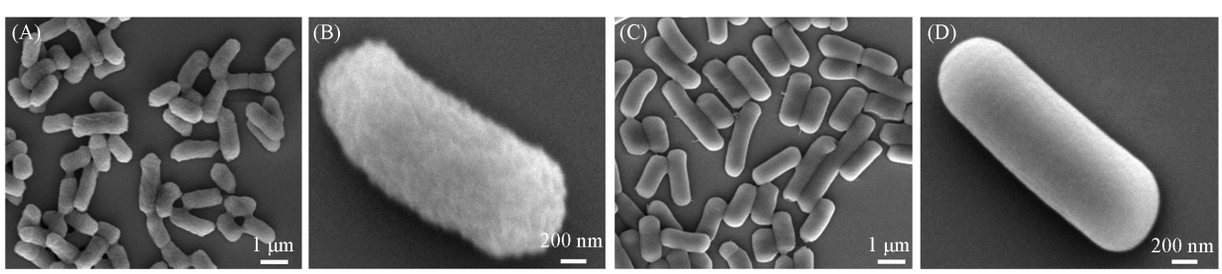


Fig. 1 SEM images of *E. coli*(A, B) and *B. subtilis*(C, D) with different magnifications

compared as detailed below.

3.2 BUP Enhanced Protein Coverage While MALDI Boosted Reproducibility

Among the three protein profiling approaches, bottom-up proteomic analysis (BUP), as expected, gave the most comprehensive protein information, with 1867 and 1914 proteins (average from four repetitions) identified for *E. coli* and *B. subtilis*, respectively [Fig. 2(A) and (B)]. The MALDI protein fingerprinting approach (MALDI) detected 78 and 83 well-resolved mass peaks (average from nine repetitions, *S*/*N* higher than 3, TIC-normalized intensity higher than 0.00005) from the two species, respectively, in the mass range of *m*/*z* 3000—20000 [Fig. 2(A) and (B)]. The peaks should be proteins or protein fragments within this mass range. Similar to the MALDI fingerprinting, the top-down proteomic analysis (TDP) identified 86 and 83 proteins or protein fragments (average from four repetitions) from the two species, respectively [Fig. 2(A) and (B)]. A bacterial protein typically contains tens to hundreds of amino acids, with the mass generally ranging from several to hundreds of kilodalton. A mass spectrometry analysis is quite often to show a decrease in ion current with increase mass of the analyte, as large molecules are generally difficult to ionize. This is the case for MALDI fingerprinting and TDP, with the detected protein mass mostly smaller than 50000. Instead of analyzing un-digested native proteins, BUP, while, detect a protein by analyzing its unique peptides. The bacterial proteins were digested by trypsin into short peptides (7—16 amino acids in average) with the mass decrease to around 700—2000[13]. Decreased analyte mass led to increased ionization efficiency and thus enhanced protein detection capability. Factors like fragmentation efficiency, instrumentation limitations and data interpretation also constrain the protein composition profiling capability of TDP compared to BUP. Specifically, the fragmentation of large intact proteins during TDP is less efficient and more variable, making it challenging to generate comprehensive fragmentation patterns needed for accurate identification. Peptides

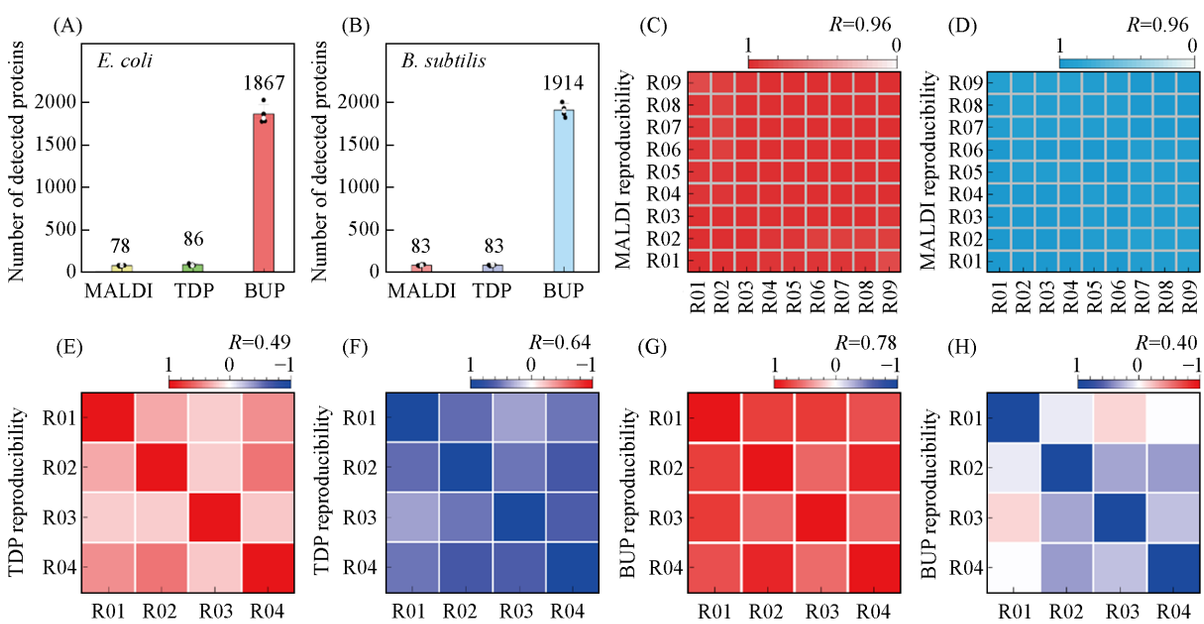


Fig. 2 Investigation of protein coverage and detection reproducibility

(A) Numbers of proteins or protein peaks detected from *E. coli* by MALDI fingerprinting (MALDI), top-down proteomics (TDP) and bottom-up proteomics (BUP); the average number was displayed for each approach; (B) numbers of proteins or protein peaks detected from *B. subtilis* by using the three protein profiling approaches; (C) detection reproducibility for nine repetition tests of *E. coli* by using MALDI approach; (D) detection reproducibility for nine repetition tests of *B. subtilis* by using MALDI approach; (E) detection reproducibility for four repetition tests of *E. coli* by using TDP approach; (F) detection reproducibility for four repetition tests of *B. subtilis* by using TDP approach; (G) detection reproducibility for four repetition tests of *E. coli* by using BUP approach; (H) detection reproducibility for four repetition tests of *B. subtilis* by using BUP approach.

fragment more predictably and efficiently in BUP, resulting in reliable and interpretable tandem MS spectra that facilitate protein identification. With regard to the instrumentation limitations, TDP typically requires highly specialized mass spectrometers with high mass accuracy and resolving power to handle the complexity of intact proteins. Many mass spectrometers struggle with the dynamic range and resolving power needed for high-confidence identification. BUP, however, often utilizes standard mass spectrometers that are well-optimized for peptide detection and fragmentation, including state-of-the-art instruments such as the Bruker's tims TOF Ultra and the Thermo Fisher Scientific's Orbitrap Astral. Moreover, data interpretation for intact proteins, especially those with various post-translational modifications, is computationally intensive and less mature compared to BUP. The algorithms, software and databases for TDP data interpretation are still under development, limiting the number of proteins that can be confidently identified. Although MALDI fingerprinting detected lowest number of proteins among the three approaches, it provided the highest detection reproducibility, with the reproducibility value reaching 0.96 for both bacteria in nine repetition tests [Fig. 2(C) and (D)]. While the detection reproducibility was much lower for the two proteomic approaches, *i. e.*, 0.49 for *E. coli* and 0.64 for *B. subtilis* in four repeated TDP analyses[Fig. 2(E) and (F)], 0.78 for *E. coli* and 0.40 for *B. subtilis* in four repeated BUP analyses[Fig. 2(G) and (H)]. The lower reproducibility in proteomic analysis could be caused by the additional protein pretreatment like chromatography separation or protein digestion before mass spectrometry measurement. Here, the reproducibility value was calculated as the average of the Spearman's rank correlation coefficient between any two repetition tests, with the value ranging from -1(totally opposite) to 1(exactly the same)[14]. Altogether, the MALDI fingerprinting approach provided the briefest but most reproducible protein profiles in a fastest manner(minutes per sample). While, BUP analysis was the most time-consuming(hours per sample) and laborious but provided the most informative protein profiles. Interestingly, the number of detected proteins or protein peaks were quite close for the two different bacteria using any of the three given approaches, consistent with the similar protein composition size in the two species(*i. e.*, 4260 known proteins in the *B. subtilis* strain 168, and 4404 known proteins in the *E. coli* strain MG1655, as collected in the UniProt database).

3.3 MALDI-TDP Integration Advanced Protein Profile Annotation and Biomarker Discovery

MALDI fingerprinting provided an intuitive visualization of the molecular composition in the bacterial cells, by profiling the mass and relative abundance of the molecules. According to the *lucky survivor model*, most of the fingerprint peaks were singly charged[15]. As shown in Fig.3(A), the MALDI fingerprints of *E. coli* and *B. subtilis* are quite different from each other, with the fingerprint Spearman similarity as low as 0.029, indicating a huge difference in their molecular composition. In order to clarify the protein identity of the fingerprint peaks, we tentatively correlated the fingerprints with the TDP data according to the measured masses. The mass tolerance was set as 1500 ppm, considering the moderate mass accuracy and resolution for MALDI-TOF MS detection of proteins at linear positive mode. Such a correlation is rational as both approaches measured native un-digested proteins or protein fragments with the signal intensity highly related to the protein abundance[16]. Accordingly, 22 and 37 of the MALDI fingerprint peaks detected from *E. coli* and *B. subtilis*, respectively, were successfully assigned to particular protein(or the protein fragment) molecules. The assignment details are listed in Tables 1 and 2, with the protein entry name selectively labeled on the highly abundant peaks[Fig.3(A)]. To be noted, none of the assigned peaks exhibited overlap between the two bacterial species, as shown in the Venn diagram in Fig.3(A). Protein properties that could influence MALDI measurement, including the subcellular location, isoelectric point, instability index, aliphatic index and grand average of hydropathy, were investigated[Fig.3(B), Tables 1 and 2]. The subcellular location of a protein is determined by the combination of UniProt annotation and a machine learning-based prediction system named

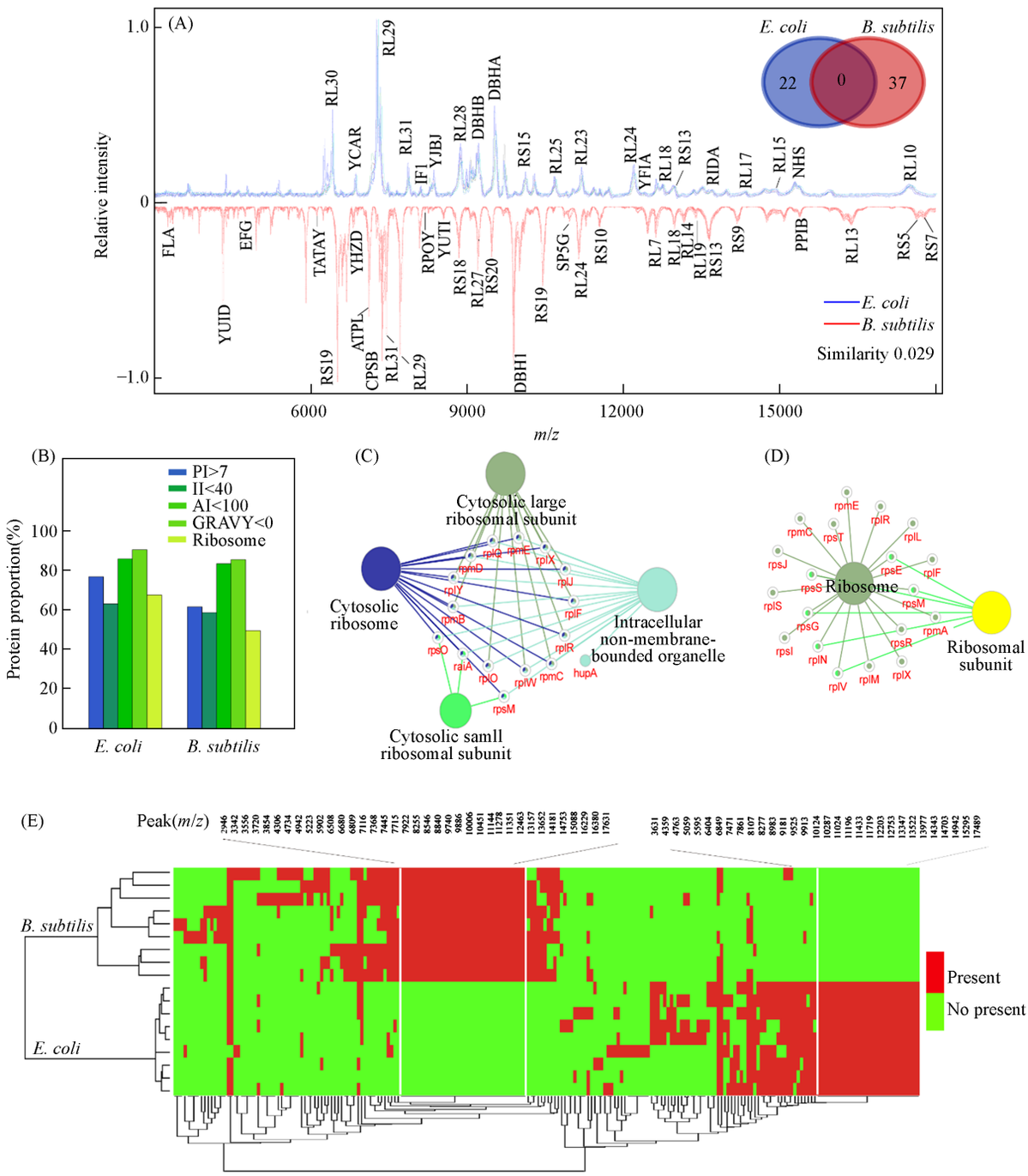


Fig. 3 Correlation between MALDI and TDP protein profiles

(A) MALDI protein fingerprints of *E. coli* and *B. subtilis*, with protein identity labeled on the main peaks. The protein identity was tentatively assigned through matching the fingerprint peak masses to the protein (or protein fragment) masses measured by TDP; insert: fingerprint peak numbers detected from *E. coli* and *B. subtilis* with protein identity assigned; (B) investigation of the properties of proteins assigned on the MALDI fingerprints of *E. coli* and *B. subtilis*. The investigated properties include isoelectric point (PI), instability index (II), aliphatic index (AI), grand average of hydropathy (GRAVY), and if the proteins were ribosomal proteins; (C) biological functional analysis of the proteins assigned for *E. coli* using ClueGO system; (D) biological functional analysis of the proteins assigned for *B. subtilis* using ClueGO system; gene ontology (GO) terms are represented as nodes, with node size representing the enrichment significance; only the most significant terms were labeled, with the exhibition of representative enriched pathway ($P < 0.05$) interactions among the main targets; (E) a deep hierarchical clustering analysis of the MALDI fingerprints using cosine correlation distance function and average linkage, correlated with the peak presence profile. The peaks exclusively detected from *E. coli* and *B. subtilis* were listed.

Table 1 Properties of proteins assigned on the MALDI fingerprints of *E. coli*

MALDI (*m/z*)	TDP precursor mass	Protein entry name	Gene ID	Subcellular location	Isoelectric point	Instability index	Aliphatic index	GRAVY	Detected by BUP
6404	6406.60	P0AG51_RL30	rpmD	Cytoplasm	10.96	38.72	102.59	−0.137	YES
6849	6848.53	P0AAZ7_YCAR	ycaR	Cytoplasm	4.31	36.22	105.39	−0.127	YES
7264	7268.98	P0A7M6_RL29	rpmC	Cytoplasm	9.98	50.27	102.22	−0.657	YES
7861	7865.92	P0A7M9_RL31	rpmE	Cytoplasm	9.46	48.68	57.00	−0.649	NO
8107	8113.28	P69222_IF1	infA	Cytoplasm	9.23	36.00	91.83	−0.354	NO
8354	8344.10	P68206_YJBJ	yjbJ	Cytoplasm/Periplasm	5.64	38.38	15.60	−2.432	NO
8871	8869.83	P0A7M2_RL28	rpmB	Cytoplasm	11.42	36.21	77.87	−0.650	YES
9213	9219.99	P0ACF4_DBHB	hupB	Cytoplasm/Periplasm	9.69	16.40	92.44	−0.042	YES
9525	9529.19	P0ACF0_DBHA	hupA	Cytoplasm/Periplasm	9.57	15.88	91.22	−0.228	NO
10124	10131.47	P0ADZ4_RS15	rpsO	Cytoplasm	10.40	46.76	90.91	−0.673	NO
10676	10685.76	P68919_RL25	rplY	Cytoplasm	9.60	31.89	91.28	−0.392	YES
11196	11192.22	P0ADZ0_RL23	rplW	Cytoplasm	9.94	21.46	97.40	−0.373	YES
12203	11208.28	P60624_RL24	rplX	Cytoplasm	10.21	9.01	89.81	−0.405	YES
12638	12645.59	P0AD49_YFIA	raiA	Cytoplasm	6.18	47.48	90.62	−0.535	YES
12753	12761.93	P0C018_RL18	rplR	Cytoplasm	10.42	32.35	86.92	−0.395	YES
12971	12960.23	P0A7S9_RS13	rpsM	Cytoplasm	10.78	38.74	96.75	−0.424	YES
13522	13518.11	P0AF93_RIDA	ridA	Cytoplasm	5.36	40.34	97.56	0.086	NO
14343	14355.66	P0AG44_RL17	rplQ	Cytoplasm	11.05	45.90	79.29	−0.565	NO
14942	14957.31	P02413_RL15	rplO	Cytoplasm	11.18	43.69	87.43	−0.252	YES
15385	15398.99	P0ACF8_HNS	hns	Cytoplasm	5.44	40.47	82.70	−0.751	NO
17547	17569.35	P0A7J3_RL10	rplJ	Cytoplasm	9.04	28.74	94.21	0.045	NO
18745	18761.12	P0AG55_RL6	rplF	Cytoplasm	9.71	17.04	91.93	−0.227	NO

Table 2 Properties of proteins assigned on the MALDI fingerprints of *B. subtilis*

MALDI (*m/z*)	TDP precursor mass	Protein entry name	Gene ID	Subcellular location	Isoelectric point	Instability index	Aliphatic index	GRAVY	Detected by BUP
3342	3342.72	P02968_FLA	hag	Cytoplasm	6.24	27.62	82.82	−0.513	YES
4376	4377.46	O32107_YUID	yuiD	Membrane	8.39	31.42	120.00	0.382	YES
4769	4774.53	P80868_EFG	fusA	Cytoplasm	8.25	48.35	76.00	−0.511	YES
5042	5039.81	O32111_YUZG	yuzG	Membrane/Secreted	9.40	27.87	129.35	0.548	NO
5431	5427.16	P24469_C550	cccA	Membrane	6.37	19.70	77.17	−0.549	YES
6099	6089.51	O05522_TATAY	tatAy	Membrane/Secreted	6.18	30.93	95.96	−0.258	YES
6508	6502.85	P21476_RS19	rpsS	Cytoplasm/Membrane	10.11	48.28	64.57	−0.740	YES
6640	6641.39	P23308_SINI	sinI	Cytoplasm/Secreted	6.26	55.83	73.68	−0.804	NO
6809	6807.46	C0H3Y1_YHZD	yhzD	Cytoplasm/Secreted	5.79	13.70	83.28	−0.461	NO
7012	7009.60	P70994_4OT	ywhB	Cytoplasm	5.37	59.42	75.32	−0.644	YES
7116	7116.98	P37815_ATPL	atpE	Membrane	6.05	14.25	145.14	1.274	YES
7368	7360.61	P32081_CSPB	cspB	Cytoplasm	4.54	18.56	66.87	−0.337	YES
7445	7438.68	O03223_RL31	rpmE	Cytoplasm/Secreted	9.21	35.10	47.27	−0.689	YES
7715	7708.26	P12873_RL29	rpmC	Cytoplasm	10.10	34.04	96.21	−0.626	YES
8255	8247.24	O31718_RPOY	rpoY	Cytoplasm	4.81	50.16	83.19	−0.616	YES
8546	8539.41	O32119_YUTI	yutI	Cytoplasm	4.34	47.47	116.96	0.200	YES
8840	8832.90	P21475_RS18	rpsR	Cytoplasm/Secreted	11.04	43.37	76.58	−0.639	YES
9209	9201.90	P05657_RL27	rpmA	Membrane/Secreted	10.32	11.28	57.79	−0.809	YES
9470	9462.29	P21477_RS20	rpsT	Cytoplasm/Secreted	10.94	30.28	77.84	−0.714	YES
9886	9878.28	P08821_DBH1	hupA	Cytoplasm	8.96	29.52	82.83	−0.485	YES
10451	10445.56	P21476_RS19	rpsS	Cytoplasm/Membrane	10.11	48.28	64.57	−0.740	YES
10952	10943.59	P28015_SP5G	spoVG	Cytoplasm	5.25	26.40	83.40	−0.479	YES
11150	11163.36	P0CI78_RL24	rplX	Cytoplasm/Membrane	10.04	16.40	77.38	−0.644	YES

Continued

MALDI (*m/z*)	TDP precursor mass	Protein entry name	Gene ID	Subcellular location	Isoelectric point	Instability index	Aliphatic index	GRAVY	Detected by BUP
11536	11527.33	P21471_RS10	rpsJ	Cytoplasm	9.79	48.41	99.41	−0.460	YES
11591	11585.06	P0C174_GPSB	gpsB	Cytoplasm	5.72	50.84	85.51	−0.853	YES
12463	12451.83	P42060_RL22	rplV	Cytoplasm	10.73	31.93	97.61	−0.244	YES
12622	12611.85	P02394_RL7	rplL	Cytoplasm	4.56	44.34	114.31	0.115	YES
12972	12960.98	P46899_RL18	rplR	Cytoplasm	10.11	19.12	87.92	−0.370	YES
13157	13146.12	P12875_RL14	rplN	Cytoplasm	9.94	31.75	96.56	−0.137	YES
13390	13378.60	O31742_RL19	rplS	Cytoplasm	10.97	41.84	98.17	−0.475	YES
13652	13647.71	P20282_RS13	rpsM	Cytoplasm	11.07	42.95	96.61	−0.720	YES
14213	14210.74	P21470_RS9	rpsI	Cytoplasm/Membrane	10.60	40.42	87.85	−0.512	YES
15286	15271.55	P35137_PPIB	ppiB	Cytoplasm	5.53	20.21	59.30	−0.364	YES
16380	16364.70	P70974_RL13	rplM	Cytoplasm	9.87	15.64	80.07	−0.632	YES
17631	17611.78	P21467_RS5	rpsE	Cytoplasm	9.92	34.39	108.61	−0.001	YES
17770	17754.49	P21469_RS7	rpsG	Cytoplasm	10.81	45.66	98.17	−0.621	YES
19385	19366.40	P46898_RL6	rplF	Cytoplasm	9.49	31.26	87.49	−0.501	YES

CELLO[17]. For both bacteria, a majority of the proteins were located in the cytoplasm, but some membrane or secretory (extracellular) proteins were also detected, especially for *B. subtilis*. This indicates that both cell surface proteins and the deeper intracellular proteins could be detected by MALDI from intact whole bacterial cells without a previous protein extraction, with the intracellular cytoplasmic proteins favorably detected. Isoelectric point (PI) of a protein, *i. e.*, the pH at which the net electrostatic charge of the protein equals zero, is often used to indicate the global basic or acidic character of the protein. A protein with PI higher than 7 is considered basic, and could be more likely to be positively charged and detected by mass spectrometry under positive mode. This is indeed the case for both *E. coli* and *B. subtilis*, as 77.3%(17/22) and 62.2%(23/37) of the proteins had the PI>7, respectively. Instability index (II) of a protein is related to the ratio of non-polar to polar amino acids in the protein, providing an estimate of the protein stability in a test tube with the value less than 40 indicating stable. For both bacteria, the proportion of proteins with II<40 is more than half, *i. e.*, 63.6%(14/22) for *E. coli* and 59.5%(22/37) for *B. subtilis*. This indicates that proteins with higher stability are more likely to be detected by MALDI fingerprinting, possibly due to the lower likelihood of the proteins to undergo aggregation, misfolding or degradation before and during the MALDI fingerprinting. Aliphatic index (AI) of a protein measures the relative volume occupied by aliphatic side chain of amino acids (alanine, valine, leucine and isoleucine), and is often used to evaluate the thermostability and hydrophobicity of a protein, with a decreased value indicating an increase of the protein hydrophilicity. Grand average of hydropathy (GRAVY) is the average hydropathy value of a protein or peptide, with a positive value indicating hydrophobic and a negative value indicating hydrophilic. In *E. coli*, 86.3%(19/22) of the proteins had the AI<100 and 90.9%(20/22) of the proteins had the GRAVY<0. The values for *B. subtilis* are 83.8%(31/37) and 86.5%(32/37), respectively. This indicates that bacterial proteins detected by MALDI tend to be hydrophilic. Hydrophilic proteins usually have polar or charged side chains that could interact favorably with the matrix molecules (typically organic acids), facilitating energy transfer and efficient desorption when exposed to laser. They could also have high proton affinity due to the presence of polar or charged groups, making them easy to acquire protons from matrix molecules and resulting in a high ionization efficiency.

In order to clarify their underlying biological functionals, those proteins were analyzed using GluoGo in Cytoscape, which integrates Gene Ontology (GO) terms as well as KEGG and BioCarta pathways to create a functionally organized GO-pathways network[18]. It turns out that the proteins detected from *E. coli* were mainly involved in the pathways of cytosolic small and large ribosomal subunits and intracellular

non-membrane-bounded organelles[Fig. 3(C)], and the proteins detected from *B. subtilis* mainly participated in the ribosome and ribosomal subunit pathways[Fig. 3(D)]. This is consistent with the data in Fig. 3(B), *i. e.*, 50%—70% of the proteins were ribosomal proteins(RPs) for both bacteria. Bacterial ribosome, responsible for translating mRNA into protein, is composed of the large subunit(50S) and the small subunit(30S). It contains approximately 55%—60% ribosomal RNA (rRNA) and 40%—45% RPs. RPs control the co-transcriptional rRNA folding during ribosome assembly and engage in the formation of the ribosome functional sites like the mRNA-binding sites, tRNA-binding sites, the peptidyl transferase center, and the protein exit tunnel[19]. A deep clustering analysis of the MALDI fingerprints was further conducted to figure out protein peak markers to distinguish between the two species[heatmap in Fig. 3(E)]. It picked out 31 peaks exclusively detected from *E. coli* and 37 peaks exclusively detected from *B. subtilis*, all of which were 100% presence in the repetitions[peak list in Fig. 3(E)]. A marker should exhibit not only extremely high specificity and reproducibility but also substantial signal intensity with clear molecular identity. Taking those into consideration, eight peaks were ultimately selected as the marker for *E. coli*, including *m/z* 6404 from protein RL30, *m/z* 6849 from protein YCAR, *m/z* 9525 from DBHA, *m/z* 10124 from RS15, *m/z* 11196 from RL23, *m/z* 12203 from RL24, *m/z* 12753 from RL18, and *m/z* 14343 from RL17. And eleven peaks were selected as the marker for *B. subtilis*, including *m/z* 6508 and *m/z* 10451 from protein RS19(two different fragments of RS19), *m/z* 7116 from ATPL, *m/z* 7715 from RL29, *m/z* 8840 from RS18, *m/z* 9886 from DHB1, *m/z* 12436 from RL22, *m/z* 13157 from RL14, *m/z* 13652 from RS13, *m/z* 16380 from RL13, and *m/z* 17631 from RS5. This marker discovery strategy presents a promising tool for the detection and identification of proteins associated with antimicrobial resistance. Mass peaks specific to bacterial strains resistant against an antibiotic could be readily found by comparing the MALDI protein fingerprints to those of susceptible counterparts. Subsequently, the protein identity of the resistance-specific peaks could be elucidated by correlating the MALDI fingerprints with the TDP data of the same bacterial strains. Furthermore, signaling pathways associated with the resistance can be clarified through subjecting those resistance-specific proteins to GO-pathway enrichment analysis, thereby aiding in the understanding of the molecular mechanism underlying the resistance.

3.4 Protein Preference of TDP and BUP

BUP is able to identify a significant number of proteins, often more than TDP currently, but we are curious to know if the two approaches have their respective preferred proteins. We performed GO enrichment analysis on the proteins identified from *E. coli* and *B. subtilis* to investigate the protein distribution. The top ten enriched categories for cellular component(CC), molecular functions(MF) and biological process(BP) are detailed below(Fig. 4).

According to the fold enrichment (FE) and statistical significance (P<0.05), the *E. coli* proteins obtained by TDP had a notable enrichment in CC categories like cytosolic ribosome, ribosome subunit, intracellular non-membrane-bounded organelle and intracellular organelle[Fig. 4(A)]. The enrichment for *B. subtilis* were mainly focused on the categories of ribonucleoprotein complex, intracellular organelle and intracellular non-membrane-bounded organelle[Fig. 4(C)]. Both are closely associated with the ribosome, an intracellular non-membrane-bounded organelle composed of RNA-protein complexes that are responsible for the protein synthesis. Nevertheless, the CC enrichment for proteins obtained by BUP from both bacteria showed low FE values for all the enriched categories, predominantly ranging from 0.7 to 1.2[Fig. 4(B) and (D)]. These values were substantially lower than those observed for TDP proteins, which mostly ranged from 10 to 20. And these categories were randomly distributed among various cellular categories, including the cell envelope, cell periphery, cytosol and intracellular organelles. It indicates that BUP exhibited no apparent

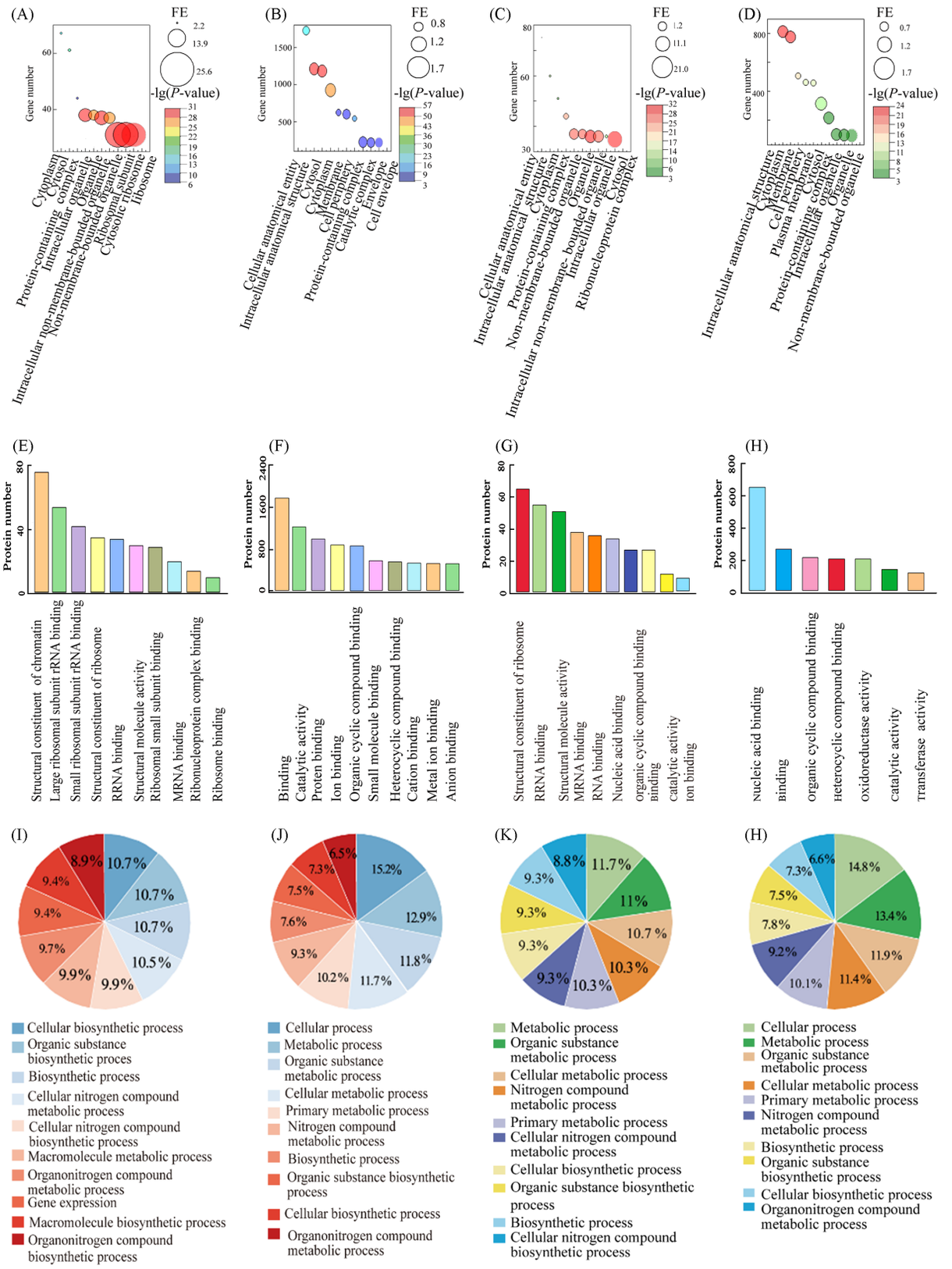


Fig. 4 Protein preference of TDP and BUP revealed by GO enrichment analysis

(A—D) Enrichment results of the cellular component (CC) for *E. coli* (A, B) and *B. subtilis* (C, D); (E—H) enrichment results of the molecular function (MF) for *E. coli* (E, F) and *B. subtilis* (G, H); (I—L) enrichment results of the biological process (BP) for *E. coli* (I, J) and *B. subtilis* (K, L). The enrichment analysis was conducted according to the instruction on the Gene Ontology platform, with the top ten enriched categories displayed in each category. *P*-value lower than 0.05 indicated a significant enrichment.

preference for any specific cellular components in bacteria. In contrast, TDP preferred intracellular ribosomal proteins, akin to the selective behavior observed in MALDI fingerprinting. Ribosomal proteins present in high abundance in bacterial cells, for instance, representing around 40% of the top abundant proteins in *E. coli*[20]. In addition, the proteins usually possess high stability and hydrophilicity. The stability is essential for the structure integrity and proper protein synthesis function of the ribosome[21]. While the hydrophilicity helps to maintain the stability of ribosome by interacting with the surrounding aqueous cytosol environment[22]. Those properties promote the ionization of the intact un-digested ribosomal proteins during MALDI and TDP measurement.

The MF terms describe activities that occur at the molecular level. The GO enrichment analysis exhibited that the *E. coli* proteins obtained by TDP had a significant enrichment in MF categories including structural constituent of chromatin, ribosomal subunit rRNA binding, structural constituent of ribosome, structure molecular activity, rRNA and mRNA binding[Fig.4(E)]. The enriched categories for *B. subtilis* were mainly structural constitute of ribosome, rRNA and mRNA binding, structural molecule activity[Fig.4(G)]. Both are highly relevant to the protein synthesis and structural integrity of the bacterial cells. In the case of BUP, the identified proteins mainly exhibited an enrichment in the MF categories of anion binding, metal ion binding, cation binding, heterocyclic compound binding, and organic cyclic compound binding for *E. coli* [Fig.4(F)]. And the categories for *B. subtilis* were mainly nucleic acid binding, organic cyclic compound binding, and heterocyclic compound binding[Fig.4(H)]. Those categories refer to the biochemical activities of metabolism, enzymatic reactions, cellular signaling and genetic information propagation, each of which plays a critical role in the life cycle and survival of bacteria. In contrast to the CC and MF enrichment results, the BP enrichment reveals no significant difference between the TDP and BUP protein profiles. The proteins were not preferentially distributed across various cellular biosynthesis processes and metabolic processes for both bacteria[Fig.4(I—L)].

3.5 BUP Profile Exhibited the Highest Overlap Between Bacterial Species

E. coli and *B. subtilis* belong to distinct bacterial phyla, *i. e.*, *Pseudomonadota* and *Bacillota*, and exhibit different cellular structures, being gram-negative and gram-positive, respectively. Consequently, their protein profiles are expected to be significantly different from each other. As a verification, the number of proteins or protein peaks detected by each of the three mass spectrometry profiling approaches was quantified for the two bacterial species respectively. The counts of species-common and species-specific proteins or protein peaks for each profiling approach are displayed in the Venn diagram in Fig.5(A). Notably, MALDI fingerprinting turned out to be the most effective in distinguishing between the two species, as the shared protein peak number accounted for only 4.4%(23/526) and 4.6%(23/498) of the total distinct peaks from *E. coli* and *B. subtilis*, respectively. The overlap of protein profiles obtained from BUP was the most substantial, slighter larger than that from TDP. The two species shared 644 common proteins in their BUP profiles, constituting 29.1%(644/2213) and 26.2%(644/2454) of the total distinct proteins detected from *E. coli* and *B. subtilis*, respectively(summary from four repeated tests).

To clarify the underlying functions of those species-common proteins, their protein-protein interaction network was mapped using STRING database and Cytoscape(gene IDs of the proteins were used for analysis)[23]. As displayed in Fig.5(B), the protein with gene ID *tsf* was revealed to be the central hub protein, exhibiting an indegree value of 156. Following *tsf*, proteins with the gene ID of *rpoB*, *tig*, *trmD* and *rpsL* showed indegree value of 148, 147, 141 and 131, respectively. The indegree value quantifies the number of interactions a protein has with other proteins, or the number of proteins that bind or interact with it. These five proteins have been reported to play key roles in the cellular processes of transcription and translation(protein

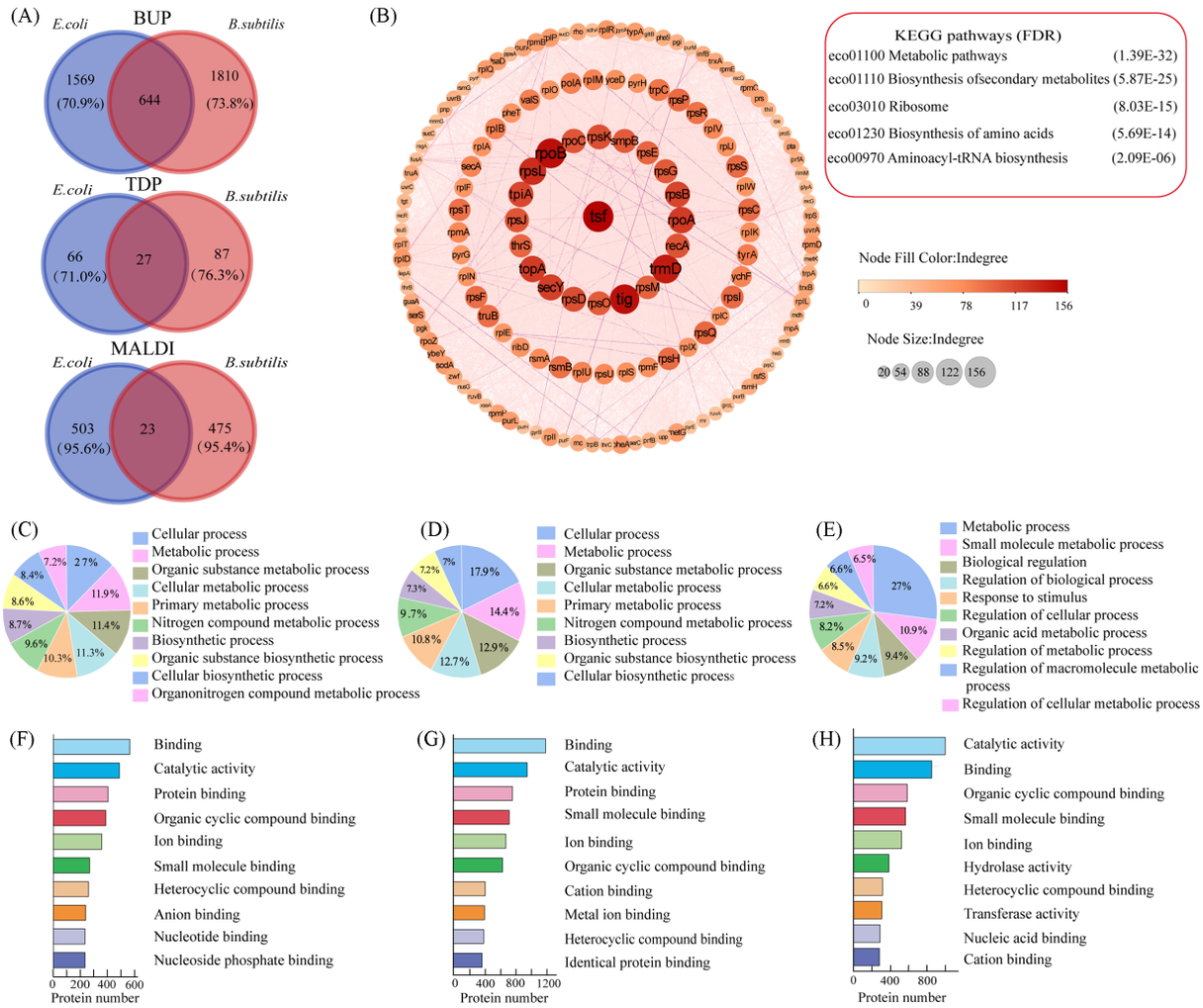


Fig. 5　BUP profile exhibited the largest overlap between bacterial species

(A) Counts of species-common and species-specific proteins or protein peaks detected from *E. coli* and *B. subtilis* using MALDI, TDP and BUP respectively; (B) the protein-protein interaction network of the BUP species-common proteins, constructed using the String database and Cytoscape with an indegree score threshold of 30; (C—E) top ten significantly enriched biological processes (BPs) from the species-common proteins (C), *E. coli*-specific proteins (D) and *B. subtilis*-specific proteins (E) detected by BUP; (F—H) top ten significantly enriched molecular functions (MFs) from the species-common proteins (F), *E. coli*-specific proteins (G) and *B. subtilis*-specific proteins (H) detected by BUP. *P*-value lower than 0.05 indicated a significant enrichment.

synthesis). Specifically, *tsf* (protein elongation factor Ts, EF-TS) is critical for protein biosynthesis[24]. Highly conserved across species, this protein assists in the binding of aminoacyl-transfer RNA (tRNA) to the ribosomal A site during translation elongation, a vital function for the translation elongation cycle in bacteria. The *tig* (protein trigger factor, TIG) is also involved in the translation process in various organisms. It binds to the ribosome during protein synthesis, and assists in the release and protection of unfolded proteins[25]. The *rpoB* (protein DNA-directed RNA polymerase subunit beta, RPOB) is an essential component of the bacterial RNA polymerase complex, responsible for transcription initiation by recognizing specific DNA sequences and recruiting the RNA polymerase holoenzyme. It is also highly conserved across bacteria, with similar sequences found in *E. coli* and *B. subtilis*[26]. The *trmD* {protein tRNA [guanine-*N*(1)-]-methyltransferase, TRMD} is an enzyme crucial for the modification of tRNA in bacteria like *E. coli* and *B. subtilis*[27]. It methylates guanine at position 1 of the tRNA molecule. This modification is essential for tRNA stability and proper functioning, deficiency in which can disrupt protein synthesis. Lastly, *rpsL* (protein small ribosomal

subunit protein uS12, RS12) is a component of the small subunit of the ribosome in *E. coli* and *B. subtilis*. It is involved in the decoding process of mRNA during translation initiation, essential for protein synthesis. A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was further performed for the 644 species-common proteins. The most significant pathways were indeed associated with the processes of translation and protein synthesis, as well as the cellular metabolic processes like the biosynthesis of secondary metabolites.

The species-common proteins, as well as the species-specific proteins, further underwent GO enrichment process to investigate their associated BPs and MFs. The analysis revealed that the species-common proteins were significantly implicated in the cellular biosynthetic processes like the organic substance biosynthesis and the cellular metabolic processes like the metabolism of organic substance and nitrogen compound[Fig.5(C)]. And the enriched MF categories are mainly related to cellular metabolism like the catalytic activity and cyclic compound binding, and also are relevant to cellular synthesis like protein binding[Fig.5(F)]. These are consistent with the protein-protein interaction profiles, emphasizing the existence of similar molecular pathways related to cellular biosynthesis and metabolism across bacterial species. The analysis of the species-specific proteins revealed that *E. coli* showed a higher enrichment in the BP category of biosynthetic process[7.3% in *E. coli*, not enriched in *B. subtilis*, Fig.5(D)] and in the MF category of protein binding[ranked 3rd in *E. coli*, not enriched in *B. subtilis*, Fig.5(G)]. Conversely, *B. subtills* exhibited greater enrichment of proteins in the BP category of metabolic process[27.0% in *B. subtilis*, 14.4% in *E. coli*, Fig.5(E)] and in the MF category of enzymatic activities like hydrolase activity and transferase activity[ranked 6th and 8th in *B. subtilis* respectively, not enriched in *E. coli*, Fig.5(H)]. These disparities reflect differences in protein synthesis and metabolic processes related to nutrition cycling, energy production and environmental adaption between the two bacterial species. Gram-negative bacteria like *E. coli* possess a periplasmic space between the inner and outer membranes, allowing for additional translational machinery and enables protein synthesis to occur in this region. The complex cell wall structure also requires specific enzymes for protein synthesis. In contrast, gram-positive bacteria like *B. subtilis* lack a periplasmic space, and thus protein synthesis is confined to the cytoplasm. Furthermore, *B. subtilis* is known for its ability to form endospores in response to nutrient limitation and environmental stress, a feature not found *E. coli*. Their respiratory systems also exhibit notable difference. *E. coli*, being facultatively anaerobic, can thrive in both aerobic and anaerobic conditions, whereas *B. subtilis* is an obligate aerobe, requiring oxygen for respiration. Those physiological and structural differences significantly contribute to their distinct metabolic properties.

4 Conclusions

In this study, we utilized three mass spectrometry methodologies——protein fingerprinting *via* MALDI-TOF MS, top-down proteomics using LC-MS/MS, and bottom-up proteomics using LC-MS/MS to profile the protein composition of gram-negative bacteria *E. coli* and gram-positive bacteria *B. subtilis*. The two proteomic approaches provided detailed information on protein identity and abundance within the bacteria, while MALDI fingerprinting facilitated rapid and high-throughput mapping of bacterial protein distribution based on the mass and abundance. Our comparative analysis revealed that bottom-up proteomics offered the highest protein coverage for both bacterial species, and also exhibited the greatest overlap in protein profiles between bacterial species. Conversely, MALDI protein fingerprinting produced the most differentiated protein profiles for distinct bacterial species, demonstrating superior detection reproducibility and effectiveness in distinguishing between bacterial species. Both MALDI fingerprinting and top-down proteomics were adept at detecting intact, undigested proteins in their native states, with a preference for abundant, stable,

and hydrophilic bacterial ribosomal proteins. Due to technology limitations, top-down proteomics identified much fewer proteins compared to the bottom-up approach; however, it provided crucial information on protein sequences and masses, facilitating the discovery of bacterial protein markers when combined with MALDI fingerprinting. The marker discovery could be essential not only for bacterial identification but also for antimicrobial resistance detection. This comprehensive study offers valuable insights that can guide the selection of protein profiling strategies for bacterial analysis.

References

[1] Zhao X., Bi H. Y., *J. Agric. Food Chem.*, **2022**, *70*(24), 7525—7534
[2] Yang K. B., Cameranesi M., Gowder M., Martinez C., Shamovsky Y., Epshtein V., Hao Z. T., Nguyen T., Nirenstein E., Shamovsky I., Rasouly A., Nudler E., *Nature*, **2023**, *622*(7981), 180—187
[3] Chirania P., Holwerda E. K., Giannone R. J., Liang X. Y., Poudel S., Ellis J. C., Bomble Y. J., Hettich R. L., Lynd L. R., *Nat. Commun.*, **2022**, *13*, 3870
[4] Schwenzer A. K., Kruse L., Jooss K., Neususs C., *Proteomics*, **2024**, *24*(3-4), e2300135
[5] Zhou F., Johnston M. V., *Electrophoresis*, **2005**, *26*(7/8), 1383—1388
[6] Lee P. Y., Saraygord-Afshari N., Low T. Y., *J. Chromatogr. A*, **2020**, *1615*, 460763
[7] Rebecca W., Scheller C., Krebs F., Wätzig H., Oltmann-Norden I., *Electrophoresis*, **2020**, *42*(3), 206—218
[8] Zhu Y. D., Qiao L., Prudent M., Bondarenko A., Gasilova N., Möller S. B., Lion N., Pick H., Gong T. Q., Chen Z. X., Yang P. Y., Lovey L. T., Girault H. H., *Chem. Sci.*, **2016**, *7*(5), 2987—2995
[9] Toby T. K., Fornelli L., Srzentic K., DeHart C. J., Levitsky J., Friedewald J., Kelleher N. L., *Nat. Protoc.*, **2019**, *14*(1), 119—152
[10] Dupree E. J., Jayathirtha M., Yorkey H., Mihasan M., Petre B. A., Darie C. C., *Proteomes*, **2020**, *8*(3), 8030014
[11] Gibson B., Wilson D. J., Feil E., Eyre-Walker A., *Proc. Biol. Sci.*, **2018**, *285*(1880), 20180789
[12] Du M. G., Yuan Z. N., Werneburg G. T., Henderson N. S., Chauhan H., Kovach A., Zhao G. P., Johl J., Li H. L., Thanassi D. G., *Nat. Commun.*, **2021**, *12*(1), 5207
[13] Nefedov A. V., Mitra I., Brasier A. R., Sadygov R. G., *J. Proteome Res.*, **2011**, *10*(9), 4150—4157
[14] van Doorn J., Ly A., Marsman M., Wagenmakers E. J., *J. Appl. Statist.*, **2020**, *47*(16), 2984 —3006
[15] Jaskolla T. W., Karas M., *J. Am. Soc. Mass Spectrom.*, **2011**, *22*(6), 976—988
[16] Zhu Y. D., Pick H., Gasilova N., Li X. Y., Lin T. E., Laeubli H. P., Zippelius A., Ho P. C., Girault H. H., *Chem*, **2019**, *5*(5), 1318—1336
[17] Yu C. S., Lin C. J., Hwang J. K., *Prot. Sci.*, **2004**, *13*(5), 1402—1406
[18] Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., Kirilovsky A., Fridman W. H., Pagès F., Trajanoski Z., Galon J., *Bioinformatics*, **2009**, *25*(8), 1091—1093
[19] Fromm S. A., O'Connor K. M., Purdy M., Bhatt P. R., Loughran G., Atkins J. F., Jomaa A., Mattei S., *Nat. Commun.*, **2023**, *14*(1), 1095
[20] Ishihama Y., Schmidt T., Rappsilber J., Mann M., Hartl F. U., Kerner M. J., Frishman D., *BMC Genom.*, **2008**, *9*, 102
[21] Zundel M. A., Basturea G. N., Deutscher M. P., *RNA*, **2009**, *15*(5), 977—983
[22] Fedyukina D. V., Jennaro T. S., Cavagnero S., *J. Biol. Chem.*, **2014**, *289*(10), 6740—6750
[23] Shannon P., Markiel A., Ozier O., Baliga N. S., Wang J. T., Ramage D., Amin N., Schwikowski B., Ideker T., *Genom. Res.*, **2003**, *13*(11), 2498—2504
[24] Burnett B. J., Altman R. B., Ferrao R., Alejo J. L., Kaur N., Kanji J., Blanchard S. C., *J. Biol. Chem.*, **2013**, *288*(19), 13917—13928
[25] Houry W., *Protein Sci.*, **2021**, *30*, 24—25
[26] Rungsirivanich P., Inta A., Tragoolpua Y., Thongwai N., *Sci. Rep.*, **2019**, *9*, 16561
[27] de Crécy-Lagard V., Ross R. L., Jaroch M., Marchand V., Eisenhart C., Brégeon D., Motorin Y., Limbach P. A., *Biomolecules*, **2020**, *10*(7), 977

(Ed.: L, W, K)