Original Article

Verification of Ribosomal Proteins of *Aspergillus fumigatus* for Use as Biomarkers in MALDI-TOF MS Identification

Sayaka Nakamura¹, Hiroaki Sato^{*,1},
Reiko Tanaka², and Takashi Yaguchi²

¹Environmental Management Research Institute, National Institute of Advanced Industrial Science and Technology (AIST),
16-1 Onogawa, Tsukuba, Ibaraki 305-8569, Japan

²Medical Mycology Research Center, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8673, Japan

We have previously proposed a rapid identification method for bacterial strains based on the profiles of their ribosomal subunit proteins (RSPs), observed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). This method can perform phylogenetic characterization based on the mass of housekeeping RSP biomarkers, ideally calculated from amino acid sequence information registered in public protein databases. With the aim of extending its field of application to medical mycology, this study investigates the actual state of information of RSPs of eukaryotic fungi registered in public protein databases through the characterization of ribosomal protein fractions extracted from genome-sequenced *Aspergillus fumigatus* strains Af293 and A1163 as a model. In this process, we have found that the public protein databases harbor problems. The RSP names are in confusion, so we have provisionally unified them using the yeast naming system. The most serious problem is that many incorrect sequences are registered in the public protein databases. Surprisingly, more than half of the sequences are incorrect, due chiefly to mis-annotation of exon/intron structures. These errors could be corrected by a combination of *in silico* inspection by sequence homology analysis and MALDI-TOF MS measurements. We were also able to confirm conserved post-translational modifications in eleven RSPs. After these verifications, the masses of 31 expressed RSPs under 20,000 Da could be accurately confirmed. These RSPs have a potential to be useful biomarkers for identifying clinical isolates of *A. fumigatus*.

Copyright © 2016 Sayaka Nakamura, Hiroaki Sato, Reiko Tanaka, and Takashi Yaguchi. This is an open access article distributed under the terms of Creative Commons Attribution License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Please cite this article as: Mass Spectrom (Tokyo) 2016; 5(1): A0049

Keywords: *Aspergillus fumigatus*, ribosomal protein, MALDI-TOF MS, protein database, sequence correction

(Received August 18, 2016; Accepted September 26, 2016)

INTRODUCTION

Aspergillus is a diverse genus of very common fungi that have high economic and social impact.¹⁾ Some strains are used industrially for microbial fermentation and production of organic compounds and enzymes. Several *Aspergillus* species are also known to be causative agents for mycoses, which has been shown to cause aspergilloses, including allergic bronchopulmonary aspergillosis, aspergilloma, and invasive aspergillosis.²⁾ Because susceptibilities to antifungal agents vary according to *Aspergillus* species, accurate identification of unknown *Aspergillus* clinical isolates is the key to selecting an appropriate antifungal agent.

Identification of *Aspergillus* species has been tradition-

ally performed based on the morphology of the conidia and conidiogeneses.^{1,2)} However, morphological discrimination is subjective and requires special skills and experience. This has led to the increasing use of DNA-based characterizations to determine *Aspergillus* species. Identification of *Aspergillus* species has been reported using the internal transcribed spacer (ITS) region between the 18S, 5.8S, and 28S ribosomal RNA (rRNA) genes,³⁾ the D1/D2 region of the 28S rRNA gene⁴⁾ and the housekeeping genes such as β -tubulin⁵⁾ and calmodulin⁶⁾ genes.

On the other hand, we have proposed a ribosomal protein based MALDI-TOF MS method for bacteria characterization.⁷⁻¹⁴⁾ Our method can identify the species of a bacteria based on the profiles of its ribosomal subunit proteins (RSPs), which are highly abundant house-keeping proteins

*Correspondence to: Hiroaki Sato, Environmental Management Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 16-1 Onogawa, Tsukuba, Ibaraki 305-8569, Japan, e-mail: sato-hiroaki@aist.go.jp

Abbreviations: RSP, ribosomal subunit protein; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; NCBI, the National Center for Biotechnology Information; PDB, potato dextrose broth; TFA, trifluoroacetic acid; CDS, coding DNA sequence; ORF, open reading frame

and easily observed by MALDI-TOF MS. The results of identification at species level and discrimination at strain level are correlated with the molecular evolution of these house-keeping proteins. Prokaryotic (bacterial) ribosomal proteins consist of more than 50 subunits, so equivalent results as analyzing many genes are obtained by using RSPs as biomarkers. The key of the RSP based method is the reliability of the reference mass list of RSP biomarkers. The preparation of the reference mass list of RSP biomarkers is supported by bioinformatics. The theoretical mass of RSP biomarkers can be calculated from their amino acid sequences registered in the public protein databases such as the National Center for Biotechnology Information (NCBI) database and UniProt Knowledgebase (UniProtKB). Therefore, this method has a potential for universal use, since it is not circumscribed by commercial databases.

To extend this ribosomal protein based method to the identification of eukaryotic *Aspergillus* species, we have first attempted to characterize RSPs of various genome-sequenced *Aspergillus* strains by MALDI-TOF MS. However, most RSPs in every strains were hard to be assigned. Here, we have found that the difficulty is mainly caused by two problems in the public protein databases.

The first problem is originated from the confusion of the nomenclature in fungi. Prokaryotic (bacterial) ribosomes consist of 57 kinds of RSPs, whereas eukaryotic ribosomes typically consist of 78 RSPs. The difference of numbers induces disagreements in the names of RSPs. So far, the nomenclature are proposed based on *Escherichia coli* in prokaryotes, while the two nomenclatures are proposed based on yeast and rats in eukaryotes. Various names based on the different nomenclatures are muddled now. Therefore, it is difficult to search information from databases and references based on RSPs' names. Although a unified naming system for RSPs has also been proposed,¹⁵⁾ this proposal is not employed in the public protein databases at this time.

The second problem is that many amino acid sequences on databases seem to be incorrect. Different from prokaryotes genes, the genes of eukaryotes including *Aspergillus* fungi have intron sequences. We have performed the homology analysis of RSPs of *Aspergillus* species, and found that there were low homology parts in amino acid sequences. Because the house-keeping ribosomal proteins should be highly conserved, we have speculated that the intron sequences may be mis-annotated. Therefore, the sequence correction of RSPs would be accomplished by combining *in silico* inspection by sequence homology analysis and the verification of expressed mass of RSPs by MALDI-TOF MS measurements.

In this paper, we have described the detailed procedures concerning the verification and correction of information of RSPs (*i.e.*, protein names, intron sequences, amino acid sequences, and post-translational modifications) using two genome-sequenced strains of *A. fumigatus* as a model.

EXPERIMENTAL

Cell culture and preparation of ribosomal protein samples

The genome-sequenced strains of *A. fumigatus* Af293 (=IFM 54229) and A1163 (=IFM 53842), the neotype strain IFM 57323^{NT}, and a clinical isolate of IFM 62104 were provided by Chiba University's Medical Mycology Research

Center. The genome-sequenced strains and IFM 57323^{NT} were grown in potato dextrose broth (PDB) medium at 25°C for three days. The IFM 62104 strain was grown in PDB medium at 37°C for four days.

After incubation, the growing medium was centrifuged at 5,800 *g* for 10 min. Fungus bodies were harvested by centrifugation, and ground (twice, for 20 s each time, at 7,000 rpm) between zirconia silica beads (*ca.* 1,300 mg, 0.1 mm in diameter) in a MagNA Lyser (Roche). After removing the beads and cell debris by centrifugation, the fungus lysates were subjected to ultra-centrifugation at 73,400 *g* for 1 h to isolate the ribosome fraction as precipitates. The resulting ribosome fraction was solubilized in 20–50 μ L 50% acetonitrile containing 1% trifluoroacetic acid (TFA), and then subjected to MALDI-TOF MS measurements.

MALDI-TOF MS measurements

Sample preparation, apparatus, and MALDI-TOF MS data acquisition methods were similar to those described in our previous papers.^{7–14)} The ribosomal protein sample solution (approx. 1 μ L) was spotted onto the MALDI target. Approx. 1 μ L sinapinic acid matrix solution at a concentration of 20 mg/mL in 50% acetonitrile with 1% trifluoroacetic acid was then overlaid and dried in air. The MALDI-TOF MS measurements were performed using an AXIMA CFR-plus time-of-flight mass spectrometer (Shimadzu/Kratos, Kyoto, Japan) in positive linear mode. More than three mass spectra for each sample were collected from more than three sample spots. External mass calibration was carried out using three peaks of ACTH (human, 1–24) ($[M+H]^+$, *m/z* 2932.6) and myoglobin ($[M+H]^+$, *m/z* 16952.6 and $[M+2H]^{2+}$, *m/z* 8476.8) as references.

Calculation of the theoretical mass of RSPs

The amino acid sequence of each RSP was obtained from the UniProtKB (<http://www.uniprot.org/>). The sequence mass of each RSP was predicted using a Compute pI/Mw tool on the ExPASy proteomics server (http://www.expasy.org/tools/pi_tool.html), with N-terminal methionine loss considered first as a possible post-translational modification. The possibilities of other modifications will be discussed below in Results and Discussion section. The theoretical mass of each expressed RSP was calculated as $[M+H]^+$ ion.

RESULTS AND DISCUSSION

Unification of the RSP name system

The nomenclature of RSPs is in a state of confusion. Names are typically composed of an alphabetical letter (L for large subunit proteins and S for small subunit proteins) and a digit, in which the numbering rule is different for each species. The first nomenclature of RSPs was proposed for bacterial (*Escherichia coli*) RSPs in 1971.¹⁶⁾ For eukaryotic RSPs, mammalian (rat) RSPs were the first to be characterized and named,¹⁷⁾ and the proposal for the yeast (*Saccharomyces cerevisiae*) RSP naming system¹⁸⁾ was followed. To solve the nomenclatural confusion, a unified naming system for RSPs has been discussed, in which homologous RSPs are assigned with the same name, independent of organism species. The first proposal was based on a protein family,¹⁹⁾ and it was further modified to a new sys-

Table 1. Correct names of ribosomal proteins of *A. fumigatus* strains and their accession No. in UniProt.

Protein name		<i>A. fumigatus</i> Af293		<i>A. fumigatus</i> A1163	
Yeast name	Unified name	Designation in UniProt	Accession No. in UniProt	Designation in UniProt	Accession No. in UniProt
Large subunit proteins					
L1	uL1	Ribosomal protein	E9QU85	Ribosomal protein	B0XQU0
L2	uL2	L8, putative	Q4WTW7	L8, putative	B0Y3E2
L3	uL3	L3	Q8NKF4	L3	B0XSL2
L4	uL4	L4, putative	Q4WEH4	L4	B0Y2P9
L5	eL18	L5, putative	Q4WSG1	L5	B0XR75
L6	eL6	L6	Q4WSZ2	L6	B0XQN2
L7	uL30	L7	Q4W9S6	L7	B0YEG9
L8	eL8	L7A	Q4WLM5	L7A	B0XM24
L9	uL6	L9, putative	Q4WTJ3	L9, putative	B0XQ32
L10	uL16	L10	Q4X1P8	L10	B0XRW7
L11	uL5	L11	Q4WP20	L11	B0Y5W6
L12	uL11	L12	Q4WK81	L12	B0XMZ1
L13	eL13	L13	Q4W9L9	L13	B0YEB1
L14	eL14	L14	Q4WD82	L14	B0YD67
L15	eL15	L15	Q4WJV5	L15	B0XNP4
L16	uL13	L16a	Q4WJH1	L16a	B0XPG3
L17	uL22	L17	Q6MY48	L17	B0XMS0
L18	eL18	L18	Q4X279	L18	B0XW36
L19	eL19	L19	Q4X220	L19	B0XW91
L20	eL20	L20	Q4WJW9	L20	B0XNN1
L21	eL21	L21, putative	Q4WWT1	L21, putative	B0XYU3
L22	eL22	L22, putative	Q4WYA0	L22, putative	B0XWY6
L23	uL14	Alkaline serine protease	Q4WI20	Alkaline serine protease	B0XUE5
L24	eL24	L24a	Q4WCU3	L24a	B0YDK4
L25	uL23	L23	Q4WTP5	L23	B0Y372
L26	uL24	L26	Q4WM42	L26	B0Y8G7
L27	eL27	L27	Q4WJD7	L27e	B0XJP7
L28	uL15	L27a, putative	Q4WWF0	L27a, putative	B0XZ73
L29	eL29	L29, putative	Q4WKA9	L29, putative	B0XMW3
L30	eL30	L30, putative	Q4X1P9	L30, putative	B0XRW6
L31	eL31	L31e	Q4WLL1	L31e	B0XLZ0
L32	eL32	L32	Q4WZN0	L32	B0XV02
L33	eL33	L35Ae	Q4WX73	L35Ae	B0XYE5
L34	eL34	L34, putative	Q4WI54	L34 protein, putative	B0XUB0
L35	uL29	L35	Q4WT53	L35	B0XQH1
L36	eL36	L36	Q4WNZ0	L36	B0Y5T6
L37	eL37	L37	Q4WWR1	L37	B0XYW1
L38	eL38	L38, putative	Q4WP31	Rpl38, putative	B0Y5X8
L39	—	—	—	—	—
L40	eL40	Ubiquitin UbiA, putative	A4D9S6	Ubiquitin UbiA, putative	B0XNB9
L42	eL42	L44	Q4X205	Uncharacterized protein	B0XWA6
L43	eL43	L37a	Q4WZH8	L37a	B0XVB7
P0	uL10	P0	Q4WJR3	—	—
P1	P1/P2	P1	Q9HGV0	P1	B0XPQ5
P2	P1/P2	P2	Q9UUZ6	P2/allergen Asp F 8	B0XS47
Small subunit proteins					
S0	uS2	S0	Q4WYK1	S0	B0XWG9
S1	eS1	S1	Q4WTM9	S1	B0Y356
S2	uS5	S5	Q4WAI8	S5	B0YBW2
S3	uS3	S3, putative	Q4WJK8	S3, putative	B0XP55
S4	eS4	S4	Q4WWR9	S4	B0XYV4
S5	uS7	S5, putative	Q4WRU9	S5, putative	B0XN49
S6	eS6	S6	Q4WPX5	S6	B0Y6R5
S7	eS7	S7e	Q4WXU5	S7e	B0XXS8
S8	eS8	S8	Q4WJZ0	S8	B0XNE5
S9	uS4	S9	Q4WWT2	S9	B0XYU2
S10	eS10	S10b	Q4WLQ8	S10b	B0Y8V2
S11	uS17	S11	Q4WHU8	S11	B0XUT5
S12	eS12	S12	Q4WJM1	S12	B0XP41
S13	uS15	S13	Q4WJG9	S13	B0YCP0
S14	uS11	S11	Q4X1C6	S11	B0XS79
S15	uS19	S15, putative	Q4X1G1	S15, putative	B0XS46
S16	uS9	Rpl16, putative	Q4X1C0	S9	B0XS84
S17	eS17	S17, putative	Q4X1E0	S17, putative	B0XS66
S18	uS13	S13p/S18e	Q4WLH1	S13p/S18e	B0XM75
S19	eS19	S19	Q4WJN7	S19	B0XP26
S20	uS10	S10a	Q4WIE3	S10a	B0XTV5
S21	eS21	S21	Q4WI01	S21	B0XUN2
S22	uS8	S22	Q4WRN1	S22	B0XN14
S23	uS12	S23	Q873W8	S23 (S12)	B0XQ66
S24	eS24	S24	Q4WAQ6	S24	B0YC29
S25	eS25	S25, putative	Q4WRF2	—	—
S26	eS26	S26	Q4WJ94	S26	B0XPP9
S27	eS27	S27	Q4WWP9	S27	B0XYX4
S28	eS28	S28e	Q4WGB8	S28e	B0YCF7
S29	uS14	S29, putative	Q4WLQ2	S29, putative	B0Y8V8
S30	eS30	S30/ubiquitin fusion	Q4WCU4	S30/ubiquitin fusion	B0YDK3
S31	eS31	Ubiquitin (UbiC), putative	Q4WXZ8	Ubiquitin (UbiC), putative	B0XXM3

tem for naming RSPs proposed in 2014.¹⁵⁾ Unfortunately, the new unified naming system¹⁵⁾ is not employed in the public protein databases at this time. This paper therefore provisionally adopts the yeast name system¹⁸⁾ for convenience of homology search, since *Aspergillus* and *Saccharomyces* are related organisms.

To unify the name of each *A. fumigatus* RSP into the yeast name, a homology search of *A. fumigatus* RSPs was performed using the NCBI blastp program (<http://blast.ncbi.nlm.nih.gov/>) to seek the RSPs of *S. cerevisiae*. Table 1 summarizes the data on *A. fumigatus* RSPs, such as the accession number and registered name in UniProtKB, the name using the yeast name system, and the name employing the unified naming system as a reference for the future. Most of the RSPs of *A. fumigatus* registered in UniProtKB were named using the yeast name system. The remaining RSPs, named using another naming system, were renamed to the yeast name in the following manner. For example, L37a of *A. fumigatus* Af293 registered in UniProtKB as Q4WZH8, showed high homology with *S. cerevisiae* L43A (where A means one of the duplicate genes). Because L37a is based on the mammalian ribosome name, it is renamed to L43 in line with the yeast name (incidentally, it corresponds to eL43 in the unified name¹⁵⁾). This L43 protein showed more than 95% similarity to L43 of *A. clavatus* NRRL1, *A. terreus* NIH2624, and *A. niger* CBS513.88. These homologs of another *Aspergillus* species are registered using the yeast name. To prevent such confusion, all RSPs of *A. fumigatus* Af293 and A1163 were unified to the yeast name.

Ribosomal proteins L40, S30, and S31 are synthesized as fusion proteins with ubiquitin^{20,21)} (note that S31 is assigned as S27a in ref. 20). There are several different types of ubiquitin, all of which are highly conserved and well characterized, so identification of the ubiquitin part in a fusion protein sequence is an easy task. In UniProtKB, L40 is registered as “Ubiquitin UbiA” (accession numbers: A4D9S6 for Af293 and B0XNB9 for A1163). In this fusion protein, ubiquitin forms a part of the N-terminal-side 76 amino acids, whereas L40 is the remaining part of C-terminal-side 52 amino acids.²⁰⁾ In the case of S31 registered as “Ubiquitin (UbiC)” (Q4WXZ8 and B0XXM3), since the N-terminal side 76 amino acid is ubiquitin, the remaining C-terminal side chain is S31. To increase the confusion, S30, which is registered as “S30/ubiquitin fusion” (Q4WCU4 and B0YDK3), is not a fusion protein, and the full length of the registered amino acid sequence corresponds to S30.

The page for alkaline serine protease in UniProtKB (Q4WI20) includes the “ribosomal protein L14P family” in the Family & Domains field. L14P is the bacterial RSP name, which corresponds to yeast L23. The amino acid sequence of this protein showed a high homology with L23 of *S. cerevisiae*, so the name of this protein was changed to L23. All the names of *A. fumigatus* RSPs were verified and changed to the yeast name using this procedure.

Observation of MALDI-TOF mass spectra and peak assignment

The next step is the calculation of the theoretical mass of each RSP based on the corresponding amino acid sequences obtained from UniProtKB. The theoretical mass was then compared with the observed mass. Figure 1 shows the mass spectra of the ribosomal protein fraction prepared from

A. fumigatus Af293 and A1163, with the peaks under m/z 20,000 assigned. Finally, we were able to assign 31 RSPs, but at this stage only eight peaks could be assigned for each strain when using the registered amino acid sequences in UniProtKB and only if taking N-terminal methionine loss into account. These peaks are indicated as the boxed protein names in Fig. 1. In our previous studies of bacterial RSPs,^{7–14)} most could be assigned by referring to the theoretical mass calculated from the registered amino acid sequences while only considering N-terminal methionine loss. The main reasons why only eight RSPs could be assigned might be speculated as (1) many incorrect amino acid sequences are registered in the protein databases and (2) post-translational modifications occur, other than N-terminal methionine losses. The following section discusses the actual state of the registered information and how to correct erroneous sequences and speculate post-translational modifications.

Correction of registered amino acid sequences

Incorrectly registered amino acid sequences in bacterial RSPs were mainly caused by mis-annotation of start codons.^{9,12)} In this study, we found that incorrect sequences of eukaryotic RSPs of *A. fumigatus* were caused by mis-annotation of the exon/intron structure. Accurate coding DNA sequence (CDS) was determined by a combination of informatics procedures involving a homology search and a manual inspection of the DNA sequence of the corresponding genes, followed by confirmation of the correct mass of the expressed RSPs by MALDI-TOF MS measurements. The details of the correction procedures are described below.

The amino acid sequences of RSPs tend to be highly preserved, and show high homology with other species' proteins. However, RSPs not assigned at the beginning tended to have different sequence lengths registered in the database. For example, Fig. 2 shows the multiple alignment of S29 of *A. fumigatus*, for which the peak could not be observed at the calculated mass, and other *Aspergillus* species such as *A. clavatus* NRRL1, *A. nidulans* FGSC A4, and *A. niger* CBS513.88. The amino acid sequences between 1 and 54 are highly conserved between these strains, while the homology and length of C-terminal side are markedly different. Eukaryotic S29 is highly conserved from yeast to humans,²²⁾ and has 56 amino acids containing a specific zinc finger-like motif (C-x-x-C).²³⁾ Since S29 of *A. niger* CBS513.88 and *A. nidulans* FGSC A4 have the zinc finger-like motif and 56 amino acid sequences, these sequences are more likely to be right. The DNA sequence of the S29 gene (*rps29*) of *A. fumigatus* Af293 was therefore compared to that of *A. niger* CBS513.88.

The *rps29* gene of *A. niger* CBS513.88 is located on c482296-481588 (708 bp) of supercontig An06 (NT_166522.1 in NCBI) and consists of 5 exons and 4 introns. The *rps29* gene of *A. fumigatus* Af293 is located on c3211760-3211177 (583 bp) of chromosome 6 (NC_007199.1 in NCBI) and consists of 5 exons and 4 introns. Figure 3 shows the sequence alignment of these genes, with exon regions underlined. In spite of the high sequence similarity of exon-1 to exon-3, the length of exon-4 is different: it is 57 bp for *A. niger* CBS513.88 and 61 bp for *A. fumigatus* Af293. Thus, the differences of 4 bp indicated by the box in Fig. 3 seems to be a redundancy. If these 4 bp are assigned as an intron, as they

Fig. 1. MALDI mass spectra of ribosomal protein fractions obtained from (a) *A. fumigatus* Af293 and (b) *A. fumigatus* A1163. The peak assignments with box indicates the RSPs assigned when using the registered amino acid sequences in UniProtKB and without considering any post-translational modifications except for N-terminal methionine loss. The peak labels with +Ac, +Me, and +Hyd indicate acetylation, methylation, and hydroxylation as the post-translational modifications.

are in *A. niger* S29, a frame shift occurs at exon-5, resulting in a shift in the stop-codon (*i.e.*, removal of the redundant italic sequence at the 3'-side in Fig. 3). The numbers of base pairs now match, with the correct amino acid sequence being 56 aa, which is common to a wide range of eukaryotes. The correct amino acid sequence of S29 showed more than 90% similarity to that of *A. clavatus* and *A. nidulans*. The correct mass of S29 ion ($[M+H]^+$) was calculated as 6646.7 Da, and the corresponding peak was clearly observed in the mass spectra, as shown in Fig. 1. The same procedure

was performed for S29 of *A. fumigatus* A1163, revealing the same sequence and mass as those of the Af293 strain.

The sequence information of L39 of both Af293 and A1163 strains was not registered in the protein databases. We tried to find the open reading frame (ORF) of the L39 gene (*rpl39*) in the genome sequence of Af293 and A1163 strains using the *rpl39* gene sequence of other *Aspergillus* species by manual inspection. As a result of a blast search performed using known *rpl39* gene sequences, highly homologous sequences of *rpl39* gene

	1	30	60
<i>A. fumigatus</i> Af293	MTHEVWYSRPRKF	GKGSRECRVCAHRAGL	IRKYGMDICRQCFREKAQDIGFYKVPLNL
<i>A. fumigatus</i> A1163	MTHEVWYSRPRKF	GKGSRECRVCAHRAGL	IRKYGMDICRQCFREKAQDIGFYKVPLNL
<i>A. clavatus</i> NRRL1	MTHEVWYSRPRKF	GKGSRECRVCAHRAGL	IRKYGMNIDCRQCFREKAQDIGFYKHLANE
<i>A. nidulans</i> FGSC A4	MTHEVWYSRPRKF	GKGSRCRVCAHRAGL	IRKYGMDICRQCFREKASDIGFFKFR
<i>A. niger</i> CBS 513. 88	MTHEVWYSRPRKF	GKGSRCRVCSHRAGL	IRKYGMNIDCRQCFREKSSDIGFHKYR
	*****	*****	*****
		78	
<i>A. fumigatus</i> Af293	FSRSISCF	TNNRYVLGAR	
<i>A. fumigatus</i> A1163	FSRSISCF	TNNRYVLGAR	
<i>A. clavatus</i> NRRL1	EQFTNNL	FPNRQFR	
<i>A. nidulans</i> FGSC A4			
<i>A. niger</i> CBS 513. 88			

Fig. 2. Comparison of amino acid sequences of S29. Bold sequences (C-x-x-C) identify the zinc finger-like motif.

(1) Data acquisition of Ribosomal protein S29 gene (*rps29*) sequence from NCBI.

A. fumigatus Af293: NC_007199.1, c3211760-3211177

A. niger CBS513.88: NT_166522.1, c482296-481588 (reference)

(2) Sequence alignment of *rps29* gene of *A. fumigatus* Af293 and *A. niger* CBS513.88.

	1	30	60	90
<i>A. fumigatus</i>	<u>ATGACTCAGAGTCCGTTTGGTACAGCCGG</u>	<u>CCCCGCAAGTAAG</u> -----GATATTCCTC	AATTCTGGAATGCGTCGATACGATTGGCT	
<i>A. niger</i>	<u>ATGACTCAGAGTCCGTTTGGTACAGCCGG</u>	<u>CCCCGCAAGTAAGTAAACGATGTCCTC</u>	GAATTGATATGTATGTGTGGGA---GGCT	
	*****	*****	*****	*****
		120	150	180
<i>A. fumigatus</i>	AATTGTATCTC-TCTATAGGTTCCGCAA	<u>GGGTTCCCGTGAATG</u> GTGTGAACG----	AAATCCCAGGCG-ATACCAGGAGGATAG	
<i>A. niger</i>	AATT-GATGTGTATTACAGGTTCCGCAA	<u>GGGTTCCCGTCTTGT</u> GTCGTATATGCGAC	ATATTCCTGAGGAGGAGATTAGGGGATTG	
	**** *	*****	*****	*****
		210	240	270
<i>A. fumigatus</i>	-----CGTCAAATATTA-T	GGA----TTGCGAGCAGGAAGGGCTAATCA	GATACT---TTGCGTTGAAA-----CAGCC	
<i>A. niger</i>	GGGGGATTTTGAAGACGAAGAGAAACAAT	GGAGGAATTACGAAATGTGAAGGAAGTTA	TGTGCTGATTGTTTGAATAATACAGCC	
	** *	*** *	*** *	*** *
		300	330	360
<i>A. fumigatus</i>	<u>GTGTTTGCGCCACCGTGTGCTGATCC</u>	<u>GCAAGTACGGGGTATGGAATGAGCCAAT</u>	AT-----CGAGATCA-ACAAGCATTG	
<i>A. niger</i>	<u>GTGTTTGCCTCCACCGCGCGGTCTGATCC</u>	<u>GCAAGTACGGGGTATGGAATCACCAGATT</u>	CTTTGCGATGCGAAATTATAGAGGGGCG	
	**** *	*****	*****	*****
		390	420	450
<i>A. fumigatus</i>	GATATGCGACTGATAG-----TGGT	TATAGATGGACATCTGCGACAAATGCTTCC	<u>GTGAGAAGGCTCAGGACATCGGTTTCTACA</u>	
<i>A. niger</i>	GAACAAATGGCTAACTGGTTTCTTCTTGGC	TACAGATGAACATCTGCGGTGATGCTTCC	<u>GTGAGAAGTCTCTGACATCGGTTTCCACA</u>	
	** *	*****	*****	*****
		480	510	540
<i>A. fumigatus</i>	<u>AGGTTG</u> TTGGTTCCACATAAGCCG----	-----CAGAGTCGGAA-----	---AAC-----CATGGCCAGG-----	
<i>A. niger</i>	<u>AGGTTG</u> TTGGTTTCCACATATCCGTATA	CACATTCCAGCAGCAGCAGCAGCATCATATC	CGCAATGGCATTGGCATGGCGAGGAGGATT	
	*****	*** *	*** *	*** *
		570	600	630
<i>A. fumigatus</i>	CTC-AACCCGACGACACTCTCGG-----	-----AGAACTGG-----	-----TTGGCA--GAAGGA--	
<i>A. niger</i>	TTCGAAGAGATGGGACTTGCACAAAAA	GAAAAGCAATGAAGAACGGGAATGGAAT	GAATTGAAATGCCACTCGACATCGAAAGAG	
	** *	*****	*** *	*** *
		660	690	720
<i>A. fumigatus</i>	--AGAGCG-----GATCAC----	----TAACAGAG-----	--TTACCTTTCTCTACAGT <u>ACCGTTAAATC</u>	
<i>A. niger</i>	CCAGAACACACATACACAGATCACGCCA	CACATACCAGAACAGAACCTAACTAA	CACCACACCCCAACAGT <u>ACCGTTAA</u> ---	
	*** *	*** *	*** *	*** *
		750	780	
<i>A. fumigatus</i>	<u>TATTTTCTCGATCCATTCTCTGTTTCACGA</u>	<u>CGAATCGTTACGTGCTCGGCGCTCGGTAG</u>		
<i>A. niger</i>	-----	-----		

Underlined sequences indicate exon region and others are intron according to the registered information on the NCBI database. Box sequence should be corrected as intron. As a result, frameshift occurred and the *italic sequence* is unnecessary.

(3) Correct CDS:

ATGACTCAGAGTCCGTTTGGTACAGCCGGCCCGCAAGTTCGGCAAGGGTTCCTGTGAA
TGCCGTGTTTGCGCCACCGTGTGCTGATCCGCAAGTACGGGATGGACATCTGCAGA
CAATGCTTCGGTGAAGAGGCTCAGGACATCGGTTTCTACAAGTACCGTTAA

(4) Correct amino acid sequence:

THEVWYSRPRKF GKGSRECRVCAHRAGL IRKYGMDICRQCFREKAQDIGFYKYR
* N-terminal methionine is lost.

(5) Correct calculated mass of $[M+H]^+$ ion: m/z 6646.7

Fig. 3. Procedures for sequence correction of S29 of *A. fumigatus* Af293.

were found in chromosome 5 of *A. fumigatus* Af293 (NC_007198.1, c1443605-1444119) and ctg_000043 of A1163 (ABDB01000043.1, c422041-421524). An alignment analysis of the putative *rpl39* gene sequences with those of several *Aspergillus* species gave the exon/intron structure and a total of 156 bp of CDS. The resulting amino acid sequences were the same between the Af293 and A1163 strains, and also the same as L39 of *A. oryzae* RIB40 and *A. flavus* AF70. The theoretical mass of L39 ion ($[M+H]^+$) was determined as 6151.2 Da, and the corresponding peak was observed as shown in Fig. 1. These results strongly support the speculated sequence and expressed mass of L39 of the *A. fumigatus* strains.

In this manner, the verification of *A. fumigatus* RSPs under 20,000 Da could be performed by a combination of manual sequence inspections and MALDI-TOF MS measurements. Surprisingly, more than half (17 of 31) of the RSPs were incorrectly registered in the public protein databases, mainly due to erroneous annotations of exon/intron structures. In addition, two RSPs were registered as fusion proteins, and L39 was absent. The corrected CDS and amino acid sequences of these 17 RSPs are summarized in the supporting information Table SI-1 for *A. fumigatus* Af293 and Table SI-2 for *A. fumigatus* A1163.

The automatic annotation of exon/intron structures after whole-genome sequencing is likely to be imperfect, since the only clue to determining introns applied is the GT-AG rule (most introns start with GT and end with AG). Because accurate determination of cDNA by mRNA sequencing is both expensive and time-consuming, a full set of experimental cDNA sequence data of *Aspergillus* RSPs has not yet been reported. Our approach appears to be a simple and effective method of speculating accurate amino acid sequences of RSPs.

Post-translational modifications

Unidentified RSPs still remained after sequence correction, suggesting the presence of post-translational modification. In this study, post-translational modifications could be speculated for 11 RSPs, as described in this section. These modifications appear to be conserved in eukaryotes.

Acetylation, especially at the N-terminus, seems to be a common post-translational modification in eukaryotic RSPs. Nine RSPs (L31, L35, S11, S15, S16, S18, S21, S24, and S28) showed clear peaks at +42 Da over the calculated sequence mass, suggesting acetylation. For example, although the amino acid sequence of S21 is slightly different between Af293 and A1163 strains, clear peaks are seen in the +42 Da position for both samples, as shown in Fig. 4.

In yeast RSPs, when the penultimate amino acid residue is serine, N-terminal methionine loss followed by N-terminal acetylation is likely to occur.^{24,25)} Among probably acetylated nine RSPs, L31, L35, and S18 have an MS- sequence at the N-terminal side. In yeast RSPs, S21 with ME- and S28 with MD- are acetylated.²⁵⁾ This information strongly suggests the acetylation of S21 and S28 of *A. fumigatus* strains with the same N-terminal sequences. Yeast S11, S15, S16, and S24 with MS- sequences are N-acetylated.²⁵⁾ However, rat S11 (in UniProtKB, P62282) and S15²⁶⁾ with MA- would also be N-acetylated. Therefore, S11 and S15 (and also probably S16) with MA- are likely to be N-acetylated.

Methylation is another possible post-translational modifi-

Fig. 4. Peak shift of +42 Da from sequence mass of S21. (a) *A. fumigatus* Af293, and (b) *A. fumigatus* A1163. Although the amino acid sequences and sequence masses of S21 are different between Af293 and A1163 strains, the observations of a common peak shift of +42 Da suggest common acetylation.

cation of RSPs. In methylation of L42 at Lys-55 is evolutionally conserved among eukaryotes.²⁷⁾ Because sequence homology around Lys-55 is high (yeast Lys-55 corresponds to Lys-50 of *A. fumigatus* by similarity), methylation is likely to be a post-translational modification of L42 of *A. fumigatus*. A clear peak could in fact be observed around m/z 12028.3, taking account of +14 Da added to the calculated sequence mass.

Prolyl dihydroxylation of eukaryotic S23 is known as an evolutionarily conserved modification,²⁸⁾ and Pro-64 is hydroxylated in yeast S23. High sequence homology around Pro-64 of S23 suggests S23 of *A. fumigatus* strains to also be hydroxylated, resulting in a +32 Da shift. The corresponding peaks could be clearly observed around m/z 15802.5.

List of ribosomal protein biomarkers and its applicability

In this way, we could finally confirm the mass of 31 of 50 expressed RSPs under 20,000 Da. Most of the intense peaks observed under m/z 20,000 could be identified, as shown in Fig. 1. Unidentified RSPs are probably caused by low ionization efficiency due to the acidic properties and unclear post-translational modifications (we found more putative methylated and acetylated RSPs, but they are omitted in this paper due to a lack of supporting references). Tables 2 and 3 summarize the assigned ribosomal proteins of *A. fumigatus* Af293 and A1163 strains, together with calculated masses and possible post-translational modifications. Almost all identified RSPs have the same sequence and mass except for S21 with only one amino acid difference.

To confirm the applicability of the reference mass list, RSPs of the neotype strain IFM 57323^{NT} and a clinical isolate IFM 62104 were further characterized. Because the criteria of species identification is the similarity to the type strain, the characterization of RSPs of IFM 57323^{NT} would be important to establish the reliable biomarker list for the identification of *A. fumigatus*. The characterization of the

Table 2. Assigned ribosomal subunit proteins of *A. fumigatus* Af293.

Protein name		Accession No. in UniProt	pI	Calculated mass as [M+H] ⁺	Observed mass	Sequence correction	Modifications ^a
Yeast name	New system						
L40	eL40	A4D9S6	9.5	6002.3	6001.7	yes	
L39	eL39	—	12.6	6151.2	6151.4	yes	–Met
S29	uS14	Q4WLQ2	10.1	6646.7	6646.5	yes	–Met
S30	eS30	Q4WCU4	11.5	6789.1	6789.5		–Met
L29	eL29	Q4WKA9	11.6	7456.6	7457.0		–Met
S28	eS28	Q4WGB8	10.9	7710.0	7710.0	yes	+Ac
S31	eS31	Q4WXZ8	9.8	9134.9	9133.9	yes	
L38	eL38	Q4WP31	10.3	9153.8	9154.5		–Met
L43	eL43	Q4WZH8	10.5	10025.8	10024.5		–Met
S21	eS21	Q4WI01	8.5	10052.2	10053.2		+Ac
L37	eL37	Q4WWR1	11.0	10386.9	10386.4	yes	–Met
L30	eL30	Q4X1P9	9.9	11171.1	11170.9		–Met
L36	eL36	Q4WNZ0	11.9	11869.8	11869.4	yes	–Met
L42	eL42	Q4X205	10.5	12028.3	12028.1	yes	–Met, +Me
L33	eL33	Q4WX73	10.2	12215.1	12215.2	yes	–Met
L34	eL34	Q4WI54	9.6	13164.5	13164.4	yes	–Met
S26	eS26	Q4WJ94	10.9	13338.7	13337.7		–Met
L31	eL31	Q4WLK1	10.5	13919.1	13918.5		–Met, +Ac
L35	uL29	Q4WT53	11.1	14532.0	14533.0	yes	–Met, +Ac
L32	eL32	Q4WZN0	11.3	14836.6	14835.9		–Met
L26	uL24	Q4WM42	10.9	14979.4	14979.1	yes	–Met
S24	eS24	Q4WAQ6	10.7	15226.6	15226.0		–Met, +Ac
L27	eL27	Q4WJD7	10.5	15682.6	15683.5	yes	
S23	uS12	Q873W8	10.5	15802.5	15801.9		–Met, +Hyd (2)
S16	uS9	Q4X1C0	10.2	15883.4	15881.9		–Met, +Ac
S17	eS17	Q4X1E0	10.0	16089.5	16087.9	yes	–Met
S19	eS19	Q4WJN7	9.6	16351.4	16350.6	yes	–Met
L28	uL15	Q4WWF0	10.4	16631.1	16630.9		–Met
S15	uS19	Q4X1G1	10.1	17626.5	17626.5		–Met, +Ac
S18	uS13	Q4WLH1	10.5	17779.5	17780.5	yes	–Met, +Ac
S11	uS17	Q4WHU8	10.8	18478.6	18480.7	yes	–Met, +Ac

^a –Met: N-Methionine loss, +Ac: acetylation, +Me: methylation, +Hyd: hydroxylation.Table 3. Assigned ribosomal subunit proteins of *A. fumigatus* A1163.

Protein name		Accession No. in UniProt	pI	Calculated mass as [M+H] ⁺	Observed mass	Sequence correction	Modifications ^a
Yeast name	New system						
L40	eL40	B0XNB9	9.5	6002.3	6001.9	yes	
L39	eL39	—	12.6	6151.2	6151.5	yes	–Met
S29	uS14	B0Y8V8	10.1	6646.7	6646.0	yes	–Met
S30	eS30	B0YDK3	11.5	6789.1	6790.8 ^b		–Met
L29	eL29	B0XMW3	11.6	7456.6	7457.5		–Met
S28	eS28	B0YCF7	10.9	7710.0	7709.7	yes	+Ac
S31	eS31	B0XXM3	9.8	9134.9	9135.0	yes	
L38	eL38	B0Y5X8	10.3	9153.8	9154.8		–Met
L43	eL43	B0XVB7	10.5	10025.8	10025.2 ^b		–Met
S21	eS21	B0XUN2	8.5	10038.1	10037.7		+Ac
L37	eL37	B0XYW1	11.0	10386.9	10385.9	yes	–Met
L30	eL30	B0XRW6	9.9	11171.1	11170.7		–Met
L36	eL36	B0Y5T6	11.9	11869.8	11869.3	yes	–Met
L42	eL42	B0XWA6	10.5	12028.3	12027.9 ^b	yes	–Met, +Me
L33	eL33	B0XYE5	10.2	12215.1	12214.8	yes	–Met
L34	eL34	B0XUB0	9.6	13164.5	13166.1	yes	–Met
S26	eS26	B0XPP9	10.9	13338.7	13336.9		–Met
L31	eL31	B0XLZ0	10.5	13919.1	13918.5		–Met, +Ac
L35	uL29	B0XQH1	11.1	14532.0	14531.4	yes	–Met, +Ac
L32	eL32	B0XV02	11.3	14836.6	14836.3		–Met
L26	uL24	B0Y8G7	10.9	14979.4	14979.3	yes	–Met
S24	eS24	B0YC29	10.7	15226.6	15225.9		–Met, +Ac
L27	eL27	B0XPJ7	10.5	15682.6	15682.3	yes	
S23	uS12	B0XQ66	10.5	15802.5	15802.2		–Met, +Hyd (2)
S16	uS9	B0XS84	10.2	15883.4	15884.1 ^b		–Met, +Ac
S17	eS17	B0XS66	10.0	16089.5	16089.2	yes	–Met
S19	eS19	B0XP26	9.6	16351.4	16350.7	yes	–Met
L28	uL15	B0XZ73	10.4	16631.1	16630.4		–Met
S15	uS19	B0XS46	10.1	17626.5	17626.0		–Met, +Ac
S18	uS13	B0XM75	10.5	17779.5	17779.2	yes	–Met, +Ac
S11	uS17	B0XUT5	10.8	18478.6	18479.8	yes	–Met, +Ac

^a –Met: N-Methionine loss, +Ac: acetylation, +Me: methylation, +Hyd: hydroxylation. ^b Shoulder peak.

Fig. 5. Partial MALDI mass spectra of ribosomal protein fractions obtained from (a) clinical isolate IFM 62104, compared with those from *A. fumigatus* (b) Af293 and (c) A1163 strains. The right mass spectra are expanded between m/z 9,950–10,150.

clinical isolate IFM 62104, which have been already identified as *A. fumigatus*, was performed as a demonstration for the analysis of real samples.

Figure 5 shows the partial mass spectra of ribosomal protein fractions obtained from (a) the Af293, (b) A1163, (c) IFM 57323^{NT}, and (d) IFM 62104 (whole mass spectra of IFM 57323^{NT} and IFM 62104 are shown in Figs. SI-1 and SI-2 in the supporting information). In this mass range, seven identified RSPs (S31, L38, L43, S21, L37, L30, and L36) are commonly observed. Here, of two types of S21, the peak for IFM 57323^{NT} and IFM 62104 appeared the same as S21 of A1163. In the entire mass spectra, all 31 RSP biomarkers could be observed for the IFM 57323^{NT} and IFM 62104 strains. These results suggest that the reference mass list can be used as a clue for the species identification of *A. fumigatus*.

CONCLUSION

In this study, we have investigated the actual state of RSPs in the public protein databases by characterizing the RSPs of genome-sequenced strains of *A. fumigatus* Af293 and A1163. As a result, we could solve the problems of the registered information of RSPs in the public protein databases.

As for the problem concerning the confusion of the nomenclature, all the RSPs' names were verified and unified to the names based on yeast which is most prevalent in the public protein databases (also listed under the new unified naming system¹⁵). As for the second problem originated from incorrect sequence information, we have pointed out that more than half of the *A. fumigatus* RSPs are incorrect mainly due to mis-annotation of exon/intron structures. Because RSPs are highly conserved, we could easily find out the candidates of the correct sequences, and verify them by comparing the theoretical mass with the observed mass. In addition, the post translational modifications such as acetylation and methylation could also be confirmed.

By solving these problems, we have successfully completed the reference mass list of two genome-sequenced strains of *A. fumigatus*. By using the completed sequence information of the RSPs of *A. fumigatus* as a reference, information on the RSPs of other related fungal strains can be more easily verified by combining *in silico* inspection with MALDI-TOF MS measurements. We are proceeding with the characterization of RSPs of other *Aspergillus* genome-sequenced strains to make reliable lists of biomarker RSPs

for identification of *Aspergillus* species. Once the *Aspergillus* RSP biomarker lists have been compiled, ribosomal protein-based MALDI-TOF MS is anticipated to be a powerful and reliable tool in the field of clinical microbiology.

Acknowledgements

This work was supported in part by a research grant from the Institute for Fermentation, Osaka (IFO), JSPS Kakenhi Grant Number 25430198, and the National Bioresource Project (Pathogenic Microbes) in Japan (<http://www.nbrp.jp/>).

Supporting information

Wide range of MALDI mass spectra of ribosomal protein fractions of the neotype strain IFM 57323^{NT} and a clinical isolate IFM 62104. Correct CDS and amino acid sequences of RSPs of *A. fumigatus* Af293 and A1163.

REFERENCES

- 1) R. A. Samson, C. M. Visagie, J. Houbraken, S. B. Hong, V. Hubka, C. H. W. Klaassen, G. Perrone, K. A. Seifert, A. Susca, J. B. Tancey, J. Varga, S. Kocsube, G. Szigeti, T. Yaguchi, J. C. Frisvad. Phylogeny, identification and nomenclature of the genus *Aspergillus*. *Stud. Mycol.* 78: 141–173, 2014.
- 2) J. P. Latge. *Aspergillus fumigatus* and aspergillosis. *Clin. Microbiol. Rev.* 12: 310–354, 1999.
- 3) T. Henry, P. C. Iwen, S. H. Hinrichs. Identification of *Aspergillus* species using internal transcribed spacer regions 1 and 2. *J. Clin. Microbiol.* 38: 1510–1515, 2000.
- 4) H. P. Hinrikson, S. F. Hurst, T. J. Lott, D. W. Warnock, C. J. Morrison. Assessment of ribosomal large-subunit D1–D2, internal transcribed spacer 1, and internal transcribed spacer 2 regions as targets for molecular identification of medically important *Aspergillus* species. *J. Clin. Microbiol.* 43: 2092–2103, 2005.
- 5) N. L. Glass, G. C. Donaldson. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Appl. Environ. Microbiol.* 61: 1323–1330, 1995.
- 6) S. B. Hong, S. J. Go, H. D. Shin, J. C. Frisvad, R. A. Samson. Polyphasic taxonomy of *Aspergillus fumigatus* and related species. *Mycologia* 97: 1316–1329, 2005.
- 7) L. Sun, K. Teramoto, H. Sato, M. Torimura, H. Tao, T. Shintani. Characterization of ribosomal proteins as biomarkers for matrix-assisted laser desorption/ionization mass spectral identification of *Lactobacillus plantarum*. *Rapid Commun. Mass Spectrom.* 20: 3789–3798, 2006.
- 8) K. Teramoto, H. Sato, L. Sun, M. Torimura, H. Tao, H. Yoshikawa, Y. Hotta, A. Hosoda, H. Tamura. Phylogenetic classification of *Pseudomonas putida* by MALDI-MS using ribosomal proteins as biomarkers. *Anal. Chem.* 79: 8712–8719, 2007.
- 9) K. Teramoto, H. Sato, L. Sun, M. Torimura, H. Tao. A simple intact protein analysis by MALDI-MS for characterization of ribosomal proteins of two genome-sequenced lactic acid bacteria and verification of their amino acid sequences. *J. Proteome Res.* 6: 3899–3907, 2007.
- 10) Y. Hotta, K. Teramoto, H. Sato, H. Yoshikawa, A. Hosoda, H. Tamura. Classification of genus *Pseudomonas* by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* operon at strain level. *J. Proteome Res.* 9: 6722–6728, 2010.
- 11) Y. Hotta, J. Sato, H. Sato, A. Hosoda, H. Tamura. Classification of the genus *Bacillus* based on MALDI-TOF MS analysis of ribosomal proteins coded in *S10* and *spc* operons. *J. Agric. Food Chem.* 59: 5222–5230, 2011.

- 12) H. Sato, K. Teramoto, Y. Ishii, K. Watanabe, Y. Benno. Ribosomal protein profiling by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for phylogeny-based subspecies resolution of *Bifidobacterium longum*. *Syst. Appl. Microbiol.* 34: 76–80, 2011.
- 13) Y. Hotta, H. Sato, A. Hosoda, H. Tamura. MALDI-TOF MS analysis of ribosomal proteins coded in *S10* and *spc* operons rapidly classified the *Sphingomonadaceae* as alkylphenol polyethoxylate-degrading bacteria from the environment. *FEMS Microbiol. Lett.* 330: 23–29, 2012.
- 14) H. Sato, M. Torimura, M. Kitahara, M. Ohkuma, Y. Hotta, H. Tamura. Characterization of the *Lactobacillus casei* group based on the profiling of ribosomal proteins coded in *S10-spc-alpha* operons as observed by MALDI-TOF MS. *Syst. Appl. Microbiol.* 35: 447–454, 2012.
- 15) N. Ban, R. Beckmann, J. H. D. Cate, J. D. Dinman, F. Dragon, S. R. Ellis, D. L. J. Lafontaine, L. Lindahl, A. Liljas, J. M. Lipton, M. A. McAlear, P. B. Moore, H. F. Noller, J. Ortega, V. G. Panse, V. Ramakrishnan, C. M. T. Spahn, T. A. Steitz, M. Tchorzewski, D. Tollervey, A. J. Warren, J. R. Williamson, D. Wilson, A. Yonath, M. Yusupov. A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* 24: 165–169, 2014.
- 16) H. G. Wittmann, G. Stoffler, I. Hindenna, C. G. Kurland, L. Randallh, E. A. Birge, M. Nomura, E. Kaltschm, S. Mizushim, R. R. Traut, T. A. Bickle. Correlation of 30S ribosomal proteins of *Escherichia coli* isolated in different laboratories. *Mol. Gen. Genet.* 111: 327–333, 1971.
- 17) I. G. Wool, Y. L. Chan, A. Gluck. Structure and evolution of mammalian ribosomal proteins. *Biochem. Cell Biol.* 73: 933–947, 1995.
- 18) W. H. Mager, R. J. Planta, J. P. G. Ballesta, J. C. Lee, K. Mizuta, K. Suzuki, J. R. Warner, J. Woolford. A new nomenclature for the cytoplasmic ribosomal proteins *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 25: 4872–4875, 1997.
- 19) L. Jenner, S. Melnikov, N. G. de Loubresse, A. Ben-Shem, M. Isakova, A. Urzhumtsev, A. Meskauskas, J. Dinman, G. Yusupova, M. Yusupov. Crystal structure of the 80S yeast ribosome. *Curr. Opin. Struct. Biol.* 22: 759–767, 2012.
- 20) Y. L. Chan, K. Suzuki, I. G. Wool. The carboxyl extensions of 2 rat ubiquitin fusion proteins are ribosomal-proteins S27a and L40. *Biochem. Biophys. Res. Commun.* 215: 682–690, 1995.
- 21) J. Olvera, I. G. Wool. The carboxyl extension of a ubiquitin-like protein is rat ribosomal-protein S30. *J. Biol. Chem.* 268: 17967–17974, 1993.
- 22) L. N. Liu, S. C. Zhang, Z. H. Liu, H. Y. Li, M. Liu, Y. J. Wang, L. F. Ma. Ribosomal proteins L34 and S29 of amphioxus *Branchiostoma belcheri tsingtauense*: cDNAs cloning and gene copy number. *Acta Biochim. Pol.* 52: 857–862, 2005.
- 23) Y. L. Chan, K. Suzuki, J. Olvera, I. G. Wool. Zinc finger-like motifs in rat ribosomal-proteins S27 and S29. *Nucleic Acids Res.* 21: 649–655, 1993.
- 24) H. Takakura, S. Tsunasawa, M. Miyagi, J. R. Warner. NH₂-terminal acetylation of ribosomal-proteins of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 267: 5442–5445, 1992.
- 25) R. J. Arnold, B. Polevoda, J. P. Reilly, F. Sherman. The action of N-terminal acetyltransferases on yeast ribosomal proteins. *J. Biol. Chem.* 274: 37035–37040, 1999.
- 26) M. Kitagawa, S. Takasawa, N. Kikuchi, T. Itoh, H. Teraoka, H. Yamamoto, H. Okamoto. *rig* encodes ribosomal protein S15 The primary structure of mammalian ribosomal protein S15. *FEBS Lett.* 283: 210–214, 1991.
- 27) A. Shirai, M. Sadaie, K. Shinmyozu, J. Nakayama. Methylation of ribosomal protein L42 regulates ribosomal function and stress-adapted cell growth. *J. Biol. Chem.* 285: 22448–22460, 2010.
- 28) C. Loenarz, R. Sekirnik, A. Thallhammer, W. Ge, E. Spivakovsky, M. M. Mackeen, M. A. McDonough, M. E. Cockman, B. M. Kessler, P. J. Ratcliffe, A. Wolf, C. J. Schofield. Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. *Proc. Natl. Acad. Sci. U.S.A.* 111: 4019–4024, 2014.