



Cell-free synthesis of stable isotope-labeled internal standards for targeted quantitative proteomics



Ryohei Narumi ^a, Keiko Masuda ^b, Takeshi Tomonaga ^a, Jun Adachi ^a, Hiroki R. Ueda ^{c,d}, Yoshihiro Shimizu ^{b,e,*}

^a Laboratory of Proteome Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8, Satio-Asagi, Ibaraki, Osaka 567-0085, Japan

^b Laboratory for Single Cell Mass Spectrometry, RIKEN Quantitative Biology Center (QBiC), 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan

^c Department of Systems Pharmacology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

^d Laboratory for Synthetic Biology, RIKEN Quantitative Biology Center (QBiC), 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan

^e Laboratory for Cell-Free Protein Synthesis, RIKEN Quantitative Biology Center (QBiC), 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan

ARTICLE INFO

Article history:

Received 26 December 2017

Received in revised form

16 February 2018

Accepted 18 February 2018

Keywords:

Absolute quantification

Mass spectrometry

Cell-free protein synthesis system

In vitro translation

Targeted quantitative proteomics

PURE system

ABSTRACT

High-sensitivity mass spectrometry approaches using selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) methods are powerful tools for targeted quantitative proteomics-based investigation of dynamics in specific biological systems. Both high-sensitivity detection of low-abundance proteins and their quantification using this technique employ stable isotope-labeled peptide internal standards. Currently, there are various ways for preparing standards, including chemical peptide synthesis, cellular protein expression, and cell-free protein or peptide synthesis. Cell-free protein synthesis (CFPS) or in vitro translation (IVT) systems in particular provide high-throughput and low-cost preparation methods, and various cell types and reconstituted forms are now commercially available. Herein, we review the use of such systems for precise and reliable protein quantification.

© 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	97
2. Chemical peptide synthesis and cellular protein synthesis	98
3. Cell-free protein synthesis systems	98
4. The protein synthesis Using Recombinant Elements (PURE) system for direct cell-free peptide synthesis	101
5. Future perspectives for obtaining more peptides for high-sensitivity quantification	102
Declaration of interest	102
Acknowledgements	102
References	102

1. Introduction

Cells survive by constituting very complicated molecular networks involving various small molecules and macromolecules

including nucleic acids, proteins, carbohydrates and lipids. A wide variety of complex biological functions such as migration, predation and metabolism are orchestrated by functioning networks that contribute to prevent entropy from increasing in non-equilibrium systems, as discussed by Schrödinger [1], which essentially forms the basis of all life activities. While individual reactions in a network operate via simple mechanisms governed by thermodynamics, complexity is generated by the enormous number of reactions in the network. Therefore, the most intuitive way to

* Corresponding author. Laboratory for Cell-Free Protein Synthesis, RIKEN Quantitative Biology Center (QBiC), 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan.
E-mail address: yshimizu@riken.jp (Y. Shimizu).

Peer review under responsibility of KeAi Communications Co., Ltd.

understand this complexity is to accurately measure the information present among the individual molecules over time, and to clarify the characteristics of the entire reaction network by data assimilation with the model of an entire cellular network [2] or a partial molecular network [3]. Among the molecules involved, proteins, the expression products of genetic information, are the major cellular components responsible for almost all biochemical processes. Thus, a systems-level biological approach based on quantitative proteomics [4] is crucial for understanding life.

Traditionally, quantification of proteins has been conducted using radioactive isotopes or antigen-antibody interactions. Recently, methods using intracellular detection of protein dynamics by fluorescence or luminescence imaging coupled with genetic engineering have also been adopted [5–7]. However, for detecting and quantifying proteins as they occur in biological samples without using special procedures employing antibodies and/or genetic engineering, mass spectrometry (MS) is most suitable method, in terms of both throughput and sensitivity. MS-based quantitative global proteomics researches have revealed whole proteome maps of yeast, mouse and human cells [8–13], detecting more than 10,000 proteins in the process, demonstrating its high-throughput capabilities.

In standard liquid chromatography (LC)-MS approaches, biological samples are processed with proteases such as trypsin, which cleaves peptide bonds between the carboxyl group of arginine or lysine and the amino group of the adjacent amino acid. The resultant proteotypic peptides, specific to their parent proteins, are separated by LC and individually subjected to MS analysis, which separates each peptide peak into two dimensions based on their individual chemical properties and masses. The addition of a pre-fractionation step such as strong cation exchange (SCX) column chromatography can further increase the sensitivity by allowing three- or more-dimensional separation [14,15]. The introduction of tandem mass spectrometry, also termed MS/MS, can improve the detection resolution by determining the peptide sequences through the generation of fragment ions from precursor ions. MS/MS is an indispensable technique for both global and targeted proteomics. Sequence determination by MS/MS is necessary for shotgun analysis of various peptides in global proteomics approaches. For selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) methods that are frequently employed in targeted proteomics experiments, selection of both precursor and fragment ions by a triple quadrupole mass spectrometer is necessary to limit the number of peptides observed for each target, which enables highly sensitive analysis of target peptides in samples.

Since the intensity of peptide peaks depends on their physical properties, absolute quantification of peptides of interest by MS is achieved by utilizing a stable isotope-labeled peptide internal standard with the same sequence at a known concentration [16]. Apart from the difference in mass, all other chemical and physical properties of the two peptides are identical; hence retention times in LC, ionization efficiencies and fragmentation patterns in MS/MS are all identical. Thus, quantitative determination of peptides of interest can be carried out only by comparing the intensities of the two peaks (Fig. 1). It is over three decades since this concept was first proposed [17], and it is now possible to prepare peptide standards in various ways. In recent years, targeted quantitative proteomics using internal standards synthesized by cell-free protein synthesis (CFPS) or in vitro translation (IVT) systems have gradually attracted attention. This review focuses on the cell-free synthesis of stable isotope-labeled internal standards for targeted quantitative proteomics studies.

2. Chemical peptide synthesis and cellular protein synthesis

Various approaches currently employed for preparing stable isotope-labeled peptides are shown in Fig. 2. The methods can be roughly divided into two types, chemical synthesis from the C-terminus, and biological synthesis from the N-terminus using ribosomes, among which chemically synthesized peptides represented by Absolute Quantification (AQUA) is the most widely used method [18]. AQUA peptides are generated by solid-phase synthesis [19], in which peptides bound to a solid matrix with a free N-terminal amino group are reacted with an N-terminally protected amino acid, followed by deprotection of the amino group and washing of the solid phase. Peptides can be elongated at their N-terminus by consecutive reactions. Stable isotope labels can be introduced by replacing one residue with a stable isotope-labeled amino acid. Moreover, by using modified amino acid such as phosphorylated serine as a substrate, modified peptides that mimic post-translational modification can be produced, demonstrating the versatility of this approach.

However, there are some limitations associated with this method. It requires large amounts of isotopically labeled amino acids as substrates. Synthesis can be difficult depending on the peptide sequence due to sequence-dependent side reactions. These limitations make it difficult to reduce the cost per peptide, which is a barrier to multiplexed synthesis. To overcome these drawbacks, multiplexed chemical peptide synthesis, termed SPOT synthesis [20], is applied to establish SRM assays in a high-throughput manner using crude (impure) peptides without laborious purification [21]. A stable isotope dimethyl labeling method [22,23] where the chemically synthesized peptides are labeled with stable isotope-labeled formaldehyde is another cost-effective alternative for MS-based quantification [24,25].

Biological synthesis can be further divided into cellular and cell-free synthesis approaches (Fig. 2). Stable Isotope Labeling by Amino acids in Cell culture (SILAC) is a representative cellular synthesis method [26]. In this method, the composition of the medium is adjusted so that a stable isotope-labeled essential amino acid is necessarily introduced into the protein as it is, so that a specific amino acid of the intracellular protein is completely replaced with a labeled residue. Unlike chemical synthesis methods, stable isotope labeling can be introduced nonspecifically to intact cell proteins, enabling various peptides to be synthesized simultaneously, which is more effective for global proteomics analysis. In recent years, SILAC-based global quantitative proteomics has been applied at the tissue, organism and individual level by growing on a diet containing isotope-labeled amino acids [27–31].

Development towards targeted quantitative proteomics has also progressed. In the Absolute SILAC method, targeted analysis can be carried out by constructing expression systems for specific proteins in cells, and tryptic digests of the produced proteins are used as internal standards [32]. Cellular protein expression systems are also used in the QconCAT method, which uses tryptic digests of artificial proteins comprising a concatenation of proteotypic peptides [33,34].

3. Cell-free protein synthesis systems

In general, to prepare a specific protein sample, a recombinant protein expression system is employed in which the gene encoding the protein of interest is inserted into an expression vector and protein is produced in *Escherichia coli*, yeast or other cells. Once the method is established, it has an advantage that the target protein samples can be obtained reproducibly, and therefore, it is a method indispensable for preparation of protein samples. However, to perform protein production using this system, it is necessary to go

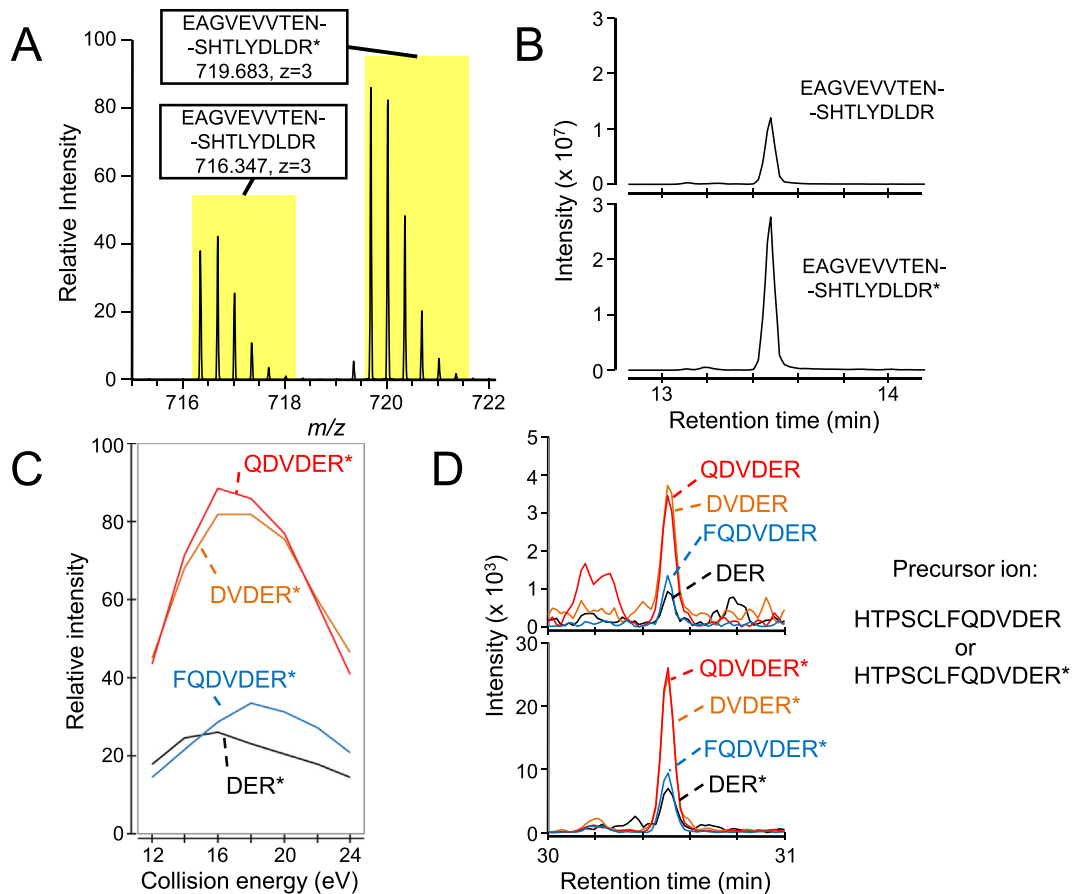


Fig. 1. Quantification of peptides in samples using stable isotope-labeled peptide internal standards.

(A) and (B) LC-MS analysis of the EAGVEVVTENSHTLYDLDR peptide. Mass spectrum (A) and mass chromatograms (B) of a peptide in the sample with a trivalent ion valency ($m/z = 716.347$) and its isotope-labeled form synthesized for quantification ($m/z = 719.683$). The retention time of the two peptides is identical, and the quantification can be performed by comparison of the peak areas in each mass chromatogram. (C) and (D) SRM analysis of the HTPSCLFQDVDER peptide. In this case, transitions of four fragment ions (FQDVDER, QDVDER, DVDER and DER) from the precursor ion (HTPSCLFQDVDER) are monitored. The collision energy for each transition is optimized using the synthesized isotope-labeled peptide (C), and both transitions from the non-labeled and labeled precursor ions are monitored (D). Quantification can be performed by comparison of peak areas in each SRM transition. The asterisk (*) indicates the isotope-labeled amino acid. MS analyses were performed according to the previous study [69].

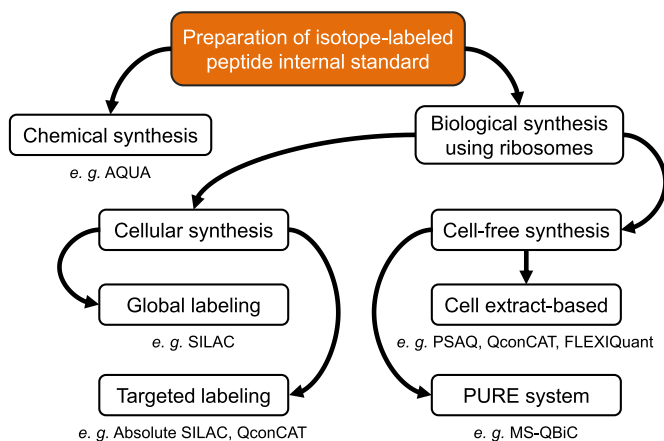


Fig. 2. Approaches currently employed for preparing stable isotope-labeled peptide internal standards.

through laborious processes such as DNA cloning, cell transformation, culturing and disruption, and protein purification from cell lysates. Additionally, proteins that are toxic to host cells are difficult to produce using this approach.

CFPS and IVT methods have been developed to replace cellular protein production systems [35–38]. These systems are not limited by the need to maintain living organisms, and can therefore synthesize proteins that are toxic to cells. Proteins can be synthesized only by adding template DNA or RNA in a high-throughput manner. Artificial manipulation of the system is easy; hence proteins containing isotope-labeled amino acids or non-natural amino acids can be synthesized by changing the components of the system. Thus, there are many advantages compared with protein expression systems using living cells.

The CFPS/IVT system was originally developed by Nirenberg and Matthaei [39] for polyphenylalanine synthesis by adding polyuridylic acid to a cell extract obtained from *E. coli* cells. This experiment revealed that UUU (TTT in the DNA sequence) encodes phenylalanine in the genetic code. This was also the first example of template-dependent synthesis of a polypeptide with a specific amino acid sequence, which underpinned the subsequent development of CFPS/IVT systems. The prototype of currently commercially available CFPS/IVT systems, in which template mRNA transcription and protein synthesis is performed simultaneously in the same reaction mixture, was developed by Zubay et al. [40]. Using a cell extract termed S30, which represents a supernatant derived from centrifugation at 30,000 g, RNA polymerases, transcription substrates (ATP, GTP, CTP and UTP), translation substrates

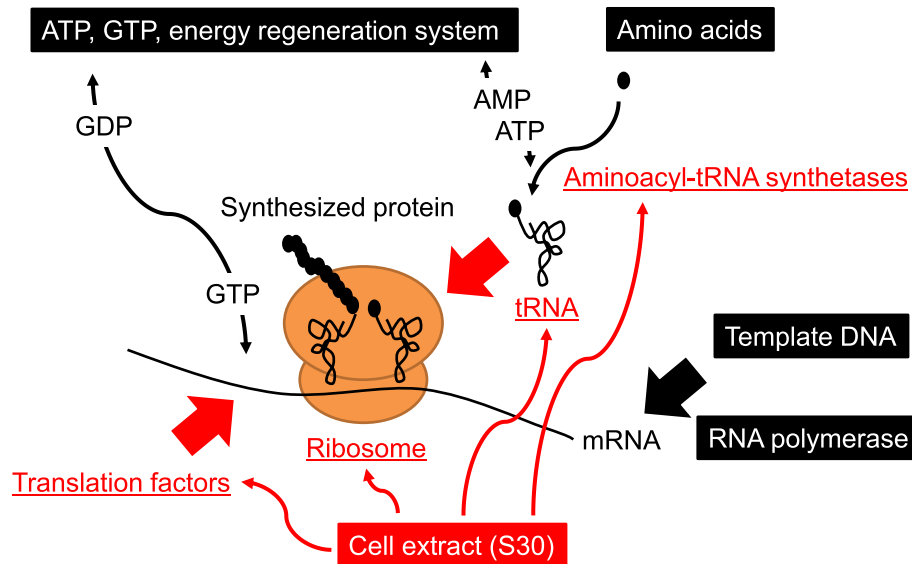


Fig. 3. Cell-free protein synthesis systems based on the S30 cell extract. Ribosomes, translation factors, tRNA and aminoacyl tRNA synthetases necessary for protein synthesis are supplied by the S30 extract. Template mRNA is synthesized in the same reaction mixture by introducing template DNA and RNA polymerase into the system. Efficient protein synthesis is performed by adding substrate amino acids, ATP, GTP and an energy regeneration system.

(amino acids), and energy substrates such as creatine phosphates and phosphoenolpyruvates are mixed, and template DNA is added to initiate transcription and translation (Fig. 3). For synthesis of internal standards for quantitative proteomics using CFPS/IVT, it is easy to use stable isotope-labeled amino acids as a substitute for non-labeled one where their concentration is sufficient under millimolar range. Furthermore, the reaction volume can be scaled from several microliters to several tens of microliters, making the method extremely cost-effective.

Methods employing cellular protein expression systems such as Absolute SILAC [32] and QconCAT [33,41,42] can also adopt CFPS/IVT systems to take advantage of these benefits. The Protein Standard Absolute Quantification (PSAQ) method, similar to Absolute SILAC but specialized for CFPS/IVT use, has also been developed for targeted quantitative proteomics [43,44]. In these methods, intact proteins or artificial proteins comprising a concatenation of proteotypic peptides are expressed using *E. coli* cell extracts in the presence of a stable isotope-labeled amino acid. The products are subjected to tryptic digestion and analyzed simultaneously with samples of interest, as occurs in methods using cellular protein expression systems. It should be noted that, in methods using cellular expression systems, the homogeneity of isotope-labeled peptides is reduced due to conversion of labeled amino acids into other molecules, or *vice versa*, by cellular metabolic enzymes, which leads to inaccuracies in quantification [32,45]. Although commercially available CFPS/IVT kits can be used in proteomics studies, a similar problem can also occur when using CFPS/IVT systems [46,47]. Ideally, systems should be constructed using cell extracts prepared from strains deficient in specific metabolic enzymes such as those employed in Absolute SILAC [32], but where this is impractical or impossible, the degree of amino acid conversion in the MS analysis should be carefully considered.

One major advantage of CFPS/IVT systems is the large number of reactions that can be performed in a small reaction volume, and various studies have capitalized on this in recent years. CFPS/IVT-based high-throughput protein synthesis is performed to efficiently select highly sensitive proteotypic peptides necessary for establishing the SRM/MRM method [48]. Other studies also used CFPS/IVT systems for selecting sensitive proteotypic peptides to

detect low-abundance proteins such as transcription factors [49] or “missing” proteins not detected by global proteomics studies [50]. Thus, approaches using CFPS/IVT systems are regularly adopted for high-sensitivity analysis using SRM/MRM.

CFPS/IVT systems can also use cell extracts from organisms other than *E. coli*, and systems based on archaeal, protozoan, fungal, plant, insect and mammalian cell extracts have been reported [38]. Among them, a wheat germ cell extract system is now widely used for quantitative proteomics research on eukaryotic proteins due to its high stability and translation efficiency. As originally reported, wheat germ embryos are washed several times to prepare cell extract solutions in which endosperm-derived enzymes that inhibit the translation system are removed [51]. Template DNA amplified by PCR can be used for transcription and translation coupled systems, and it is now possible to conduct high-throughput proteomics research using the wheat germ system [52].

Several reports using the wheat germ system for quantitative targeted proteomics have been reported. In the Full-Length Expressed Stable Isotope-labeled Proteins for Quantification (FLEXIQuant) method, full-length eukaryotic proteins are expressed in the wheat germ system, directly labeled with a stable isotope-labeled amino acid and used for MS quantitative analysis of post-translational modifications [53]. Using this strategy, a workflow was developed to monitor quantitative information about protein phosphorylation (FLEXIQinase) [54]. A similar approach was also utilized in a quantitative study of transmembrane proteins where 263 proteins were analyzed from cDNA clones [55]. The same group also integrated the QconCAT method with the wheat germ system to prepare a reference stable isotope-labeled peptide library for the development of a large-scale SRM/MRM assay for targeted proteomics [56]. Using mTRAQ labeling [57], which originated from Isobaric Tags for Relative and Absolute Quantification (iTRAQ) labeling [58], rather than direct labeling using a stable isotope-labeled amino acid, over 18,000 human recombinant proteins were expressed from cDNAs using the wheat germ system, and used to develop experimentally verified MRM assays of mass tag-labeled peptides, resulting in the In vitro Proteome-assisted Multiple Reaction Monitoring for Protein Absolute Quantification (iMPAQT) platform [59].

4. The protein synthesis Using Recombinant Elements (PURE) system for direct cell-free peptide synthesis

CFPS/IVT systems using the S30 cell extract contain all the necessary elements for protein synthesis, such as ribosomes, tRNA, translation factors and aminoacyl tRNA synthetases (Fig. 3). However, these systems also contain protein factors or enzymes unrelated to the protein synthesis process that can make fine control difficult because it is impossible to accurately grasp what their impact may be, and what kind of reactions are occurring in the reaction mixtures. To overcome this disadvantage, the Protein synthesis Using Recombinant Elements (PURE) system, a reconstituted CFPS system from individual components, was constructed [60]. Unlike the S30-based system, the components included in the PURE system are known, and their exact composition is controllable. Additionally, stable protein synthesis can be performed since factors or enzymes that are inhibitory to the translation system, such as nucleases or proteases, are much less abundant [61]. Furthermore, like cell extract-based systems, this system is commercially available, but the protocol for preparation is also in the public domain, allowing researchers to prepare it by themselves [62,63]. More simplified preparation method that enables co-purification of PURE system components by inserting hexahistidine sequences into these genes by applying Multiplex Automated Genome Engineering (MAGE) [64,65], by *de novo* designing large cistron consists of these genes [66], and by constructing synthetic microbial consortia [67] are also developed and available.

Recently, two research groups reported that features of the PURE system can be utilized for targeted quantitative proteomics studies [68,69]. Both schemes are similar but different in detail, synthesizing a short peptide in the form of a purification tag and a target peptide concatenated. Xian et al. [68] synthesized a peptide in which a fixed five-residue peptide (MGAGR), a variable target peptide and a fixed 11-residue peptide (WSHPQFEKGGD), including *Strep*-tag (WSHPQFEK) for purification and quantification, are concatenated in this order (Fig. 4A). In the MS-Quantification By isotope-labeled Cell-free products (MS-QBiC) workflow reported by Narumi et al. [69], a peptide with a fixed FLAG-tag with an N-terminal methionine (MDYKDDDDK) for purification, a fixed five-residue spacer peptide (LLLLK), a fixed seven-residue peptide (LVTDLTK) derived from bovine serum albumin (BSA) for quantification and a variable target peptide are concatenated in this order (Fig. 4B).

In both cases, a PCR-amplified DNA template is used for peptide synthesis using the PURE system (Fig. 4C). The template contains all elements required for peptide synthesis using an *E. coli*-based translation system, including a T7 promoter, a Shine-Dalgarno sequence and an open reading frame. PCR amplification is completed within two steps, and amplified DNA is directly used in the PURE system where the translation reaction is carried out in the presence of isotope-labeled arginine and lysine (Fig. 4C). The synthesized peptide is purified using magnetic beads and used for LC-MS or SRM/MRM analyses. For quantification of the synthesized peptide, spiking with unlabeled *Strep*-tag peptide (WSHPQFEK) or

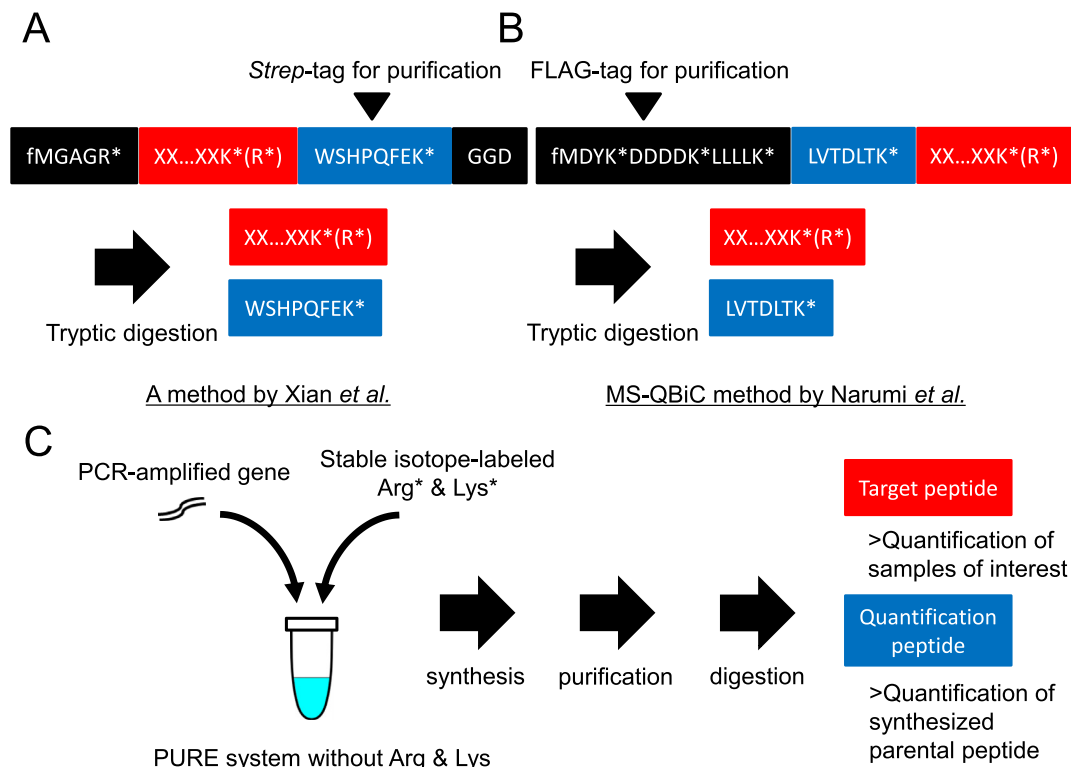


Fig. 4. The PURE system for targeted quantitative proteomics studies.

(A) Peptide synthesis using the method of Xian et al. [68]. A target peptide ending with a stable isotope-labeled arginine or lysine (colored red) and a fixed peptide with the sequence WSHPQFEK* (colored in blue) are produced by tryptic digestion. (B) Peptide synthesis using the MS-QBiC method of Narumi et al. [69]. A target peptide ending with a stable isotope-labeled arginine or lysine (colored red) and a fixed peptide with the sequence LVTDLTK* (colored blue) are produced by tryptic digestion. (C) Typical schemes for both methods. PCR-amplified DNA template for peptide synthesis and stable isotope-labeled arginine and lysine are added to the PURE system without non-labeled arginine and lysine to synthesize the parental peptide. The peptide is affinity-purified, and target and quantification peptides are produced by tryptic digestion. The quantification peptide is used for quantification of the synthesized parental peptide, and the target peptide is used for quantification of the sample of interest. The character 'f' indicates the formyl group of the synthesized peptide, and the asterisk (*) indicates the isotope-labeled amino acid.

quantification peptide (LVTDLTK) and tryptic digestion are performed to separate the parental peptide into individual peptides. In the case of MS-QBiC, spiking of full-length BSA at a known concentration can also be utilized. The determined concentration of the synthesized peptide is used for quantification of target peptides in the sample of interest (Fig. 4C).

Three features arising from the fact that the PURE system is a reconstituted system are exploited in these studies. Firstly, contamination of metabolic enzymes is highly suppressed, and the homogeneity of isotope-labeled peptides is therefore ensured, resulting in reliable quantification. Secondly, DNA degradation by nucleases is unlikely to occur, which enables the use of very short PCR-amplified DNA templates for peptide expression. It is not necessary to perform the laborious cDNA cloning or gene synthesis required for conventional methods employing cellular or cell extract-based protein expression systems. Peptide synthesis can be achieved only by designing appropriate oligo DNA and performing PCR, which enables immediate synthesis of peptides conducive to high-throughput experiments. Thirdly, DNA, RNA and peptide degradation by nucleases and proteases is unlikely to occur; hence stoichiometric equivalence of quantification peptides and target peptides in the synthesized parent peptide is guaranteed. When linear DNA templates are used in cell extract-based systems, DNA and transcribed RNA can be severely degraded, and various precautions must be taken in terms of template sequence or the use of specific strains for preparing cell extracts [70–72]. In addition, it is difficult to prepare peptides using cell extract-based systems because it is not possible to exclude peptidases such as oligopeptidase [73]. By contrast, the parental peptide is intact in the PURE system, and the molar ratio between quantification peptides and target peptides is preserved, enabling a more reliable quantification.

High-throughput peptide synthesis using the PURE system is exploited in the MS-QBiC workflow. The developed scheme has been applied to the quantification of 20 circadian clock proteins using 120 target peptides via SRM/MRM analysis. Using 47 peptides with a signal-to-noise ratio sufficient for the detection of endogenous targets, the circadian dynamics of 16 proteins expressed in the mouse liver were illuminated [69].

5. Future perspectives for obtaining more peptides for high-sensitivity quantification

Analysis of the expression dynamics of circadian clock proteins using the MS-QBiC workflow [69] demonstrates the necessity to prepare and synthesize as many peptides as possible from parental proteins. Despite the successful synthesis of numerous designed peptides, more than half (73/120) were not suitable for quantification, mainly due to their low signal-to-noise ratio for the detection of target proteins. Thus, because theoretically designed peptides are often unsuitable for quantification due to their physical properties, it is desirable to prepare as many peptides for each target protein as possible, which is one of the reasons why CFPS/IVT systems are useful for proteomics studies. In addition, a variety of factors, including the presence of unknown protein isoforms and unknown post-translational modifications, and inefficient or missed cleavages by tryptic digestions, can affect the precise quantification of proteins. To overcome this problem, it is important to use many peptides for the analysis of each target protein.

It is also necessary to improve data interpretation. Because the preparation of many peptide candidates as internal standards has been difficult, a generalized scheme for quantifying one protein with multiple peptides has not yet been proposed. While some reports do not discuss the correctness of the obtained quantitative values, others do address this issue. For example, a recent study, in

which over 1800 *Saccharomyces cerevisiae* proteins were absolutely quantified using the QconCAT method [74], did not include peptides with low quantitative values based on certain criteria. However, there are currently no clear criteria for selection, and it is necessary to establish a standard method for uniform protein quantification. Detailed analyses using many peptides for each target protein are also important for this purpose.

Although trypsin-based protein cleavage is currently considered the gold standard in the field of MS-based proteomics [75], a major problem arises because it is difficult to obtain a large number of candidate peptides for internal standards, especially for basic proteins containing many Lys and Arg residues. Although other proteases such as chymotrypsin and glutamyl endopeptidase are also used in some cases [76,77], it is important to explore other proteases that are versatile in terms of specificity and efficiency. Increasing the number of internal standards obtained for each target protein can be achieved using new proteases via metagenomics [78,79] or by designing *de novo* proteases via computational methods [80]. Such approaches are likely to lead to more accurate and quantitative proteomics approaches in the future.

Declaration of interest

None.

Acknowledgements

This work was supported by a Grant-in-Aid in number 17H05680 (YS) from Japan Society for the Promotion of Science (JSPS), the strategic programs for R&D (President's discretionary fund) of RIKEN (YS), and an intramural Grant-in-Aid from the RIKEN Quantitative Biology Center (YS).

References

- [1] Schrödinger E. *What is Life? The physical aspect of the living cell*. Cambridge: University Press; 1945.
- [2] Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival Jr B, Assad-Garcia N, Glass JI, Covert MW. A whole-cell computational model predicts phenotype from genotype. *Cell* 2012;150:389–401.
- [3] Matsuura T, Tanimura N, Hosoda K, Yomo T, Shimizu Y. Reaction dynamics analysis of a reconstituted *Escherichia coli* protein translation system by computational modeling. *Proc Natl Acad Sci U S A* 2017;114:E1336–44.
- [4] Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* 2011;80:273–99.
- [5] Sato TK, Yamada RG, Ukai H, Baggs JE, Miraglia LJ, Kobayashi TJ, Welsh DK, Kay SA, Ueda HR, Hogenesch JB. Feedback repression is required for mammalian circadian clock function. *Nat Genet* 2006;38:312–9.
- [6] Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 2010;329:533–8.
- [7] Saito K, Chang YF, Horikawa K, Hatsugai N, Higuchi Y, Hashida M, Yoshida Y, Matsuda T, Arai Y, Nagai T. Luminescent proteins for high-speed single-cell and whole-body imaging. *Nat Commun* 2012;3:1262.
- [8] de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Fröhlich F, Walther TC, Mann M. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 2008;455:1251–4.
- [9] Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Pääbo S, Mann M. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* 2011;7:548.
- [10] Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R. The quantitative proteome of a human cell line. *Mol Syst Biol* 2011;7:549.
- [11] Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabudhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathie GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS, Hruban RH, Kerr CL,

- Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A. A draft map of the human proteome. *Nature* 2014;509:575–81.
- [12] Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese JH, Bantscheff M, Gerstmaier A, Faerber F, Kuster B. Mass-spectrometry-based draft of the human proteome. *Nature* 2014;509:582–7.
- [13] Deshmukh AS, Murgia M, Nagaraj N, Treebak JT, Cox J, Mann M. Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Mol Cell Proteomics* 2015;14:841–53.
- [14] Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* 2007;6:2212–29.
- [15] Di Palma S, Mohammed S, Heck AJ. ZIC-CHILIC as a fractionation method for sensitive and powerful shotgun proteomics. *Nat Protoc* 2012;7:2041–55.
- [16] Brun V, Masselon C, Garin J, Dupuis A. Isotope dilution strategies for absolute quantitative proteomics. *J Proteomics* 2009;72:740–9.
- [17] Desiderio DM, Kai M. Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed Mass Spectrom* 1983;10:471–9.
- [18] Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 2003;100:6940–5.
- [19] Merrifield RB. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc* 1963;85:2149–54.
- [20] Hilpert K, Winkler DF, Hancock RE. Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nat Protoc* 2007;2:1333–49.
- [21] Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, Wenschuh H, Aebersold R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* 2010;7:43–6.
- [22] Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 2003;75:6843–52.
- [23] Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* 2009;4:484–94.
- [24] Tsujino K, Narumi R, Masumoto KH, Susaki EA, Shinohara Y, Abe T, Iigo M, Wada A, Nagano M, Shigeyoshi Y, Ueda HR. Establishment of TSH β real-time monitoring system in mammalian photoperiodism. *Genes Cells* 2013;18:575–88.
- [25] Ode KL, Ukai H, Susaki EA, Narumi R, Matsumoto K, Hara J, Koide N, Abe T, Kanemaki MT, Kiyonari H, Ueda HR. Knockout-rescue embryonic stem cell-derived mouse reveals circadian-period control by quality and quantity of CRY1. *Mol Cell* 2017;65:176–90.
- [26] Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002;1:376–86.
- [27] Kruger M, Moser M, Ussar S, Thievesen I, Lubner CA, Forner F, Schmidt S, Zanivan S, Fässler R, Mann M. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* 2008;134:353–64.
- [28] Gouw JW, Krijgsvelde J, Heck AJ. Quantitative proteomics by metabolic labeling of model organisms. *Mol Cell Proteomics* 2010;9:11–24.
- [29] Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* 2010;7:383–5.
- [30] Larance M, Bailly AP, Pourkarimi E, Hay RT, Buchanan G, Coulthurst S, Xirodimas DP, Gartner A, Lamond AI. Stable-isotope labeling with amino acids in nematodes. *Nat Methods* 2011;8:849–51.
- [31] Geiger T, Velic A, Macek B, Lundberg E, Kampf C, Nagaraj N, Uhlen M, Cox J, Mann M. Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. *Mol Cell Proteomics* 2013;12:1709–22.
- [32] Hanke S, Besir H, Oesterheld D, Mann M. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J Proteome Res* 2008;7:1118–30.
- [33] Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* 2005;2:587–9.
- [34] Brownridge PJ, Harman VM, Simpson DM, Beynon RJ. Absolute multiplexed protein quantification using QconCAT technology. *Methods Mol Biol* 2012;893:267–93.
- [35] Shimizu Y, Kuruma Y, Ying BW, Umekage S, Ueda T. Cell-free translation systems for protein engineering. *FEBS J* 2006;273:4133–40.
- [36] Harris DC, Jewett MC. Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. *Curr Opin Biotechnol* 2012;23:672–8.
- [37] Rosenblum G, Cooperman BS. Engine out of the chassis: cell-free protein synthesis and its uses. *FEBS Lett* 2014;588:261–8.
- [38] Zemella A, Thoring L, Hoffmeister C, Kubick S. Cell-free protein synthesis: pros and cons of prokaryotic and eukaryotic systems. *ChemBiochem* 2015;16:2420–31.
- [39] Nirenberg MW, Matthaei JH. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polynucleotides. *Proc Natl Acad Sci U S A* 1961;47:1588–602.
- [40] Foster DA. In memoriam: geoffrey Louis Zubay, 1931–2014. Pioneer in cell-free gene expression studies and molecular genetics. *Trends Biochem Sci* 2014;39:505–6.
- [41] Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ, Beynon RJ. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat Protoc* 2006;1:1029–43.
- [42] Mirzaei H, McBee JK, Watts J, Aebersold R. Comparative evaluation of current peptide production platforms used in absolute quantification in proteomics. *Mol Cell Proteomics* 2008;7:813–23.
- [43] Brun V, Dupuis A, Adrait A, Marcellin M, Thomas D, Court M, Vandenesch F, Garin J. Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics* 2007;6:2139–49.
- [44] Dupuis A, Hennekinne JA, Garin J, Brun V. Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics* 2008;8:4633–6.
- [45] Janecki DJ, Bemis KG, Tegeler TJ, Sanghani PC, Zhai L, Hurley TD, Bosron WF, Wang M. A multiple reaction monitoring method for absolute quantification of the human liver alcohol dehydrogenase ADH1C1 isoenzyme. *Anal Biochem* 2007;369:18–26.
- [46] Ozawa K, Headlam MJ, Schaeffer PM, Henderson BR, Dixon NE, Otting G. Optimization of an *Escherichia coli* system for cell-free synthesis of selectively N-labelled proteins for rapid analysis by NMR spectroscopy. *Eur J Biochem* 2004;271:4084–93.
- [47] Yokoyama J, Matsuda T, Koshiha S, Tochio N, Kigawa T. A practical method for cell-free protein synthesis to avoid stable isotope scrambling and dilution. *Anal Biochem* 2011;411:223–9.
- [48] Stergachis AB, MacLean B, Lee K, Stamatoyannopoulos JA, MacCoss MJ. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nat Methods* 2011;8:1041–3.
- [49] Simicevic J, Schmid AW, Gilardoni PA, Zoller B, Raghav SK, Krier I, Gubelmann C, Lisacek F, Naef F, Moniatte M, Deplancke B. Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. *Nat Methods* 2013;10:570–6.
- [50] Horvatovich P, Vegvari A, Saul J, Park JG, Qiu J, Syring M, Pirrotte P, Petritis K, Tegeler TJ, Aziz M, Fuentes M, Diez P, Gonzalez-Gonzalez M, Ibarrola N, Droste C, De Las Rivas J, Gil C, Clemente F, Hernaez ML, Corrales FJ, Nilsson CL, Berven FS, Bischoff R, Fehniger TE, LaBaer J, Marko-Varga G. In vitro transcription/translation system: a versatile tool in the search for missing proteins. *J Proteome Res* 2015;14:3441–51.
- [51] Madin K, Sawasaki T, Ogasawara T, Endo Y. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc Natl Acad Sci U S A* 2000;97:559–64.
- [52] Sawasaki T, Ogasawara T, Morishita R, Endo Y. A cell-free protein synthesis system for high-throughput proteomics. *Proc Natl Acad Sci U S A* 2002;99:14652–7.
- [53] Singh S, Springer M, Steen J, Kirschner MW, Steen H. FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J Proteome Res* 2009;8:2201–10.
- [54] Singh SA, Winter D, Bilimoria PM, Bonni A, Steen H, Steen JA. FLEXIQinase, a mass spectrometry-based assay, to unveil multikinase mechanisms. *Nat Methods* 2012;9:504–8.
- [55] Takemori N, Takemori A, Matsuoka K, Morishita R, Matsushita N, Aoshima M, Takeda H, Sawasaki T, Endo Y, Higashiyama S. High-throughput synthesis of stable isotope-labeled transmembrane proteins for targeted transmembrane proteomics using a wheat germ cell-free protein synthesis system. *Mol Biosyst* 2015;11:361–5.
- [56] Takemori N, Takemori A, Tanaka Y, Ishizaki J, Hasegawa H, Shiraishi A, Ohashi Y. High-throughput production of a stable isotope-labeled peptide library for targeted proteomics using a wheat germ cell-free synthesis system. *Mol Biosyst* 2016;12:2389–93.
- [57] DeSouza LV, Taylor AM, Li W, Minkoff MS, Romaschin AD, Colgan TJ, Siu KW. Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *J Proteome Res* 2008;7:3525–34.
- [58] Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 2004;3:1154–69.
- [59] Matsumoto M, Matsuzaki F, Oshikawa K, Goshima N, Mori M, Kawamura Y, Ogawa K, Fukuda E, Nakatsumi H, Natsume T, Fukui K, Horimoto K, Nagashima T, Funayama R, Nakayama K, Nakayama KI. A large-scale targeted proteomics assay resource based on an in vitro human proteome. *Nat Methods* 2017;14:251–8.
- [60] Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T. Cell-free translation reconstituted with purified components. *Nat Biotechnol* 2001;19:751–5.
- [61] Ohashi H, Shimizu Y, Ying BW, Ueda T. Efficient protein selection based on ribosome display system with purified components. *Biochem Biophys Res Commun* 2007;352:270–6.
- [62] Shimizu Y, Ueda T. PURE technology. *Methods Mol Biol* 2010;607:11–21.
- [63] Shimizu Y, Kuruma Y, Kanamori T, Ueda T. The PURE system for protein production. *Methods Mol Biol* 2014;1118:275–84.

- [64] Wang HH, Huang PY, Xu G, Haas W, Marblestone A, Li J, Gygi SP, Forster AC, Jewett MC, Church GM. Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. *ACS Synth Biol* 2012;1:43–52.
- [65] Li J, Gu L, Aach J, Church GM. Improved cell-free RNA and protein synthesis system. *PLoS One* 2014;9:e106232.
- [66] Shepherd TR, Du L, Liljeruhm J, Samudiyata Wang J, Sjodin MOD, Wetterhall M, Yomo T, Forster AC. De novo design and synthesis of a 30-cistron translation-factor module. *Nucleic Acids Res* 2017;45:10895–905.
- [67] Villarreal F, Contreras-Llano LE, Chavez M, Ding Y, Fan J, Pan T, Tan C. Synthetic microbial consortia enable rapid assembly of pure translation machinery. *Nat Chem Biol* 2018;14:29–35.
- [68] Xian F, Zi J, Wang Q, Lou X, Sun H, Lin L, Hou G, Rao W, Yin C, Wu L, Li S, Liu S. Peptide biosynthesis with stable isotope labeling from a cell-free expression system for targeted proteomics with absolute quantification. *Mol Cell Proteomics* 2016;15:2819–28.
- [69] Narumi R, Shimizu Y, Ukai-Tadenuma M, Ode KL, Kanda GN, Shinohara Y, Sato A, Matsumoto K, Ueda HR. Mass spectrometry-based absolute quantification reveals rhythmic variation of mouse circadian clock proteins. *Proc Natl Acad Sci U S A* 2016;113:E3461–7.
- [70] Ahn JH, Chu HS, Kim TW, Oh IS, Choi CY, Hahn GH, Park CG, Kim DM. Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochem Biophys Res Commun* 2005;338:1346–52.
- [71] Jun SY, Kang SH, Lee KH. Continuous-exchange cell-free protein synthesis using PCR-generated DNA and an RNase E-deficient extract. *Biotechniques* 2008;44:387–91.
- [72] Katsura K, Matsuda T, Tomabechi Y, Yonemochi M, Hanada K, Ohsawa N, Sakamoto K, Takemoto C, Shirouzu M. A reproducible and scalable procedure for preparing bacterial extracts for cell-free protein synthesis. *J Biochem* 2017;162:357–69.
- [73] Jiang X, Zhang M, Ding Y, Yao J, Chen H, Zhu D, Muramatsu M. *Escherichia coli* prfC gene encodes a trypsin-like proteinase regulating the cell cycle. *J Biochem* 1998;124:980–5.
- [74] Lawless C, Holman SW, Brownridge P, Lanthaler K, Harman VM, Watkins R, Hammond DE, Miller RL, Sims PF, Grant CM, Evers CE, Beynon RJ, Hubbard SJ. Direct and absolute quantification of over 1800 yeast proteins via selected reaction monitoring. *Mol Cell Proteomics* 2016;15:1309–22.
- [75] Glatter T, Ludwig C, Ahrné E, Aebersold R, Heck AJ, Schmidt A. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* 2012;11:5145–56.
- [76] Tsiatsiani L, Heck AJ. Proteomics beyond trypsin. *FEBS J* 2015;282:2612–26.
- [77] Giansanti P, Tsiatsiani L, Low TY, Heck AJ. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* 2016;11:993–1006.
- [78] Ufarté L, Potocki-Veronese G, Laville É. Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology. *Front Microbiol* 2015;6:563.
- [79] DeCastro ME, Rodríguez-Belmonte E, González-Siso MI. Metagenomics of thermophiles with a focus on discovery of novel thermozyms. *Front Microbiol* 2016;7:1521.
- [80] Huang PS, Boyken SE, Baker D. The coming of age of *de novo* protein design. *Nature* 2016;537:320–7.