Peptide Biosynthesis with Stable Isotope Labeling from a Cell-free Expression System for Targeted Proteomics with Absolute Quantification*

Feng Xian‡§¶, Jin Zi§, Quanhui Wang‡§, Xiaomin Lou‡, Haidan Sun‡, Liang Lin§, Guixue Hou‡§, Weiqiao Rao§, Changcheng Yin**, Lin Wu‡, Shuwei Li‖‡, and Siqi Liu‡§¶

Because of its specificity and sensitivity, targeted proteomics using mass spectrometry for multiple reaction monitoring is a powerful tool to detect and quantify pre-selected peptides from a complex background and facilitates the absolute quantification of peptides using isotope-labeled forms as internal standards. How to generate isotope-labeled peptides remains an urgent challenge for accurately quantitative targeted proteomics on a large scale. Herein, we propose that isotope-labeled peptides fused with a quantitative tag could be synthesized through an expression system in vitro, and the homemade peptides could be enriched by magnetic beads with tag-affinity and globally quantified based on the corresponding multiple reaction monitoring signals provided by the fused tag. An *Escherichia coli* cell-free protein expression system, protein synthesis using recombinant elements, was adopted for the synthesis of isotope-labeled peptides fused with Strep-tag. Through a series of optimizations, we enabled efficient expression of the labeled peptides such that, after Strep-Tactin affinity enrichment, the peptide yield was acceptable in scale for quantification, and the peptides could be completely digested by trypsin to release the Strep-tag for quantification. Moreover, these recombinant peptides could be employed in the same way as synthetic peptides for multiple reaction monitoring applications and are likely more economical and useful in a laboratory for the scale of targeted proteomics. As an application, we synthesized four isotope-labeled glutathione S-transferase (GST) peptides and added them to mouse sera pre-treated with GST affinity resin as internal standards. A quantitative assay of the synthesized GST peptides confirmed the absolute GST quantification in mouse sera to be measurable and reproducible. Molecular & Cellular Proteomics 15:10.1074/mcp.O115.056507, 2819–2828, 2016.

Targeted proteomics is a powerful technology that can quantify a collection of pre-selected peptides from complex biological samples using multiple reaction monitoring (MRM)\(^1\) (1). Such analysis is usually performed in a triple-quadrupole mass spectrometer (QQQ), in which a peptide of interest is isolated in Q1 from a mixture of peptides, fragmented in Q2 to yield a series of product ions, and detected in Q3 based on its characteristic fragments. A pair consisting of a precursor and one of its characteristic product ions is referred to as a transition. Three or more transitions from the same precursor usually allow the high-confidence identification of this peptide from a highly complex sample (2). A modern triple-quadrupole mass spectrometer is capable of scanning hundreds of transitions in a single run to detect multiple peptides concurrently (3). Thus, when stable isotope-labeled peptides of known quantity are spiked as internal standards (a.k.a. reference peptides), the absolute quantification of the selected peptides can be determined with high precision (4).

The success of targeted proteomics experiments clearly relies on the availability of isotope-labeled peptides, which are usually prepared by standard solid-phase peptide synthesis

---

\(^1\) The abbreviations used are: MRM, multiple reaction monitoring; PURE, protein synthesis using recombinant elements; GST, glutathione S-transferase; SPPS, standard solid-phase peptide synthesis; QconCAT, quantification concatamers; DTT, dithiothreitol; IAM, iodoacetamide; IgJ, immunoglobulin J chain; ANOVA, analysis of variance.
(SPPS) (5). However, each synthetic peptide must be purified chromatographically and quantified by tedious amino acid analysis (6). SPPS is also sequence-dependent, and some peptides are extremely difficult to synthesize and purify (e.g. amyloid beta-peptide) (7). It is also expensive to make multiple high-purity isotope-labeled peptides. To overcome these challenges, biological methods based on protein expression have recently become more popular. For example, one widely used method begins with the assembly of oligonucleotides encoding multiple peptides into a concatenated gene (QconCAT), which is then expressed in *Escherichia coli* grown in a minimal medium with $^{13}$N-labeled ammonium chloride as the only nitrogen source (8). The resulting protein is purified, quantified by Bradford assay, and digested into separate peptides at equimolar concentration (9). Despite the many benefits offered by the QconCAT platform, it also suffers from several limitations. It is difficult to predict whether an artificial QconCAT gene can be expressed at all, or can generate a protein resistant to nonspecific protease degradation in *Escherichia coli* (7). In addition, because one QconCAT gene yields reference peptides for multiple proteins at an equimolar ratio, quantitative analysis is not reliable if these proteins have significantly different concentrations *in vivo*. This QconCAT gene can also be expressed by a conventional cell-free coupled transcript/translation system (e.g. wheat germ extract, rabbit reticulocyte lysate) (10, 11). However, the same problems remain inadequately solved.

With rapid advances in the discovery of optimal peptides for targeted proteomics (12, 13), there is an urgent need to generate isotope-labeled peptides for each protein in the human proteome, especially for low-abundance proteins (14). Neither SPPS nor QconCAT has the necessary ease and versatility for such large-scale endeavors (7). Here, we introduce a novel approach based on the PURE (Protein synthesis Using Recombinant Elements) system for the preparation of stable isotope-labeled reference peptides (15). The basic procedures of this approach are illustrated in supplemental Fig. S1. Through the technique improvement described below, we have demonstrated that the new approach is feasible for accurate quantification in large-scale targeted proteomics.

**EXPERIMENTAL PROCEDURES**

**Materials and Reagents—**Acetonitrile (HPLC-grade) and formic acid (MS-grade) were obtained from Fisher Scientific (Waltham, MA). The PUREExpress® $\Delta$ (aa, tRNA) Kit was purchased from New England Biolabs (Ipswich, MA). Sequence-grade trypsin was obtained from Promega (Madison, WI). Stable isotope-labeled U-$^{13}$C$_6$U-$^{15}$N$_2$ arginine and U-$^{13}$C$_6$U-$^{15}$N$_2$ lysine were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). All of the other native amino acids, dethiobiotin and four AQUA peptides (AQ0118, AQ0239, AQ0229, and AQ0043) were purchased from Sigma-Aldrich (St. Louis, MO). Eight isotopic peptides (25 nmol/each) used in this study, CKM-1 (LNKYPEEPYDLSK*), CKM-2 (LGGSVEQVQLVQVGYK*), MP-1 (FGQALGDAADIR*), MP-2 (FGVTQPDVQLTLYQ*), MP-3 (SEEFLIAQK*), CKB-1 (DLDFPIHEDR*), CYS-1 (ASNDMYHSSR*), and CYS-2 (ALDFVAGEYNK*), were gifts from Professor Borchers (University of Victoria, Canada). Streptavidin magnetic and agarose beads were obtained from IBA (Göttingen, Germany). Oligonucleotides were synthesized by BGI (Shenzhen, China). All PCR reagents were purchased from TransGen Biotech (Beijing, China), and DNA purification kits were purchased from GalaxyBio (Jinan, Shandong, China).

**Preparation of DNA Templates Encoding Peptides—**A database written in-house was used to calculate DNA templates that encode peptides based on the codon usage of *Escherichia coli*. For peptide lengths over 9 amino acids, 2 DNA templates were generated to keep each template less than 60 residues, whereas only 1 DNA template was used for peptide lengths shorter than nine residues. The PCR reaction mixture (10× Pfu buffer 5 μl, 10 mM dNTP 1.25 μl, 10 mM DNA template one and two 1.25 μl) was assembled on ice in a clear tube, briefly vortexed and centrifuged, and placed in a thermocycler (BioRad) for 5 cycles first (94 °C for 30 s, 58 °C for 30 s, and 72 °C for 10 s). Then, 1.25 μl universal primer mixture (10 μM) was added, and the reaction mixture was subjected to another 35 cycles (94 °C for 30 s, 58 °C for 30 s, and 72 °C for 20 s). The PCR product was purified by following the standard protocol provided by the manufacturer.

**Peptide Synthesis with PURE System—**The reaction conditions basically followed the instruction of PUREExpress® $\Delta$ (aa, tRNA) Kit, but with optimization of the quantity of DNA template used per reaction and the reaction time. A PCR thermocycler was used to maintain the reaction temperature at 37 °C. The peptide expression was terminated by putting the reaction mixture on ice. For expression, all arginine and lysine residues were isotopically labeled by supplementing the PURE system with U-$^{13}$C$_6$U-$^{15}$N$_2$ arginine (R*) and U-$^{13}$C$_6$U-$^{15}$N$_2$ lysine (K*), together with the 18 other native amino acids unlabeled. Thus, an isotope-labeled peptide is expected to have a 10 Da difference in molecular mass for R* or 8 Da for K* from the corresponding unlabeled peptide.

**Peptide Purification and Enrichment—**For Strep-Tactin agarose beads, 100 μl washing buffer (50 mM NH$_4$HCO$_3$, pH 8.0) was used to equilibrate 50 μl beads twice. Thirty microliters of the PURE expression solution was mixed with 70 μl washing buffer and incubated on a shaker at 4 °C for 1 h. After the mixture was briefly centrifuged and the supernatant carefully removed, 100 μl washing buffer was added, and then the beads were vortexed and centrifuged again to remove the supernatant. This washing was repeated three times. Then, 50 μl elution buffer (50 mM NH$_4$HCO$_3$, 10 mM dethiobiotin, pH 8.0) was added, and the suspension was gently stirred on a shaker for 5 min. The mixture was then centrifuged, and the supernatant was collected. For Strep-Tactin magnet beads, the entire procedure was the same as for the Strep-Tactin agarose beads, except using a magnetic separator instead of centrifugation.

**Peptide Digestion—**If the enriched peptides contained cysteine residues, they were reduced with 10 μl dithiothreitol (DTT) at 56 °C and then alkylated by 55 mM iodoacetamide (IAM) in the dark at room temperature before trypsin digestion. Otherwise, the enriched peptides were digested by trypsin directly. The enriched peptides were first treated with 2 μg trypsin in elution buffer for 4 h at 37 °C, and then another 2 μg trypsin was added for a further 8 h of incubation at 37 °C. The solvent was removed by Speedvac after digestion.

**Quantification of GSTs in mouse serum—**Whole blood was taken from the eyes of C57BL/6J mice purchased from a commercial source and directly subjected to centrifugation at 2000 RCF and 4 °C for 15 min. The serum was further diluted for affinity enrichment or stored at −20 °C for further use. GST-affinity magnetic beads were used to enrich GSTs in the serum. Approximately 200 μl serum was diluted with 1 ml binding buffer (125 mM Tris, 150 mM NaCl, pH 8.0) and incubated at 4 °C for 2 h with 30 μl of glutathione resin-coated magnetic beads (Thermo Scientific). After three washes with binding buffer (300 μl each time) to remove nonspecific binding proteins, the...
bound GSTs were reduced with 10 mM dithiothreitol (DTT) at 56 °C and alkylated by 55 mM iodoacetamide (IAM) in the dark at room temperature, then digested with 2 μg trypsin in 50 μl NH4HCO3 (25 mM, pH 8). For quantitative evaluation of the GSTs in mouse serum, isotope-labeled GST peptides were used for calibration and spiking, according to the method developed in this study.

**Peptide Separation by LC**—Tryptic peptides were dissolved in 30 μl 0.1% formic acid solution (2% acetonitrile, 98% H2O) and delivered into a reverse phase nano-HPLC column (75 μm ID × 150 mm length, filled with 3 μm C18, resin) mounted in a Prominance Nano HPLC system (Shimadzu, Nakagyo-ku, Kyoto, Japan) or an Eksigent NanoLC ultra (AB SCIEX, Framingham, MA). For each analysis, the sample was first loaded into a trap column (2 cm × 75 μm, C18, 3 μm) at 5 μl/min with aqueous solution containing 2% (v/v) acetonitrile and 0.1% formic acid for 8 min. Peptide elution was performed by applying a mixture of solvents A/B. Solvent A was HPLC grade water with 0.1% (v/v) formic acid and 2% (v/v) acetonitrile, and solvent B was HPLC grade acetonitrile with 0.1% (v/v) formic acid and 2% (v/v) water. The peptide separations were performed using an acetonitrile gradient from 5% to 80% for 35 min at 300 nl/min.

**Peptide Analysis by Mass Spectrometry**—The eluent from nano-HPLC was directly delivered into a Triple-TOF 5600 MS (AB SCIEX, Framingham, MA) for full mass scanning or a QTRAP 5500 MS (AB SCIEX, Framingham, MA) for MRM quantification. In all experiments, the nano-HPLC eluent was introduced into the mass spectrometer via an integrated electrospray emitter (10 μm inner diameter, New Objective) operating at 2.1—2.4 kV and coupled to a nano-ESI source. The peptide analysis conducted in the Triple-TOF 5600, the full scan event employed an m/z 350—1800 mass selection. At most, 40 precursors (2—5 charged, intensity over 150 cps) were selected for fragmentation (mass tolerance 50 mDa). The high-sensitivity mode was used to detect the fragmented ions, and MS2 scans were acquired with an m/z range of 100—1500. Exclude former target ions was set to 15 s. The transitions used in the MRM experiments are provided in supplemental Table S1. The Skyline software (version 3.1) designed by MacCoss Lab was used to help to build the MRM-MS methods, such as the prediction of theoretical transitions, de-cluster potential and collision energy. The dwell time for each transition was set to 50 ms, and the resolution for both Q1 and Q3 was unit mode.

**Peptide Quantification**—The peptide quantification was performed based on the internal standard quantification method in QTRAP 5500. In detail, synthesized and purified peptide was digested with trypsin and re-dissolved in 0.1% formic acid solution (2% acetonitrile, 98% H2O). Several equal volume aliquots of the recombinant peptide with heavy labeling were spiked into different concentrations of light peptide, and the MRM signals generated from the mixture were used to construct a calibration curve. This curve was plotted using the MRM peak area ratio of (light/heavy) against the quantity of unlabeled Strep-tag.

**Quantification Evaluation**—The quantification was evaluated in two ways. (1) Three tryptic peptides from the immunoglobulin J chain (IgJ) were selected for synthesis by PURE in isotope-labeled forms: Pep1, LVPLNRR*; Pep2, SSEDNPNDER*; and Pep3, ENISDPTPSLR*. After the recombinant peptides were quantified by Strep-tag calibration, known amounts of the three peptides were spiked into the serial dilutions of the IgJ peptides generated by trypsin digestion. The calibration curves for these peptides were generated by plotting the MRM signal ratios of L versus H against the sample dilution factors. The quantification accuracy was evaluated by comparing the quantified results achieved from the three recombinant peptides. (2) The quantification results were compared with the quantified peptides. Four peptides (100 pmol/each) were purchased from Sigma-Aldrich (5): AQ0229 (LGADSCAAPPTSK*), AQ0118 (ESVLATATSNLPIVKK*), AQ0239 (YGIPYFETSATGQNVK*), and AQ0043 (YGLIPK*). The commercial peptides and gifted peptides were both dissolved in 20% (v/v) methanol to final concentrations of 1 pmol/μl and 250 pmol/μl, respectively, then diluted with 0.1% formic acid and 2% acetonitrile solution to working concentrations. Based on the peptide sequences provided by Sigma-Aldrich and Professor Borchers, the Strept-peptides were also synthesized by the PURE system. The two sets of peptides generated by the different approaches were delivered to MRM-MS and quantified using the method described above.

**Data Analysis**—The raw data from Triple-TOF MS was converted into MGF format using a program written in-house. The MGF MS2 data were analyzed using an in-house Mascot server (version 2.3, Matrix Science, Boston, MA) with the enzyme specificity set to Trypsin, allowing one missed cleavage. The peptide tolerance and fragment tolerance were set to 0.03 Da and 0.07 Da, respectively. The searches were allowed to include fixed modifications (carbamidomethyl on cysteine). Oxidation on methionine and stable isotope-labeled lysine (13C6, 15N2) and arginine (13C6, 15N4) were used as variable modifications. The MRM data were analyzed using MultiQuant (version 2.3, AB SCIEX). The Gaussian smooth width and baseline subtraction window were set to 1 point and 1 min, respectively, and the other parameters were set to default values.

**RESULTS AND DISCUSSION**

**Synthesis of Isotope-Labeled Peptides by PURE**—PURE is an Escherichia coli cell-free protein expression system first reported in 2001 (15). Unlike other conventional in vitro protein translation techniques, PURE consists of purified enzymes and chemically defined factors necessary for protein expression, such as 20 aminoacyl tRNA synthetases, ribosomes, tRNA molecules, and amino acids. Therefore, PURE lacks nonspecific nuclease and protease activity. This expression system has been generally used for the expression of recombinant proteins when large quantities are not required (16).

Herein, we postulated that if PURE could be used for the expression of individual short reference peptides on a large scale, a long QconCAT gene expressing multiple peptides is then no longer needed. Such individually expressed peptides make it much easier to measure multiple proteins with disparate abundance in which we can spike reference peptides close to their respective endogenous concentrations. In addition, one or more amino acids can be completely replaced with their isotope-labeled counterparts in the PURE system, allowing more flexible peptide quantification.

To employ PURE for peptide expression, we can obtain a double-stranded DNA template by PCR amplification from synthetic oligonucleotides. This DNA template contains, in order, a T7 promoter, a ribosome binding site, and the sequence encoding the target peptide (Fig. 1A). As PURE already includes T7 RNA polymerase and the necessary components for RNA transcription, the double-stranded DNA template can be added directly without being transcribed to mRNA first. The resulting parent peptide starts from a constant tripeptide (MDR and ends with a constant 11-residue peptide (WSHPOFEKGGD) including a Strep-tag (WSHPQFEK)). Here, the N-terminal M is a formylmethionine that is required for the initiation of protein translation in Escherichia coli, whereas the extra three residues (GDD) at the C-
A variable reference peptide sequence (for instance, the peptide AEVVTPMGIPAEEIPK in Fig. 1B) is located between these two constant sequences. This approach is basically illustrated as a flowchart in Fig. 1C. The coding DNA templates of a reference peptide are first calculated using an in-house database according to the codon preference of *Escherichia coli*. Following PCR amplification, a double-stranded DNA template for peptide synthesis is generated from synthetic oligos. This DNA template contains all elements for transcription and translation. After evaluating the quantity of PCR products, the double-stranded template is added to the PURE expression system to synthesize a parent peptide. In this design, the C terminus Strep-tag has dual functions. First, because each purified peptide generates an equimolar mixture of the labeled Strep-tag (WSHPQFEK*) and a labeled reference peptide after trypsin digestion, the reference peptide can be quantified indirectly based on the amount of the labeled Strep-tag, which is determined by spiking unlabeled Strep-tag (WSHPQFEK) with a known concentration. Second, it allows the full-length parent peptide to be purified with Strep-Tactin beads, which is a streptavidin mutant with higher affinity for the eight-residue Strep-tag than the wild type (17, 18). As an example, we designed the DNA template to express the peptide fMDR*AEVVTPMGIPAEEIPK*WSHPQFEK*GGD. After Strep-Tactin purification and trypsin digestion, we performed a quick MALDI-MS analysis and detected two expected peaks corresponding to labeled AEVVTPMGIPAEEIPK* and WSHPQFEK*, which represented 8 Da mass above the theoretical molecular masses of the two peptides (Fig. 1B). This result strongly supported the success of our strategy. The question may arise whether the Strep-tag beads could nonspecifically bind some proteins and cause data misinterpretation. Although the limited proteins are assumed to exhibit nonspecific binding to the affinity beads, the results in Fig. 1 suggested that the recombinant peptide remained the source of the dominant MS signals in the peptide mixture. Therefore, it is a reasonable deduction that the limited nonspecific peptides have less influence on the peptide identifi-
cation and quantification. Additionally, the quantification principle in this method was based on the labeled peptides quantified by the calibration curve of Strep-tag, which were unique signals in a peptide mixture in the PURE system. Although the recombinant peptide was not of high purity despite Strep-tag affinity enrichment, the impurity without Strep-tag would not significantly affect the quantification.

Optimization of Reaction Conditions—To maximize peptide yield, we optimized a few conditions using a commercially available PURExpress® /H9004 (aa, tRNA) Kit. A randomly selected peptide, called A3-pep1 (AILNYIASK), was used for this optimization. We first tested different concentrations of the DNA template and found that 400 ng DNA generated the highest amount of the peptide (Fig. 2 A). We also tested different reaction times and discovered that a period of 6 hours was optimal (Fig. 2 B). However, it was surprising that the yield decreased beyond six hours, rather than reaching a plateau. We believe that residual protease activity in the commercial PURE system may cause the peptide to be slowly degraded. In addition, we tested whether different types of Strep-Tactin beads affect the yield of the enriched Strep-peptides. Magnetic beads coated with Strep-Tactin gave more than twofold peptide enrichment over Strep-Tactin agarose (Fig. 2 C), probably because the agarose beads underwent a relatively high loss during centrifugation. Finally, we found that some unlabeled amino acids still remained in the recombinant peptides.

To ensure the incorporation ratio of isotope-labeled arginine and lysine was above 98%, we examined the isotope R* and K* concentrations from 3 mM to 30 mM in a reaction mixture (Fig. 2 D) and determined 12 mM R* and K* to be appropriate concentrations to exclude unlabeled R and K from a recombinant peptide.

Using this optimized protocol, we were able to routinely generate 1–3 pmol of labeled reference peptides from a standard 30 µl reaction mixture, which might be enough to support a general MRM experiment. The entire procedure, from PCR amplification to the quantification of reference peptides, only took eighteen to twenty hours. More importantly, every step involved in this procedure could be performed in 96-well plates, which can greatly facilitate the effort to acquire reference peptides for proteomics study on a large scale.

Optimization of N terminus Leading Sequence—After purification and digestion, a peptide synthesized by the PURE system was analyzed by LC-MS (Triple TOF 5600), and the MS2 signals were searched against a home-made database using MASCOT. The search results suggested that some minor peptide species resulted from incomplete trypsin digestion or digestion by other proteases contaminating PURE system, so we extracted the ion chromatograph according to the theoretical mass of the fragments of incomplete digestion. For example, when the parent peptide "MDR*VTYYDFLVDQHR*WSHPQFEK"GGD was digested by trypsin, the...
Biosynthesis of Stable Isotope-Labeled Peptides for MRM

Fig. 3. Impact of different N terminus constant sequences on trypsin digestion efficiency. T1, 'MVTYDFLVDLDQHR*; T2, 'MDRTYDFLVDLDQHR*; and T3, 'MGAGR'VTVYDFLVDLDQHR*. The light gray indicates the peak area of the completely digested target peptide VTVYDFLVDLDQHR*, and the peak areas of two incompletely digested forms are marked in dark gray ('MVTYDFLVDLDQHR') and black (DR'VTVYDFLVDLDQHR*). The peak areas were calculated using the PeakView software.

tryptic peptide DR'VTVYDFLVDLDQHR* was generated at ~4.53% of the total peptide products (Fig. 3). Based on MASCOT estimation, this peptide was apparently derived from an incomplete trypsin digestion, but the loss of M at this peptide was unexpected. Based on our knowledge of proteases, methionine aminopeptidase is the enzyme that cleaves M. If the PURE system were slightly contaminated by methionine aminopeptidase, the incomplete digestion result would be explainable. To test this hypothesis, another parent peptide, 'MVTVDFLVDLDQHR*WSHPQFEKGGD, was synthesized by PURE system, in which the N terminus valine was the neighboring residue of M and might exert a steric effect on the activity of methionine aminopeptidase (19). After peptide synthesis and identification, MS analysis indicated that more than 30% of this peptide contained M at the N terminus. This result demonstrated that methionine aminopeptidase was indeed a contaminant in the PURE system. To optimize a short leading peptide immediately before the N terminus of the target peptide, there were three factors to be considered: (1) starting from a methionine and ending with a lysine or arginine; (2) the neighboring residue next to methionine at the N terminus with a limiting effect on the activity of methionine aminopeptidase; and (3) a length between 5 and ~6 amino residues to ensure complete trypsic digestion. Based on this principle, the peptide MGAGR was chosen as the leading sequence. For instance, in Fig. 3, the target peptide designed was MGAGR'VTVYDFLVDLDQHR*WSHPQFEKGGD, and the synthesis and MS results revealed that the identified peptide VTVYDFLVDLDQHR*, which was consistent with our expectation. The first residue in the target peptide may also affect the trypsin digestion efficiency, especially when it is a charged residue such as D/E/R/K (20), but we can avoid using these peptides as internal standards. If it is necessary to use a peptide starting with these residues, we can still push the digestion to completion by using a large amount of trypsin and incubating for a longer time.

Quantification of Reference Peptides—To quantify a target peptide in a complex biological sample, the corresponding reference peptide needs to be added at a known concentration. The peak ratios of the target and reference peptides are then measured and used to estimate the target peptide concentrations. As a result, it is necessary to determine the reference peptides accurately. Solid-phase synthesized AQUA peptides are usually quantified by amino acid analysis, a tedious multi-step procedure in which a peptide needs to be hydrolyzed completely into free amino acids, which are then labeled with a UV absorbance or fluorescence tag and separated by LC. Although amino acid analysis can provide rather accurate quantification of reference peptides, it is inconvenient in daily experiments, especially when multiple reference peptides are needed for large-scale analysis. In our design, all the recombinant peptides contain a fixed Strep-tag sequence, which is at a 1:1 ratio to the reference peptide derived from complete trypsin digestion. Therefore, we can quantify this Strep-tag instead to deduce the concentration of the accompanying reference peptide.

To demonstrate this approach, we expressed and enriched the peptide MGAGR'AILNYIASK*WSHPQFEKGGD. After this peptide was digested with trypsin, 5 μl of the 30 μl mixture of tryptic peptides was taken out and diluted 10 times with 0.1% formic acid solution (2% acetonitrile, 98% H2O). Then, several 1 μl aliquots of diluted isotope-labeled peptides were added to the serial diluted Strep-tag without isotope labeling, and the mixtures were detected by LC-MRM-MS to generate the calibration curve. As shown in Fig. 4A, the light (unlabeled Strep-tag) versus heavy (labeled Strep-tag) MRM signal ratios were correlated with the concentration of Strep-tag with R2>0.99, indicating that the linear regression line was qualified for the quantitative evaluation of the synthesized reference peptide (AILNYIASK*). Based on this equation, the diluted recombinant peptide was estimated to be at a concentration of 1.985 fmol/μl, which was the reading on the x axis when the signal ratio (y axis) was equal to 1.

The accuracy of the PURE-synthesized and quantified peptides for quantification was further evaluated in an internal calibration experiment. Regarding protein quantification by MRM, the quantitative result supported by multiple peptides is considered to be acceptable because several peptides derived from the same protein should maintain similar molar ratios after a completely tryptic digestion. Based on this principle, three peptides derived from the recombinant protein of the immunoglobulin J chain (IgJ) were selected. After synthesis and quantification, known amounts of three recombinant peptides were spiked into serial dilutions of the peptides of recombinant immunoglobulin J chain (expressed in E. coli and purified using His-tag affinity resin) generated by trypsin digestion. With the IgJ MRM signals, the ratios of the MRM peak areas for unlabeled peptides against the areas for la-
beled ones were obtained (y axis) and plotted with the dilutions of the IgJ digestion (x axis). In Fig. 4B, if the ratio of unlabeled to labeled peptides equals 1, then the added IgJ peptide is equal in quantity to the digested IgJ peptide. We then calculated the concentrations of the three peptides to be 305.8, 268.6, and 253.2 fmol/μL, respectively (Fig. 4C). Therefore, the absolute quantification of the immunoglobulin J chain protein was determined to be 275.8 fmol/μL with a CV of 10% from the three individual peptides, and the lower CV values thus led to the conclusion that the quantitative result of IgJ as determined by the three IgJ peptides was quantitatively acceptable. This experiment indicated that the PURE-expressed isotope peptides can be used for protein quantification in targeted proteomics experiments.

The Strep-tag-based quantitative results obtained from our approach were further evaluated using the synthetic peptides purchased from a commercial source or gifted from Professor Brochers. The quantities of all of the synthetic peptides were measured by amino acid analysis. Of 12 peptides tested, two commercial peptides, AQ0229 and AQ0043, were not detected in the MRM signals even at the 300 fmol/injection based on the commercial recommendation; however, the recombinant peptides showed strong MRM signals at 10 fmol/injection. It is well known that the detection sensitivity of MRM is generally in the femtomole range, so we questioned whether the low MRM signals for AQ0229 and AQ0043 might be attributable to quantification error or a low redissolving efficiency of the two peptides. We performed the MRM assays on two sets of peptides with the same amino acid sequences, synthetic and recombinant, and compared the calibration curves with the MRM signals against the peptide quantities achieved by either amino acid analysis or Strep-tag. As shown in supplemental Fig. S2, all 10 recombinant peptides exhibited higher MRM signals than were generated from synthetic peptides at similar peptide concentrations; however, the calibration curves of the two sets of peptides were basically comparable and parallel even though the concentration axes (x-axes) shifted by ~7 to 30-fold (synthetic versus recombinant peptides). Therefore, the MRM results derived from the synthetic peptides offered evidence supporting the quantification results produced by our approach. The concentration axis shifts were likely because of the different methods used to determine the peptide concentrations because within a certain range of MRM intensities, the concentrations of the synthetic peptides were globally larger than the concentrations of the recombinant peptides.

Effect of Peptide Properties on the Yields of Synthesized Peptides—Synthetic peptides are often limited by their size and sequence. Even though advanced technologies, such as microwave acceleration, have been introduced into several brands of peptide synthesizers to increase the purity and yield of longer peptides, sequence dependence remains an unpredictable challenge. For example, neighboring tryptophan and arginine residues are usually difficult to synthesize because of the side reaction between the two amino acid precursors used in solid phase synthesis. It is also common to have issues when synthesizing hydrophobic peptides because they
Biosynthesis of Stable Isotope-Labeled Peptides for MRM

Fig. 5. Peptide yield from standard 30 μl PURE expression is not dependent on peptide properties. A, Box plot between peptide yield and size. B, Box plot between peptide yield and theoretical hydropathicity. A total of 28 peptides were included.

tend to aggregate. In contrast to the traditional approaches, we believe biosynthetic peptides produced from PURE should be immune from these problems. To confirm this hypothesis, we first tested the correlation between the synthetic peptide yields and the corresponding sizes of peptides. Because of the detection limit of the triple quadrupole mass spectrometer, the length of the peptides used in targeted proteomics experiments usually ranges from six to 20 residues. We collected most of the peptides synthesized in this study and divided them into four groups based on their lengths (Group 1, 6 ~ 8 residues; Group 2, 9 ~ 11 residues; Group 3, 12 ~ 15 residues; and Group 4, 15 ~ 20 residues). Fig. 5A shows that the yields of the synthesized peptides spanned from 20 fmol/μl to 80 fmol/μl. The median yields in these groups differed, although statistical evaluation by ANOVA indicated no significant differences for the peptides in all of the groups (p value = 0.01). Because a limited number of peptides (n = 28) were taken for statistic, however, an influence of peptide length on the yield through PURE could not be excluded.

Next, we tested whether the peptide hydrophobicity affected the yield of peptide synthesis in PURE. Most of the peptides synthesized and quantified in the study were divided into four groups (Group 1, −1 ~ −0.6; Group 2, −0.6 ~ −0.1, Group 3, −0.1 ~ 0.3, and Group 4, 0.3 ~ 0.9) based on their hydrophobicity (GRAVY score) as calculated by the online tool at http://web.expasy.org PROTParam. As shown in Fig. 5B, in which the same statistical treatment was used to evaluate the yield difference in the four groups, the ANOVA analysis suggested the yield variations for the designed sequences independent from hydrophobicity (p value = 0.34). Nevertheless, this assessment was limited by the small number of peptides studied.

The evidence described above reveals significant advantages of producing biosynthetic peptides using the PURE system. Unlike solid-phase synthesis, biosynthesis is performed in a mild aqueous condition without the requirement of a harsh condensation step, so all building block molecules are native amino acids, not the protected counterparts used in chemical synthesis. The yield and purity of biosynthetic peptides are independent on their physical properties such as size and sequence. As a result, it is much more convenient and cost-effective to generate isotope-labeled reference peptides for targeted proteomics projects.

Quantification of the Peptide in Complex Biological Samples—One of the most remarkable features of QQQ is its ability to recognize specific peptides from complex biological backgrounds, such as digested peptides from serum proteins. The recombinant peptides in our approach are expected to be useful for applications involving serum. Glutathione S-transferases (GSTs) are a protein family widely involved in detoxification and anti-oxidation (21–24). The GST protein abundance in serum is reported in response to carcinogenesis, neurological disease and drug resistance (25, 26). For a long time, GST quantitative measurement has been mainly performed by ELISA or enzyme activity. Restricted by antibody and substrate specificity, these methods are still questioned because of the unsatisfactory quality of the GST quantification. The development of a new approach to accurately quantify GST is badly needed in both research and clinical applications (27). We herein propose that the isotope GSTs synthesized by our method could serve as internal standards in serum, and the endogenous GSTs in the serum could be measured at the level of absolute quantification.

A total of four GST peptides were selected for expression in vitro using PURE: YIATPIFSK*(GSTM1), VTYVDFLVYDVLDQHR*(GSTM2), QFSLFLGK*(GSTM5), and AILNYIASK*(GSTA3). First, the MS2 features of endogenous and synthesized GSTs were carefully examined by QTOF 5600 mass spectrometry. For instance, the synthesized GSTM1 was spiked into the mouse serum peptide mixture after tryptic digestion as a reference peptide, and the MS2 spectra and retention time for the two peptides were monitored. The evidence described above reveals significant advantages of producing biosynthetic peptides using the PURE system:

1. Unlike solid-phase synthesis, biosynthesis is performed in a mild aqueous condition without the requirement of a harsh condensation step.
2. All building block molecules are native amino acids, not the protected counterparts used in chemical synthesis.
3. The yield and purity of biosynthetic peptides are independent on their physical properties such as size and sequence.

As a result, it is much more convenient and cost-effective to generate isotope-labeled reference peptides for targeted proteomics projects.

The evidence described above reveals significant advantages of producing biosynthetic peptides using the PURE system.
599.3660, 700.4126, 884.5383, and 771.4515 as the y3, y5, y6, y8, and y7 ions of the synthesized GSTM1. It is clear that the m/z values for all of the y ions of the native GSTM1 were globally less than the corresponding y ions of the synthesized GSTM1 by 8 Da. This result demonstrated that the synthesized GSTM1 truly represented its endogenous form and that the transitions from this peptide could be selected for further MRM assays. Based on the MS2 spectra for all of the synthesized GST peptides, their MRM features are presented in supplemental Fig. S3A. Next, we sought to evaluate the absolute abundance of the GSTs in mouse serum. After the mouse sera were treated by GST affinity chromatography and complete tryptic digestion, the pool of PURE synthesized GST peptides was added to each sample, followed by MRM-MS. A total of 8 individual mouse sera were used for the GST quantification with duplicated injections for each sample. The quantitative curves were plotted by the concentrations of isotopic peptides against their MRM signals (peak area), and all of the linear regressions for the tested GST peptides were satisfactory for the quantitative evaluation of the endogenous GSTs in mouse serum (supplemental Fig. S3B). The absolute abundance of the four GSTs in the 8 mouse sera were estimated asGSTM1, 4.08 ± 1.87 pmol/L; GSTM2, 98.07 ± 26.50 pmol/L; GSTM5, 2.48 ± 1.02 pmol/L; and GSTA3, 4.28 ± 1.03 pmol/L (supplemental Table S2). Although the CVs obtained from the duplication of each sample were well accepted, less than 10%, the relatively large S.D. values of the average GST abundance implied the presence of individual errors for the GST isoforms in the mouse sera (n = 8). In short, the GSTs synthesized by PURE were proven as typical internal standards for the absolute quantification of GSTs in body fluids. Importantly, the GST concentrations measured by MRM indicated that the superfamily members exhibited a widely dynamic range in their abundance, which might benefit the further functional investigation of GSTs.

CONCLUSIONS

PURE, a system of reconstituted Escherichia coli translation machinery, can offer a general and rapid approach to produce many stable isotope-labeled peptides in a 96-well plate within a short period. Our strategy using PURE exhibits several unique features ideal for the production of MRM reference peptides. First, as a reconstituted mixture from purified components, the synthesis system is protease-free and lacks enzymes for post-translational protein modifications, so peptides can be expressed in their native form and without degradation. Second, the isotope labeling of peptides can be almost 100% isotopically pure based on the observation of MS spectra. Third, hundreds of peptides, regardless of their lengths and sequences, can be expressed, purified, and quantified. Fourth, the recombinant peptides can be accurately and globally quantified based on the MRM Strep-tag signals. Finally, this method was successfully employed to measure the abundance of GSTs in mouse serum at absolute quantification. Compared with QconCAT, our approach can generate isotopic peptides within 1–2 days from the target gene design to final peptide quantification, whereas the QconCAT approach takes longer. Meanwhile, we realize that this method still faces several technical limitations. Because of the lack of modification enzymes involved in PURE system, it is limited to synthesizing post-translational peptides, and when a large quantity of peptide is required, it is restricted by
Biosynthesis of Stable Isotope-Labeled Peptides for MRM

synthesis yield or batch errors caused by multiple syntheses. In summary, the isotope-labeled peptides that are expressed and quantified in parallel are quite suitable for MRM studies with many protein candidates. Our approach will therefore certainly find broad applications in both basic science and translational medicine.

Acknowledgments—We thank Professor Borchers for help in providing high-quality peptides with stable isotope labeling and quantification. The mass spectrometry data have been uploaded to PeptideAtlas with the identifier PASS00757 (http://www.peptideatlas.org/).

* This work was supported by the National High Technology Research and Development Program of China (2012AA020206, 2014AA020906) and the National Basic Research Program of China (2011CB910704).

† † This article contains supplemental material.

‡‡ To whom correspondence should be addressed: Beijing Institute of Genomics, NO. 1 Beichen West Road, Chaoyang District, Beijing 100101, China. Tel.: 0086-10-8409-7465; Fax: 0086-10-8409-7465; E-mail: siqiliu@genomics.cn; Department of Chemistry and Biology, University of Maryland, College Park, MD 20742. Tel.: (240) 314-6330; Fax: (240) 314-6330; E-mail: siil@umd.edu (S. Li).

REFERENCES
1. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat. Methods 9, 555–566
2. Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., Domon, B., Deutsch, E. W., Grant, R. P., Hoofnagle, A. N., Huttenhain, R., Koome, J. M., Liebler, D. C., Liu, T., MacLean, B., Mani, D. R., Mansfield, E., Neubert, H., Paulovich, A. G., Reiter, L., Vitek, O., Aebersold, R., Anderson, L., Bethem, R., Blonder, J., Boja, E., Botelho, J., Boyne, M., Bradshaw, R. A., Burlage, A. L. Chan, D., Keshishian, H., Kuhn, E., Kisinger, C., Lee, J. S., Lee, S. W., Mortiz, R., Oses-Prieto, J., Ritali, N., Ritchie, J., Rodriguez, H., Srinivas, P. R., Townsend, R. R., Van Eyk, J., Whiteley, G., Wilta, and Weintraub, S. (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Mol. Cell. Proteomics 13, 907–917
3. Meng, Z., and Veenstra, T. D. (2011) Targeted mass spectrometry approaches for protein biomarker verification. J. Proteomics 74, 2950–2968
4. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. U.S.A. 100, 6940–6945
5. Stemmann, O., Zou, H., Gerber, S. A., Gygi, S. P., and Kirschner, M. W. (2001) Dual inhibition of sister chromatid separation at metaphase. Cell 107, 715–726
6. Pan, S., Zhang, H., Rush, J., Eng, J., Zhang, N., Patterson, D., Comb, M. J., and Aebersold, R. (2005) High throughput proteome screening for biomarker detection. Mol. Cell. Proteomics 4, 182–190
7. Mirzaei, H., McBee, J. K., Watts, J., and Aebersold, R. (2008) Comparative evaluation of current peptide production platforms used in absolute quantification in proteomics. Mol. Cell. Proteomics 7, 619–623
8. Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J., and Beynon, R. J. (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat. Protocols 1, 1029–1043
9. Beynon, R. J., Doherty, M. K., Pratt, J. M., and Gaskell, S. J. (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nat. Methods 2, 587–589
10. Hino, M., Kataoka, M., Kajimoto, K., Yamamoto, T., Kido, J.-I., Shinohara, Y., and Baba, Y. (2008) Efficiency of cell-free protein synthesis based on a crude cell extract from Escherichia coli, wheat germ, and rabbit reticulocytes. J. Biotechnol. 133, 183–189
11. Anderson, L., and Hunter, C. L. (2008) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol. Cell. Proteomics 7, 573–588
12. Mohammed, Y., Domanski, D., Jackson, A. M., Smith, D. S., Deelder, A. M., Palmblad, M., and Borchers, C. H. (2014) PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. J. Proteomics 106, 151–161
13. Qi, E., Omasits, U., Goetze, S., Stekhoven, D. J., Frey, J. E., Basler, K., Wollscheid, B., Brunner, E., and Ahrens, C. H. (2014) Improved prediction of peptide detectability for targeted proteomics using a rank-based algorithm and organism-specific data. J. Proteomics 108, 269–283
14. Stergachis, A. B., MacLean, B., Lee, K., Stamatoyanopoulos, J. A., and MacCoss, M. J. (2011) Rapid empirical discovery of optimal peptides for targeted proteomics. Nat. Methods 8, 1041–1043
15. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. Nat. Biotechnol. 19, 751–755
16. Ge, X., Luo, D., and Xu, J. (2011) Cell-free protein expression using macromolecular crowding conditions. PLoS ONE 6, e28707
17. Pahler, A., Hendrickson, W. A., Kolks, M. A., Argarana, C. E., and Cantor, C. R. (1987) Characterization and crystallization of core streptavidin. J. Biol. Chem. 262, 13933–13937
18. Schmidt, T. G., and Skerra, A. (1993) The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. Protein Engineering 6, 109–122
19. Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G., and Blanquet, S. (1989) Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. U.S.A. 86, 8247–8251
20. Pan, Y., Cheng, K., Mao, J., Liu, F., Liu, J., Ye, M., and Zou, H. (2014) Quantitative proteomics reveals the kinetics of trypsin-catalyzed protein digestion. Anal. Bioanal. Chem. 406, 6247–6256
21. Hayes, J. D., Flanagan, J. U., and Jowsey, I. R. (2005) Glutathione transferases. Annu. Rev. Pharmacol. Toxicol. 45, 51–88
22. Farkas, M. H., Berry, J. O., and Aga, D. S. (2007) Chlortetracycline detoxification in maize via induction of glutathione S-transferases after antibiotic exposure. Environ. Sci. Technol. 41, 1450–1456
23. Balogh, L. M., Roberts, A. G., Shireman, L. M., Greene, R. J., and Atkins, W. M. (2008) The stereochemical course of 4-hydroxy-2-nonenal metabolism by glutathione S-transferases. J. Biol. Chem. 283, 16702–16710
24. Lapatinina, E. A., Sudnikovich, E. J., Maksimchik, J. Z., Zabrodskaya, S. V., Zavodnik, L. B., Kubyshev, V. L., Nocun, M., Kazmierczak, P., Dobaczewski, W., Watala, C., and Zavodnik, I. B. (2006) Antioxidative enzyme and glutathione S-transferase activities in diabetic rats exposed to long-term ASA treatment. Life Sci. 79, 1804–1811
25. Gupta, R. K., Kesari, A. N., Divakar, S., Tyagi, A., Tandon, V., Chandra, R., and Watal, G. (2008) In vivo evaluation of anti-oxidant and anti-lipidemic potential of Annona squamosa aqueous extract in Type 2 diabetic models. J. Ethnopharmacol. 118, 21–25
26. Kalinina, E. V., Chernov, N. N., Saprin, A. N., Kotova, Y. N., Remizov, V. I., and Shcherbner, B. N. (2007) Expression of genes for reduct-dependent glutathione S-transferase isofoms GSTP1-1 and GSTA4-4 in tumor cell during the development doxorubicin resistance. Bull. Exp. Biol. Med. 143, 328–330
27. Thomson, R. E., Bigley, A. L., Foster, J. P., Jowsey, I. R., Elcombe, C. R., Onrot, T. C., and Hayes, J. D. (2004) Tissue-specific expression and subcellular distribution of murine glutathione S-transferase class kappa. J. Histochem. Cytochem. 52, 653–662