

## High Performance Cell-Free Wheat Germ Protein Expression System

### Application Note

#### MEERCAT: Enabling large-scale quantitative proteome analysis

##### Abstract

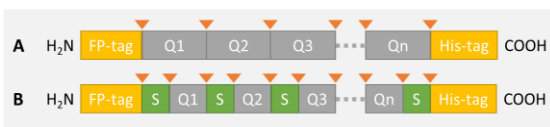
Protein quantification by SRM (Selected Reaction Monitoring) analysis of peptides is the gold standard for quantitative proteomics in research and clinical applications. While initially focusing on few selected proteins, recent technologies allow for proteome-wide studies using large sets of isotope-labeled reference peptides. Instead of chemically synthesizing each peptide, multiple peptides can be concatenated into a “QconCAT”. After expression of the QconCAT gene, individual peptides are then released from the QconCAT in a co-digestion step with the analyte. MEERCAT now extends the QconCAT concept by shifting to CFS’ wheat germ cell-free expression system which enables highly parallel preparation of tens of thousands of standard peptides for studies on complex proteomes.

##### Introduction

Accurate protein quantification is important for understanding protein complexes, biological processes or for diagnoses from clinical samples [1]. Targeted mass spectrometry (MS) is the most universal and accurate approach, and different strategies have been developed to determine relative and absolute protein amounts e.g. by label-free methods, comparing samples by SILAC, or using an isotope-labeled reference protein (PSAQ) or peptide (AQUA). Stable-isotope-labeled reference peptides are preferred for proteomics as such reference peptides can be made to distinguish many different proteins in the same sample [2]. However, prior information on each peptide is needed before it can be used in MS experiments. Since peptide detectability is not fully predictable [3], each peptide must be experimentally verified and characterized. Reference databases like PeptideAtlas [4], ProteomicsDB [5], or ProteomeTools [6] provide information on thousands of peptides to help selecting suitable reference peptides for each

experiment. These peptides are commonly obtained by chemical synthesis, which is costly when many labeled peptides are required. Therefore, Beynon et al. developed **QconCATs** (Quantification **C**oncatamers) to make the use of many peptides affordable [7, 8]. A QconCAT is an artificial protein made by gene design to encode multiple reference peptides that are released during co-digestion with analyte proteins. The known amount of the isotopically-labeled QconCAT allows for accurate, multiplexed preparation of several reference peptide standards for quantification of their unlabeled counterparts within a sample preparation. QconCATs have routinely been made in *E. coli* [8]. More recently, a shift to expression *in vitro*, using our wheat germ cell-free protein expression system, overcame several limitations of the *E. coli* system and has facilitated highly-multiplexed expression reactions for the simultaneous production of tens of thousands of reference peptides. This new cell-free method was termed “MEERCAT” for **Multiplexed Efficient Expression of Recombinant QconCATs** [9].

## Application Note 4/2018: MEERCAT - scaling up QconCAT production



**Figure 1:** Schematic description of QconCAT proteins: In a QconCAT individual peptides (Q1 to Qn) are concatenated in one longer protein. The peptide sequences may be linked directly (A) or, more commonly, are separated by flanking sequences (S) that mimic digestion sites (orange triangles) in the native proteins (B). QconCATs have a His-tag at the C-terminus to assure the purification of the full-length protein. In MEERCAT an additional “mass-coded” FP-tag had been added to the N-terminus to barcode each QconCAT in multiplexed expression reactions.

### Use of cell-free protein expression in MS

The wheat germ cell-free protein expression system has a long record of use in protein research, including the preparation of reference proteins for MS [10]. The universal power of our system for large-scale protein production was demonstrated by expressing over 18,000 human recombinant proteins from the AIST and ORFeome clone collections [11]. Those proteins were digested and then chemically labeled with mTRAQ  $\Delta$  4 to prepare a genome-wide peptide resource. The resulting database comprises 216,476 unique peptides for 17,973 proteins, representing about 86.3% of all (20,819) human protein-coding genes. However, chemical labeling can be fully avoided by direct incorporation of <sup>13</sup>C/<sup>15</sup>N labeled lysine and arginine during a cell-free protein expression reaction [12]. Direct protein labeling in our wheat germ system was demonstrated in several studies e.g. working on transmembrane proteins [13] or using the FLEXIQuant method to prepare full-length reference proteins for MS experiments [14, 15]. Takemori et al. were the first to use this approach for high-throughput production of a stable-isotope-labeled library comprising 162 QconCATs covering 2201 selected peptides achieving a 25-fold gain in efficiency over previous QconCAT experiments [16]. In their protein expression and labeling experiments, they could achieve up to 99% labeling efficiency for incorporating [<sup>13</sup>C, <sup>15</sup>N]-L-Lys and [<sup>13</sup>C, <sup>15</sup>N]-L-Arg. This greatly increases the sensitivity of MS analysis and provides for a wider dynamic range than commonly

possible when using *in vivo* labeling methods. Thus, the wheat germ cell-free protein expression system is a preferable method to prepare stable-isotope-labeled reference proteins.

### MEERCAT: Enhancing QconCAT Production

Following up on their first publication, the Takemori group at Ehime University and the Beynon group at the University of Liverpool [9] tested together whether the wheat germ cell-free system could overcome problems experienced while preparing QconCATs from *E. coli* cells. Based on the experience from working on over 100 individual QconCATs, about 1 out of 10 QconCATs could not be expressed at high levels in *E. coli*. Moreover, some of the QconCATs that were expressed in *E. coli* were proteolyzed during expression and subsequent purification, reducing their usefulness.

New expression vectors for the wheat germ system were prepared by PCR cloning starting from existing expression vectors to prepare a set of 12 QconCATs that could not be obtained from *E. coli* (11 cases failed to express, and one had been degraded). Therefore, this study kept on using inserts codon-optimized for expression in *E. coli*. Using the new expression vectors in small-scale 240  $\mu$ l cell-free expression reactions, the team could obtain all QconCATs at an average concentration of about 0.1  $\mu$ g/ $\mu$ l with an incorporation rate for the isotope-labeled lysine and arginine of 99.6%, even though the templates had not been optimized for the wheat germ system. For one QconCAT, they could further show that the proteins made in the wheat germ system are less likely to be degraded as compared to working in *E. coli*. Obtaining some 20 to 30  $\mu$ g for each QconCAT in a simple bilayer reaction enables hundreds of SRM assays, making this reaction scale suitable for routinely preparing reference proteins for use in MS studies within a day.

The open nature of a cell-free protein expression reaction offers great flexibility over the reaction conditions. This includes the use of several expression

## Application Note 4/2018: MEERCAT - scaling up QconCAT production

vectors in the same translation reaction for the simultaneous coexpression of multiple proteins. The groups used the same set of 12 QconCATs from the forgoing experiment to demonstrate the coexpression of all 12 proteins in one translation reaction. In this experiment, they obtained even better protein yields for each QconCAT as compared to the individual expression reactions described above. To quantify each QconCAT individually in complex reaction mixtures, sequence variants of glu fibrinopeptide (FP) differing by a single amino acid, were encoded in each QconCAT, introducing “mass-coded” tags at the N-terminus of each of the different QconCATs for independent quantification in a 12-plex expression reaction. After coexpression, each of the variant tags gave sharp chromatographic peaks and yielded the expected fragmentation spectra, confirming that mass-coded tags could be used to quantify each QconCAT in the 12-plex mixture.

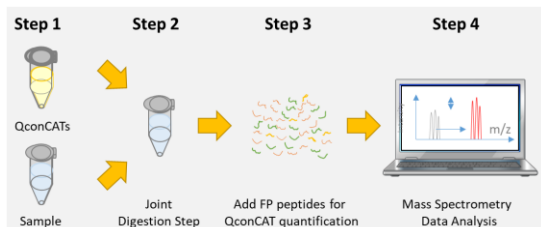


Figure 2: Outline of MEERCAT experiment: 1. A pool of isotope-labeled QconCATs is prepared in a multiplexed expression reaction, 2. QconCATs and the sample are mixed and digested together, 3. Unlabeled FP peptides are added to the digestion products to quantify individual QconCATs, and 4. Labeled peptides derived from QconCATs and unlabeled FP peptides are analyzed by MS along with peptides derived from the sample.

The groups then attempted independent coexpression of 76 proteins, each about 70 to 75 kDa long and together comprising about 4,000 standard peptides for quantification of a significant part of the *S. cerevisiae* proteome [17]. Out of this set, 71 proteins could be expressed individually and later combined in one joint protein expression reaction for detection by MS in an SRM experiment. In their experiments, the authors saw a weak correlation between the protein yields and the amount of template DNA used in the expression reactions.

Therefore, normalizing the DNA amounts could be a way to achieve more equal protein yields from complex protein expression reactions. In fact, the authors were able to obtain 149 out of 150 small QconCATs of about 25 kDa from a single protein expression reaction pushing even further on the possible complexity of such reactions. To keep protein yields up even working with very complex reactions, the volumes of their bilayer translation reactions were increased up to a 6 ml scale.

The successful coexpression of many QconCATs in a single cell-free protein expression reaction encouraged the authors to suggest the preparation of a QconCAT reference set of 25,000 proteins, enough to cover the human proteome. Using twelve 96-well plates, 1152 expression templates could be made to cover every human protein with at least two independent peptides. This resource, which is also infinitely reproducible, could be prepared at a much lower cost than using chemical synthesis to make some 56,000 labeled peptides.

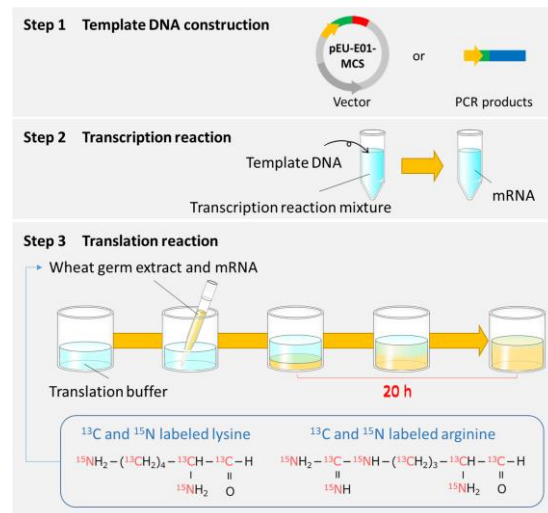


Figure 3: Outline of the workflow to prepare  $^{13}\text{C}$  and  $^{15}\text{N}$  lysine and arginine labeled QconCATs, or other protein standards, using the Premium PLUS Expression Kit for MS: The three-step process includes 1. Preparation of an expression template by PCR or using the pEU-E01-MCS expression vector included in the kit, 2. Performing transcription reaction to obtain RNA from expression template, and 3. Using RNA in bilayer translation reaction to synthesize the labeled protein in the wheat germ system. The labeled amino acids are already included in the wheat germ extract provided with the kit.

## Application Note 4/2018: MEERCAT - scaling up QconCAT production

### Conclusion

The MEERCAT method demonstrates the benefits of the wheat germ cell-free protein expression system in the preparation of isotope-labeled standards for MS experiments. With the Premium PLUS Expression Kit for MS (CFS-PLUS-MS), CFS offers ready-to-use premixed reagents to perform protein labeling reactions in a simple three-step process. <sup>13</sup>C and <sup>15</sup>N-labeled lysine and arginine are already added to the reagents, and reaction conditions have been optimized to achieve about 99% labeling efficiency. The Premium PLUS Expression Kit for MS can be directly used to prepare QconCATs on a 226 µl scale as described in the MEERCAT publication.

Using cell-free protein expression can make it easier to obtain more QconCAT standards for protein MS or other applications like RePLiCAL, QconCAT proteins used for retention time standardization in proteomics [18]. We hope that the MEERCAT approach and individual QconCATs will pave the way to extend the scope of protein quantification experiments.

### References

1. Gillette, M.A. and S.A. Carr, *Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry*. Nat Methods, 2013. **10**(1): p. 28-34.
2. Hoofnagle, A.N., et al., *Recommendations for the Generation, Quantification, Storage, and Handling of Peptides Used for Mass Spectrometry-Based Assays*. Clin Chem, 2016. **62**(1): p. 48-69.
3. Muntel, J., et al., *Abundance-based classifier for the prediction of mass spectrometric peptide detectability upon enrichment (PPA)*. Mol Cell Proteomics, 2015. **14**(2): p. 430-40.
4. Deutsch, E.W., *The PeptideAtlas Project*. Methods Mol Biol, 2010. **604**: p. 285-96.
5. Schmidt, T., et al., *ProteomicsDB*. Nucleic Acids Res, 2018. **46**(D1): p. D1271-D1281.
6. Zolg, D.P., et al., *Building ProteomeTools based on a complete synthetic human proteome*. Nat Methods, 2017. **14**(3): p. 259-262.
7. Beynon, R.J., et al., *Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides*. Nat Methods, 2005. **2**(8): p. 587-9.
8. Simpson, D.M. and R.J. Beynon, *QconCATs: design and expression of concatenated protein standards for multiplexed protein quantification*. Anal Bioanal Chem, 2012. **404**(4): p. 977-89.
9. Takemori, N., et al., *MEERCAT: Multiplexed Efficient Cell Free Expression of Recombinant QconCATs For Large Scale Absolute Proteome Quantification*. Mol Cell Proteomics, 2017. **16**(12): p. 2169-2183.
10. Harbers, M., *Wheat germ systems for cell-free protein expression*. FEBS letters, 2014. **588**(17): p. 2762-73.
11. Matsumoto, M., et al., *A large-scale targeted proteomics assay resource based on an in vitro human proteome*. Nat Methods, 2017. **14**(3): p. 251-258.
12. Narumi, R., et al., *Cell-free synthesis of stable isotope-labeled internal standards for targeted quantitative proteomics*. Synth Syst Biotechnol, 2018. **3**(2): p. 97-104.
13. Takemori, N., et al., *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**(2): p. 361-5.
14. Singh, S., et al., *A practical guide to the FLEXIQuant method*. Methods in molecular biology, 2012. **893**: p. 295-319.
15. Singh, S., et al., *FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides*. Journal of proteome research, 2009. **8**(5): p. 2201-10.
16. Takemori, N., et al., *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**(8): p. 2389-93.
17. Lawless, C., et al., *Direct and Absolute Quantification of over 1800 Yeast Proteins via Selected Reaction Monitoring*. Mol Cell Proteomics, 2016. **15**(4): p. 1309-22.
18. Holman, S.W., L. McLean, and C.E. Eyers, *RePLiCal: A QconCAT Protein for Retention Time Standardization in Proteomics Studies*. J Proteome Res, 2016. **15**(3): p. 1090-102.

### Acknowledgement

We are very grateful for the support of Prof Robert J. Beynon at the University of Liverpool and Dr Nobuaki Takemori at Ehime University.

**Follow us on Twitter (@CFSciences) for up-to-date information on new publications, products and the use of our expression system.**

CFS is certified under ISO 9001:2015. - CFS products and services are for research purpose only. - Refer to our homepage at <http://www.cfsciences.com/eg/> for more information on our products and services, or contact us directly at [tech-sales@cfsciences.com](mailto:tech-sales@cfsciences.com).