


METHODOLOGY ARTICLE

Open Access

Construction of à la carte QconCAT protein standards for multiplexed quantification of user-specified target proteins



James Johnson¹, Victoria M. Harman², Catarina Franco², Edward Emmott², Nichola Rockliffe¹, Yaqi Sun³, Lu-Ning Liu³, Ayako Takemori⁴, Nobuaki Takemori⁴ and Robert J. Beynon^{2*} 

Abstract

Background: QconCATs are quantitative concatamers for proteomic applications that yield stoichiometric quantities of sets of stable isotope-labelled internal standards. However, changing a QconCAT design, for example, to replace poorly performing peptide standards has been a protracted process.

Results: We report a new approach to the assembly and construction of QconCATs, based on synthetic biology precepts of biobricks, making use of loop assembly to construct larger entities from individual biobricks. The basic building block (a Qbrick) is a segment of DNA that encodes two or more quantification peptides for a single protein, readily held in a repository as a library resource. These Qbricks are then assembled in a one tube ligation reaction that enforces the order of assembly, to yield short QconCATs that are useable for small quantification products. However, the DNA context of the short construct also allows a second cycle of loop assembly such that five different short QconCATs can be assembled into a longer QconCAT in a second, single tube ligation. From a library of Qbricks, a bespoke QconCAT can be assembled quickly and efficiently in a form suitable for expression and labelling in vivo or in vitro.

Conclusions: We refer to this approach as the ALACAT strategy as it permits *à la carte* design of quantification standards. ALACAT methodology is a major gain in flexibility of QconCAT implementation as it supports rapid editing and improvement of QconCATs and permits, for example, substitution of one peptide by another.

Keywords: Synthetic biology, Standards, Absolute quantification, Stable isotopes, QconCAT, Quantitative proteomics, Loop assembly

Background

Absolute quantification of proteins by mass spectrometry is typically based on the use of accurately quantified stable isotope labelled internal standards, usually peptides, as surrogates for the protein quantification. There are many ways to generate these labelled peptides, including direct chemical synthesis (AQUA peptides [1,

2]), from full length labelled proteins (PSAQ [3–5]) or shorter epitopic fragments (PrEST [6, 7]). An additional approach is the use of QconCAT technology [4, 8, 9]. QconCATs are multiplexed protein standards for proteomics, products of artificial genes designed to encode concatamers of peptides, wherein each peptide or more commonly, a pair of peptides, is chosen to act as mass spectrometric standard(s) for absolute quantification of multiple peptides. The initial publications on QconCATs [10–12] have received over 1000 citations and the methodology is well known and embedded in the community.

* Correspondence: rbeynon@liverpool.ac.uk

²Centre for Proteome Research, Institute of Systems and Integrative Biology, University of Liverpool, Crown Street, Liverpool L697ZB, UK
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

At a typical size of about 60–70 kDa, a QconCAT encodes approximately 50 tryptic peptides, permitting the quantification of around 25 proteins at a ratio of two peptides per protein. The genes are then expressed as recombinant proteins in bacteria grown in the presence of SIL amino acids, usually lysine and arginine, ensuring a single labelling position for every standard tryptic peptide. Because the genes are designed *de novo*, it is feasible to introduce additional features, such as purification tags, sacrificial peptides to protect the QconCAT from exoproteolysis and peptide sequences, common to each QconCAT as a quantification standard, permitting absolute quantification of each standard within the proteomics workflow—in effect, an ‘internally standardised standard’. We have demonstrated that the QconCAT approach can be used successfully in large scale proteome quantification studies and have reported the absolute quantification of approx. 1800 proteins in the *Saccharomyces cerevisiae* proteome [13], by far the largest absolute quantification study conducted to date. Because each peptide derived by complete excision from a QconCAT are present in equal quantities, QconCATs also have utility in the determination of subunit stoichiometry of multiprotein complexes, such as the proteinaceous bacterial metabolosomes for propanediol degradation in *Salmonella* [14].

Although QconCATs have been widely adopted, their broader deployment can be challenging. First, QconCAT expression requires skills in molecular biology and facilities for bacterial expression of heterologous proteins. We have addressed this in part through the introduction of cell-free synthesis of QconCATs, which brings added advantages of concurrent, single tube synthesis that we have extended to over 100 QconCATs simultaneously, a strategy we have dubbed MEERCAT [15, 16]. Secondly, QconCATs cover a set of target proteins based on the needs of one research group, which may not always match the requirements of subsequent research groups. Thirdly, the choice of peptides is often obliged to be made without knowledge of the performance of these peptides in absolute quantification. Lastly, editing of any QconCAT, for example, the removal or addition of a single protein, has required complete resynthesis and expression of the gene.

To overcome these complications, we now introduce the concept of ‘ALACATS’ - ‘à la carte’ QconCATs, the term reflecting the ability to design a QconCAT of any length that encodes peptides for a user-specified set of target proteins. ALACATS are assembled from ‘Qbricks’, oligonucleotides that encode (typically) two quantotypic peptides for a single target protein, together with short flanking peptides to recapitulate the correct primary sequence context, and thus normalise digestion rates. Each Qbrick (one for each target protein in the proteome) is a

discrete entity, a double-stranded DNA construct that can be readily synthesised, stored, catalogued and accessed to enable the synthesis of an ALACAT to order. These are the fundamental building blocks in the ALACAT workflow.

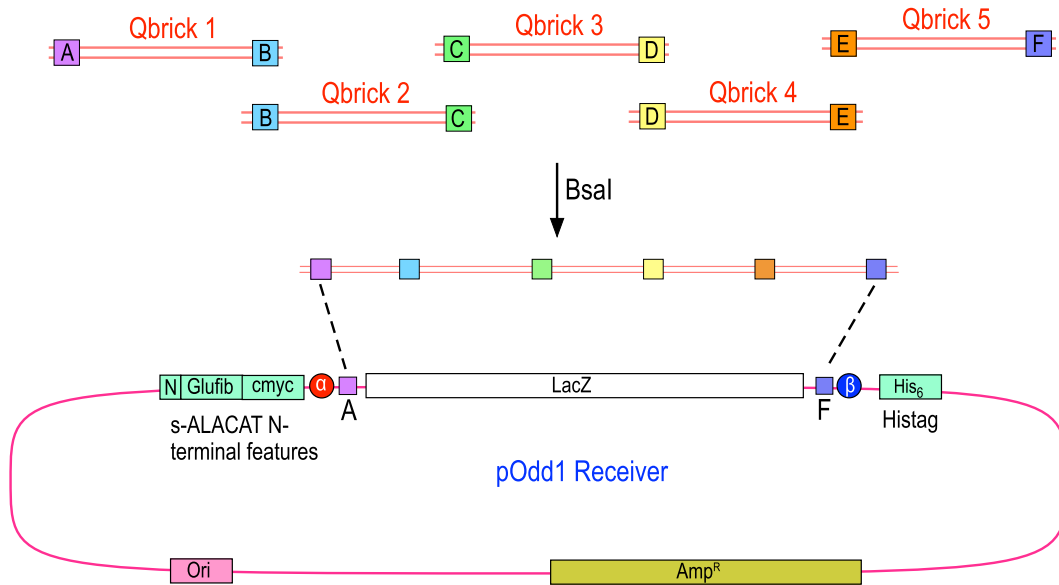
Results

Design, synthesis and assembly of Qbricks

A Qbrick (‘quantification brick’, a type of biobrick [17]) is defined as a short, double-stranded oligonucleotide that encodes two or more Qpeptides that are quantotypic for a single protein and is thus the smallest building block (Fig. 1). The Qbrick also encodes interspersed peptide sequences that recapitulate the primary sequence context of the two peptides, thus equalising digestion rates of standard and analyte. Each Qbrick has asymmetric overhangs at each end, creating sticky ended DNA molecules that permit assembly by a strategy called ‘Loop Assembly’, driven by sequential use of two Type IIS restriction endonucleases [18]. Different unique overhang sequences (A, B, C, D, E and F) flank the Qbricks. Five Qbricks are assembled in a single reaction—the ‘odd’ cycle [18]. Annealing during assembly maintains the reading frame through the QconCAT, adding two amino acids to the interspersed linker with no effect on the peptides generated from the QconCAT (Fig. 1a). These *short QconCATs*, assemblies of five Qbricks that encode 10 peptides, are perfectly useable when expressed as a five-target protein standard suitable for small, focused studies. A short QconCAT, containing 10 Qpeptides (from five Qbricks), interspersed linker peptides, quantification and purification peptides as well as suitable sacrificial sequences at either end, totals approximately 170–220 amino acids, of a size suitable for expression and deployment.

For more wide-reaching quantification studies, individual short ALACATS are subsequently concatenated in a second reaction. The initial 5-Qbrick constructs are cloned into plasmids that introduce a second set of six overhanging linker sequences, distinct from those used in the ‘odd’ cycle. These linkers (‘even cycle’, α , β , γ , δ , ϵ , λ ; Fig. 1b) allow assembly of the five short ALACATS into a complete, ‘*long ALACAT*’, capable of encoding Qpeptides for quantification of approximately 25 proteins. These QconCATs would be 75–90 kDa (relatively shorter because of the single instances of N-terminal and C-terminal features), typical for cell-free or bacterial expression. Of course, any variant, from two to five short ALACATS, encoding quantification standards for any number of proteins between 5 and approximately 30 is possible using this approach. This greatly expands the flexibility of the QconCAT approach. The sequences of the constructs used in this paper, and the cloning syntaxes, are provided in Additional file 1.

a) Odd cycle to create short ALACATs



b) Even cycle to create full length, long ALACATs

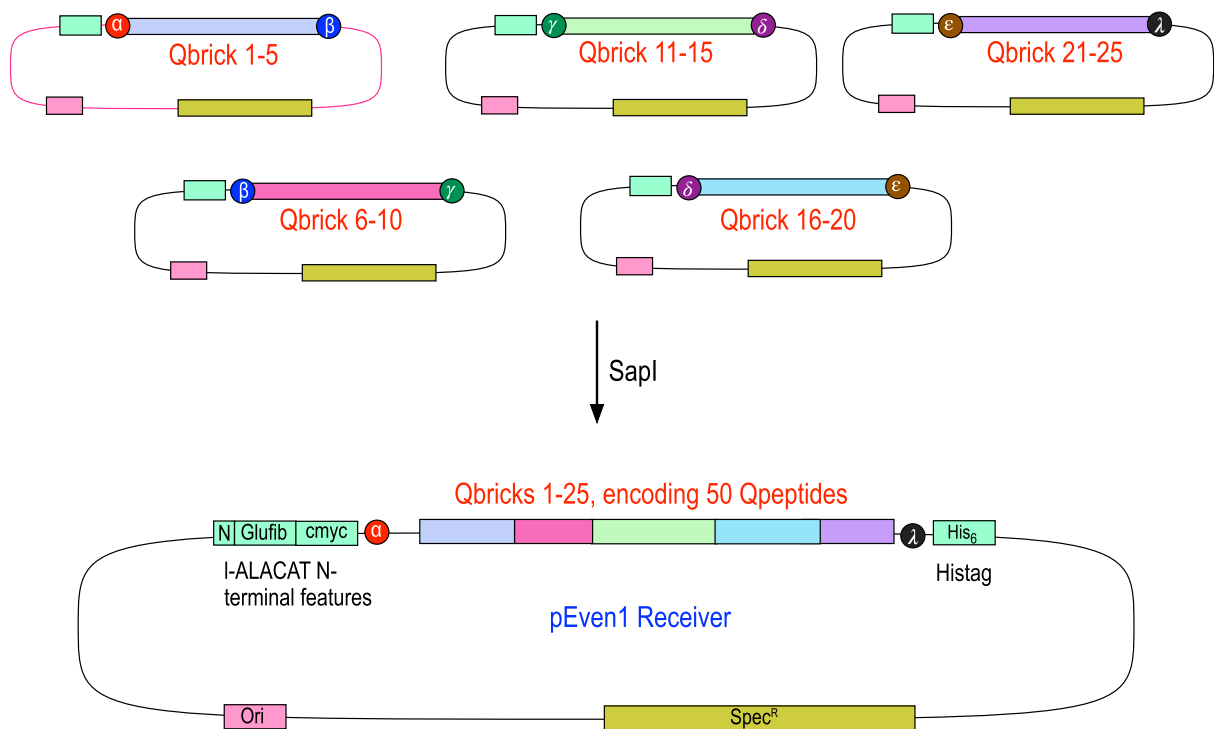


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Overall strategy for building block assembly of short and long ALACATs. The smallest unit of an ALACAT is a single double-stranded oligonucleotide encoding peptides (one, two, more) for quantification of a single target protein, flanked on either side by tripeptides that preserve the natural primary sequence context. These oligonucleotides include linker regions compatible with a Type IIS restriction enzyme (*Bsal*) that allows all five Qbricks to be assembled in the correct order in a single ligation reaction (**a**, odd cycle assembly) to form short ALACATs. In turn, these short ALACATs contain DNA sequences that are compatible with a second Type IIS restriction enzyme (*SapI*) and can be similarly assembled in a one tube reaction to create a long ALACAT (**b**, even cycle assembly). The vectors for the odd and even cycles both include in-frame fusions to glu-fibrinopeptide and c-myc encoding regions (quantification) and a hexahistidine tag (purification)

As proof of concept, we built an ALACAT from a series of Qbricks encoding standards for 25 human plasma proteins (Additional file 2: Table S1). We first assembled five short ALACATs and then, in turn, assembled these into a long ALACAT. Each short ALACAT was expressed independently in a wheat germ cell-free system (CFS), as well as the long ALACAT and yields of all were high (Fig. 2a). Typically, yields were of the order of 500 pmol, which is a substantial quantity for LC-MS/MS-based quantification (typically, a single LC-MSMS run would require 50 fmol on column). The expressed short and long QconCATs were then digested with trypsin and analysed by LC-MS/MS (Fig. 2b). All peptides were readily detected (Fig. 2c), and whether derived from a short or long ALACAT, the relative peptide intensities remained the same (Fig. 2d). Further information on the analysis of the short and long ALACATs is provided in Additional file 2: Figures S1-S6.

Even after purification, QconCAT preparations also contained several proteins derived from the expression system. To assess these, we searched digests of a different set (C) of purified QconCATs against a wheat database (*Triticum aestivum*, UniProt UP000019116). The proteins were very consistent across six constructs (five short ALACATs and one long ALACAT), and the most abundant proteins were ribosomal proteins and seed storage proteins; other proteins were one to two orders of magnitude lower (Additional file 2: Figure S7). However, the QconCATs were the most abundant proteins on SDS-PAGE, and because they are deployed using specific MS assays, contaminant peptides would not be an issue.

Each QconCAT contained two peptides common to every construct—the glu fibrinopeptide (EGVNDNEEGF FSAR) that we have used previously for quantification of the QconCAT [8, 13] and a second peptide derived from the common c-myc peptide (LISEEDLGGR) to give a tag for monitoring expression by western blotting if necessary. We were able to use these two peptides to assess the consistency of the intensity of the quantification peptide, whether in long or short ALACATs (Fig. 2c). The correlation was extremely high, confirming that the peptides were cleaved and released similarly, irrespective of the nature of the ALACAT.

We also demonstrated the synthesis of ALACAT in a *E. coli* cell-free system. The *E. coli* system couples

transcription and translation in a single tube, which allows us to skip the *in vitro* transcription reaction required in the wheat cell-free system. In this study, we set up a small-scale reaction system using a microdialysis device (Fig. 3). All prepared ALACAT genes were successfully synthesized in this system, and the efficiency of $^{13}\text{C}/^{15}\text{N}$ incorporation into their lysine and arginine residues was more than 99%.

Editability of ALACATs

One of the advantages of the ALACAT approach is the introduction of straightforward editability of the construct. Previously, there was no simple route to exchange one peptide for another without extensive resynthesis of the gene. However, with ALACATs, the editability simplifies the introduction of changes in the sequence and embedded peptides. This editing process can take place at two levels. First, individual peptides can be replaced in Qbricks, and a new short ALACAT could be constructed. The only new DNA required would be the sequence of the Qbrick. Alternatively, an entire short ALACAT could be exchanged, replacing multiple peptides in a single process. This might be of value, for example, in a multi-species construct, if some short ALACATs contained species-specific peptide sequences, and others contained sequences that were identical in both species. A simple switch from species A to species B would only require an exchange of the relevant short ALACAT in the one-step, even cycle ligation reaction. Further, about 10% of all traditionally synthesised QconCATs failed to express in bacteria [16] and the ability to quickly create a large set of rearrangements of different Qbricks or short ALACATs would be able to deliver a library of ALACATs, with equivalent function, that could be quickly screened for expression potential. Alternatively, this type of combinatorial synthesis could be used to explore adjacency and proximity effects. To test this possibility *in extremis*, we therefore initiated a ‘one pot’ combinatorial ligation of two families of Qbricks, or, in a separate experiment, two families of short ALACATs.

We tested both levels of editability using the two ALACAT series (B, plasma and H for analysis of the stoichiometry of a metabolic compartment; Additional file 1) described above. First, we demonstrated the ease of exchange of short ALACATs by building a combinatorial

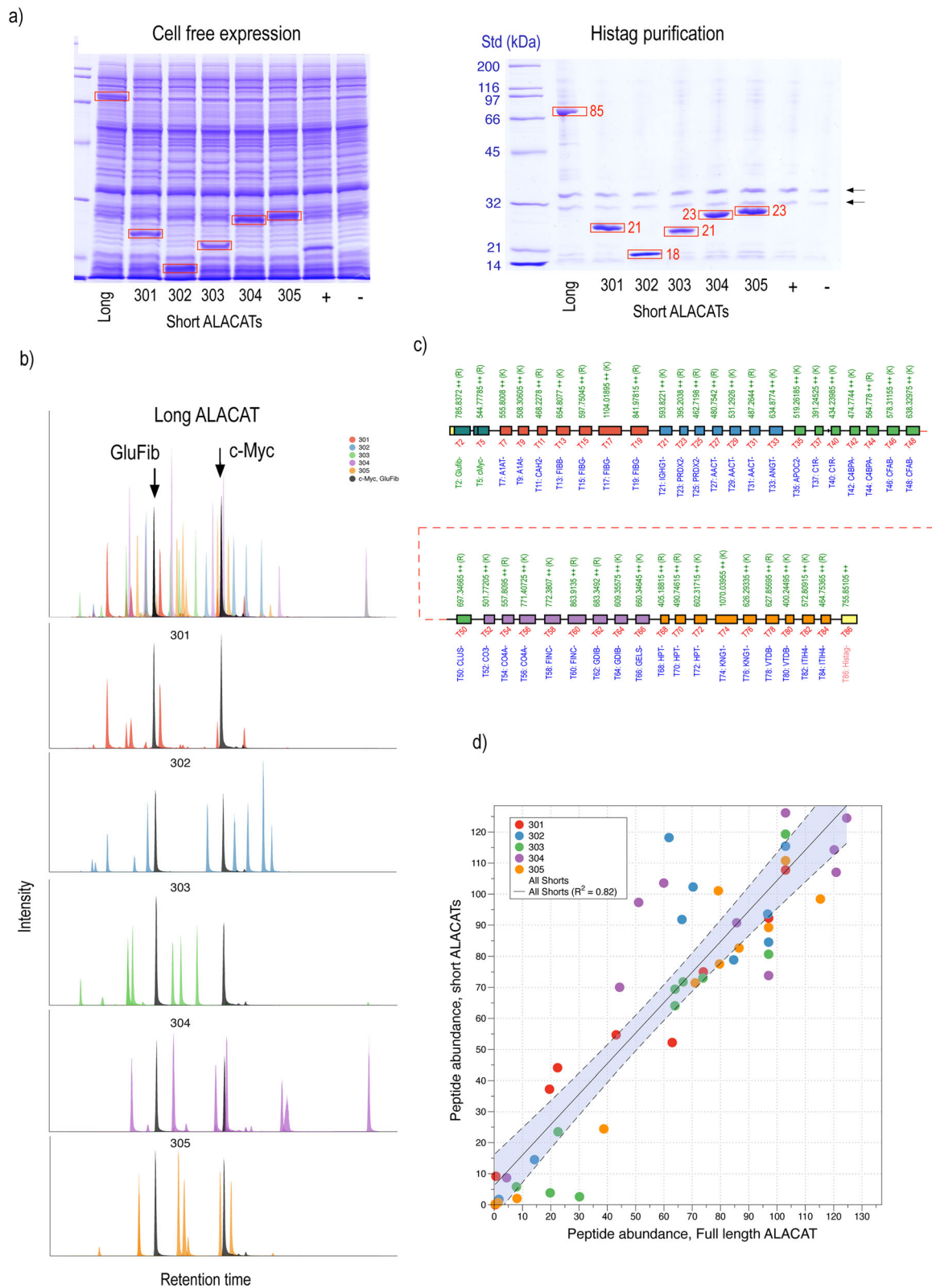


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Construction of a human plasma protein QconCAT using the ALACAT strategy. A series of Qbricks were designed, each encoding two peptides for each of 25 plasma proteins. Groups of five were assembled in an odd cycle reaction to short ALACATs (301-305) that were expressed in vitro using wheat germ lysate and purified by virtue of their hexahistidine tag. In addition, the short ALACATs were assembled in an even cycle reaction into a single long ALACAT that was also expressed and purified (a). Each of the ALACATs was then digested and analysed by LC-MS/MS; all peptides were detected (c, infilled peptides are those visible by LC-MS/MS (b); different colours define the short ALACAT origins of the peptides in the long ALACAT). Using common peptides (glu fibrinopeptide and c-myc epitope) as normalisation controls, the peak areas for peptides in short ALACATs were compared to the areas of the same peptides in the long ALACATs (d)

series of long ALACATs created from random introduction of appropriate short ALACATs—the ‘even’ cycle. Each position in the long ALACAT could contain a short ALACAT from either the B or H series. Rather than create one editing reaction to prove the swap of one short ALACAT for another, we took a different approach and set up a single reaction, in which we mixed ten short ALACATs derived from the two different families, prefixed B and H, such that B1 and H1 would share common *SapI* overhang sequences and similarly, the other four pairs (H2/B2 to H5/B5). Thus, short ALACATs H1 and B1 would represent a binary choice at position one. In this assembly, a random ligation process would generate $2^5 = 32$ different combinations, from H1:H2:H3:H4:H5 through, for example, H1:H2:B3:B4:B5 etc. to B1:B2:B3:B4:B5. After the single tube ligation, 81 colonies were picked and the ligation product DNA was sequenced. Of the 32 combinations that could have been synthesised, we detected 26 different short ALACATs (80% of all possible different combinations, Fig. 4, Additional file 3), confirming the ease of editing and reconstruction of new short ALACATs. There was no indication of any systematic bias in the selection of one or the other sets of Qbricks, establishing the ease of random ligation.

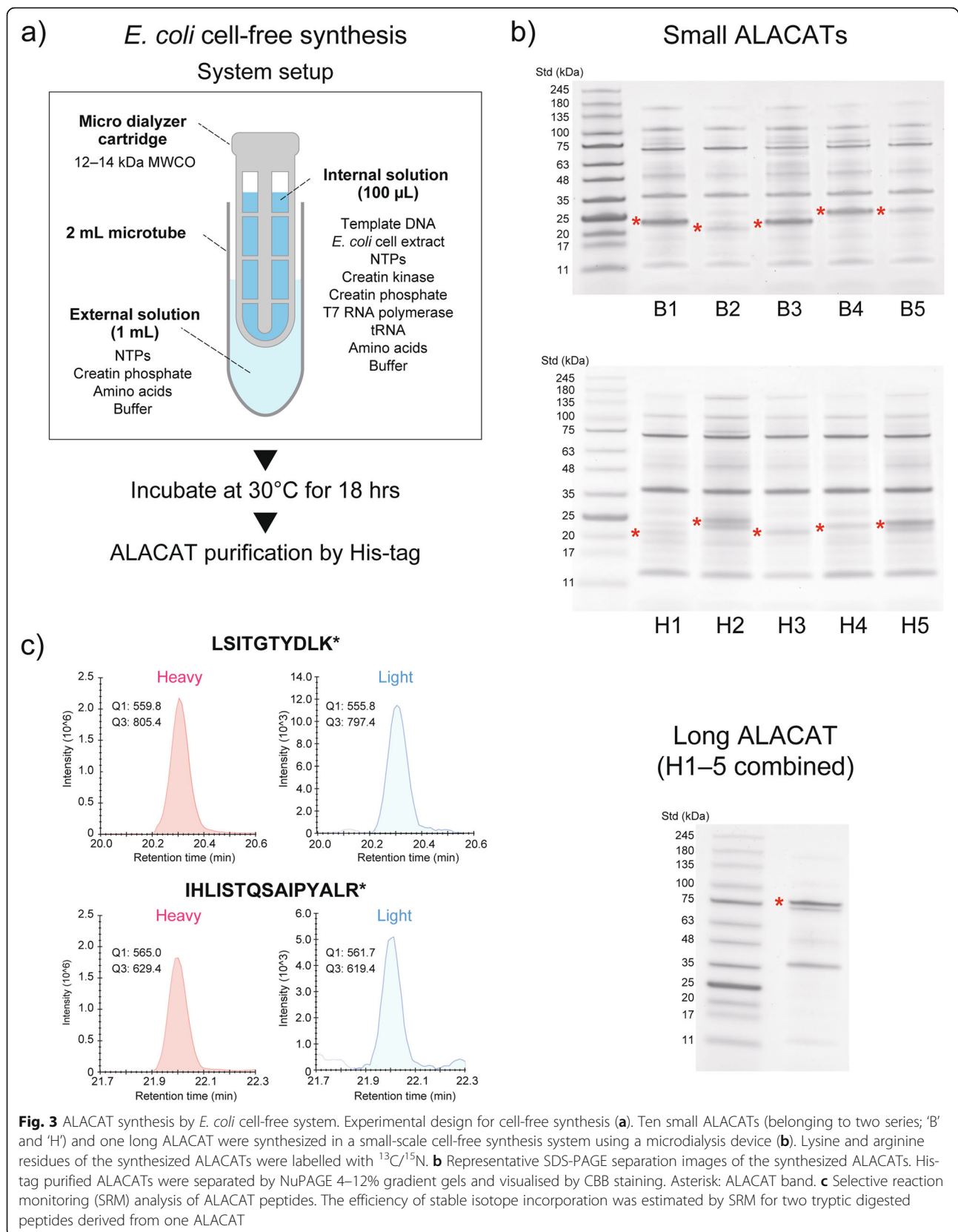
To assess the equivalent combinatorial substitution of Qbricks, we performed essentially the same experiment in an ‘even’ cycle but with two sets of Qbricks from the two families (B and H), again picking multiple clones from a single tube ligation reaction. To increase complexity, we created further potential by providing H4 and B5 with two assembly contexts (Fig. 5, Additional file 4). After assembly, multiple ALACAT clones were picked and sequenced. From this experiment, 26 unique short ALACATs were constructed, spread across 65 sequences that were sampled. Of these 65, seven were long variants of five Qbricks (made possible by our construction strategy) but the majority comprised assemblies of four Qbricks, a total of 19 combinations from a set of 24 possibilities were recovered. Further, 16 were unique, two were replicated once, three occurred thrice, two were four-fold, up to one assembly that was sequenced in 16 (approx. 25%) of the clones. It is possible that this bias reflected differences in the relative concentrations of the input DNA sequences, which would allow tuning of the system to preferred assemblies.

Discussion

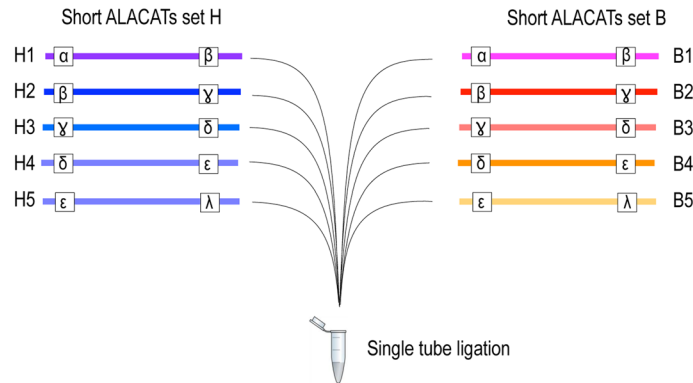
Although the QconCAT approach is well recognised, there are undoubtedly barriers to widespread adoption. The selection of peptides is an early commitment, followed by the synthesis of a gene, embedded in a suitable expression vector, and finally, expression and labelling by biosynthesis in vivo. This requires routine skills in molecular biology that may not be present in a typical proteomics team. Moreover, QconCAT expression in vivo is no different to expression of other heterologous proteins; sometimes, the expression fails. However, we have demonstrated that synthesis in vitro overcomes this issue—so far, every QconCAT we have made by cell-free synthesis has not only expressed well, but also incorporates a higher degree of labelling [16].

The ALACAT concept has multiple advantages over traditional QconCAT gene synthesis and expression. The synthesis and storage of individual Qbrick oligonucleotides is straightforward and in time, these Qbricks could be drawn from an ever-expanding library, stored at the point of synthesis. Once a set of Qbricks are available, assembly through the odd cycle is a single tube reaction, creating the possibility of expression in vitro of a short ALACAT; useful for a quick check of the suitability of the encoded peptides. When the short ALACATs have been evaluated, a second, single tube reaction leads to the even cycle assembly of the full length QconCAT. A primary advantage of the ALACAT approach is therefore that the clustering of Qpeptides into QconCATs becomes a late decision, driven by the interests of specific users and/or research programmes. If peptides are sub-optimal for a particular mass spectrometric approach, often an unknown factor before the construct is made, it would be trivial to replace one Qbrick, build a new short ALACAT, and if required, subsequently assemble the new short ALACAT into the full length ALACAT, both steps being single-tube reactions.

Further, ALACATs can be designed and delivered at any size (although we recommend an upper limit of 50 to 60 target proteins), according to the focus and depth of individual quantitative proteomics studies. For example, a single short ALACAT would be a rapidly generated resource for quantification of a few key proteins. To increase the target numbers, two or three short ALACATs could be combined to form highly efficient



a) Combinatorial ligation of short ALACATs

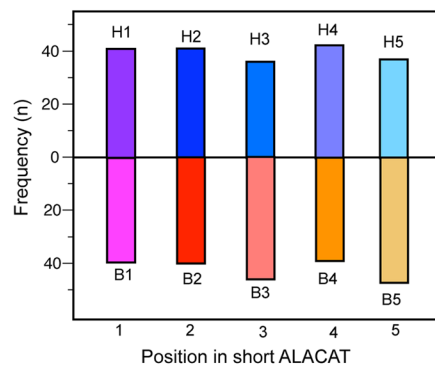


b) Colony picking, DNA sequencing

α	β	γ	δ	ϵ	λ	Count
B1	B2	B3	B4	B5		2
B1	B2	B3	B4	H5		2
B1	B2	B3	H4	B5		4
B1	B2	H3	B4	B5		2
B1	B2	H3	H4	B5		5
B1	H2	B3	B4	B5		1
B1	H2	B3	B4	H5		5
B1	H2	B3	H4	B5		5
B1	H2	H3	B4	B5		5
B1	H2	H3	B4	H5		4
B1	H2	H3	H4	B5		3
B1	H2	H3	H4	H5		2
H1	B2	B3	B4	B5		6
H1	B2	B3	B4	H5		3
H1	B2	B3	H4	B5		3
H1	B2	B3	H4	H5		5
H1	B2	H3	B4	B5		3
H1	B2	H3	H4	B5		2
H1	B2	H3	H4	H5		3
H1	H2	B3	B4	B5		3
H1	H2	B3	H4	B5		4
H1	H2	B3	H4	H5		2
H1	H2	H3	B4	B5		1
H1	H2	H3	B4	H5		2
H1	H2	H3	H4	B5		1
H1	H2	H3	H4	H5		3

Long ALACAT assemblies (81 clones)

c) Usage of short ALACATs



d) Distribution (first three short ALACAT positions)

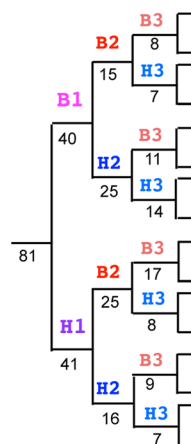


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Combinatorial assembly of short ALACATs into long ALACATs. To test the ease of editability and swapping of short ALACATs into long ALACATs, we created a ligation reaction in which there were two choices of short ALACAT at each of four or five positions (a). The reaction products were cloned and a total of 81 clones were selected for DNA sequencing, to establish the composition of the long ALACATs (from either the B or H series) were incorporated (b). There was an even representation of the short ALACATs across the entire structure (c) and the evenness of the creation of the products is evidenced by the split of the products through the first three positions (d)

QconCATs of intermediate size. Of course, multiple ALACATs can be co-expressed *in vivo*, simplifying resource expansion and providing an efficiency and flexibility which, coupled with cell-free synthesis and MEERCAT, means that large scale absolute proteome quantifications are now eminently feasible, sustainable and modifiable.

Many stages of the ALACAT workflow are suitable for delivery through laboratory automation, reducing the need for human intervention. The Qbrick approach means that it would be possible to create an ever-expanding resource of Qbrick DNA (in the form of double-stranded oligonucleotides) that could be assembled ‘to order’ in response to requests by any research group. There would be no reliance on prior clustering of peptides, and the assembly would be a simple additional step. Moreover, the ability to ‘swap out’ specific Qbricks without having to redesign and build the QconCAT from scratch means that problematic peptides will be rapidly expunged from the resource (Fig. 6). The advantages of having ‘editable QconCATs’ cannot be overstated. This added flexibility in standard design and optimisation, coupled with ever-increasing selectivity and sensitivity of LC-MS/MS platforms, makes absolute quantification of part, or even all, of a proteome increasingly feasible.

Finally, QconCATs, assemblies of peptides generated by proteolytic digestion, are a simple route to the generation of stoichiometric quantities of sets of peptides that can be used for purposes other than absolute quantification, such as instrument quality control or calibration of retention time index [19–23]. The combinatorial experiments described in this paper, for example, create the ability to build a large number of different combinations of peptides from a common library, and could be used in the understanding of local influences on ionisation, or even to test the emergent methods for prediction of precursor or product ion intensity [24–29]. The more straightforward the production methodology, the more likely tests of such predictive methods can be created.

Conclusions

- QconCATs can now be assembled from libraries of Qbricks, where each Qbrick is a short oligonucleotide that encodes quantotypic peptides for a target protein.

- The assembly of Qbricks in the ALACAT process can create QconCATs, at two or more peptides per target protein, in two successive one tube reactions, giving substantial time savings.
- The ALACAT approach allows for rapid replacement of one Qbrick for another, whether to introduce superior peptides or to replace one target protein with another.
- The ALACAT approach sets the stage for large libraries of Qbricks that can be assembled in a bespoke, *à la carte* fashion according to specific project goals.

Methods

Materials and reagents

All enzymes, competent cells and manual DNA purification kits were purchased from New England Biolabs (Hitchin, UK), all oligonucleotides were purchased desalted and lyophilised from Integrated DNA Technologies BVBA (Leuven, Belgium) or Eurofins Genomics (Ebersberg, Germany). All bacterial media and antibiotics were purchased from Formedium Ltd (Hunstanton, UK).

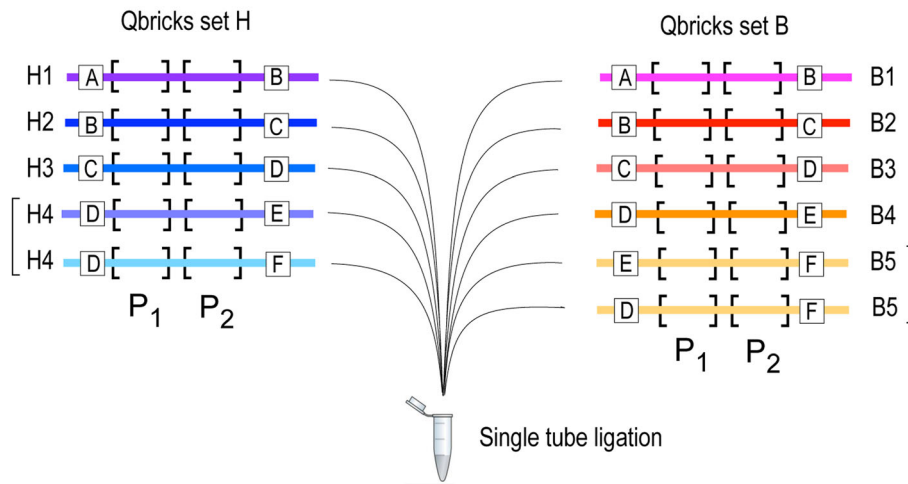
Production of pOdd and pEven acceptor vectors

Plasmid pEU01-MCS (CellFree Sciences, Ehime, Japan) was domesticated via site-directed mutagenesis to remove unwanted *Bsa*I and *Sap*I restriction sites. pOdd vectors were produced by inserting a lacZ cassette with *Sap*I and *Bsa*I sites as indicated with appropriate syntaxes flanked N-terminally by GluFib and Myc tag linkers and C-terminally with 6x His-Tag and stop codons. pEven vectors were produced similarly but the Amp^R gene was exchanged for Spec^R gene amplified from pGM134_1 and cloned via an NEBuilder (NEB, UK) reaction. All lacZ cassettes were synthesised by Twist Bioscience (San Francisco, USA) and cloned as single fragments into modified pEU01-MCS via NEBuilder, producing pOdd vectors pGM247_2 – 6 and pEven vectors pGM247_8 – 12.

Design of oligonucleotides and production of Qbrick DNA Blocks

Qbrick peptide sequences were reverse translated using Geneious software (Biomatters Ltd), set up to use the *Escherichia coli* K12 codon usage table [30] and to avoid internal restriction sites of *Bsa*I, *Sap*I, *Bbs*I and *Bsm*BI

a) Combinatorial ligation



b) Colony picking, DNA sequencing

Assemblies (65 clones)

Four Qbrick constructs

A	B	C	D	E	Count
B1	H2	B3	B4		1
H1	B2	H3	B4		1
B1	B2	B3	B5		1
H1	B2	B3	B5		3
B1	H2	B3	B5		1
H1	H2	B3	B5		1
B1	B2	H3	B5		1
H1	B2	H3	B5		1
B1	H2	H3	B5		2
H1	H2	H3	B5		1
B1	B2	B3	H4		4
H1	B2	B3	H4		1
B1	H2	B3	H4		7
H1	H2	B3	H4		16
B1	B2	H3	H4		4
H1	B2	H3	H4		3
H1	B2	H3	H4		1
B1	H2	H3	H4		3
H1	H2	H3	H4		5

Five Qbrick constructs

A	B	C	D	E	F	Count
H1	B2	B3	B4	B5		1
H1	H2	H3	B4	B5		1
B1	B2	B3	H4	B5		1
H1	B2	B3	H4	B5		1
H1	H2	B3	H4	B5		1
H1	B2	H3	H4	B5		1
B1	H2	H3	H4	B5		2

c) Usage of Qbricks

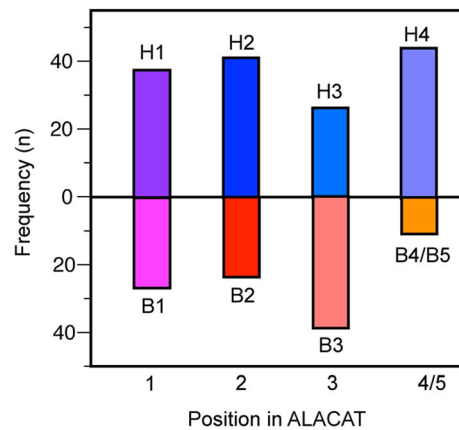


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Combinatorial assembly of Qbricks into short ALACATs. To test the ease of editability and swapping of Qbricks into short ALACATs, we created a ligation reaction in which there were two choices of Qbrick at each of four or five positions (a), encoding pairs of quantotypic peptides P₁ and P₂. The reaction products were cloned and a total of 65 clones were selected for DNA sequencing, to establish the composition of the long ALACATs (from either the B or H series) were incorporated (b). There was a reasonably even representation of the Qbricks across the entire structure (c)

and >5 nt homopolymers. These were converted to overlapping oligonucleotides (T_{ann} approx. 60 °C). Required 5' overhangs for *BsaI* or *SapI* recognition sequences and molecular syntaxes were then added. Pairs of overlapping oligonucleotides were mixed at 2.5 μM ea. (final conc.) in Q5 2x mastermix in 20 μl total reaction. These were annealed and extended using the following thermocycler parameters: 98 °C for 60 s followed by five cycles of 98 °C for 10 s, 60 °C for 30 s, 72 °C for 15 s and a final incubation at 72 °C for 60 s. These reactions were diluted 1:100 in water before added to cloning reactions below (approx. 25 fmol/μl).

Odd level cloning reactions (short ALACATs)

Required QBrick blocks (~ 25 fmol μL⁻¹) were reacted with empty pOdd vector in 10 μL total as follows: 0.5 μL each QBrick block, 10 ng pOdd vector, 1 μL T4 DNA Ligase Buffer, 0.5 μL T4 DNA Ligase (400 U μL⁻¹), 0.5 μL *BsaI* (10 U μL⁻¹). These were incubated with the following conditions: 37 °C for 10 min, followed by 50 cycles of 37 °C for 1 min and 16 °C for 1.5 min before final 50 °C for 5 min. 1 μL of the reaction was transformed into 25 μL of NEB 5-alpha chemically competent cells. 20% were plated onto LB agar plates containing 100 μg mL⁻¹ Carbenicillin, 20 μg mL⁻¹ X-Gal and 100 μM IPTG for blue/white selection. White colonies were then screened by colony PCR using universal screening primers (Forward: 5'-TAACCACCTATCTACATCACC-3' and Reverse: 5'-CGAGCTCGAGAACTAGTGAT-3'). PCR products were analysed using QIAxcel DNA Screening Gel automated electrophoresis (QIAGEN, Manchester, UK). Correct PCR products were cleaned with ExoCIP and Sanger sequenced (Source Bioscience Ltd, Nottingham, UK) to confirm the sequence.

Even level cloning reactions (Long ALACATs)

Required pOdd clones (10 ng μL⁻¹) were reacted with empty pEven vector in 10 μL total as follows: 0.5 μL each pOdd clone, 10 ng pEven vector, 1 μL 10x T4 Ligase Buffer, 0.5 μL T4 DNA Ligase (400 U μL⁻¹), 0.5 μL *SapI* (10 U μL⁻¹). These were incubated at 37 °C for 120 min. 1 μL of the reaction was transformed into 25 μL of NEB 5-alpha chemically competent cells. 20% were plated onto LB agar plates containing 50 μg mL⁻¹ Spectinomycin, 20 μg mL⁻¹ X-Gal and 100 μM IPTG. White colonies were then screened by colony PCR using universal screening primers (Forward: 5'-TAACCACCTATCTA

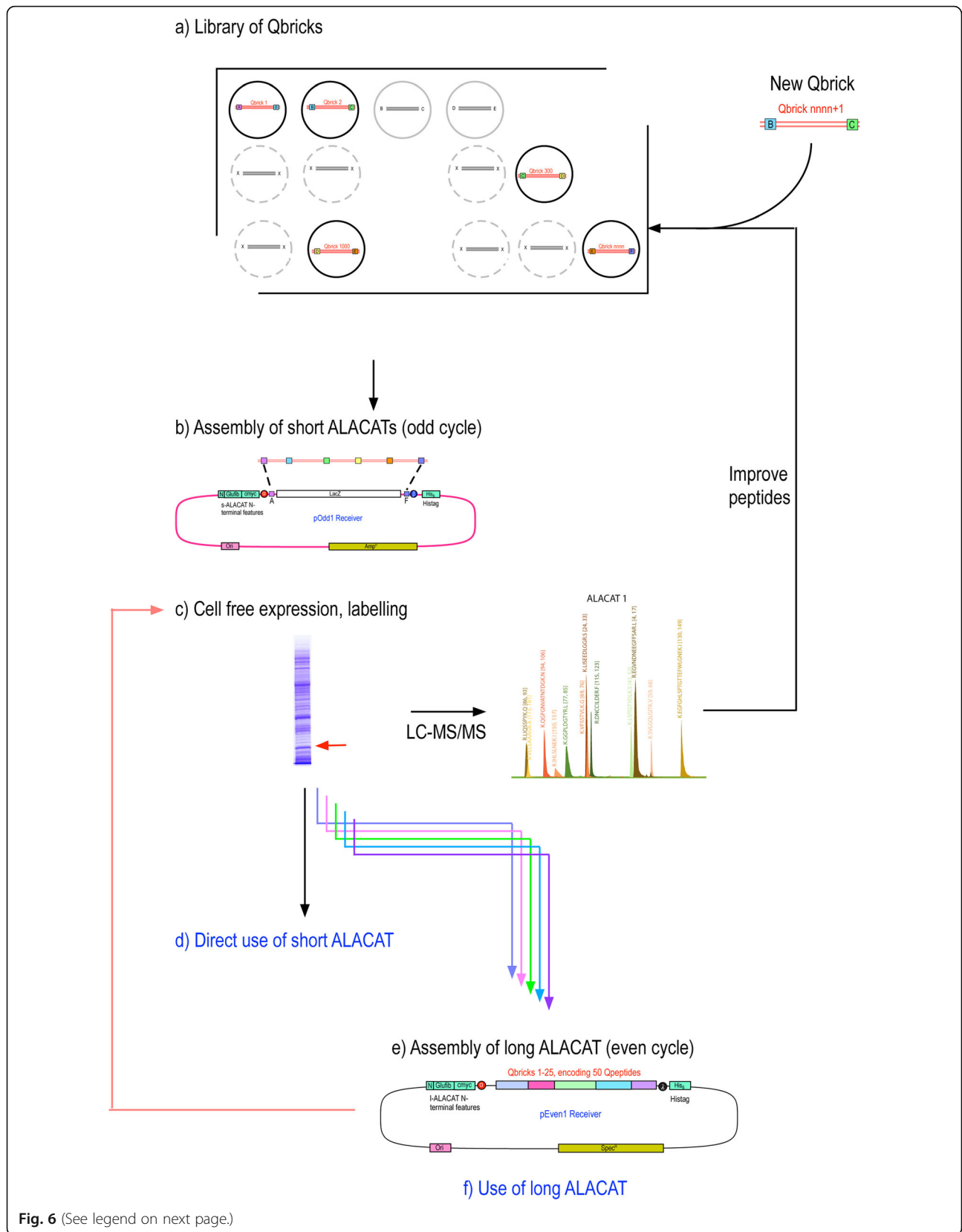
CATCACC-3' and Reverse: 5'-CGAGCTCGAGAACTAGTGAT-3'). PCR products were analysed using QIAxcel DNA Screening Gel electrophoresis.

Cell-free expression of short and long ALACATs

For each ALACAT, 2 μg DNA in pEU-E01 vector (Cell-Free Sciences Co., Ltd, Japan) was used for a single expression reaction. Synthesis was completed in 240 μL scale using WEPR8240H full Expression kit (2BScientific Ltd, UK). A positive control (pEU-E01-DHFR coding dihydrofolate reductase gene derived from *E. coli*) and negative control (pEU-E01-MCS empty vector) were used, both supplied with the kit. Full kit instructions were followed, including preparation of WEPRO8240H aliquots and 2 x SUB AMIX reagent. The Transcription Mix for each expression was prepared with 20 U RNase inhibitor, 20 U SP6 RNA Polymerase, 50 nmol NTP mix and a 0.2 x dilution Transcription Buffer. DNA for the ALACAT or controls, and nuclease-free water, were added to a final volume of 20 μL. The transcription reaction occurred over 6 h at 37 °C and the resulting mRNA was stored briefly at room temperature before transcription.

A 1 x SUB AMIX was prepared with a 0.5 x dilution of 2 x SUB AMIX into nuclease-free water and 60 nmol of each of the standard 20 amino acids (R, K, A, N, D, C, E, Q, G, H, I, L, M, F, P, S, T, W, Y, V), with substituted stable isotope labelled [¹³C₆],[¹⁵N₄] arginine and [¹³C₆],[¹⁵N₂] lysine (CK Isotopes Ltd, UK). The Translation Mix for each expression was prepared with 12 nmol of each of the same standard 20 amino acids, including ¹³C₆ ¹⁵N Arg and Lys, combined with 0.8 μg creatinine kinase, 10 μL WEPRO8240H wheat germ lysate, 0.5 x dilution of 2 x SUB AMIX, and 10 μL of mRNA for each ALACAT or standard. A 96-well plate was prepared with 200 μL of 1 x SUB AMIX in each well. The Translation Mixture was carefully pipetted beneath the SUB AMIX in each well to form a bilayer. The plate was sealed and incubated at 16 °C for 16 h.

E. coli cell-free synthesis was performed using a Musaibou-Kun protein synthesis kit (Catalog #A183-0242, Taiyo Nippon Sanso Corporation, Tokyo, Japan). For ALACAT synthesis, an amino acid cocktail with lysine and arginine universally labelled with ¹³C and ¹⁵N (Catalog # A91-0128, Taiyo Nippon Sanso Corporation) was used. All synthetic reactions were performed using an Xpress micro-dialyzer MD100 with molecular weight



(See figure on previous page.)

Fig. 6 A model for the creation of an ALACAT resource. From a library of Qbricks (that may include redundant sequences to increase the choice of quantotypic peptides for specific proteins) assembly into short ALACATs would provide a test system to assess the suitability of peptides for quantification. Once the short ALACATs are optimised, sets of them could subsequently be assembled into long ALACATs. Long ALACAT DNA can be readily used to drive protein synthesis in the end user laboratory, labelled as appropriate

cut-off of 12–14 kDa (Scienova, Spitzweidenweg, Germany) inserted into a 2 mL microtube. Before synthesis, 825 μ L of the outer solution was mixed with 75 μ L amino acid cocktail and 100 μ L distilled water, incubated at 30 °C, and added to the outside of the dialysis unit at the start of synthesis. Then, 77.5 μ L of the internal solution for synthesis was mixed with 10 μ L template DNA (50 ng/ μ L), 7.5 μ L amino acid cocktail, and 5 μ L distilled water, and added to the dialysis unit. The synthesis reaction was carried out at 30 °C for 18 h. After the synthesis was completed, all the solution in the dialysis unit was collected into a new tube.

ALACAT purification

Note that the positive control used for expression does not have a hexa-histidine tag and therefore both controls were used as negative controls in this next stage of the protocol. The 240 μ L contents of each individual well of the 96-well plate was transferred to a low binding tube (Biotix Inc., USA). This was then combined with 400 μ L Bind Buffer pH 7.4 (20 mM sodium phosphate, 0.5 M sodium chloride, 20 mM imidazole, 6 M guanidine hydrochloride), and incubated at room temperature for 1 h using a rotor mixer, before the addition of 10 μ L Ni Sepharose suspension (GE Healthcare Ltd, UK) and a further 1 h incubation. Centrifuge filters (Corning Costar Spin-X 0.45 μ m pore size cellulose acetate membrane, Merck, UK) were washed once with 750 μ L Bind Buffer and centrifuged, before the addition of the sample and Ni Sepharose, and further centrifugation; all centrifuge steps used 6000 \times g 2 min 4 °C. This was followed by three further washes by centrifugation with Bind Buffer; two 400 μ L washes and one 200 μ L wash. Sample was eluted by centrifugation from the resin with two additions of 15 μ L Elution Buffer pH 7.4 (20 mM sodium phosphate, 0.5 M sodium chloride, 1 M imidazole, 6 M guanidine hydrochloride), after each addition the resin and buffer were agitated to mix before centrifugation.

The final 30 μ L elution was transferred to a low binding tube for protein precipitation. To each tube, 600 μ L HPLC grade methanol (Fisher Scientific Ltd, UK) was added and mixed well before the addition of 150 μ L chloroform and 400 μ L HPLC grade water (VWR International, UK) to precipitate proteins. Following centrifugation at 13,000 \times g for 3 min a bilayer was formed, the uppermost layer of which was carefully removed. A further 600 μ L methanol was added and gently mixed by inversion. After a second centrifugation step the majority

of the liquid was removed and discarded, with the remaining liquid allowed to evaporate. The precipitate was resuspended in 30 μ L 25 mM ammonium bicarbonate, with 0.1% (w/v) RapiGestTM SF surfactant (Waters, UK) and protease inhibitors (Roche cOmpleteTM, Mini, EDTA-free Protease Inhibitor Cocktail, Merck, UK). Before tryptic digestion, the protein concentration of each sample was estimated using a NanoDrop Spectrophotometer (ThermoFisher Scientific, UK). All uncropped images of gels used in this publication are presented in Additional file 2: Figure S8.

Tryptic digestion

For digestion, 0.5 μ g protein for each was treated with 0.05% (w/v) RapiGestTM SF surfactant at 80 °C for 10 min, reduced with 4 mM dithiothreitol (Melford Laboratories Ltd., UK) at 60 °C for 10 min and subsequently alkylated with 14 mM iodoacetamide at room temperature for 30 min. Proteins were digested with 0.01 μ g Trypsin Gold, Mass Spectrometry Grade (Promega, USA) at 37 °C overnight. Digests were acidified by the addition of trifluoroacetic acid (Greyhound Chromatography and Allied Chemicals, UK) to a final concentration of 0.5% (v/v) and incubated at 37 °C for 45 min before centrifugation at 13,000 \times g 4 °C to remove insoluble non-peptidic material.

LC-MS/MS

Samples were analysed using an UltiMateTM 3000 RSLCnano system coupled to a Q ExactiveTM HF Hybrid Quadrupole-OrbitrapTM Mass Spectrometer (ThermoFisher Scientific, UK). Protein digests were loaded onto a trapping column (Acclaim PepMap 100 C18, 75 μ m \times 2 cm, 3 μ m packing material, 100 Å) using 0.1% (v/v) trifluoroacetic acid, 2% (v/v) acetonitrile in water at a flow rate of 12 μ L min⁻¹ for 7 min. For samples 301, 302 and 304, 5 ng was loaded, and for the Long ALACAT, 303 and 305, 10 ng was loaded. The peptides were eluted onto the analytical column (EASY-Spray PepMap RSLC C18, 75 μ m \times 50 cm, 2 μ m packing material, 100 Å) at 30 °C using a linear gradient of 30 min rising from 3% (v/v) acetonitrile/0.1% (v/v) formic acid (Fisher Scientific, UK) to 40% (v/v) acetonitrile/0.1% (v/v) formic acid at a flow rate of 300 nL min⁻¹. The column was then washed with 79% (v/v) acetonitrile/0.1% (v/v) formic acid for 5 min, and re-equilibrated to starting conditions. The nano-liquid chromatograph was operated under the control of Dionex Chromatography MS Link 2.14.

The nano-electrospray ionisation source was operated in positive polarity under the control of QExactive HF Tune (version 2.5.0.2042), with a spray voltage of 1.8 kV and a capillary temperature of 250 °C. The mass spectrometer was operated in data-dependent acquisition mode. Full MS survey scans between m/z 350–2000 were acquired at a mass resolution of 60,000 (full width at half maximum at m/z 200). For MS, the automatic gain control target was set to $3e^6$, and the maximum injection time was 100 ms. The 16 most intense precursor ions with charge states of 2–5 were selected for MS/MS with an isolation window of 2 m/z units. Product ion spectra were recorded between m/z 200–2000 at a mass resolution of 30,000 (full width at half maximum at m/z 200). For MS/MS, the automatic gain control target was set to $1e^5$, and the maximum injection time was 45 ms. Higher-energy collisional dissociation was performed to fragment the selected precursor ions using a normalised collision energy of 30%. Dynamic exclusion was set to 30 s.

The efficiency of stable-isotope incorporation of ALACAT synthesized in *E. coli* cell-free system was determined using LC-SRM analysis. Prior to SRM analysis, ALACAT protein B1 separated by SDS-PAGE (shown in Fig. 3b) was digested with trypsin in a gel and purified with a STAGE tip containing an Empore SDB-XC disc as described previously [16]. The purified peptide sample was dissolved in 20 μ L of 0.1% (v/v) TFA solution, of which 1 μ L was injected into an Eksigent nanoLC system coupled to a SCIEX QTRAP 5500 mass spectrometer. For LC separation, solvent A was 0.1% (v/v) formic acid, and solvent B was 0.1% (v/v) formic acid/80% (v/v) acetonitrile. The sample was desalted using a 200 μ m i.d. \times 0.5 mm cHiPLC trap column (SCIEX) at a flow rate of 5 μ L/minute for 10 min using 0.1% (v/v) TFA and then transported to a fused-silica capillary column packed with C18 resin (12.5 cm \times 75 μ m i.d.; Nikkyo Technos) at a flow rate of 300 nL/minute according to the following gradient schedule: 0–15 min, 2–50% B; 15–18 min, 50–90% B; hold at 90% B for 6 min; and re-equilibrate at 2% B for 20 min. SRM analysis was conducted in positive ion mode with the following parameters: ion spray voltage = 2300 V; curtain gas = 20; ion source gas1 = 20; collision gas = 12; interface heater temperature = 150; entrance potential = 10; collision cell exit potential = 9; and Q1/Q3 = low resolution. Details of SRM transitions and quantitative data are accessible via the Panorama-Web server [31] (<https://panoramaweb.org/alacat.url>).

The raw MS data files were loaded into Thermo Proteome Discoverer v.1.4 (ThermoFisher Scientific, UK) and searched against a custom ALACATs database using Mascot v.2.7 (Matrix Science London, UK) with trypsin as the specified enzyme, one missed cleavage allowed, carbamidomethylation of cysteine, label

[$^{13}\text{C}_6$][$^{15}\text{N}_2$]lysine and [$^{13}\text{C}_6$][$^{15}\text{N}_4$]arginine set as fixed modifications and oxidation of methionine set as a variable modification. A precursor mass tolerance of 10 ppm and a fragment ion mass tolerance of 0.01 Da were applied. To obtain a semi-quantitative profile of wheat proteins that might be present in purified ALACATs, we searched a series of ALACAT expressions against the Uniprot *Triticum aestivum* database (UP000019116) and imported the data into Progenesis QI (Waters Corp, Wilmslow, UK). The raw MS/MS files were also imported into the PEAKS software (Bioinformatics Solutions Inc, Waterloo, Canada) to obtain peptide coverage maps. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [32] partner repository with the dataset identifier PXD027521

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-021-01135-9>.

Additional file 1. DNA and protein sequences of construct, translated sequences and cloning syntaxes.

Additional file 2: Table 1. Proteins/peptides used in ALACAT B. **Figure S1.** Extracted ion chromatogram and peptide coverage map for short ALACAT 301. **Figure S2.** Extracted ion chromatogram and peptide coverage map for short ALACAT 302. **Figure S3.** Extracted ion chromatogram and peptide coverage map for short ALACAT 301. **Figure S4.** Extracted ion chromatogram and peptide coverage map for short ALACAT 304. **Figure S5.** Extracted ion chromatogram and peptide coverage map for short ALACAT 305. **Figure S6.** Extracted ion chromatogram and peptide coverage map for long ALACAT B. **Figure S7.** Wheat germ proteins present in purified QconCATs. **Figure S8.** Uncropped SDS-PAGE images.

Additional file 3. DNA products from long combinatorial study.

Additional file 4. DNA products from long combinatorial study.

Acknowledgements

We are grateful to Dr Philip Brownridge for excellent instrumentation support.

Authors' contributions

RJB and JJ conceived the approach and the study and obtained funding. Molecular cloning was by NR and JJ. JJ, EE, LL, YS and RJB contributed to ALACAT design. Mass spectrometric analysis was by CF, VMH, NT and AT. Wheat germ expression was by VMH, and bacterial expression by NT and AT. All authors contributed to the development of the manuscript and approved the final version.

Authors' information

Twitter handles: @johnson.jamesuk (James Johnson); @vickymharman (Victoria M Harman); @catfranco (Catarina Franco); @edemmott (Edward Emmott); @luningliu (Lu-Ning Liu); @takemoriayako (Ayako Takemori); @nobu_takemori (Nobuaki Takemori); @astacus (Robert J. Beynon).

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council (BB/S020241/1, RJB, and BB/R003890/1, BB/M024202/1, BB/V009729/1, L-NL) and the Royal Society (URF\R\180030, RGF\EA\181061, RGF\EA\180233, L-NL).

Availability of data and materials

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [32] partner repository with the dataset identifier PXD027521. Details of SRM transitions and quantitative data are accessible via the PanoramaWeb server (<https://panoramaweb.org/alacat.url>). The plasmids for the two full length (long ALACATs) have been deposited at Addgene.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The ALACAT construction strategy has been patented by the University of Liverpool (preliminary filing GB 21 03126.5).

Author details

¹GeneMill, Institute of Systems Molecular and Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK. ²Centre for Proteome Research, Institute of Systems and Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK. ³Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK. ⁴Division of Analytical Bio-Medicine, Advanced Research Support Center, Ehime University, Shitsukawa, Toon, Ehime, Japan.

Received: 3 May 2021 Accepted: 28 August 2021

Published online: 08 September 2021

References

- 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(12):6940–5. <https://doi.org/10.1073/pnas.0832254100>.
- Kirkpatrick DS, Gerber SA, Gygi SP. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods*. 2005;35(3):265–73. <https://doi.org/10.1016/j.jymeth.2004.08.018>.
- Dupuis A, Hennekinne JA, Garin J, Brun V. Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics*. 2008;8(22):4633–6. <https://doi.org/10.1002/pmic.200800326>.
- Brun V, Masselon C, Garin J, Dupuis A. Isotope dilution strategies for absolute quantitative proteomics. *J Proteomics*. 2009;72(5):740–9. <https://doi.org/10.1016/j.jprot.2009.03.007>.
- Picard G, Lebert D, Louwagie M, Adrait A, Huillet C, Vandenesch F, et al. PSAQ™ standards for accurate MS-based quantification of proteins: from the concept to biomedical applications. *J Mass Spectrom*. 2012;47(10):1353–63. <https://doi.org/10.1002/jms.3106>.
- Persson A, Hober S, Uhlén M. A human protein atlas based on antibody proteomics. *Curr Opin Mol Ther*. 2006;8:185–90.
- Zeiler M, Straube WL, Lundberg E, Uhlen M, Mann M. A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol Cell Proteomics*. 2012;11(3):O111.009613. <https://doi.org/10.1074/mcp.O111.009613>.
- Brownridge P, Holman SW, Gaskell SJ, Grant CM, Harman VM, Hubbard SJ, et al. Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. *Proteomics*. 2011;11(15):2957–70. <https://doi.org/10.1002/pmic.201100039>.
- Simpson DM, Beynon RJ. QconCATs: design and expression of concatenated protein standards for multiplexed protein quantification. *Anal Bioanal Chem*. 2012;404(4):977–89. <https://doi.org/10.1007/s00216-012-6230-1>.
- 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(8):587–9. <https://doi.org/10.1038/nmeth774>.
- 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(2):1029–43. <https://doi.org/10.1038/nprot.2006.129>.
- Rivers J, Simpson DM, Robertson DH, Gaskell SJ, Beynon RJ. Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol Cell Proteomics*. 2007;6(8):1416–27. <https://doi.org/10.1074/mcp.M600456-MCP200>.
- Lawless C, Holman SW, Brownridge P, Lanthaler K, Harman VM, Watkins R, et al. Direct and Absolute Quantification of over 1800 Yeast Proteins via Selected Reaction Monitoring. *Mol Cell Proteomics*. 2016;15:1309–22. <https://doi.org/10.1074/mcp.M115.054288>.
- Yang M, Simpson DM, Wenner N, Brownridge P, Harman VM, Hinton JCD, et al. Decoding the stoichiometric composition and organisation of bacterial metabolosomes. *Nat Commun*. 2020;11(1):1976. <https://doi.org/10.1038/s41467-020-15888-4>.
- Takemori N, Takemori A, Matsuoka K, Morishita R, Matsushita N, Aoshima M, 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):361–5. <https://doi.org/10.1039/c4mb00556b>.
- Takemori N, Takemori A, Tanaka Y, Endo Y, Hurst JL, Gómez-Baena G, et al. MEERCAT: Multiplexed Efficient Cell Free Expression of Recombinant QconCATs For Large Scale Absolute Proteome Quantification. *Mol Cell Proteomics*. 2017;16(12):2169–83. <https://doi.org/10.1074/mcp.RA117.000284>.
- Shetty RP, Endy D, Knight TF. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng*. 2008;2(1):5. <https://doi.org/10.1186/1754-1611-2-5>.
- Pollak B, Cerda A, Delmans M, Álamos S, Moyano T, West A, et al. Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol*. 2019;222(1):628–40. <https://doi.org/10.1111/nph.15625>.
- Eyers CE, Simpson DM, Wong SC, Beynon RJ, Gaskell SJ. QCAL—a novel standard for assessing instrument conditions for proteome analysis. *J Am Soc Mass Spectrom*. 2008;19(9):1275–80. <https://doi.org/10.1016/j.jasms.2008.05.019>.
- Shinoda K, Tomita M, Ishihama Y. Aligning LC peaks by converting gradient retention times to retention index of peptides in proteomic experiments. *Bioinformatics*. 2008;24(14):1590–5. <https://doi.org/10.1093/bioinformatics/btn240>.
- Moruz L, Tomazela D, Käll L. Training, selection, and robust calibration of retention time models for targeted proteomics. *J Proteome Res*. 2010;9(10):5209–16. <https://doi.org/10.1021/pr1005058>.
- Holman SW, Mclean L, Eyers CE. RePLiCal: A QconCAT Protein for Retention Time Standardization in Proteomics Studies. *J Proteome Res*. 2016;15(3):1090–102. <https://doi.org/10.1021/acs.jproteome.5b00988>.
- Zolg DP, Wilhelm M, Yu P, Knaute T, Zerweck J, Wenschuh H, et al. PROCAL: A Set of 40 Peptide Standards for Retention Time Indexing, Column Performance Monitoring, and Collision Energy Calibration. *Proteomics*. 2017;17. <https://doi.org/10.1002/pmic.201700263>.
- Frank AM. Predicting intensity ranks of peptide fragment ions. *J Proteome Res*. 2009;8(5):2226–40. <https://doi.org/10.1021/pr800677f>.
- Eyers CE, Lawless C, Wedge DC, Lau KW, Gaskell SJ, Hubbard SJ. CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. *Mol Cell Proteomics*. 2011;10(11):M110.003384. <https://doi.org/10.1074/mcp.M110.003384>.
- Sun S, Yang F, Yang Q, Zhang H, Wang Y, Bu D, et al. MS-Simulator: predicting y-ion intensities for peptides with two charges based on the intensity ratio of neighboring ions. *J Proteome Res*. 2012;11(9):4509–16. <https://doi.org/10.1021/pr300235v>.
- Jarnuczak AF, Lee DC, Lawless C, Holman SW, Eyers CE, Hubbard SJ. Analysis of Intrinsic Peptide Detectability via Integrated Label-Free and SRM-Based Absolute Quantitative Proteomics. *J Proteome Res*. 2016;15(9):2945–59. <https://doi.org/10.1021/acs.jproteome.6b00048>.
- Wichmann C, Meier F, Virreira Winter S, Brunner AD, Cox J, Mann M. MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides. *Mol Cell Proteomics*. 2019;18(5):982–94. <https://doi.org/10.1074/mcp.TIR118.001131>.
- Gessulat S, Schmidt T, Zolg DP, Samaras P, Schnatbaum K, Zerweck J, et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods*. 2019;16(6):509–18. <https://doi.org/10.1038/s41592-019-0426-7>.

30. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 2000;28(1):292. <https://doi.org/10.1093/nar/28.1.292>.
31. Sharma V, Eckels J, Taylor GK, Shulman NJ, Stergachis AB, Joyner SA, et al. Panorama: a targeted proteomics knowledge base. *J Proteome Res.* 2014; 13(9):4205–10. <https://doi.org/10.1021/pr5006636>.
32. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 2019;47:D442–50.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

