Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides

Robert J Beynon[1], Mary K Doherty[1], Julie M Pratt[1] & Simon J Gaskell[2]

Absolute quantification in proteomics usually involves simultaneous determination of representative proteolytic peptides and stable isotope–labeled analogs. The principal limitation to widespread implementation of this approach is the availability of standard signature peptides in accurately known amounts. We report the successful design and construction of an artificial gene encoding a concatenation of tryptic peptides (QCAT protein) from several chick (*Gallus gallus*) skeletal muscle proteins and features for quantification and purification.

For proteomics to support the emergent field of systems biology, there is a pressing need for facilitation of absolute quantification[1]. Most comparative proteomics studies deliver relative quantification, expressing the changes in amount of a protein in the context of a second cellular state. Absolute quantification draws on well-established precepts in analytical chemistry and relies on internal standardization, based on mass spectrometry (MS). Highly selective detection of ions (or ion fragmentations) characteristic of the analytes of interest is combined with the use of internal standards through the principle of surrogacy, to quantify indirectly by reference to a proteolytic peptide derived from the protein of interest. Analyses based on these principles, which have been

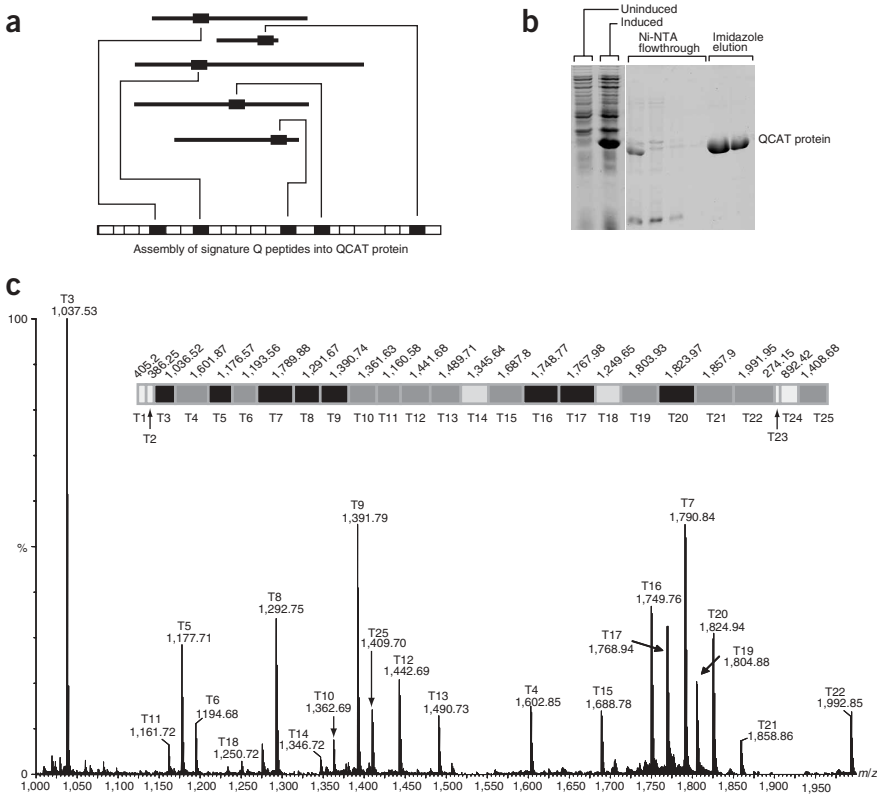

Figure 1 | Design and implementation of the QCAT strategy. (**a**) For each protein of interest, a unique Q peptide was selected. The sequences of each of these peptides were concatenated *in silico*, and used to design a gene that expresses this artificial protein (QCAT) in a heterologous expression system, permitting metabolic labeling with stable isotopes if so desired (see **Supplementary Methods**). (**b**) A QCAT was designed to monitor protein expression during development of chick skeletal muscle. The pET21a/QCAT plasmid was transformed into *E. coli* DE3 cells, and after a period of exponential growth, the expression of the QCAT was induced. The cell lysates from uninduced and induced cells were compared by SDS-PAGE. After solubilization of the pellet, and affinity chromatography on a Ni-NTA column, the purified QCAT protein was homogeneous. (**c**) After digestion of the QCAT by trypsin, the peptides were analyzed by MALDI-ToF MS. The tryptic digestion map (top) is shaded to indicate the relative intensities of signals corresponding to each peptide in the mass spectrum; peptides smaller than 900 Da, derived from the 'sacrificial' parts of the QCAT are less readily detected in this type of mass spectrometric analysis owing to interfering ions.

[1]Department of Veterinary Preclinical Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZJ, UK. [2]Michael Barber Centre for Mass Spectrometry, School of Chemistry, University of Manchester, Manchester M13 9PL, UK. Correspondence should be addressed to R.J.B. (r.beynon@liv.ac.uk).

PUBLISHED ONLINE 21 JULY 2005; DOI:10.1038/NMETH774

dubbed AQUA (absolute quantification)[2], use internal standards synthesized *de novo* by chemical methods. This approach, however, is more complex when applied to large numbers of proteins, as each standard peptide (Q peptide) would need to be chemically synthesized in stable isotope–labeled form and independently quantified[3].

We describe here the design, expression and use of artificial proteins that are concatemers of tryptic Q peptides for several proteins (**Fig. 1a**), generated by gene design *de novo*. The artificial protein, a concatemer of Q peptides (QCAT), was designed to include both amino-terminal and carboxy-terminal extensions to protect true Q peptides from exoproteolysis, to introduce a His tag for purification and a single cysteine residue for quantification (**Supplementary Methods** online). The gene encoding QCAT was inserted into a high-level expression vector and expressed in *Escherichia coli*. Within the QCAT, the Q peptides (each derived from a naturally occurring tryptic peptide in the parent protein) are in strict 1:1 stoichiometry; the entire set of concatenated Q peptides can be quantified in molar terms by determination of the QCAT protein. Moreover, labeled QCAT protein is readily produced upon expression in medium containing the selected label.

Some of our major interests are in proteome dynamics[4], and in changes in protein expression during muscle development[5,6]. A system that undergoes dramatic developmental changes in protein expression is the chicken pectoralis muscle. For the QCAT, we chose 20 chicken proteins that we had previously identified as changing in expression in developing skeletal muscle[5]. A single tryptic peptide was chosen to represent each protein (a Q peptide) according to several criteria. First, the Q peptide should not have cysteine residues, as this would prevent formation of complex intra- and intermolecular disulphide bonds in the expressed protein, and introduction of a single cysteine residue in the entire QCAT could be used for quantification. Second, the peptide chosen should be unique within the set of Q peptides. Third, the Q peptides should have masses between 1,000 Da and 2,000 Da, corresponding to the region of matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) mass spectra in which sensitivity of detection is typically high and interfering signals are low. Finally, we added an operational criterion, in as much as we selected peptides that were already demonstrated to give a strong signal on MALDI-ToF MS and 75% (15 out of 20) of which were arginine-terminated tryptic peptides—the propensity of such peptides to give stronger signals on MALDI-ToF MS is well documented[7].

The Q peptides were assembled *in silico*, and a gene was constructed, which encoded the assembled Q peptides using codons that would yield maximal expression in *E. coli*. At the C terminus, an extension was added containing a cysteine residue and a His-tag motif for purification (the latter encoded in the vector pET21a). Additional amino acids were appended to the N-terminus to provide an initiator methionine residue and a sacrificial peptide, which when cleaved would expose a true Q

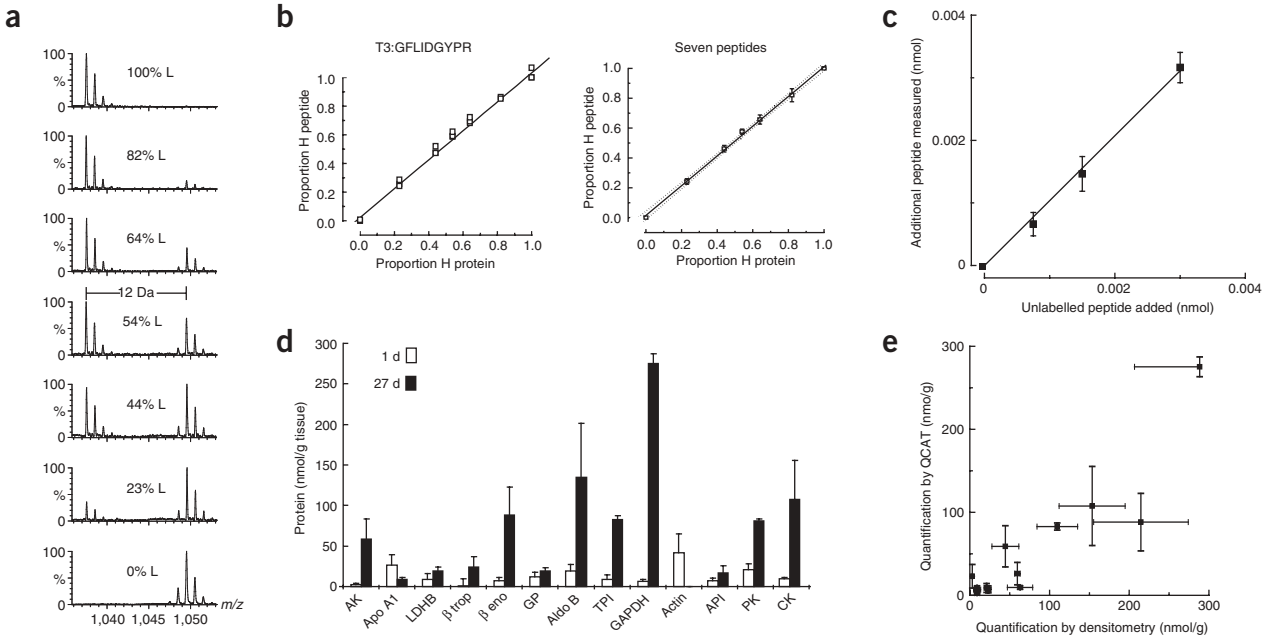

Figure 2 | Use of the QCAT in quantification. (**a**–**e**) Unlabeled (L, 'light') and uniformly labeled with ^{15}N (H, 'heavy') QCAT proteins were separately purified, quantified and mixed in different ratios, before tryptic digestion and measurement of peptide intensities by MALDI-ToF MS. (**a**) Mass spectra for peptide T3 (GFLIDGYPR, 12 nitrogen atoms); percent L QCAT protein in the mixture is indicated in each graph. (**b**) The measured proportion of H peptide plotted relative to the proportion of H protein in the mixture for one peptide, T3 (left); data points for three sets of measurements are plotted. These data were collated for three replicates of each of seven peptides (right); error bars, ±s.e.m., $n = 7$. The dotted lines define the 95% confidence limits of the fitted line. (**c**) L QCAT was added to a preparation of soluble skeletal muscle proteins in different amounts, prior to quantification with the H QCAT. The graph is the collated data for seven peptides, from three replicates of the admixture experiment; error bars, ±s.e.m. (**d**) A preparation of soluble proteins from skeletal muscle of chicks at 1 d and 27 d was mixed with ^{15}N-labeled QCAT, digested with trypsin and analyzed by MALDI-ToF MS. For a subset of proteins, the intensities of the endogenous and Q peptide were used to calculate the absolute amounts of each protein. Error bars, ±s.e.m.; $n = 3$. Protein name abbreviations are as in **Supplementary Figure 1**. (**e**) A correlation plot of quantification by densitometry after two-dimensional gel electrophoresis relative to QCAT quantification for those proteins for which both sets of data were available. Error bars in both directions are ±s.e.m.; $n = 3$.

peptide (**Supplementary Figure 1** online). This avoided complications owing to N-formylation or removal of methionine from the N terminus of the QCAT. After induction of expression, SDS-PAGE analysis confirmed high-level expression of a protein of the expected mass (~35 kDa). This protein was present in the insoluble fraction of sonicated cells, presumed to be inclusion bodies. The QCAT protein was purified by affinity chromatography using Ni-NTA resin, which resulted in a homogeneous preparation (**Fig. 1b**). The ~35 kDa gel band was subjected to in-gel digestion with trypsin and analyzed by MALDI-ToF MS. All predicted QCAT peptides were readily observed in the MALDI-ToF mass spectrum (**Fig. 1c**). The N- and C-terminal sacrificial peptidic material yielded, by design, fragments that were too small to be seen in the MALDI-ToF mass spectrum. The QCAT was digested by trypsin very effectively and there was no evidence of partial proteolytic products of the Q peptides. The protein was also expressed in minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source. When digested with trypsin, the resultant MALDI-ToF mass spectrum was of high quality, and all Q peptides were detectable at the appropriate mass shift corresponding to the number of nitrogen atoms in the peptide. Unlabeled and ^{15}N-labeled QCAT proteins were mixed in different ratios, and digested with trypsin before MALDI-ToF MS of the limit peptides. The heavy and light variants of the peptides were readily discerned, and the intensities of the corresponding signals were measured for several peptides (**Fig. 2a**). The relationship between proportion of material, and the measured heavy:light ratio was linear, with a slope of 1 (mean $\pm$ s.d. = 1.008 $\pm$ 0.008, $r^2 > 0.99$ in all instances, $n = 7$; **Fig. 2b**). We also used unlabeled QCAT to 'spike' a preparation of soluble proteins from chicken skeletal muscle to provide additional material in the context of a pre-existing unlabeled signal (slope = 1.04, $r^2 = 0.995$, $n = 7$; **Fig. 2c**).

We have applied this QCAT to protein expression analysis in chick skeletal muscle (**Fig. 2d**). Twelve proteins present in the preparation were also represented in the QCAT. MALDI-ToF data for the tryptic peptides were readily acquired, and the changes in protein levels that occur over the first 3 to 4 weeks post-hatching were determined. Because the proteins were quantified in absolute terms, we were able to express the proteins as nmol/g wet weight of tissue. We have previously measured the levels of seven of these proteins by two-dimensional gel electrophoresis and densitometry and the correlation (r^2) between the quantification using both methods was 0.82 ($P < 0.001$). Recognizing that the two methods assess different representations of the proteome, such as charge-variant isoforms or total protein complement, the correlation is good. Indeed, the overall variance was similar with the QCAT method compared with the densitometric analysis (**Fig. 2e**).

We believe that the QCAT strategy is superior to chemical synthesis of individual Q peptides in a stable isotope form. For multiplexed assays, each peptide would need to be individually quantified before use. Finally, chemically synthesized Q peptides are a finite resource whereas repeated expression of the QCAT gene is facile. To address dynamic range, QCAT genes should encode Q peptides for proteins of similar expression levels. High-quality absolute quantification may be an effective route to overcome the difficulties associated with present methods for comparative proteomics, whether based on gels or mass spectrometry. Studies of particular cellular systems, each by comparison to a QCAT-quantified reference would not only represent quantification at the individual protein level, but should be sufficiently rigorous that as the data sets grow, any pairwise comparison would be robust, transferable between individual laboratories and stable over time. The number of Q peptides that could be assembled into a single QCAT is probably limited by the ability to achieve high-level heterologous expression of large proteins. In the example given here, we chose 20 peptides with an average length of 15 amino acids and an average molecular weight of 1.5 kDa. If 100 proteins were represented in a single QCAT, the resultant recombinant protein would be 150 kDa, which should be readily expressed. The ability to quantify up to 100 proteins in a single construct invites the challenge of optimal assembly of QCATs, and new optimization algorithms will be required. Other applications for QCATs are readily envisioned, such as absolute quantification in clinical or other biomarker monitoring systems and in the assessment of stoichiometric ratios of individual proteins within a subcellular compartment or a multiprotein complex.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by Biotechnology and Biological Sciences Research Council grants to R.J.B. and S.J.G. We are grateful to M. Fischer, Entelechon for his advice. We are grateful to B. Callen for assistance with expression of the QCAT.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Received 8 April; accepted 17 June 2005
Published online at http://www.nature.com/naturemethods/

1. Righetti, P.G., Campostrini, N., Pascali, J., Hamdan, M. & Astner, H. *Eur. J. Mass Spectrom. (Chichester, Eng.)* **10**, 335–348 (2004).
2. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. *Proc. Natl. Acad. Sci. USA* **100**, 6940–6945 (2003).
3. Pan, S. *et al. Mol. Cell. Proteomics* **4**, 182–190 (2005).
4. Pratt, J.M. *et al. Mol. Cell. Proteomics* **1**, 579–591 (2002).
5. Doherty, M.K. *et al. Proteomics* **4**, 2082–2093 (2004).
6. Doherty, M.K., Whitehead, C., McCormack, H., Gaskell, S.J. & Beynon, R.J. *Proteomics* **5**, 522–533 (2005).
7. Brancia, F.L. *et al. Electrophoresis* **22**, 552–559 (2001).