Biochemical Society TRANSACTIONS
Biochemical Society Transactions www.biochemsoctrans.org

Behaviour Meets Biochemistry: Animals Making Sense of Molecules Making Scents

A joint Biochemical Society–Association for the Study of Animal Behaviour Focused Meeting held at Charles Darwin House, London, U.K., 18–20 February 2014. Organized and Edited by Rob Beynon and Jane Hurst (University of Liverpool, U.K.).

The complexity of protein semiochemistry in mammals

Robert J. Beynon*[1], Stuart D. Armstrong*, Guadalupe Gómez-Baena*, Victoria Lee*, Deborah Simpson*, Jennifer Unsworth* and Jane L. Hurst†

*Protein Function Group, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, U.K.

†Mammalian Behaviour and Evolution Group, Institute of Integrative Biology, University of Liverpool, Leahurst Campus, Neston, Cheshire CH64 7TE, U.K.

Abstract

The high degree of protein sequence similarity in the MUPs (major urinary proteins) poses considerable challenges for their individual differentiation, analysis and quantification. In the present review, we discuss MS approaches for MUP quantification, at either the protein or the peptide level. In particular, we describe an approach to multiplexed quantification based on the design and synthesis of novel proteins (QconCATs) that are concatamers of quantification standards, providing a simple route to the generation of a set of stable-isotope-labelled peptide standards. The MUPs pose a particular challenge to QconCAT design, because of their sequence similarity and the limited number of peptides that can be used to construct the standards. Such difficulties can be overcome by careful attention to the analytical workflow.

Introduction

The polymorphic and combinatorial diversity of the multiple protein families that are involved in scent communication in the mouse extends to the MHC, olfactory receptors, ABPs (androgen-binding proteins, sex-hormone-binding globulins, secretoglobins) [1,2,2a], ESPs (exocrine secretory peptides) [3–5,5a], OBPs (odorant-binding proteins) and MUPs (major urinary proteins) [6]. It is increasingly clear that the complexity of the semiochemical space is able to convey incredibly subtle information through multiple routes [7–9]. Although the power of genomic and transcriptomic analysis can be brought to bear on the complexity of these protein families, ultimately, expression has to be assessed at the protein level, and the considerable sequence similarity within families brings a substantial analytical challenge to the quantification of individual proteins in a complex biological matrix. In the present article, we discuss strategies that can be used to address the issue of multiplexed quantification of complex and highly similar groups of proteins.

The complexity of the MUP family

The gene cluster encoding MUPs in the mouse is localized to chromosome 4. Although complete sequence information for this region of the mouse genome is not yet available (probably because of the challenge of highly repetitive sequences), there are at least 21 protein-coding genes in the genome of the reference inbred mouse strain C57BL/6J [10,11]. These are arranged in the genome into a group of 15 central MUPs flanked by a total of six peripheral MUPs. Of these 21 protein-coding genes, all of the mature protein products are 162 amino acids in length, and the expression at the protein level of many of these has been confirmed by MS [12–15]. When other inbred mouse strains are included, there

Key words: major urinary protein (MUP), mass spectrometry, stable isotope, QconCAT.

Abbreviations: MUP; major urinary protein; PSAQ; protein standards for absolute quantification; QconCAT; quantification concatamer.

[1]To whom correspondence should be addressed (email r.beynon@liv.ac.uk).

is evidence for the absence or presence of new MUPs [13]. For example, some lineages of inbred mice do not express significant amounts of darcin (MGI_MUP20) in urine. When wild-caught mice are analysed, the situation becomes even more complex, and there is accumulating evidence for a very high degree of polymorphism in the MUP profile, as a consequence of allelic variation and expression levels. The extent of this variation is sufficient to create MUP patterns unique to individuals, and is an important factor in individual recognition [9,16].

The MUPs have been studied for many decades and, as genome sequencing has advanced and the resolution of protein analytical technologies has increased, different numbering systems have been proffered [10,11]; this has led to a degree of confusion. We have compiled a list (Table 1) of the protein-coding *Mup* genes using the most comprehensive annotation found at the Mouse Genome Informatics (MGI) resource (http://www.informatics.jax.org/). We recommend that all other numbering or naming conventions be abandoned in favour of this scheme, or at the very least, make accurate cross-references to this nomenclature; others have made the same suggestion [17]. It should, however, be noted that the MGI resource compiles gene annotation from VEGA [18,19] that emphasizes the longest transcript that can be identified. In a few instances, the VEGA annotation specifies a nascent protein that is longer than the 'expected' length of 180 or 181 amino acids. This might be attributable to a frameshift or alternative splice event, but the proteins predicted by these variants have not been observed in protein-level analyses of urine samples. For example, the mature proteins from MGI_MUP1 and MGI_MUP9 are predicted as having a mature mass of 24892 and 24893 Da respectively, consistent with an amino acid change at a single residue (lysine/glutamate), the same amino acid change that has been observed at position 140 in the 'usual' 162-amino-acid proteins in C57BL/6 urine. Expressed protein data suggest that these proteins are the typical MUPs, and that the extended sequences are sequencing artefacts or minor transcripts, for which no evidence yet exists at the protein level. Additionally, the presence of the 'expected' C-terminal sequence in a different reading frame is consistent with a putative sequencing error. Lastly, for all of these longer translation products, there are other transcripts in the database that do encode proteins of the same length as most MUPs. Using what we believe are the true transcripts, we derive the predicted 162-amino-acid mature protein sequences for a set of 21 MUPs, noting that groups of five (MGI_MUP9, MGI_MUP11, MGI_MUP16, MGI_MUP18 and MGI_MUP19) and two (MGI_MUP1 and MGI_MUP12) each encode identical mature MUP sequences. There are thus 16 different mature MUP sequences encoded by 21 genes. It is possible to compare these unique sequences in different ways. For example, the amino acid composition analysed by principal component analysis clearly reveals that (i) central MUPs are very homogeneous as a group, and (ii) they differ from peripheral MUPs. Moreover, peripheral MUPs in turn are far more diverse than the central MUPs (Figure 1). Alignment of these sequences [19a] confirms the extent of sequence similarity.

Characterization and quantification of MUPs

Whereas the polymorphisms and allelic variation in the MUP profile has the potential to generate considerable combinatorial diversity that can drive natural behavioural outcomes [16,20–23], it is less clear that this diversity is manifest as 'present' or 'absent' or whether more subtle variation in MUP concentration can modulate biological function. To be able to dissect such variability, there is a need for accurate quantitative profiling of MUPs in complex biological samples. Furthermore, because the role of these proteins is in scent marks or in extracellular fluids (such as nasal secretions, tears), it is not possible to use transcript profiling as a surrogate for the quantitative protein profile. Quantitative strategies must directly address the proteins in an analytical sample. Even more challengingly, the amino acid differences between individual MUP sequences are so low as to generate very few unique protein signatures that would allow individual quantification. For example, it would be extremely difficult to generate a family of antibodies that were capable of discriminating between the individual MUPs, which differ in single amino acid changes distributed throughout the sequence. At present, there are no antibodies that can discriminate between the individual MUPs, particularly the central MUPs, and, given the sequence similarity of the proteins, it is unlikely that effort will be directed to derivation of such specific antibodies.

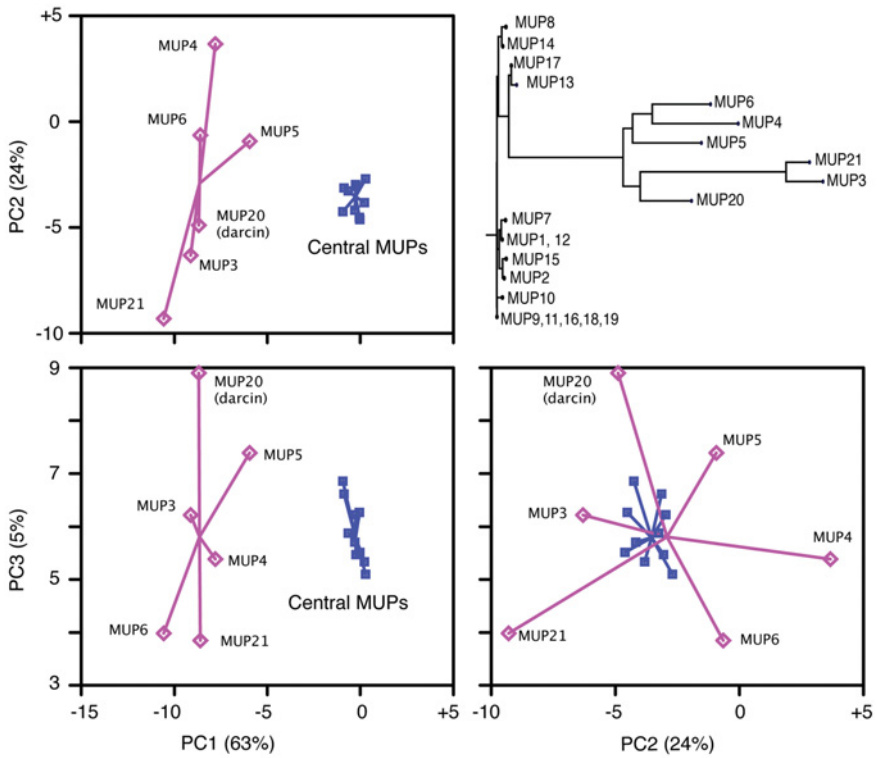
There are two achievable strategies that can be deployed for MUP quantification, both of which are MS-based. The two approaches differ in that one is based on MS of the intact protein, whereas the other is directed to peptides derived by proteolysis of the MUPs.

Protein-based quantitative strategy

The protein-based strategy uses ESI of the mixture of MUP isoforms to generate a series of multiply charged species, reflecting the binding of variable numbers of protons at different sites on the protein. The MUPs possess between 18 and 28 protonatable sites [15,24]. Typically, the charge distribution profile is centred on the $[M+12\mathrm{H}]^{12+}$ ion, although we often observe a second series that are more highly charged. Because each ion is a product of mass and charge $[M+n\mathrm{H}]^{n+}$, it is possible to use two ions in a series to calculate the protein mass, but it is preferable to use all of the profile, especially from a mixture of proteins, to generate an optimal solution in terms of a true protein mass spectrum (Figure 2). For this, the profile is deconvoluted using maximum entropy methodology (MaxEnt [15,25–27]) to generate a true mass spectrum and the MaxEnt data generated is centred to calculate the average neutral mass. MaxEnt profiles are semi-quantitative. The area under the peaks in the MaxEnt profile spectrum is a result of the summed intensities of all ions in the multiple-charge

Figure 1 | **Relationship between MUPs**

For the 21 known MUPs in the C57BL/6 genome, amino acid compositions were calculated and used to direct a principal components analysis. A clear separation between central and peripheral MUPs is evident, as is the more pronounced diversity within the peripheral MUPs compared with the central MUPs. The first component (63 % of the variance) resolves central and peripheral MUPs, whereas the second component (24 % of the variance) largely resolves individual peripheral MUPs. This compositional diversity is a consequence of sequence variation, also shown as a phylogenetic tree (top right).



state envelop spectrum. Moreover, the area under the peak is theoretically proportional to the relative concentration of the protein in the sample.

For intact mass profiling to be valuable for quantitative phenotyping, two constraints are critical. First, each protein should be individually discriminable in the mass dimension. Secondly, the deconvoluted peak area for each protein should be proportional to the amount. However, not all mass spectrometers are able to resolve proteins with, for example, masses of 18694 Da and 18693 Da (at $[M+12\mathrm{H}]^{12+}$, this would require resolution of adjacent ions separated by 0.08 Th at 1550 Th (approximately 50 p.p.m., or a resolution of 20000). Similarly, MUPs at 18708 and 18713 Da might be incompletely resolved. Higher resolution does not automatically yield a simple solution, however, as each charge state would also be resolved into the ^{13}C isotopomers, adding finely grained detail that can also compromise accurate mass measurement and quantification. Further complicating factors could arise from the different number of protonatable sites for each protein and the possibility of interaction during the electrospray process that might alter the ionization of individual proteins depending on the other proteins that are present in the sample.

Peptide-based quantitative strategy

One of the main reasons that many 'bottom-up' proteomics workflows are based on the analysis of proteolytic peptides (usually trypsin or a similar endopeptidase) is that the resultant peptides are limited in the number of protonatable sites (many tryptic peptides have two sites: the α-amino group and the C-terminal lysine or arginine residue). Moreover, the frequency of tryptic digestion is around an average of 10–12 residues, typically generating fragments of masses between 800 and 3000 Da. Peptides, predominantly $[M+2\mathrm{H}]^{2+}$ ions, have m/z values in a range that is optimal for the common mass analysers used in modern mass spectrometers. Thus a protein, or mixture of proteins, is digested to limit peptides (those in which there are no further cleavage sites) and quantification is directed to those peptides that are uniquely derived from each protein. Each protein generates a set of peptides, many of which could be unique to the parent protein.

Because MS is not intrinsically quantitative (it is not possible to relate signal intensity to the concentration of analyte), absolute quantification is optimally achieved by addition of an accurately known amount of a stable-isotope-labelled peptide as a standard. The inclusion of 'heavy' atom

Table 1 | **Summary of MUPs nomenclature, sequences and structures**

To provide a cross-reference between the different MUP numbering schemes and solved structures, the Table compiles data from multiple sources, using the MGI resource as a common and definitive labelling scheme. Two other analyses from Mudge et al. [11] and Logan et al. [10] are included to facilitate cross-reference.

Gene order	MGI gene	MGI ref	Predicted protein	Mature protein	Mudge	Logan	Mass mature (-SS-)	PDB structures	Notes
1	Mup4	MGI:97236	178	162	1	Mm1	18816.4	3KFF, 3KFG, 3KFH, 3KFI	
2	Mup6	MGI:3650962	179	162	2	Mm2	18984.5		
3	Mup7	MGI:3709615	180	162	3	Mm3	18644.8		
4	Mup2	MGI:97234	180	162	4	Mm4	18693.8		Differs by K140E to 1,12
5	Mup8	MGI:3709619	180	162	5	Mm5	18664.8		
6	Mup9	MGI:3782918	180/235*	162*	6	Mm6	18692.8	2LB6, 1MUP, 1I04, 1I05, 1I06,	Identical with 11,16,18,19. *Transcript Mup9-001 (OTTMUST00000037038) and Mup9-005 (OTTMUST00000070671) codes for observed urinary protein; 235-amino-acid protein not seen.
7	Mup1	MGI:97233	180/235*	162*	7	Mm7	18692.8		Identical with 12. *Transcript Mup1-001 (OTTMUS00000017164 codes for observed urinary protein; 235-amino-acid protein not seen.
8	Mup10	MGI:1924164	180	162	8	Mm8	18707.8	1DF3, 1JV4, 1QY0, 1QY1, 1QY2, 1ZND, 1ZNE, 1ZNG, 1ZNH, 1ZNK, 1ZNL, 2DM5, 2OZQ, 1YP6, 1YP7	
9	Mup11	MGI:3709617	181	162	9	Mm9	18693.8		Identical with 9, 16, 18, 19
10	Mup12	MGI:3780193	180	162	–	Mm10	18692.8		Identical with 1
11	Mup13	MGI:3702003	180	162	10	Mm11	18681.8		
12	Mup14	MGI:3702005	180	162	11	Mm12	18712.8		
13	Mup15	MGI:3780235	180	162	–	Mm13	18693.8		
14	Mup16	MGI:3780250	180	162	12	Mm14	18693.8		Identical with 9, 11, 18, 19
15	Mup17	MGI:3705217	180	162	13	Mm15	18682.8		
16	Mup18	MGI:3705220	181	162	14	Mm16	18693.8		Identical with 9, 11, 16, 19
17	Mup19	MGI:3705235	180/187*	162*	15	Mm17	18693.8		Identical with 9, 11, 16, 18. *Transcript Mup19-001 OTTMUST00000017265 codes for an observed protein; 187-amino-acid protein not seen
18	Mup5	MGI:104974	180	162	16	Mm18	18863.1		
19	Mup20/darcin	MGI:3651981	181	162	17	Mm24	18893.2	2L9C	
20	Mup3	MGI:97235	182	162	18	Mm25	18956.2		
21	Mup21	MGI:3650630	181	162	19	Mm26	19109.4		

centres such as ^{13}C or ^{15}N results in peptides that exhibit identical chemical, chromatographic and mass spectrometric properties that are readily resolved from the analyte by virtue of the mass shift caused by the stable isotope labelling. Stable-isotope-labelled peptides are also known as AQUA peptides [28–30].

For quantification of a single protein, one standard 'heavy' peptide that is synthesized chemically could be adequate. However, when multiple proteins must be quantified, or when the added assurance of multiple peptides to quantify one protein is required, the cost of these synthetic standards can become prohibitive. Moreover, each standard peptide must be independently quantified before use. To overcome this problem, we developed a methodology that creates artificial proteins that are concatamers [31–37] of peptides (tryptic, or from the action of other endopeptidases), prepared by gene synthesis *de novo* and cloned into standard expression vectors. When these QconCATs (quantification concatamers) are expressed in *Escherichia coli* and stable-isotope-labelled amino acids are included in the medium, it is possible to generate a 'heavy' form of the QconCAT. When the QconCAT is subsequently digested, the original peptides are recreated, but as 'heavy' standards. Assuming digestion is complete, after proteolysis, each standard peptide will be generated in stoichiometrically equal quantities. This approach to quantification standards has been applied in multiple proteomics studies [38–43] and is an efficient route to the generation of large numbers of multiplexed standards. A subtle variant of the QconCAT approach is provided by the PCS (peptide-concatenated standard) strategy, in which peptides are interspersed with hexapeptide sequences that recapitulate the native primary sequence context, theoretically balancing the rate of proteolysis of analyte and standard [44]. An even more comprehensive approach is provided by PSAQ (protein standards for absolute quantification) in which an entire protein is expressed and labelled in recombinant form to generate multiple tryptic peptides for quantification [28,45–47]. However, expression of MUPs as PSAQ standards would also generate a large number of common or shared peptides.

Not all peptides are suitable for MS-based quantification. For example, peptides containing a methionine residue can undergo partial oxidation, eliciting a 16 Da mass shift, and peptides containing dipeptide sequences such as Asn-Gly undergo side-chain deamidation at the asparagine residue. Some peptides do not readily ionize or yield good signals at the mass detector. The selection of the unique quantotypic peptides can be challenging, and it is quite feasible to reduce the number of usable peptides to zero by application of stringent criteria. Two tools that have been developed to ease peptide selection are PeptidePicker [48] and CONSeQuence [49]. Additional resources such as PeptideAtlas [50] and Passel [51] can also help, but it should be acknowledged that there is a difference between 'proteotypic' peptides (those frequently observed and representative of a specific protein) and 'quantotypic', those proteotypic peptides that can be reliably used for quantification [43,52]. Moreover,

Figure 2 | Quantification of MUPs at the protein level by MS

A mixture of MUPs generates a complex ESI mass spectrum, in which multiply charged forms of each protein create series of ions (**a**, inset is an enlarged image of one such charge state). The mass spectrum can be deconvoluted to generate a true mass spectrum, in which the intact mass of each protein is obtained (**b**). For a mixture of MUPs (in this instance, urine from a male C57BL/6 mouse, not the same sample as in **a** and **b**), the signal intensity is linearly proportional to the protein load (**c**), although absolute quantification would require knowledge of the responses of individual proteins that could lead to differences in the intensity of the signal.

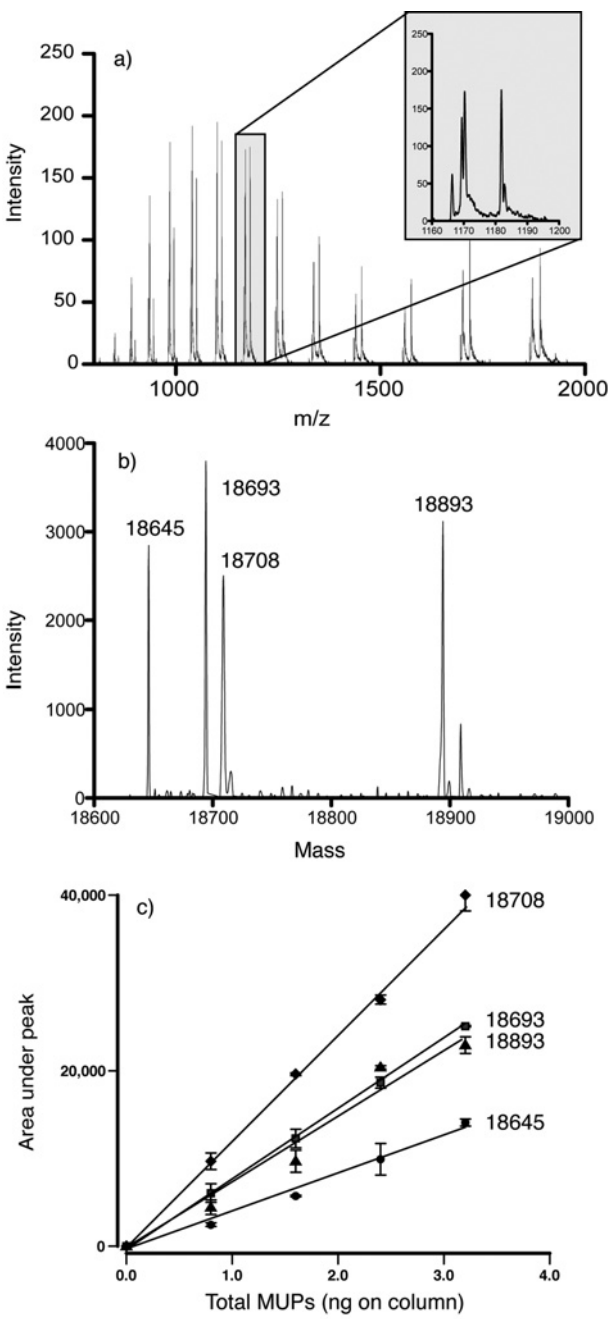
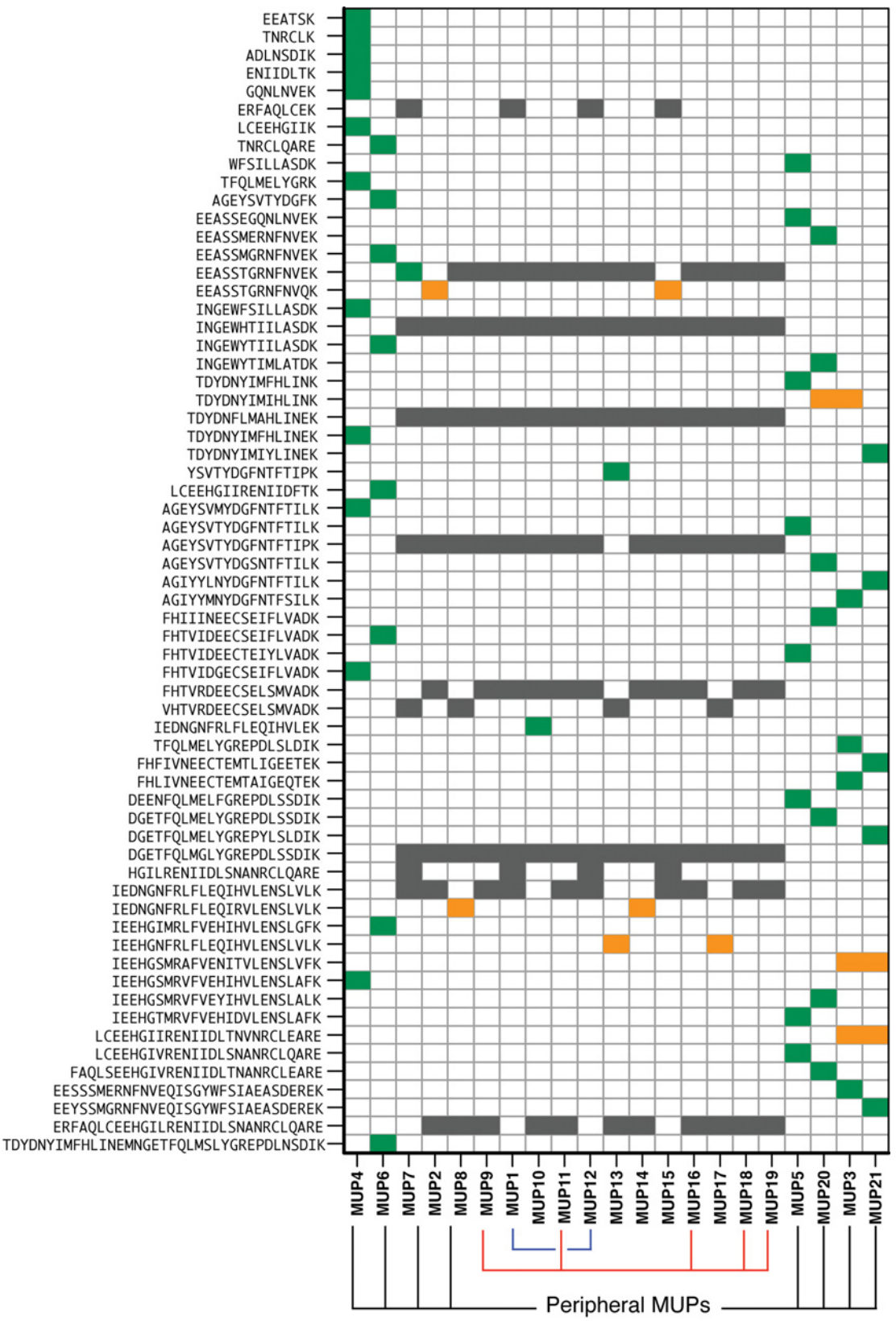


Figure 3 | **Selection of peptides for the quantification of MUPs**

In this example, the entire set of useable peptides is listed on the left-hand edge of the matrix, and a grey-shaded block indicates their presence or absence in individual MUPs. Those peptides that are unique to a single MUP are green, and peptides common to two MUPs are coloured orange. The preponderance of unique peptides in the peripheral MUPs is evident, as is the lack of the same in the central MUPs, as expected from the sequence relationships in Figure 1.



identification of quantotypic peptides for a group of dissimilar proteins is relatively straightforward, but is not as straightforward [53] when directed to the quantification of a highly homologous family, such as cytochrome P450 proteins [38] or amyloid precursors [42], or in our example, the MUPs.

The sequence identity of the MUPs is high [19a] and there are many shared tryptic peptides, such that it is impossible to find a 'one quantotypic peptide, one MUP' solution. One solution is to use an endopeptidase other than trypsin, with a lower frequency of cleavage. Using endopeptidase LysC (cleaving only at Lys–Xaa bonds), there is some movement towards a full solution, but it is still necessary to adopt an arithmetic analysis, in which the total quantification values for some peptides represent more than one MUP and a second MUP-specific peptide is then used to partition the shared signal into unique components. As a hypothetical example, Qpeptide 1 quantifies MUPA and MUPB, Qpeptide 2 quantifies MUPA only, and MUPB is quantified as the difference between Qpeptide1 and Qpeptide2. For MUP quantification, we designed a QconCAT using endopeptidase LysC as the digesting enzyme. The entire set of useable peptides, covering the entire range of MUPs is limited (63 peptides covering 21 proteins). Of these, 47 peptides are unique to one protein, but these are predominantly derived from the peripheral MUPs that show greater sequence diversity. For the central MUPs, the sequence diversity is so low that unique peptide solutions are not available (Figure 3).

It is nonetheless feasible to design a QconCAT strategy that can provide quantification data for all of the MUPs in a specific analytical space, provided that the primary sequence is known for each, or that unique peptides exist for each variant protein. In some instances, the choices of peptides are not optimal, for example being generated by digestion of basic residues with acidic residues in close proximity (14 peptides terminate in the sequence Glu-Lys, and the same number begin with an acidic residue at the N-terminus). This sequence context can reduce the rate of proteolysis, and it is critical to ensure that both standard and analyte are fully digested, as the peptides will not necessarily share the same context. There are eight endopeptidase LysC peptides that include an Asn-Gly dipeptide, prone to non-enzymic deamidation, leading to a 1 Da mass increase that can compromise quantification unless accounted for [52,54,55]. The severely limited choice of peptides thus imposes additional demands on experimental delivery. Notwithstanding such demands, a combination of intact protein analysis and peptide-mediated quantification can be combined to yield reliable quantification of the MUP isoforms in biological samples.

A final complication arises from the high rate of evolution of these proteins. Analysis of MUPs from wild caught individuals reveals both the diversity in individual MUP patterns and the existence of new allelic variants. It is possible that new isoforms or sequence variants will contain peptides that are not represented in the QconCAT standard. Under such conditions, either intact protein analysis or the redesign of additional QconCATs would be required. This would be dependent on a rapid survey strategy in which peptides derived from new MUP variants and which possess previously unseen masses were sequenced *de novo* by MS to discover new quantification candidates. Through the use of such tools, it is now possible to devise accurate quantification of proteins in biological matrixes, such as urine samples or scent marks, and explore the variation in expression of specific proteins under different behavioural conditions. Indeed, for many studies, it is possible that a strategy based on relative quantification, rather than absolute quantification, would suffice, although the requirement of unique signature peptides is not removed in such studies.

Acknowledgements

We are grateful to Dr P. Brownridge for MS support.

Funding

We gratefully acknowledge the Biotechnology and Biological Sciences Research Council [grant number BB/J002631/1] for supporting this research.

References

1 Laukaitis, C.M., Dlouhy, S.R. and Karn, R.C. (2003) The mouse salivary androgen-binding protein (*ABP*) gene cluster on chromosomes 7: characterization and evolutionary relationships. Mamm. Genome **14**, 679–691 CrossRef

2 Karn, R.C. and Laukaitis, C.M. (2003) Characterization of two forms of mouse salivary androgen-binding protein (ABP): implications for evolutionary relationships and ligand-binding function. Biochemistry **42**, 7162–7170 CrossRef PubMed

2a Karn, R.C. and Laukaitis, C.M. (2014) Selection shaped the evolution of mouse androgen-binding protein (ABP) function and promoted the duplication of *Abp* genes. Biochem. Soc. Trans. **42**, 851–860

3 Taniguchi, M., Yoshinaga, S., Haga-Yamanaka, S., Touhara, K. and Terasawa, H. (2014) Backbone and side-chain ^{1}H, ^{15}N and ^{13}C assignments of mouse peptide ESP4. Biomol. NMR Assign. **8**, 7–9 CrossRef PubMed

4 Yoshinaga, S., Sato, T., Hirakane, M., Esaki, K., Hamaguchi, T., Haga-Yamanaka, S., Tsunoda, M., Kimoto, H., Shimada, I., Touhara, K. and Terasawa, H. (2013) Structure of the mouse sex peptide pheromone ESP1 reveals a molecular basis for specific binding to the class C G-protein-coupled vomeronasal receptor. J. Biol. Chem. **288**, 16064–16072 CrossRef PubMed

5 Kimoto, H. and Touhara, K. (2013) Murine nonvolatile pheromones: isolation of exocrine-gland secreting peptide 1. Methods Mol. Biol. **1068**, 47–53 CrossRef PubMed

5a Abe, T. and Touhara, K. (2014) Structure and function of a peptide pheromone family that stimulate the vomeronasal sensory system in mice. Biochem. Soc. Trans. **42**, 873–877

6 Hurst, J.L. and Beynon, R.J. (2004) Scent wars: the chemobiology of competitive signalling in mice. BioEssays **26**, 1288–1298 CrossRef PubMed

7 Hurst, J.L. (2009) Female recognition and assessment of males through scent. Behav. Brain Res. **200**, 295–303 CrossRef PubMed

8 Roberts, S.A., Simpson, D.M., Armstrong, S.D., Davidson, A.J., Robertson, D.H., McLean, L., Beynon, R.J. and Hurst, J.L. (2010) Darcin: a male pheromone that stimulates female memory and sexual attraction to an individual male's odour. BMC Biol. **8**, 75 CrossRef PubMed

9 Cheetham, S.A., Thom, M.D., Jury, F., Ollier, W.E., Beynon, R.J. and Hurst, J.L. (2007) The genetic basis of individual-recognition signals in the mouse. Curr. Biol. **17**, 1771–1777 CrossRef PubMed

10 Logan, D.W., Marton, T.F. and Stowers, L. (2008) Species specificity in major urinary proteins by parallel evolution. PLoS ONE **3**, e3280 CrossRef PubMed

11 Mudge, J.M., Armstrong, S.D., McLaren, K., Beynon, R.J., Hurst, J.L., Nicholson, C., Robertson, D.H., Wilming, L.G. and Harrow, J.L. (2008) Dynamic instability of the major urinary protein gene family revealed by genomic and phenotypic comparisons between C57 and 129 strain mice. Genome Biol. **9**, R91 CrossRef PubMed

12 Armstrong, S.D., Robertson, D.H., Cheetham, S.A., Hurst, J.L. and Beynon, R.J. (2005) Structural and functional differences in isoforms of mouse major urinary proteins: a male-specific protein that preferentially binds a male pheromone. Biochem. J. **391**, 343–350 CrossRef PubMed

13 Cheetham, S.A., Smith, A.L., Armstrong, S.D., Beynon, R.J. and Hurst, J.L. (2009) Limited variation in the major urinary proteins of laboratory mice. Physiol. Behav. **96**, 253–261 CrossRef PubMed

14 Robertson, D.H., Hurst, J.L., Searle, J.B., Gunduz, I. and Beynon, R.J. (2007) Characterization and comparison of major urinary proteins from the house mouse, *Mus musculus domesticus*, and the aboriginal mouse, *Mus macedonicus*. J. Chem. Ecol. **33**, 613–630 CrossRef PubMed

15 Robertson, D.H., Wong, S.C., Beynon, R.J., Hurst, J.L. and Gaskell, S.J. (2008) Observation of heterogeneous gene products by FT-ICR MS. J. Am. Soc. Mass Spectrom. **19**, 103–110 CrossRef PubMed

16 Hurst, J.L., Payne, C.E., Nevison, C.M., Marie, A.D., Humphries, R.E., Robertson, D.H., Cavaggioni, A. and Beynon, R.J. (2001) Individual recognition in mice mediated by major urinary proteins. Nature **414**, 631–634 CrossRef PubMed

17 Ibarra-Soria, X., Levitin, M.O. and Logan, D.W. (2014) The genomic basis of vomeronasal-mediated behaviour. Mamm. Genome **25**, 75–86 CrossRef PubMed

18 Harrow, J.L., Steward, C.A., Frankish, A., Gilbert, J.G., Gonzalez, J.M., Loveland, J.E., Mudge, J., Sheppard, D., Thomas, M., Trevanion, S. and Wilming, L.G. (2014) The Vertebrate Genome Annotation browser 10 years on. Nucleic Acids Res. **42**, D771–D779 CrossRef PubMed

19 Loveland, J. (2005) VEGA, the genome browser with a difference. Brief. Bioinform. **6**, 189–193 CrossRef PubMed

19a Phelan, M.M., McLean, L., Hurst, J.L., Beynon, R.J. and Lian, L.-Y. (2014) Comparative study of the molecular variation between 'central' and 'peripheral' MUPs and significance for behavioural signalling. Biochem. Soc. Trans. **42**, 866–872

20 Hurst, J.L. and Beynon, R.J. (2010) Making progress in genetic kin recognition among vertebrates. J. Biol. **9**, 13 CrossRef PubMed

21 Kaur, A.W., Ackels, T., Kuo, T.H., Cichy, A., Dey, S., Hays, C., Kateri, M., Logan, D.W., Marton, T.F., Spehr, M. and Stowers, L. (2014) Murine pheromone proteins constitute a context-dependent combinatorial code governing multiple social behaviors. Cell **157**, 676–688 CrossRef PubMed

22 Sherborne, A.L., Thom, M.D., Paterson, S., Jury, F., Ollier, W.E., Stockley, P., Beynon, R.J. and Hurst, J.L. (2007) The genetic basis of inbreeding avoidance in house mice. Curr. Biol. **17**, 2061–2066 CrossRef PubMed

23 Thom, M.D., Stockley, P., Jury, F., Ollier, W.E., Beynon, R.J. and Hurst, J.L. (2008) The direct assessment of genetic heterozygosity through scent in the mouse. Curr. Biol. **18**, 619–623 CrossRef PubMed

24 Robertson, D.H., Cox, K.A., Gaskell, S.J., Evershed, R.P. and Beynon, R.J. (1996) Molecular heterogeneity in the major urinary proteins of the house mouse *Mus musculus*. Biochem. J. **316**, 265–272 PubMed

25 Evershed, R.P., Robertson, D.H., Beynon, R.J. and Green, B.N. (1993) Application of electrospray ionization mass spectrometry with maximum-entropy analysis to allelic 'fingerprinting' of major urinary proteins. Rapid Commun. Mass Spectrom. **7**, 882–886 CrossRef PubMed

26 Green, B.N., Sannes-Lowery, K.A., Loo, J.A., Satterlee, J.D., Kuchumov, A.R., Walz, D.A. and Vinogradov, S.N. (1998) Electrospray ionization mass spectrometric study of the multiple intracellular monomeric and polymeric hemoglobins of *Glycera dibranchiata*. J. Protein Chem. **17**, 85–97 CrossRef PubMed

27 Robertson, D.H., Hurst, J.L., Bolgar, M.S., Gaskell, S.J. and Beynon, R.J. (1997) Molecular heterogeneity of urinary proteins in wild house mouse populations. Rapid Commun. Mass Spectrom. **11**, 786–790 CrossRef PubMed

28 Brun, V., Masselon, C., Garin, J. and Dupuis, A. (2009) Isotope dilution strategies for absolute quantitative proteomics. J. Proteomics **72**, 740–749 CrossRef PubMed

29 Gerber, S.A., Kettenbach, A.N., Rush, J. and Gygi, S.P. (2007) The absolute quantification strategy: application to phosphorylation profiling of human separase serine 1126. Methods Mol. Biol. **359**, 71–86 CrossRef PubMed

30 Kirkpatrick, D.S., Gerber, S.A. and Gygi, S.P. (2005) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. Methods **35**, 265–273 CrossRef PubMed

31 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 CrossRef PubMed

32 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. Protoc. **1**, 1029–1043 CrossRef PubMed

33 Rivers, J., Simpson, D.M., Robertson, D.H., Gaskell, S.J. and Beynon, R.J. (2007) Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. Mol. Cell. Proteomics **6**, 1416–1427 CrossRef PubMed

34 Simpson, D.M. and Beynon, R.J. (2012) QconCATs: design and expression of concatenated protein standards for multiplexed protein quantification. Anal. Bioanal. Chem. **404**, 977–989 CrossRef PubMed

35 Bislev, S.L., Kusebauch, U., Codrea, M.C., Beynon, R.J., Harman, V.M., Rontved, C.M., Aebersold, R., Moritz, R.L. and Bendixen, E. (2012) Quantotypic properties of QconCAT peptides targeting bovine host response to *Streptococcus uberis*. J. Proteome Res. **11**, 1832–1843 CrossRef PubMed

36 Peffers, M.J., Beynon, R.J. and Clegg, P.D. (2013) Absolute quantification of selected proteins in the human osteoarthritic secretome. Int. J. Mol. Sci. **14**, 20658–20681 CrossRef PubMed

37 Brownridge, P., Lawless, C., Payapilly, A.B., Lanthaler, K., Holman, S.W., Harman, V.M., Grant, C.M., Beynon, R.J. and Hubbard, S.J. (2013) Quantitative analysis of chaperone network throughput in budding yeast. Proteomics **13**, 1276–1291 CrossRef PubMed

38 Achour, B., Russell, M.R., Barber, J. and Rostami-Hodjegan, A. (2014) Simultaneous quantification of the abundance of several cytochrome P450 and uridine 5′-diphospho-glucuronosyltransferase enzymes in human liver microsomes using multiplexed targeted proteomics. Drug Metab. Dispos. **42**, 500–510 CrossRef PubMed

39 Al-Majdoub, Z.M., Carroll, K.M., Gaskell, S.J. and Barber, J. (2014) Quantification of the proteins of the bacterial ribosome using QconCAT technology. J. Proteome Res. **13**, 1211–1222 CrossRef PubMed

40 Wei, J., Ding, C., Zhang, J., Mi, W., Zhao, Y., Liu, M., Fu, T., Zhang, Y., Ying, W., Cai, Y. et al. (2014) High-throughput absolute quantification of proteins using an improved two-dimensional reversed-phase separation and quantification concatamer (QconCAT) approach. Anal. Bioanal. Chem. **406**, 4183–4193 CrossRef

41 Zimmerman, T.A., Wang, M., Lowenthal, M.S., Turko, I.V. and Phinney, K.W. (2013) Quantification of transferrin in human serum using both QconCAT and synthetic internal standards. Anal. Chem. **85**, 10362–10368 CrossRef PubMed

42 Chen, J., Wang, M. and Turko, I.V. (2013) Quantification of amyloid precursor protein isoforms using quantification concatamer internal standard. Anal. Chem. **85**, 303–307 CrossRef PubMed

43 Carroll, K.M., Simpson, D.M., Eyers, C.E., Knight, C.G., Brownridge, P., Dunn, W.B., Winder, C.L., Lanthaler, K., Pir, P., Malys, N. et al. (2011) Absolute quantification of the glycolytic pathway in yeast: deployment of a complete QconCAT approach. Mol. Cell. Proteomics **10**, M111.007633 CrossRef PubMed

44 Kito, K., Ota, K., Fujita, T. and Ito, T. (2007) A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. J. Proteome Res. **6**, 792–800 CrossRef PubMed

45 Picard, G., Lebert, D., Louwagie, M., Adrait, A., Huillet, C., Vandenesch, F., Bruley, C., Garin, J., Jaquinod, M. and Brun, V. (2012) PSAQ standards for accurate MS-based quantification of proteins: from the concept to biomedical applications. J. Mass Spectrom. **47**, 1353–1363 CrossRef PubMed

46 Lebert, D., Dupuis, A., Garin, J., Bruley, C. and Brun, V. (2011) Production and use of stable isotope-labeled proteins for absolute quantitative proteomics. Methods Mol. Biol. **753**, 93–115 CrossRef PubMed

47 Dupuis, A., Hennekinne, J.A., Garin, J. and Brun, V. (2008) Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. Proteomics **8**, 4633–4636 CrossRef PubMed

48 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 **106C**, 151–161 CrossRef PubMed
49 Eyers, C.E., Lawless, C., Wedge, D.C., Lau, K.W., Gaskell, S.J. and Hubbard, S.J. (2011) CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. Mol. Cell. Proteomics **10**, M110.003384 CrossRef PubMed
50 Deutsch, E.W., Lam, H. and Aebersold, R. (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO Rep. **9**, 429–434 CrossRef PubMed
51 Farrah, T., Deutsch, E.W., Kreisberg, R., Sun, Z., Campbell, D.S., Mendoza, L., Kusebauch, U., Brusniak, M.Y., Huttenhain, R., Schiess, R. et al. (2012) PASSEL: the PeptideAtlas SRMexperiment library. Proteomics **12**, 1170–1175 CrossRef PubMed
52 Brownridge, P., Holman, S.W., Gaskell, S.J., Grant, C.M., Harman, V.M., Hubbard, S.J., Lanthaler, K., Lawless, C., O'Cualain, R., Sims, P. et al. (2011) Global absolute quantification of a proteome: challenges in the deployment of a QconCAT strategy. Proteomics **11**, 2957–2970 CrossRef PubMed
53 Blakeley, P., Siepen, J.A., Lawless, C. and Hubbard, S.J. (2010) Investigating protein isoforms via proteomics: a feasibility study. Proteomics **10**, 1127–1140 CrossRef PubMed
54 Brownridge, P.J., Harman, V.M., Simpson, D.M. and Beynon, R.J. (2012) Absolute multiplexed protein quantification using QconCAT technology. Methods Mol. Biol. **893**, 267–293 CrossRef PubMed
55 Rivers, J., McDonald, L., Edwards, I.J. and Beynon, R.J. (2008) Asparagine deamidation and the role of higher order protein structure. J. Proteome Res. **7**, 921–927 CrossRef PubMed

Received 12 May 2014
doi:10.1042/BST20140133