Experimental and Theoretical Modelling of the MAPK Pathway

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2012

Louise E. Maddison

Doctoral Training Centre for Integrative Systems Biology
Michael Barber Centre for Mass Spectrometry
School of Chemical Engineering and Analytical Science

Contents

Figure List

Table List

Abstract

The MAPK pathway plays a crucial role in regulating cellular response to external stimuli. Binding of growth factors and other mitogenic signals to cell surface receptors initiates a phosphorylation-dependent relay of protein activation, resulting in altered transcription, ultimately regulating cell proliferation and differentiation. Signalling through this pathway is regulated by the coordinated function of specific protein kinases and protein phosphatases. As perturbation of this signalling system is often associated with diseases such as cancer, modelling is a useful means to help understand the outcomes that may result following changes in component levels or activity. The determination of absolute quantification data, in copies per cell, for proteins of the MAPK pathway will allow the expansion of and improved accuracy within predictive models.

The strategy used within this thesis is based on the established technique of stable isotope dilution, generating isotopically labelled peptides using the QconCAT methodology. Recombinant DNA techniques were used to generate artificial concatamers of large numbers of tryptic peptides as quantification standards. A QconCAT, LM1, of 49 KDa (29 tryptic peptides), corresponding to the scaffold proteins was designed and built to encode two peptides per protein. A second QconCAT, LM2, of 58 KDa (34 tryptic peptides), encoded peptides from the dual-specificity phosphatases (DUSPs) and substrates.

Quantification was performed using ultra performance liquid chromatography coupled to mass spectrometry. A selected reaction monitoring (SRM) approach was employed where the most intense *y*-ions per peptide were selected either from experimental data or predictions *in silico*. Using the ratio of the signal for the light:heavy isotopologues, the amount of light isotopologue can be inferred, allowing copies per cell quantifications to be established. Native peptides were present below the lower limit of quantification, and therefore the upper bounds of copies per cell were obtained for the three cell lines; colon cancer cells HCT 116 (K-Ras mutant) and HT-29 (B-Raf mutant) and a control cell line of HEK-293.

Finally, mathematical modelling was undertaken to explore the mass-action kinetics of a three component scaffold signalling molecule. It was found that the optimal scaffold concentration is between the lowest and second lowest concentration of signalling protein.

Lay Abstract

An emerging research field, known as Systems Biology, has the potential to provide a vital link to new drug discoveries to combat cancer. Systems Biology combines laboratory results and specialised software to simulate human cells in the computer. By working with these virtual cells, new insights into the disease can be revealed. Cells communicate to each other across a complex network of protein to protein interactions. When a small molecule known as a growth factor attaches to the outside of the cell, a relay of signals is stimulated. These signals travel through the cell to reach the nucleus. The nucleus holds the genetic instructions for making new cells and the instructions are acted upon. During cancer these signalling pathways between proteins become confused. For instance one protein called Ras can become permanently switched on and so the cells grow uncontrollably creating a tumour. The concentrations of proteins in healthy and cancerous human cells can be measured using mass spectrometry, a type of molecular weighing scales. The computer software contains a map of all the interactions and concentrations of each protein to build a virtual model of a cell. Experiments which would be too time consuming and complex to carry out in a laboratory can be investigated in a virtual environment. Through these techniques, we can improve our understanding of diseased cells, and realise which proteins are susceptible to intervention. This provides new drug targets to the pharmaceutical industry in the fight against cancer.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on presentation of Theses.

For Annie and Bill

Acknowledgements

My thanks go to the Engineering and Physical Sciences Research Council whose funding, through the Integrative Systems Biology Doctoral Training Centre, has made this research possible. To my supervisors Hans Westerhoff and Claire Eyers. To Claire for providing the opportunity to work in the Michael Barber Centre for Mass Spectrometry. To Hans for his continued enthusiasm and support beyond the PhD. Special thanks go to Kathleen Carroll who stepped in during Claire's absence and continued to help throughout the project. To Nils Blüthgen and his laboratory, especially Ela and Franzi for providing me with cells to analyse and assistance with the modelling. To Simon Gaskell for all his support over the years and getting me started (properly) in mass spectrometry. To past and present members of the Michael Barber Centre especially Stephen, Francesco, Belen, Tarn, Nav, Hannah, Ross, Steve, Sarah, Zubida and Narciso. To my fellow members of the DTC especially Nicholas, Natalie and Stephen for impromptu tea breaks. To my boyfriend Richard for his unwavering confidence in me. And finally my thanks go to my Mother, Father and Sister for supporting my wish to become a scientist from an early age and honouring my fifth birthday present request of a microscope and a big yellow teapot.

List of Abbreviations

AC	Alternating current
AMU	Atomic mass unit
BSA	Bovine serum albumin
CI	Chemical ionisation
CID	Collision induced dissociation
CV	Coefficient of Variation
Da	Dalton
DC	Direct current
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ESI	Electrospray ionisation
ETD	Electron-transfer dissociation
FA	Formic acid
FAB	Fast atom bombardment
FT-ICR	Fourier transform ion cyclotron resonance
HCT	High capacity trap
HPLC	High-performance liquid chromatography
ICAT	Isotope-coded affinity tagging
ID	Internal Diameter
IEF	Isoelectric focusing
IMAC	Immobilised metal ion affinity chromatography
iTRAQ	Isobaric tag for relative and absolute quantification
LC	Liquid chromatography
LC-MS	Liquid chromatography coupled on-line with mass spectrometry
m/z	Mass-to-charge ratio

MALDI	Matrix-assisted laser desorption/ionisation
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
Mw	Molecular weight
nESI	Nano-electrospray
pI	Isoelectric point
PMF	Peptide mass fingerprinting
ppm	Parts per million
PTM	Post-translational modification
QconCAT	**Q**uantification **conCAT**amers
Q-TOF	Quadrupole time-of-flight
QQQ	Triple Quadrupole
RF	Radio frequency
RNA	Ribonucleic acid
RP	Reversed phase
rpm	Revolutions per minute
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis,
SILAC	Stable isotope labelling by amino acids in cell culture
SRM	Selected Reaction Monitoring
TIC	Total Ion Chromatogram
TOF	Time-of-flight
UV	Ultraviolet
2DE	Two-dimensional electrophoresis

1 Introduction

1.1 Systems Biology

Systems biology is the study of an organism, viewed as an integrated network of interacting genes, proteins and metabolites (Figure 1.1). Quantitative measurements of the proteome, transcriptome and metabolome enable mathematical models to be generated and validated based on experimental observations in a data-driven manner. These measurements are translated into mathematical models and used to generate new insights into the system. Proteomics is the identification, quantification, and functional analysis of all proteins, or subset of proteins within a given system. Absolute quantification strategies in proteomics based on mass spectrometry has developed in the last 10 years and is a major source of data for systems biology studies. In order to populate mathematical models, data can be provided on enzyme concentration *[E]*, which is directly proportional to V_{max} the maximum enzyme velocity and k_{cat} the catalytic constant as shown by equation (1), is used to populate the model. Iterative measurements of reproducible sets of proteins from differentially perturbed biological states and over a controlled time-course will lend an additional dimension to such models. The Michaelis-Menten equation is shown in (2), where *[S]* is substrate concentration and k_m the Michaelis constant. The effective V_{max} depends on the amount of active enzyme in proximity to its substrate. For example activity of the enzyme can be effected by Post Translational Modifications (PTMS) e.g. phosphorylation, arginine methylation which can cause conformational changes. Additionally, other proteins may sequester the enzyme and substrate together and therefore increase their effective concentrations within the complex.

$$V_{max} = k_{cat}\,[E] \tag{1}$$

$$v_0 = \frac{V_{max}+[S]}{k_m+[S]} \tag{2}$$

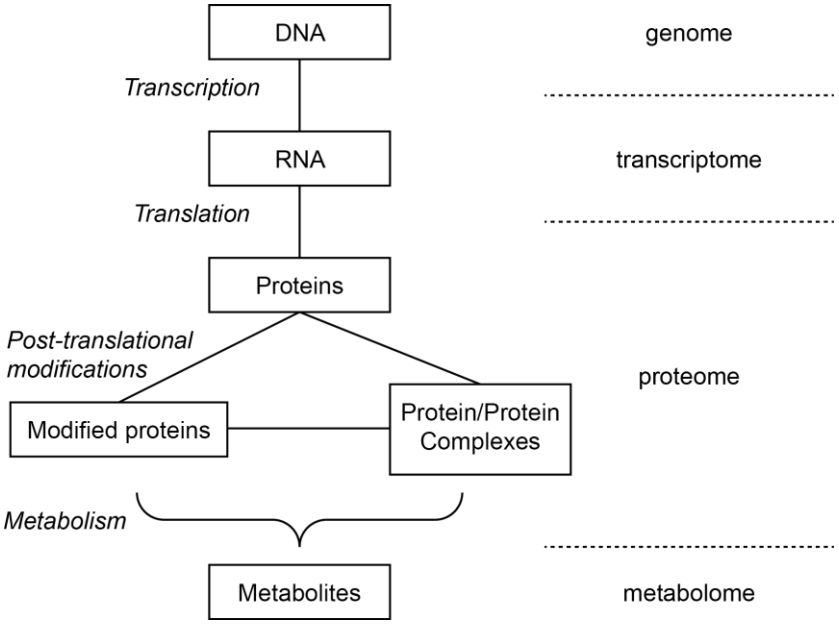


Figure 1.1 Interactions that take place between components of the cell at the levels of genome, transcriptome, proteome and metabolome.

1.2 General Structure of MAPK Pathways

Cells respond to extracellular signals by transmitting intracellular instructions to coordinate appropriate responses. The mitogen activated protein kinase (MAPK) pathway is highly conserved from yeast to humans and occupies a central role in fundamental cellular processes such as proliferation, differentiation, survival and apoptosis (Figure 1.2). There are several major groups of MAPKs that have been indentified in mammalian cells, these include the extracellular signal regulated kinase 1 and 2 (ERK1/2) cascade, which regulates cell growth and differentiation, stress activated protein kinases (SAPK) which include c-Jun N-terminal kinase (JNK) and p38 MAPK cascades these function in response to stress responses and apoptosis.[1] There is much crosstalk between components of the different cascades and to what extent this occurs is not completely known.

The MAPK pathway utilises one of the most generic signalling designs found in biological signal transduction, namely a protein phosphorylation cycle formed by a kinase phosphorylation of a target protein and one or more counter-acting phosphatases that are responsible for its dephosphorylation. This represents a fundamental mechanism by which numerous components are regulated including enzymes, receptors, transporters, docking and scaffolding proteins.

The core signalling pathway of MAPK is a three tiered cascade. At each level an activated kinase phosphorylates the kinase at the next level down. The MAP kinase (MAPK) is activated by the MAPK kinase (MAPKK) by phosphorylation at two specific amino acid residues, a threonine and a tyrosine, within the conserved sequence motif Thr-X-Tyr, and in turn the MAPKK is phosphorylated at serine and threonine residues by the MAPKK kinase (MAPKKK).[2] MAPKs are deactivated by a family of phosphatases named MKPs (MAPK phosphatase) which can inactivate the corresponding kinase at each cascade level. The MAPK pathway is thought to allow for signal amplification and additional regulatory interfaces that allow the kinetics, duration and amplitude of the activity to be tuned.[3,4]

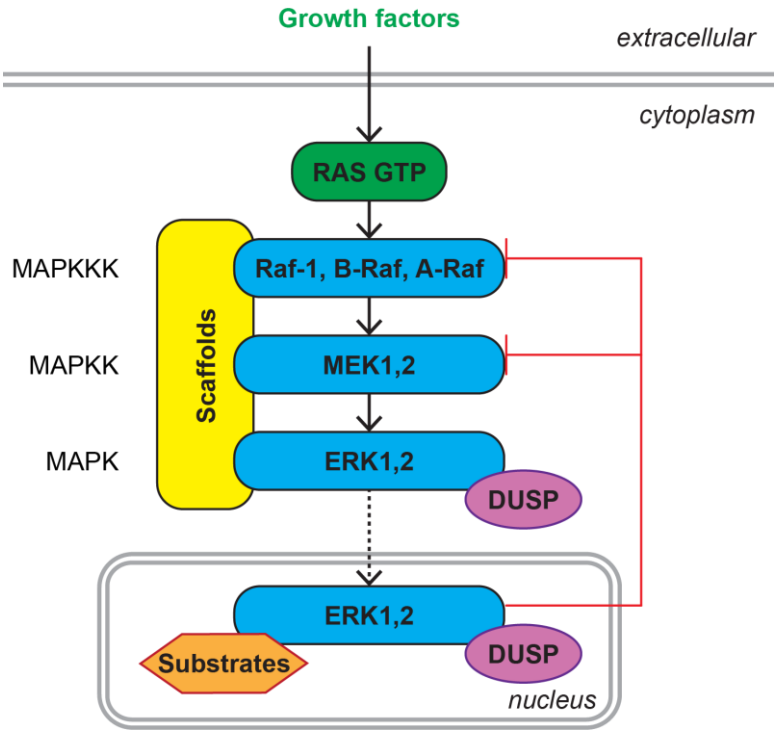


Figure 1.2 The MAPK pathway and feedback regulation. Blue arrows represent activation; red lines with blunt ends represent inhibition.

1.2.1 The Ras/Raf/MEK/ERK Module

Signal transduction along the Ras/Raf/MEK/ERK pathway begins when receptor tyrosine kinases, G protein coupled receptors and/or integrins are activated by their respective ligands. These membrane proteins recruit and activate Ras proteins by inducing the exchange of Ras bound GDP with GTP, which converts Ras into an active conformation. The process relies on the recruitment of GDP/GTP exchange factors such as SOS (son of sevenless) to the cell membrane where Ras resides. Ras has three isoforms H-Ras, N-Ras and K-Ras and each binds to Raf proteins with different binding affinities.[5] The signalling of these isoforms is partially controlled on the basis of Ras subcellular distribution.

The Raf family of MAPKKKs, comprised in mammals of A-Raf, B-Raf and C-Raf (or Raf-1) are recruited by small G-proteins to the cell membrane where Raf activation takes place. Raf activation is a complex and only partially understood process that involves high affinity binding to Ras, phosphorylation and changes in conformation and protein interactions.[6] The interaction of MEKs with Raf is dependent on a proline rich domain unique to MEKs. B-Raf activates MEK1/2 by phosphorylation of two serine residues. MEK1/2 is a dual specificity kinase that binds to inactive ERK and retains it in the cytoplasm. MEK then phosphorylates ERK1/2 at both threonine and tyrosine residues of the threonine-glutamic acid-tyrosine "TEY" motif. Active ERK is released and can dimerise and translocate to the nucleus. ERK can now phosphorylate over 150 substrates in the cytoplasm and nucleus including transcription factors.[7,8]

1.2.2 Scaffold Proteins

Scaffold proteins organise and specify MAPK cascades by binding at least two components of the signalling pathway and thereby facilitate their functional interaction e.g. enhancing kinase phosphorylation of a substrate.[9] In addition, scaffolds can enhance phosphorylation of a specific subset of downstream substrates by targeting these multi-enzyme signalling modules to different cellular locations. Lastly, it is thought that scaffolds may prevent crosstalk between similar pathways. Table 1.1 shows a summary of scaffold proteins to be quantified in this project.

1.2.3 Transcription Factor Substrates

The regulation of gene expression is a central function of MAPK signalling as the targets for the MAPK pathway are transcription factors located in the nucleus leading to altered transcription of genes. MAPK phosphorylates transcription factors that are then retained in the cytoplasm in an inactive form. Dephosphorylation of these factors is required for migration to the nucleus to activate transcriptions. The mechanism for activation induced translocation of ERK from cytoplasm to nucleus remains unclear however it has been suggested that ERK is sequestered in the cytoplasm by MEK1 and 2.[10]

Table 1.1 Selected scaffold proteins in the MAPK pathway (continued overleaf).

Scaffold	Comments
KSR1 and 2, Kinase Suppressor of Ras	Upon Ras activation, KSR translocates with MEK1/2 to the plasma membrane, which brings MEK1/2 in close proximity to its activator Raf-1 and downstream effectors of ERK1/2. These interactions result in the formation of the Raf/MEK/ERK complex.[11]
β-arrestin 1 and 2	β-arrestin 1 and 2 bind to Raf, MEK and ERK and thereby enhance ERK activity in the cytoplasm and prevent translocation into the nucleus.[12-14]
Paxillin	Located at focal adhesions, alongside other focal adhesion specific proteins such as FAK, actopaxin and PAK.[15] Focal adhesions are sites of tight adhesion between the actin cytoskeleton and the extracellular matrix and are regions of signal transduction that relate to growth control. The coordinated assembly and disassembly of protein complexes allows for cytoskeletal remodelling and thus enables cell motility. Paxillin plays a major role within this process by regulating cell spreading and migration.[16]
MP1 MEK Partner-1/MPKS1 Mitogen activated protein kinase scaffold 1	MP1 selectively binds to MEK1 and ERK1, but not MEK2. It is suggested that by bringing MEK1 and ERK1 in close proximity to one another, MP1 enhances the signalling through this pathway. This complex is targeted to late endosomes, mature organelles, by the interaction of MP1 with the endosome protein p14, where they enhance the activation of ERK1.[17] Down-regulation of MP1 or p14 expression reduced ERK/MAPK activation, whereas over-expression increased ERK/MAPK activity in an additive manner.[18] Recently it has been found that the MP1-p14 scaffold also interacts with PAK1 resulting in enhanced phosphorylation and activation of MEK during cell adhesion and spreading on fibronectin.[19]
Shoc2/Sur 8	A leucine rich repeat protein. Initial studies have shown that when over-expressed, mammalian Sur 8 enhances Raf activation by promoting its interaction with Ras.[20] Further findings have led to the hypothesis that Sur 8 acts to enhance the strength of ERK signalling by facilitating events that promote Raf activation.[21]

Scaffold	Comments
PEA-15 anti-apoptotic phosphoprotein enriched in astrocytes	Acts as a high-affinity ligand for both ERK1 and ERK2.[22] Interaction has shown to occur via two distinct regions in PEA-15, an N-terminal death effector domain (DED) and a small region of the C-terminus. PEA 15 binds to ERK/MAPK, but not other components of the ERK/MAPK pathway, in the cytoplasm and prevents the translocation of ERK/MAPK into the nucleus by anchoring these components. However the exact mechanism is unclear.[23] PEA-15 inhibits ERK/MAPK nuclear uptake by blocking ERK binding to protein subunits of the nuclear pore complex known as nucleoporins.[24] Moreover it has been demonstrated that the nuclear export sequence of PEA-15 keeps ERK/MAPK out of the nucleus.[25] It has also been reported that PEA-15 binds RSK2, and targets it to ERK thereby enhancing activation. When PEA-15 is highly expressed phosphorylation and activation of RSK2 is enhanced, whilst at low expression levels PEA-15 prevents these events.[26]
IQGAP1	Binds B-Raf, MEK and ERK and facilities ERK activation at specific levels of EGF or insulin like growth factor, which shows that optimal ERK/MAPK activation requires a balanced stoichiometry of the IQAP1-ERK2/MAPK complex.[27] IQAGP1 is over expressed in some cancers such as breast and ovarian, therefore its scaffolding functions for the ERK pathway may contribute significantly to tumuorigenesis.[28]
MVP, Major Vault Protein	The predominant protein of the large cytoplasmic ribonucleoprotein structure called vaults that are highly conserved among diverse eukaryotes.[29] MVP binds SHP2 (Src homology 2 domain containing tyrosine phosphatase) which regulates cell signalling and ERK in response to EGF.[30]
PEB1 Phosphatidyl-ethanolamine-binding protein 1/RKIP	Binds to both Raf and MEK and prevents their physical interaction, inhibiting MEK phosphorylation and activation by Raf1 and B-Raf.[31,32] Following mitogenic stimulation, RKIP dissociates from Raf to allow MEK activation. Recently it has been demonstrated that RKIP protein expression is down-regulated in metastatic cancer cells and sensitises resistant cancer cells to chemotherapy.[33-36]

1.2.4 Dual Specificity MAPK Phosphatases

Negative regulation of the MAPK pathway is mediated by MAPK phosphatases (MKPs) or dual-specificity protein phosphatases (DUSPs). Until recently research has focussed on protein kinases rather than phosphatases as important regulators of signalling cascades. DUSPs have largely been ignored during theoretical models of the MAPK pathway; however this study will involve a special focus on the DUSPs.

DUSPs de-phosphorylate both tyrosine and threonine in the activation loop of MAPKs and differ in sub cellular localisation and substrate specificity.[37] However they all share a common structure that comprises a C-terminal catalytic domain and an N-terminal non-catalytic domain. This latter region contains sequences which determine their sub cellular localisation and more importantly a conserved cluster of basic amino acids involved in MAPK recognition known as kinase interaction motif (KIM).[38]

On the basis of gene structure, sequence similarity and subcellular localisation, the DUSPs can be subdivided into three classes.[37] The first class includes DUSP1, DUSP2, DUSP4, and DUSP5, all of which localise to the nucleus and are induced by growth factors. The second group comprises DUSP6, DUSP7, and DUSP9. These localised to the cytoplasm and preferentially recognise ERK1 and ERK2 in vitro. The third class includes DUSP8, DUSP10, which preferentially recognise stress activated MAPK isoforms. Currently DUSP14 has not been allocated to a class and may require its own subgroup as it shows lack of specificity to ERK, JNK and p38 pathways and is localised in both nucleus and cytoplasm.

In addition to their role as phosphatases, there is growing evidence that DUSPs also control subcellular localisation of MAPKs through scaffolding. For example DUSP6 can anchor ERK2 in the cytoplasm by a mechanism dependent on both its nuclear export signal and KIM motif that binds to MAPK.[39] DUSP2 and DUSP4 have been shown to anchor ERK in the nucleus via their D-domains.[40]

1.3 Mathematical Models

Traditional biochemistry and molecular biology has focussed on properties of individual molecules, including proteins and their domains, substrates, ligands and their interacting partners and complexes. This reductionist approach has been successful at elucidating the structures and functions of cellular constituents. In contrast the understanding of biological processes at a wider scale such as a whole intracellular signal transduction network is far more limited.[41] However, recent technological advances such as sequencing of whole genomes and evaluation of cellular proteome by mass spectrometry have enabled high throughput data acquisition in biology. When these large scale approaches are combined with mathematical and computational methods new perspectives are gained on the systems wide level.

1.3.1 Why model?

Mathematical modelling, the process of translating a biological system into a model for subsequent computer simulation and analysis, is central to systems biology. In order for the model to be useful and answer biological questions, models should faithfully describe the biological system and make predictions about their behaviour. The basis for a model is to identify the key components and interactions and assemble a topological representation of the system. The next stage is to build a description of the system's dynamic behaviour from which its predictive power arises. Therefore the model can be used to reveal and explain unexpected behaviour, test conditions which may be difficult to study in the laboratory, check hypotheses and direct future laboratory work. Overall a model helps to improve our understanding of a biological system, identifies key regulatory components and processes and helps predict biological behaviour.

1.3.2 ODE Modelling

A large number of ERK pathway models are based on a common approach to modelling biological systems, that of the Ordinary Differential Equation (ODE). ODE modelling takes a population view of the system and requires exact knowledge of reaction rates and concentrations of proteins. A differential equation can be defined as an equation involving one or more unknown functions and their derivatives. For example to describe the concentration of A known as [A], and how this changes over time. Considering the simple reaction scheme below which depicts

the conversion of A into B, mass action kinetics can be applied, where *k* represents the rate constant of the reaction.

$$A \xrightarrow{k} B \quad (3)$$

Therefore the rate of reaction proceeds at the following rate, where the rate of reaction (v) is equal dependent on [A].

$$v = k[A] \quad (4)$$

From the above equations it is then possible to construct differential equations representing the rate of change in [A] and [B] over time:

$$\frac{d[A]}{dt} = -k[A] \quad (5)$$

$$\frac{d[B]}{dt} = k[B] \quad (6)$$

In order to simulate the above reaction and thus solve the differential equations and approximate the change in concentration of all species over time, we first require the initial (time zero) values of [A] and [B]. One of the simplest methods of numerical integration is by Euler's method; more advanced methods include Runge-Kutta, and Rosenbrock and Richardson extrapolation.[42] Although the above example is small and simple, the method can easily be expanded through the use of matrices to complex biological systems involving many more components and reaction mechanisms.

1.3.3 MAPK Models

Over last 15 years an ever increasing number of MAPK models have been developed, continually growing in size and complexity. In 2005, it was reported that there were over 30 mathematical models that incorporated the MAPK cascade.[43] These models have been used to investigate various aspects of the biological behaviour of the MAPK systems including those briefly mentioned in Table 1.2 and described further in depth in the next section. However, all of these models rely heavily on data acquired by qualitative methods such as western blotting. Previous studies have indicated that absolute quantification of MAPK pathway components

using mass spectrometry are crucial to understand the quantitative effects resulting from interactions of different components.[44]

Table 1.2 Multiple properties of the MAPK pathway described as function and mechanism. Adapted from[45]

Function	Mechanism
Ultrasensitivity	Two step distributive process
Amplification	Multiple amplification stages: receptor, dual phosphorylation steps for MAPKK and MAPK
Oscillation	Negative feedback loops
Sustained and transient activation	Activation of ERK by different growth factors

In 1981, Goldbeter and Koshland proposed model to describe the phosphorylated and unphosphorylated form of a single molecule.[4] Later, in 1996 Huang and Ferrell extended the model to a three molecule cascade of the MAPK pathway.[3] They demonstrated that the MAPK cascade exhibited ultrasensitivity, defined as a non-linear sigmoid activation curve whereby the degree of ultrasensitivity increases the further down the cascade.[46] In 1997, Ferrell and Bhatt showed that in *Xenopus oocytes* MEK phosphorylates ERK2 by a two-collision distributive mechanism rather than a single cell processive mechanism.[47] MEK phosphorylates one site, dissociates, and then rebinds to phosphorylate the second site. This experimental model provides a mechanistic basis for the understanding of how MAPK can convert graduated signal inputs into switch like outputs. Further evidence for this mechanism was provided by Burack and Sturgill.[48] Additionally, Ferrell and Mahler showed, using *Xenopus oocytes*, that the MAPK cascade is activated in an all or nothing fashion during cell maturation.[49] This behaviour is due to the MAPK cascade's intrinsic ultrasensitivity proposed previously.

In 1999, Kholodenko proposed an oscillatory mechanism for the MAPK model.[50] The study combined experimental, kinetic analysis and computational modelling of the short term pattern of cellular response to EGF in isolated hepatocytes. Kinetic parameters were based on the scientific literature and derived from basic physicochemical properties. Laboratory experiments were performed such as time courses of EGFR phosphorylation to effectively validate the model. The kinetic model predicts how the cellular response is controlled to the relative levels and

activity states of signalling proteins and under what conditions activation patterns are transient or sustained. Although the model does not include the core cascade of Raf/MEK/ERK and only goes as far as SOS protein, it was one of the first to incorporate feedback between theoretical prediction and experimental validation (Figure 1.3). Also the model considered EGFR and its associated adaptor proteins and correctly predicted a transient recruitment of SOS to EGFR at the plasma membrane, leading to a transient activation of Ras and the MAPK cascade. Therefore this model has been used as a basis for many other models of the EGFR system which extend to include the core MAPK cascade.

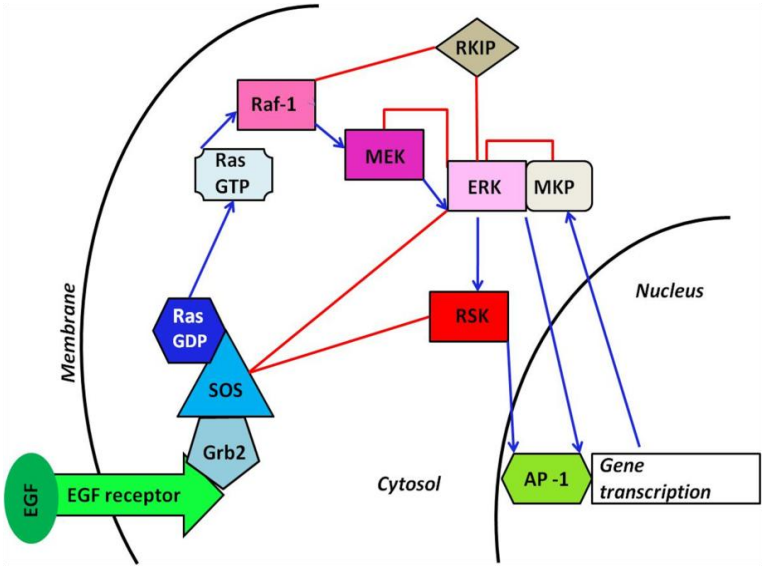


Figure 1.3 The MAPK pathway and feedback regulation. Ras and proteins involved in Ras activation are displayed in shades of blue, kinases in shades of red, inhibitory proteins in shades of grey and transcription factors in turquoise. Blue arrows represent activation; red lines with blunt ends represent inhibition. Adapted from[51].

In the same year Levchenko explored effects of scaffold proteins on the MAPK model.[52] The model proposed that scaffold proteins may affect the levels of signalling through the MAPK cascade in a biphasic manner and thereby reduce the threshold level of phosphorylated MAPK required for pathway activation. In a later theoretical study Somsen *et al* demonstrated that scaffolding could induce selectivity in different MAPK modules even if they shared the same kinases at some levels in the cascade.[53] These models highlighted the positive and negative effects that scaffold proteins exert on the MAPK pathway and showed that they amplify and attenuate signal transduction.

Brightman and Fell developed an ODE based model of the EGF induced MAPK cascade activation in PC12 cells, differentiating rat pheochromocytoma cells to investigate factors influencing the kinetics of activation.[54] The kinetic constraints of reactions and initial concentrations of species were mainly based on a range of measured or estimated values published within the literature. Analysis of the model provides quantitative evidence that feedback inhibition of the MAPK cascade is the most important factor when considering the duration of cascade activation.

In 2002, Schoeberl *et al* developed an ODE model to describe the dynamics of the EGF signal transduction pathway and is regarded as one of the most comprehensive as 125 reactions and 94 species are included (Figure 1.4).[55] The model was used to investigate the effects of receptor internalisation dynamics on Shc/Grb2/SOS adaptor complexes and Ras mediated activation of the Raf/MEK/ERK signalling cascade. It showed that EGF induced responses are stable over a 100 fold range of ligand concentration and that the critical parameter in determining signal efficacy is the initial velocity of receptor activation. The majority of the kinetic parameters were based on values published in the literature, and initial concentrations were either compiled from the literature or based on laboratory experiments. The model includes a large range of dynamic processes tested experimentally in HeLa cell culture which were found to agree well. Due to the high level of detail included in this model subsequent modellers have used this as a base to build their own.[56,57]

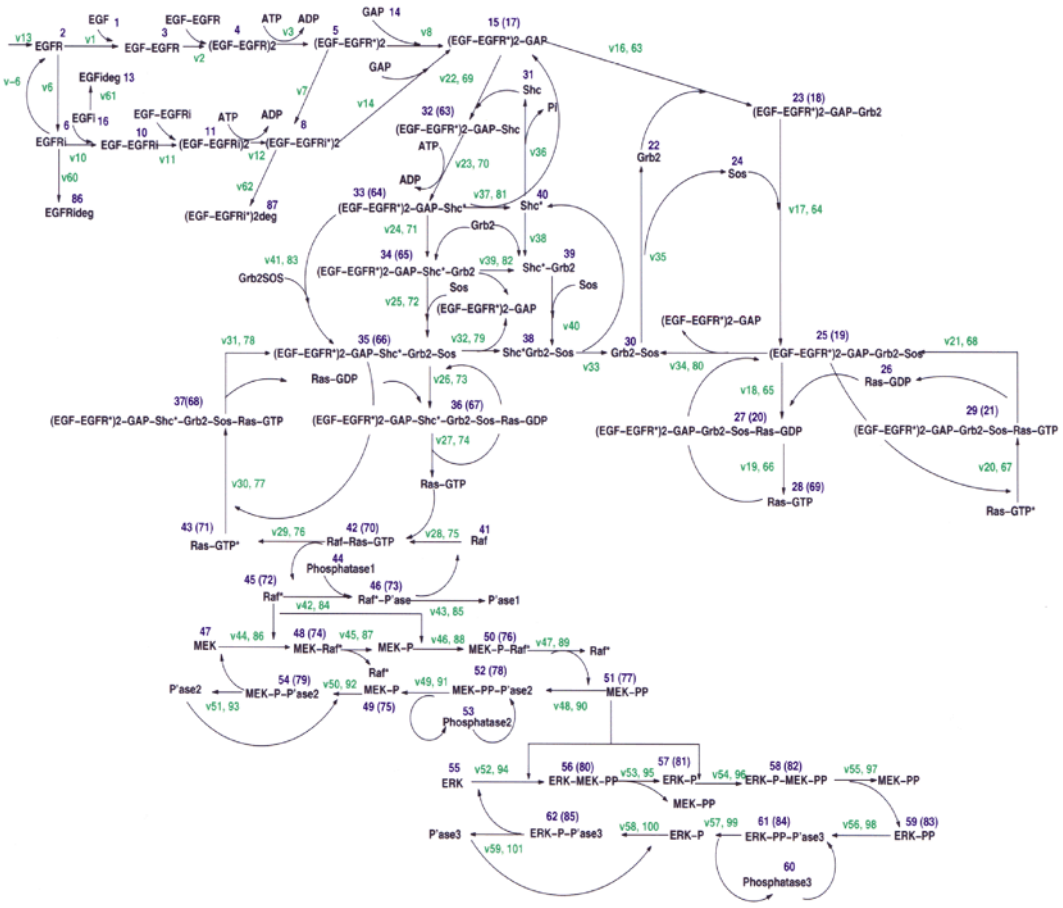


Figure 1.4 Scheme of the EGF receptor-induced MAP kinase cascade. The MAP kinase cascade can be initiated by Shc-dependent and Shc-independent pathways. Each component is identified by a specific number (blue). Blue numbers in brackets specify the components after internalization. The arrows represent reactions and reaction rates are shown by green numbers v1−v125. The second green numbers identify reaction rates after internalisation. Reprinted by permission from Macmillan Publishers Ltd.[55]

However, the Schoeberl model does have a number of errors. The most important of these errors is that there is no negative feedback loop from doubly phosphorylated ERK (ERK-PP) to SOS. This causes the dissociation of Grb2-SOS from the receptor complex. However an updated version of the model exists which includes this feedback loop (Figure 1.5).[58] Additionally, some processes have been modelled incorrectly, for example there is a build up of Ras-GTP (species 43), EGFRideg (species 87), (EGF-EGFRi*)2dg (species 86), and EGFideg (species 13). All these species continue to increase over time and a steady state is not reached. Also, the model does not account for different isoforms of proteins and some proteins remain unidentified e.g. Phosphatases 1,2 and 3 and Prot. Lastly, the scaffold proteins, DUSPs and substrates are largely ignored.

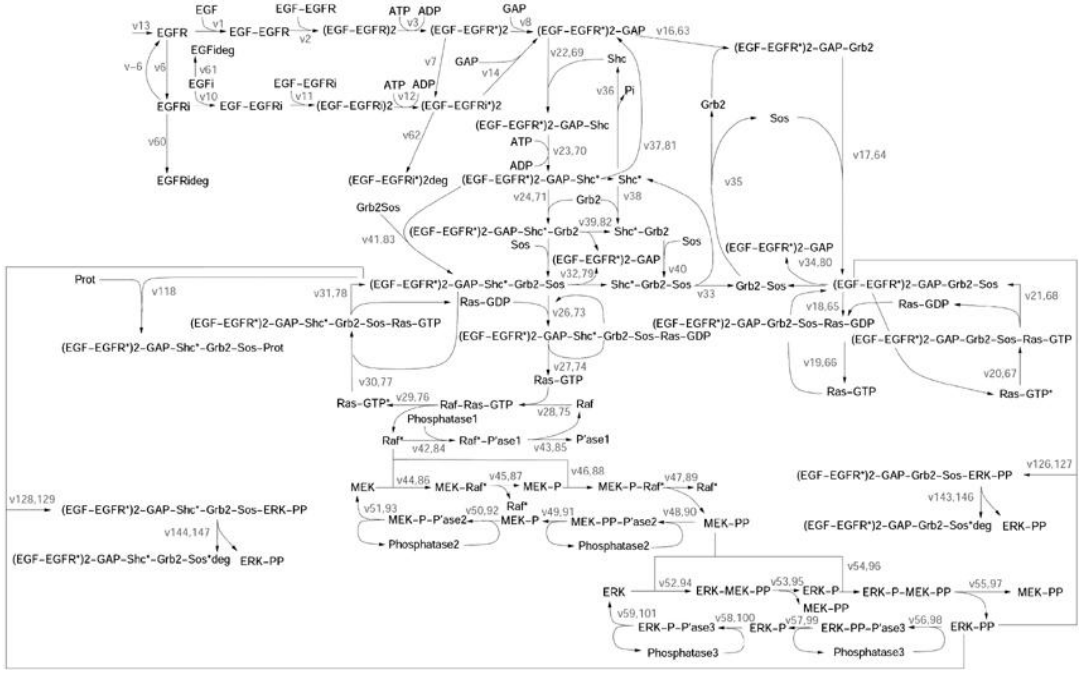


Figure 1.5 An updated Schoeberl model of the MAPK pathway which features a negative feedback loop from doubly phosphorylated ERK (ERK-PP) to Sops. Each component is identified by a specific number (blue). Blue numbers in brackets specify the components after internalization. The arrows represent reactions and reaction rates are shown by green numbers v1−v125. The second green numbers identify reaction rates after internalisation. Reprinted by permission from Macmillan Publishers Ltd.[58]

There are several other interesting features to emerge from models which will not be considered in depth here. These include the suggestion that multisite phosphorylation of MAPK acts to improve signalling specificity.[59] The proposal that bistability and hysteresis are inherent properties (see below) of MAPK cascade and do not require feedback loops.[60] Finally, that amplification, input potency and dynamic range of output maybe tuned by manipulating module components.[61]

Over recent years, the computational or mathematical modelling of MAPK pathway has led to some novel insights and predictions of how this system functions. Computational modelling is increasingly valuable and can provide useful information to understand behaviour.

1.3.4 Ultrasensitivity, Bistability and Oscillations in Signal Transduction

The stimulus response curves of the activation of the MAPK pathway shows strong sigmoidity *in vivo.* Sigmoidity of these response curves is known as ultrasensitivity which reflects the response of biological systems to changes in signals around the threshold stimulus (Figure 1.6). Sub threshold stimuli are damped and super threshold stimuli are transmitted which allows for an increasing stimulus to be converted to switch like all or nothing response.[62] Sigmoid responses can also be used to filter out noise or delay responses.[63] Sustained oscillations arise when the ultrasensitive cascades possess a negative feedback loop. Bistability (hysteresis) occurs when the ultrasensitive cascade is in combination with positive feedback.

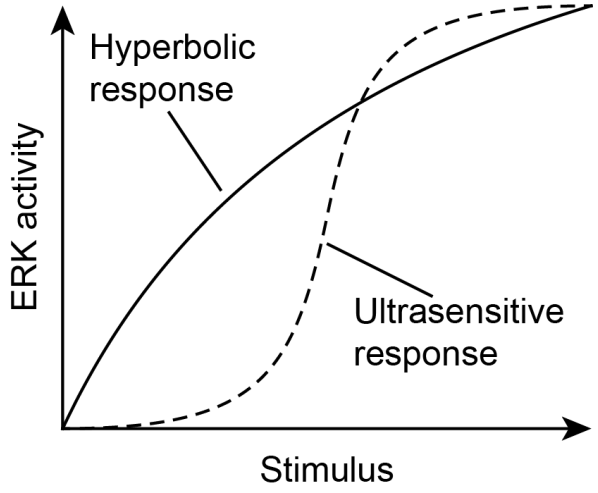


Figure 1.6 Sigmoidal stimulus-response curve representative of ultrasensitivity. Ultrasensitive responses are highly sensitive to changes around a threshold, unlike in hyperbolic responses.

Mechanisms that lead to ultrasensitive behaviour include cooperativity, multisite phosphorylation, feed forward loops, and enzymes operating at saturation. The latter mechanism is known as zero order ultrasensitivity as early theoretical work showed that phosphorylation cycles have strong ultrasensitivity if the enzymes (kinase and phosphatase) operate near saturation, that is when K_m values of enzymes are much smaller than the substrate concentration.[4] This early work by Goldbeter and Koshland highlights the key for observing ultrasensitivity is that the phosphatases must be saturated and this also relies on the amount of Raf/MEK/ERK relative to K_m of the phosphatases. If the phosphatases are not saturated then ultrasensitivity will not be observed. This prediction has not been fully tested by the scientific field and instead the measurement of phosphatases has largely been ignored. Often within the

models the enzyme is described by a Hill type equation. However, Michaelis-Menten kinetics can describe this behaviour if the phosphatase is fully saturated.

Recent experimental studies of the MAPK pathway suggest that zero order ultrasensitivity is less important in this system than previously thought due to the effects of substrate sequestration.[64] If the substrate is sequestered onto the catalysing enzyme so that enzyme and substrate concentrations are comparable then the ultrasensitivity is strongly weakened.

1.3.5 Bistability

Bistability may occur when there is a positive or double negative feedback loop involved with the ultrasensitive cascade. The system can switch at saddle node bifurcations between two stable steady states (on and off) separated by an unstable state (Figure 1.7). Depending on whether the system begins in the 'on' or 'off' state will dictate the type of stimulus response curves obtained. This behaviour is known as hysteresis and is history dependent as the positive feedback can act as a memory device. For example, when the stimulus is removed, the 'on' state is maintained indefinitely and the system shows irreversible activation.[65] Bistability is an important mechanism to filter out noise during signal transduction. In the mammalian MAPK pathway it has been proposed that bistability is due to ERK activating Raf via a positive feedback loop involving PKC (protein kinase C).[66] Recent theoretical work shows that due to enzyme sequestration effects, positive feedback and bistability arise in the Raf/MEK/ERK pathway.[60,67] Therefore filtering out noise by ultrasensitivity is a general property of mammalian MAPK pathway.

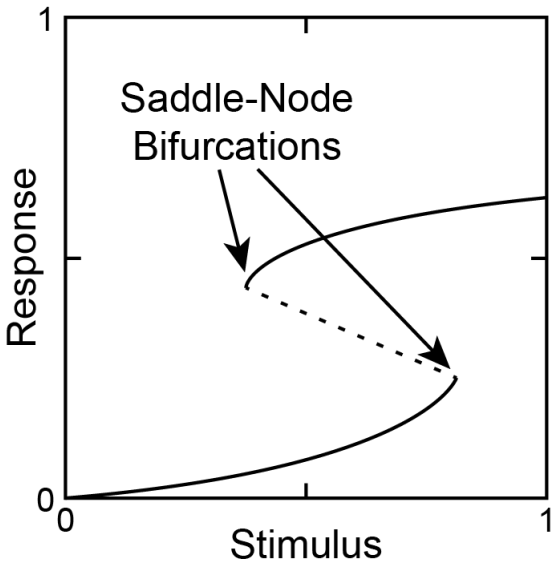


Figure 1.7 Bistability, where an ultrasensitive pathway has a positive feedback loop results in two steady states (solid line) separated by an unstable state (dotted line).

1.3.6 Oscillations

Sustained oscillations may be a property of ultrasensitive cascades with negative feedback. Kholodenko investigated this feature using a numerical analysis of the dynamics of MAPK cascade.[50] A strong negative feedback can turn off the activation of the cascade and in other cases the system steady state may lose its stability. As there is no other steady state, the phosphorylation level of cascade kinases starts to oscillate in a sustained way, and is known as a Hopf bifurcation. For oscillations to occur depends on the saturation of kinases and phosphatases, that is the ratio between their total concentrations and their K_m values and on the strength of the negative feedback.

1.4 Mass Spectrometry

Mass spectrometry is a powerful analytical tool used to assist in the identification and structural elucidation of known and unknown compounds.[68] A wide variety of analytical platforms and orthogonal separation devices have been developed, notably LC-MS, whereby liquid chromatography (LC) is coupled to mass spectrometry (MS), enabling the controlled real-time separation of the analytes prior to introduction to the mass spectrometer. The technique is able to detect compounds at very low concentrations (one part in 10^{12}) in complex mixtures. The molecular weight (MW) and structural information of the analytes can be obtained relatively quickly; in the case of peptides the primary amino acid sequence can be determined. These advantages have led to the application of mass spectrometry (MS) in a variety of fields including clinical screening, drug discovery, and environmental monitoring. MS is at the forefront of proteomics research (the analysis of proteins, their structure, concentration and post-translational modifications) by providing technological platforms for the identification, quantification and verification of proteins within complex samples, commonly following controlled enzymatic degradation to the component peptides. A mass spectrometer consists of an ionisation source, a mass analyser that measures mass-to-charge ratio (*m/z*) of the ionised analytes, and a detector that registers the number of ions at each *m/z* value (Figure 1.8).

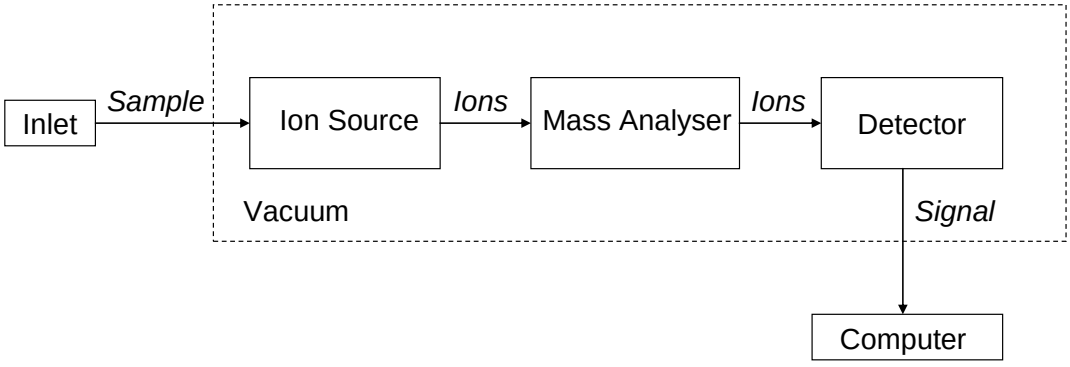


Figure 1.8 Schematic of the basic components of a mass spectrometer. For some instruments the ion source is not under vacuum.

1.4.1 Ionisation Techniques

The sample may be introduced to the ionisation source either directly or after chromatography such as high pressure liquid chromatography (HPLC), gas chromatography (GC), or capillary electrophoresis (CE). Molecules of a compound are then ionised either by ejection of an electron or by capture of a proton. There are

many methods available including electron ionisation (EI), fast atom bombardment (FAB), electrospray ionisation (ESI), matrix-assisted laser desorption ionisation (MALDI), chemical ionisation (CI). The technique selected depends on the type of sample, the mass spectrometer and the information required from the analysis. High energy methods such as EI are not suitable for the analysis of macromolecules such as proteins and polypeptides because extensive fragmentation and thermal decomposition occur during ionisation. Softer techniques, such as ESI and MALDI are commonly used to volatise and ionise macromolecules.

1.4.1.1 Matrix-Assisted Laser Desorption Ionisation (MALDI)

MALDI is generally used for simple peptide mixtures (Figure 1.9). The rate of energy transfer to the analyte is low and therefore the internal energy of the generated ions is also low. An advantage of this is that intact molecular weight can be determined; a disadvantage is that minimal structural information is obtained. However, minimal fragmentation of the ions will occur without the introduction of further internal energy by gas-phase or surface collisions of analyte molecular ions.

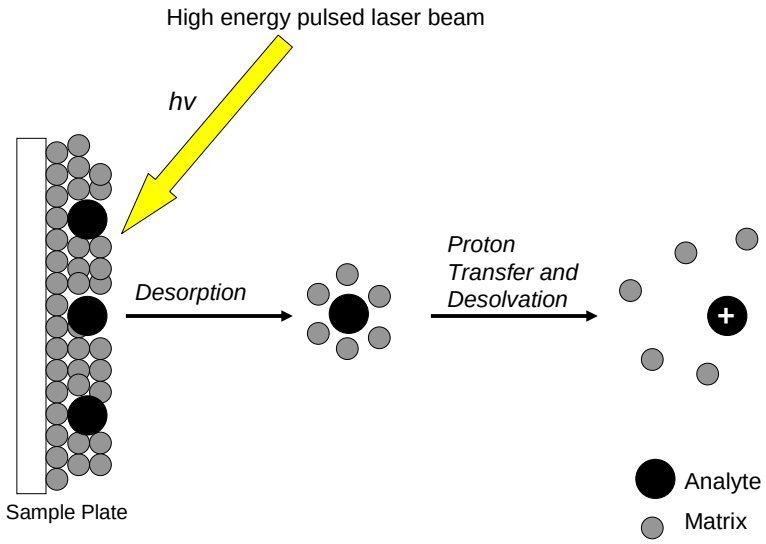


Figure 1.9 Schematic depicting MALDI (matrix-assisted laser desorption/ionisation). The analyte is generally mixed with a large molar excess (10^4) of organic matrix. A pulsed laser is used to irradiate the analyte/matrix crystals on a stainless steel surface of the sample plate in the source chamber under vacuum. The laser pulse generates a plume of gas-phase molecules from which ions are extracted under a voltage gradient into a mass analyser.

1.4.1.2 Electrospray Ionisation (ESI)

In 1968 Dole recognised the possibility that gas-phase ions of macromolecules could be generated by spraying a solution from the tip of an electrically charged capillary.[69] This work was later developed as a useable interface for mass spectrometry by Fenn and co-workers and is known as electrospray ionisation (ESI).[70]

ESI allows for the sampling of ions directly from solution at atmospheric pressure. An important benefit of direct sampling is that for macromolecules such as peptides which are insufficiently volatile or thermally stable to allow volatilisation prior to electron or chemical ionisation are instead amenable to analysis by MS (Figure 1.10).[71] Although other techniques such as fast atom bombardment (FAB) do allow for the ionisation of thermally unstable species of low volatility there are limitations in the analytical performance and ease of use for routine assays.[72]

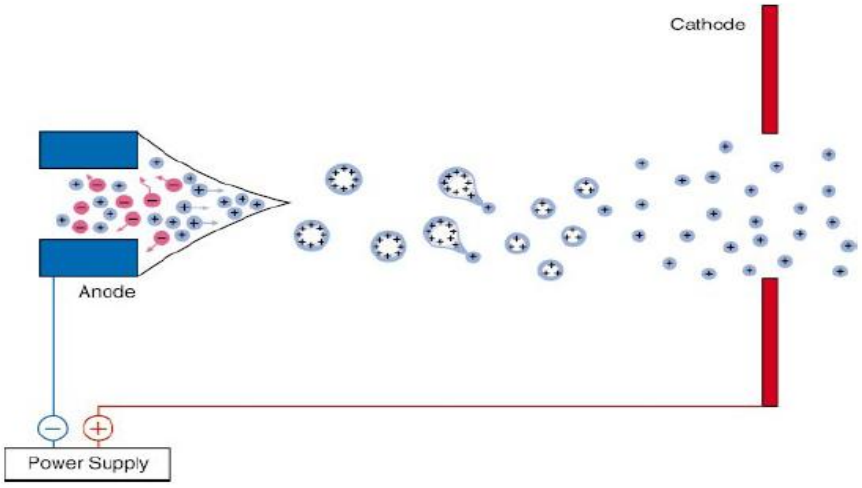


Figure 1.10 Schematic depicting ESI (electrospray ionisation). A fine spray of droplets which contain the analyte and solvent molecules is generated upon application of an electrical voltage through a narrow bore needle under vacuum. A heated capillary is placed following the electrospray needle to induce desolvation, assisted by a counter flow of nebulising and or drying gas. A positive voltage polarity will generate positive ions and the opposite is true. A Taylor cone forms when mutual repulsion between the ions at the surface becomes greater than the surface tension of the liquid. The electric field must be strong enough to cause a spray of small droplets.

In ESI a solution of the analyte is passed through a capillary which is held at high potential, typically 2-5 kV. The voltage can be positive or negative. The applied voltage provides an electric field gradient and as the solution emerges charge separation occurs at the surface of the liquid. As a result the liquid protrudes from the

capillary tip and a “Taylor Cone” is established. Evaporation of solvent causes the charge density on the surface of the Taylor cone to increase until the Rayleigh limit is reached which is the point when Columbic repulsion of the surface charge is equal to the surface tension of the solution. Fission ('Coulomb explosion') occurs and smaller droplets result which pass along a pressure gradient towards the analyser of the mass spectrometer. Ions are generated through repeated evaporation and fission and it has been proposed that gas-phase ions are generated *via* one of two models. The charged residue model (CRM) suggested by Dole and co-workers assumes that repeated solvent evaporation and fission will result in the formation of droplets containing a single ion (Figure 1.11).[69-73]

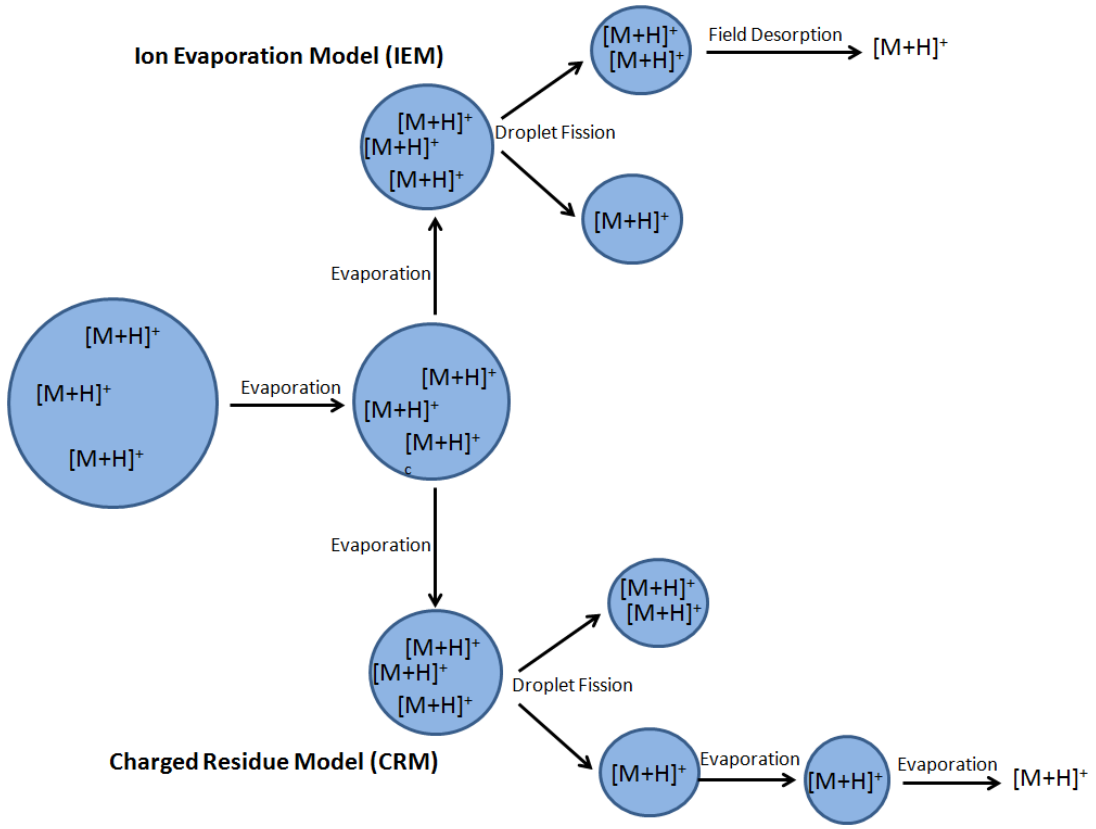


Figure 1.11 Charged Residue Model (CRM) and Ion Evaporation Model (IEM) involved in electrospray ionisation.

A second mechanism proposed by Iribarne and Thomson is known as the ion evaporation model (IEM).[74] IEM begins the same as CRM, with repeated solvent evaporation and droplet fission. However, when the droplets are very small and highly charged then the electrostatic force on the surface is strong enough to

overcome the solvation forces and as a result an ion is thus "lifted" from the solvent's surface. Within the scientific literature there is much debate over which model applies.[75] However it is widely accepted that multiply-charged species are formed via the CRM,[76] whilst small ions are formed either by CRM[77] or IEM.[78]

1.4.2 Mass Analysers

Following ionisation, gas phase ions pass through a series of ion optics to focus the ion beam to the mass analyser which is central to the instrument's sensitivity, resolution, mass accuracy and therefore ability to generate information rich mass spectra. The mass analyser resolves ions based on their structure or charge. Resolution of charged particles within a selected *m/z* range can be gained by the application of electrical or magnetic fields, or by measuring the time it takes for an ion to travel a fixed distance. There are four basic types of mass analyser used in proteomics research; ion trap, Time of flight (TOF), quadrupole and Orbitrap analysers each with different features as described below.

1.4.2.1 Time of Flight (TOF)

Time of flight mass spectrometers determine the *m/z* ratio of an analyte by measuring the time it takes for ions to traverse the length of a field-free flight tube The ions are introduced either directly from the source of the instrument or as a pulse from a previous analyser e.g. from the quadruple of the hybrid instrument Q-TOF (described in section 1.4.2.3). This results in all ions receiving the same initial kinetic energy. The ions then pass through the field free drift zone and are separated by their masses with lighter ions travelling faster. All ions of the same *m/z* should arrive at the detector at the same time and results in high sensitivity. The equation for TOF separation is described below, where *m/z* is mass-to charge ratio of the ion, *E* is the extraction potential, *s* is the length of the tube over which *E* is applied, *d* is the length of the filed free drift zone, *t* is the measured time-of-flight of the ion and *e* is the electron charge.

$$\frac{m}{z} = 2eEs\left(\frac{t}{d}\right)^2 \qquad (7)$$

However, in reality there is a kinetic energy distribution for each *m/z* which needs to be corrected to maintain the resolution. The correction is via a reflectron (in the "V" mode, Figure 1.12), a series of electric fields, which the ions are accelerated towards. The more energetic ions penetrate the reflectron more deeply and take a longer path to the detector compared with those that are less energetic. Therefore ions of the same *m/z* value are refocused on the reflectron detector and this is crucial for achieving a high resolution, typically 10,000 at full-width-half-maximum (FWHM) in "V" mode, for the TOF analyser.[79]

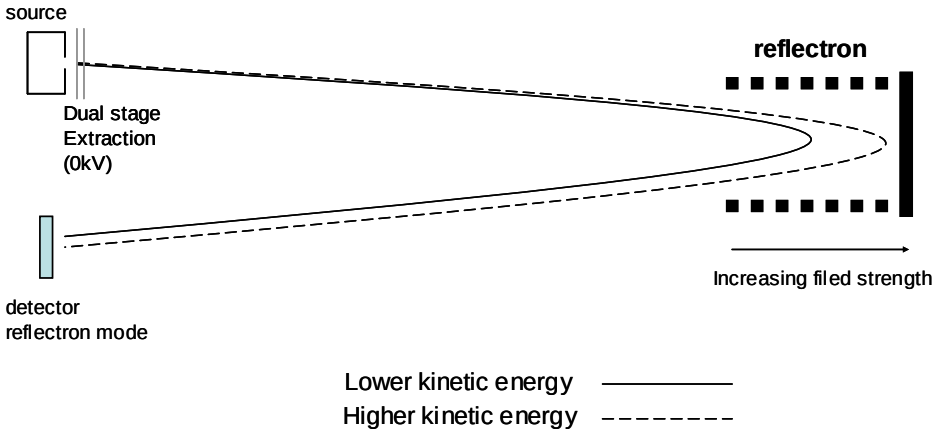


Figure 1.12 Schematic showing the principles of a reflectron time of flight (TOF) mass spectrometer when operated in "V" mode. A source generates gas-phase ions, which are accelerated in a high-voltage region into the drift-free region. The reflectron behaves as a high-voltage ion "mirror" and reflects ions onto the detector, allowing increased energy focusing and therefore mass resolution of ions. An ion's kinetic energy determines the depth to which ions penetrate the electric field.

1.4.2.2 Quadrupole Mass Filter

Quadrupole mass filters are one of the most commonly used mass analysers (Figure 1.13). A quadrupole mass filter consists of four parallel metal rods where each pair of rods has either a negative or positive applied potential according to the equations:

$$\emptyset = \left(U + Vcos(\omega t)\right) \quad (8)$$

$$-\emptyset = \left(U - Vcos(\omega t)\right) \quad (9)$$

Where θ is the electrical potential, U is a constant DC voltage, $Vcos(\omega t)$ is an AC voltage.

Figure 1.13 Schematic of a quadrupole mass analyser. Ions are introduced into the mass analyzer from the ion source. High-voltage and radio frequency are applied to the paired quadrupole rods. Ions have stable trajectories at specific DC/RF values. The analyser can therefore be employed at different DC/RF values to scan over a mass range to analyse ions of various *m/z* value of interest.

Applied voltages to these pairs of rods create a hyperbolic field within the rods which results in a force on the ions. For given DC and AC voltages only ions of specific mass to charge ratio pass through the quadrupole filer whilst all other ions collide with the rods. As voltages on the quadrupoles are varied, different ions pass through the filter and these are monitored to produce a mass spectrum. The voltages can be varied by changing ω and keeping U and Vstatic, or by varying U and V and keeping ω constant. When V is larger that U low mass ions are able to follow the rapidly alternating RF potential. In the *x*- direction these ions will stay in phase with the RF and will oscillate with increasing amplitude until they collide with the rods. The heavier ions will pass through to the other end of the quadrupoles without colliding with the quadrupoles and therefore the *x*- direction is a high pass mass filter. At the same time in the *y*-direction, heavy ions will be lost because of their inertia towards the fast altering RF field. However, lighter ions will be transmitted to the other end of the quadrupoles without striking the *y* electrodes and therefore the *y*-direction is a low mass filter. The combination of both provides a stability window, defined by the frequency ω of V and the ratio of V/U, allowing ions to be resolved.

When the function on an electric filed is substituted into the equations of motion the Mathieu equations are obtained. These describe the motion of an ion along each coordinate axes and are described below where r_0 is half the distance between the

rods and e is the magnitude of the electron charge (assuming singly charged ions). Motion along the z-axis is unaffected by the potential on the quadrupole electrodes.

$$\frac{d^2x}{dt^2} + \frac{e}{mr_0^2}(U + Vcos\omega t)x = 0 \quad (10)$$

$$\frac{d^2x}{dt^2} + \frac{e}{mr_0^2}(U + Vcos\omega t)y = 0 \quad (11)$$

These equations can be rewritten dimensionless to obtain the parameters a and q which describe the effect of the application of a DC voltage on the motion of the ions and the application of AC voltage respectively.

$$\frac{d^2x}{d\tau^2} + (a_x + 2q_x cos2\tau)x = 0 \quad (12)$$

$$\frac{d^2y}{d\tau^2} + \left(a_y + 2q_y cos2\tau\right)y = 0 \quad (13)$$

$$a = \frac{4eV}{mr_0^2\omega^2} \quad (14)$$

$$q = \frac{2eU}{mr_0^2\omega^2} \quad (15)$$

$$\tau = \frac{\omega t}{2} \quad (16)$$

A stability diagram, Figure 1.14, can be plotted as a function of U and V for ions of different masses where $m_1<m_2<m_3$. By maintaining a constant U/V ratio a scan line is obtained which allows the observation of these masses. For greater resolution the scan line can be adjusted to have a higher slope, as long as it still passes through the stability areas. As the slope of the line decreases the ion signal intensity increases but at the expense of mass resolution. The more RF cycles an ion is subjected to, then the lower the ion transmission and the greater the loss of signal at the selected *m/z* value. Quadrupoles are normally operated at unit resolution which is sufficient to separate two peaks one mass unit apart.

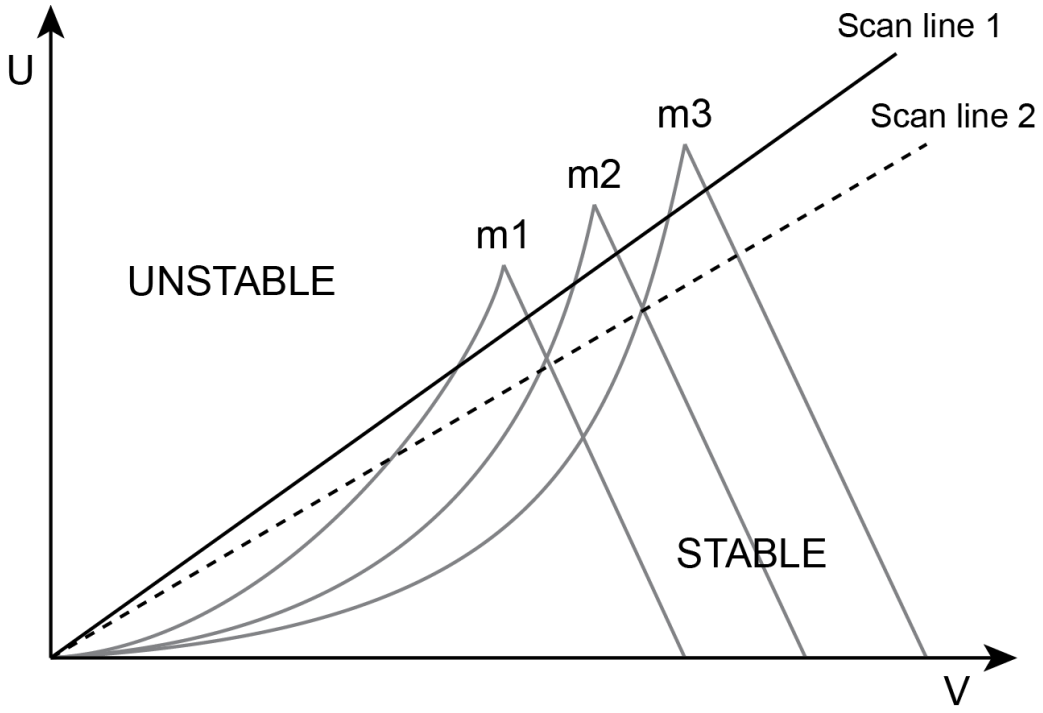


Figure 1.14 Stability diagram of a quadrupole filter. All ions whose parameters are located in the triangle above the scan lines will reach the detector. Scan line 1 offers greater resolution compared to scan line 2.

1.4.2.3 Quadrupole time-of-flight mass spectrometer (Q-TOF)

The Q-TOF is a hybrid instrument which combines the use of two different mass analysers, the quadrupole and the TOF which are coupled via a collision cell and are arranged in series. This tandem "in space" instrument allows for the quadrupole to isolate precursor ions with unit mass resolution and for these ions to be analysed by the TOF which provides higher mass resolution (10,000 FWHM in "V" mode) and good mass accuracy (< 5ppm) for the analysis of product ions.[80]

Glish and Goeringer reported the first Q-TOF mass spectrometer.[81] However, it was not until nearly 10 years later that Q-TOF instruments became commercially available driven in part by advances in ionisation techniques of biological species.[82] The initial design used thermal ionisation following desorption from a probe, a pulsed ionisation technique which allowed packets of ions to be generated. These ion packets are compatible with the pulsed operation of the acceleration plates of the TOF which facilitate transfer of ions along the flight tube. Therefore it would not

have been possible to couple the same kind of time of flight analyser to a continuous ion source, such as electrospray ionisation.

Dawson and Guilhaus developed a method that allows the conversion of ions from a continuous source, such as electrospray ionisation, into a pulsed source known as orthogonal acceleration.[83] The TOF is arranged orthogonally to the quadrupole rather than co-linearly (Figure 1.15). The ion beam is collimated by a charged lens and then enters the orthogonal accelerator. To extract the ions from the relatively slow moving ion beam, a pulsed electrostatic field with a potential necessary to push the ions is applied and thus the ions experience a force orthogonal to the ion beam axis. The original ion beam velocity is maintained but is independent of the orthogonal velocity component and produces a spontaneous drift trajectory.[84]

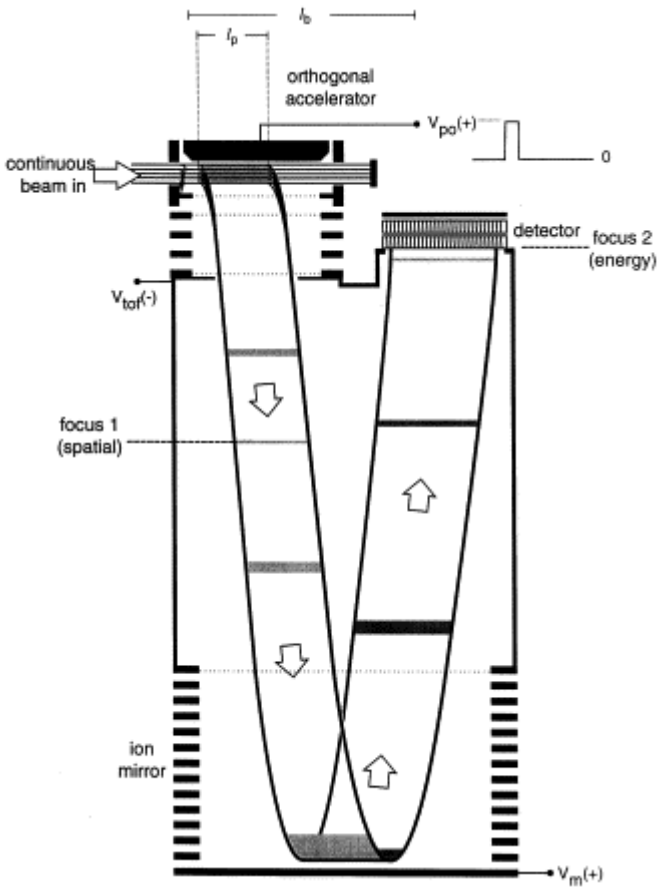


Figure 1.15 Schematic of an orthogonal acceleration time of flight (TOF) mass spectrometer. The beam "fills" the orthogonal accelerator which and a pulsed electrostatic field is applied. A packet of ions of length l_p is sampled and accelerated to enter the drift region. Reflecting TOF optics are used to focus the ions onto the detector. Whilst the ions travel the drift region and reach the detector the orthogonal accelerator is refilled. Reproduced with kind permission from Elsevier Ltd.[83]

In addition, the orthogonal acceleration approach provides a number of advantages. There is a high duty cycle as the time taken to fill the orthogonal accelerator is approximately equal to the time taken for ions to travel along the drift region and reach the detector. Space charge effects are reduced as the ions that are sampled are dispersed along the beam axis therefore additional increases in mass resolving power result. Lastly, there is improved sensitivity of the TOF as any neutral species that are produced by the ion source will be unaffected by the electrostatic field and will not enter the TOF.[85] Commercially available Q-TOFs, such as the Waters Q-TOF Ultima Global used in this study use orthogonal acceleration described above (Figure 1.16).

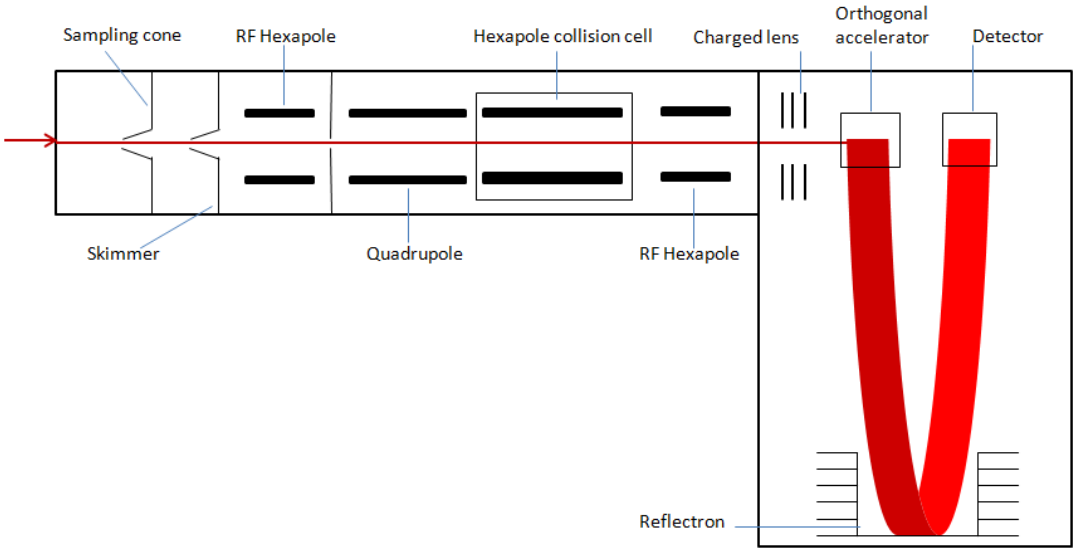


Figure 1.16 Schematic of the Q-TOF.

1.4.2.4 Quadrupole Ion Trap

The quadrupole ion trap (QIT) was developed alongside the quadrupole mass filter by Paul and co workers in 1953.[86] Ions enter the QIT through an inlet and are then trapped by three hyperbolic electrodes: the ring electrode and two end cap electrodes located at the entrance and exit. To the ring electrode, RF and DC voltages are ramped at a constant RF/DC ratio to create a 3D quadrupolar potential field which taps the ions in a stable oscillating trajectory. By gradually changing the potential the ion motions, which depend on the voltage applied and *m/z* ratio, are destabilised resulting in ejection through the exit endcap and then subsequent detection. The first commercial ion traps were not developed until 1983, when Stafford and co workers introduced the use of helium gas (1 mTorr) within the trap which acts to damp the

kinetic energy of the ions, and so ions are ejected more efficiently from the tap resulting in improved resolution.[87] A more efficient ion ejection results in a greater sensitivity as more ions reach the detector. A second development was the mass instability operation mode whereby all ions created over a given time period were trapped and then ejected consecutively from the ion trap onto an electron multiplier detector by RF ramping.

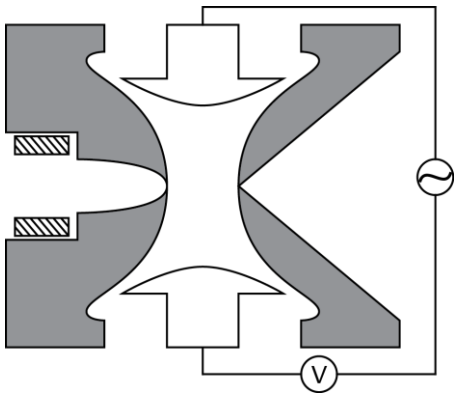


Figure 1.17 Schematic to depict a Quadrupole Ion Trap.

The stability of ions within the ion trap is also dependent upon the Mathieu Equations, as described above. A stability diagram for ions within an ion trap is reproduced below. If the amplitude of the voltage placed on the ring electrode allows for an ion of given *m/z* to have a qz value that falls within the boundaries of stability then it will be trapped. If however the qz value at that voltage falls outside the boundaries then the ion will hit the electrodes and will be lost (Figure 1.18).

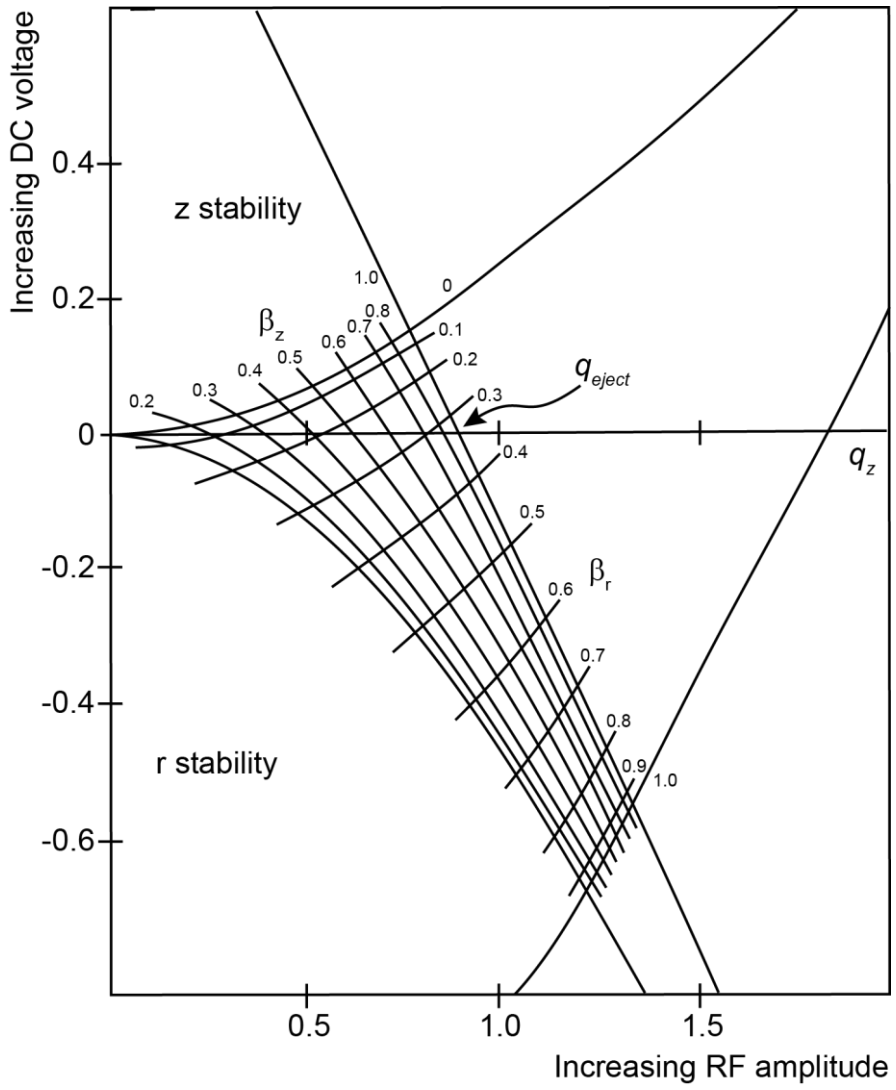


Figure 1.18 Stability diagram for ions within a quadrupole ion trap. Adapted from[88].

1.4.2.5 Linear Ion Trap

In a linear ion trap (LIT) ions are confined radially by a 2D RF field and axially by potentials applied to the electrodes. Compared to Paul traps LITs have higher injection efficiencies and higher ion storage capacities.[89] Mass selective ejection of ions from the trap can be either axial or radial depending on the apparatus. The ThermoFisher manufactured LIT, known as the LTQ, uses radial ejection and is composed of four parallel hyperbolic shaped rods cut into three sections with stopping potentials applied.[90]

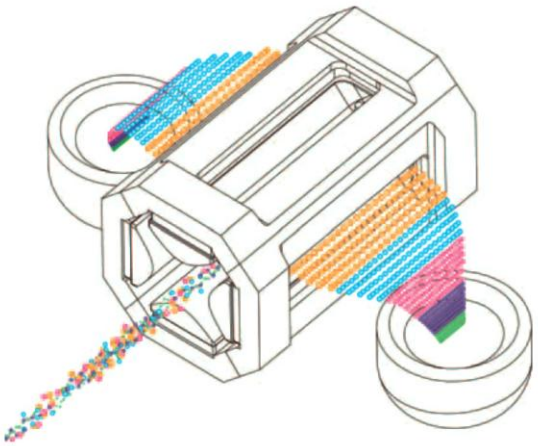


Figure 1.19 Central electrodes of linear quadrupole ion trap with mass selective radial ion ejection. Reproduced by kind permission from ThermoFisher Ltd.

During mass analysis trapped ions are ejected in the radial direction through two parallel slots in the centre of the linear ion trap. A detector is placed either side of the trap to double the number of ions detected and to maximise sensitivity.[91] The LIT can be operated as standalone mass spectrometer or combined with other mass analysers at the front end to form hybrid instruments such as LTQ-Orbitrap and LTQ-FTICR, where it is used for trapping, ion selection, and reactions.[92] The LTQ-Orbitrap XL produced by ThermoFisher (Figure 1.20) is used in this study and for high resolution mass spectrometry ions are accumulated in the linear ion trap and are then passed on to the Orbitrap analyser. At the same time as acquisition of this signal, the major peaks are isolated, fragmented and fragment ions recorded at high sensitivity in the linear ion trap.[93]

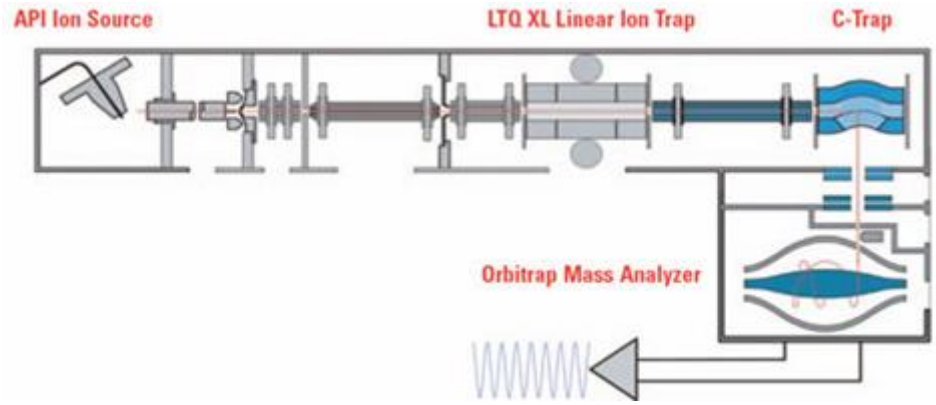


Figure 1.20 A schematic of the LTQ- Orbitrap XL. (Reproduced from http://www.rigaslabs.gr/rigaslabs/products/show.html?pid=15 accessed 1st March 2012)

A LIT has been also been developed that allows ions to be ejected axially from the device.[94] Four quadrupole cells are arranged linearly and ions can be trapped by applying stopping potentials to aperture plates at the end of the third or fourth quadrupole. Ions excited at their resonant frequencies increase their kinetic energy towards the exit as the fringing fields influence ion motion in the three directions. Ions gain sufficient kinetic energy to overcome the stopping potential at the exit.

1.4.2.6 Orbitrap

The most recent development in trapping devices has been the hybrid instrument combing linear trap with a Fourier Transform (FT) device developed by Makarov and commercialised by ThermoFisher.[95,96] The Orbitrap (Figure 1.21) can be thought as a modified Knight-style Kingdon trap with an inner spindle shaped axial shaped electrode, a barrel shaped surface, and outer coaxial electrodes. Following the specialised dynamic injection pulse a constant electric potential is applied between these two axis symmetric electrodes. Ions are injected at right angles from a C trap as a discrete packet with a kinetic energy matching the opposing potential energy of the radial electric field. The ions orbit around the central electrode but also undergo harmonic oscillations along the z-axis at the same time. The frequencies of these oscillations are inversely proportional to the square root of the *m/z* ratio and induce an image current between the two halves of the trap which is detected and FT used to convert the time domain data into frequency domain.

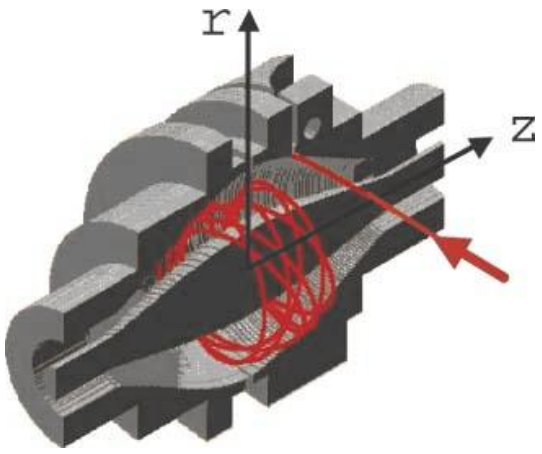


Figure 1.21 A cutaway view of the Orbitrap mass analyzer. The red arrow indicates where the ions are injected into the orbitrap with a with a velocity perpendicular to *z*-axis of the Orbitrap. Reproduced by kind permission by John Wiley and Sons Ltd.[96]

1.4.2.7 Ion Mobility Mass Spectrometry

Ion mobility can be described as a gas-phase electrophoretic technique that allows species to be separated on the basis of their charge, mass, and mobility in a given buffer gas. Mobility is related to the rotationally averaged collision cross section that is the size and shape of the analyte.[97] Early 20th century investigations on the movements of ions in gases laid the foundation of the technique which involves the measurement of an ion's velocity through a drift region. The ions are under the influence of an electric field gradient and experience collisions with a buffer gas. Reports of combining ion mobility with mass spectrometry first emerged in the 1960s.[98,99] This hybrid technique is referred to as ion mobility-mass spectrometry (IM-MS) and allows for the mass spectrometric analysis of ion mobility separated ions based on their interactions with the buffer gas as they travel through the ion mobility device. There are several types of ion mobility spectrometers including drift time,[100] field asymmetric ion mobility spectrometry (FAIMS)[101] and travelling wave.[102] These ion mobility mass spectrometers have been successfully coupled to many different types of MS analyser including TOF,[103,104] quadrupole,[105] and FT-ICR.[106]

Despite differences mentioned above for configurations in general all IM-MS experiments share a number of similarities. After sample introduction ions are generated and steered towards the drift region by gas flow and optics. The ions then reach an ion shutter or gate which pulses the ions into the drift cell, a chamber filled with a known gas at a known pressure across which an electric field is applied. The ions experience an electrostatic force which pulls them through the cell but they collide with buffer gas which delays their progress towards the detector. Larger ions will experience more collisions with the gas, and as result take longer to traverse the drift cell. An ion mobility experiment essentially consists of measuring the arrival time distributions of ion mobility separated gas phase ions.[107]

An example of an instrument capable of IM-MS is the Waters Synapt HDMS (high definition mass spectrometer). When ion mobility is invoked ions are separated using a travelling voltage wave which is incorporated in an RF ion guide (Figure 1.22) which is the central of three guides arranged orthogonally. The first guide stores the ions before injection into the ion mobility separator and the last guide transfers the

ions to the TOF. The RF only ion guide is the stacked ring ion guide (SRIG) whereby washer-like ring electrodes are arranged in sequence with opposite phases of RF applied to consecutive rings. This provides a potential well which keeps ions radially confined.

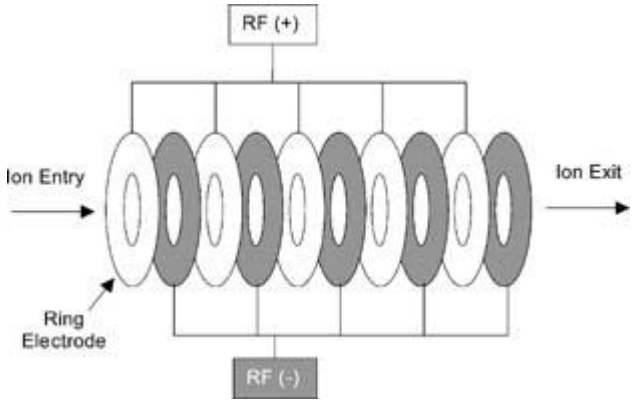


Figure 1.22 An RF–only stacked ring ion guide. Reproduced with kind permission from John Wiley and Sons Ltd.[102]

Ions are propelled through the mobility separator by a travelling wave which is superimposed on the radially confining RF voltage in the SRIG that causes the ions to move along the wave, reducing their transit time. This device is often referred to as a travelling wave ion guide (TWIG).[108] One advantage of the Synapt is that collision-induced dissociation can be performed before and after ions enter the mobility cell. Within this context, ion mobility-separation of ions is exploited to resolve the population of ions submitted to CID and therefore increase the proteome coverage. This is because chimera CID spectra, where two or more isobaric (identical *m/z*) precursor ions with similar retention time are co-sequenced by tandem MS. These ions would be otherwise unresolved by conventional MS, but can actually be separated by IM-MS based on their collisional cross section, therefore resulting in two distinct precursor ions (and two distinct tandem MS submitted to the search engine).[109] In this study a first-generation Synapt HDMS was used and ion mobility was not invoked during the LC-MS^E experiments. However, the second-generation Synapt HDMS system does invoke ion mobility during LC-MS^E experiments.[110] In conclusion, IM-MS is a powerful technique for biological investigations that allows the separation of and analysis of complex mixtures, as well as to observe and obtain information on species conformation.

1.4.3 Principles of Tandem Mass Spectrometry

Tandem mass spectrometry is a method involving the dissociation of a particular ion within a mass spectrometer to yield product ions. This can occur from the formation of metastable ions or from an activated dissociation process. The precursor ion (m_p^+) will generate product ions (m_f^+) and neutral ions (m_n), (Equation 9).

$$m_p^+ \rightarrow m_f^+ + m_n \qquad (9)$$

Tandem mass spectrometry can be defined with respect to either time or space. Tandem in space refers to mass spectrometers that are separated physically and are coupled together e.g. Quadrupole-TOF (Q-TOF) instrument. Tandem in time refers to instruments that have an ion storage device where fragmentation and mass analysis are performed sequentially.

The MS/MS product ion spectrum of a peptide can enable determination of its amino acid sequence. The various types of ion products resulting from cleavage along the peptide backbone is dependent on a variety of factors including the protease enzyme used to digest the protein, ionisation method, amino acid composition, and coupling method of the ion source into the mass spectrometer. Trypsin digested peptides subjected to electrospray ionisation are commonly doubly protonated resulting in an $(M+2H)^+$ species, where M is the mass of the peptide and H^+ is a proton.[111] As all tryptic peptide fragments contain a basic lysine or arginine residue at the C-terminal they readily ionise to form doubly charged molecular ions. Higher charge states are possible if the peptide contains basic residues e.g. histidine which can also become protonated. Peptide backbone fragmentation is through the lowest energy bonds, the amides, which produce b-ions if charge stays on the *N*-terminal fragment or *y*-ions if charge stays on the *C*-terminal fragment (Figure 1.23).[112] Fragmentation of doubly charged precursor ions is aided by the presence of a mobile ionising proton on its peptide backbone.[113] Depending on the instrumentation, the resulting product ion spectrum is predominately composed of b- and *y*- ions series where the mass difference between adjacent ions in the series corresponds to the molecular weight of the amino acid residue (minus 18 as water is lost in the formation of an amidic bond) which is characteristic except for the isobaric leucine/isoleucine pair. Amino acids

such as glutamine/lysine which have the same nominal mass may be differentiated on the basis of their exact mass using high mass accuracy instruments.

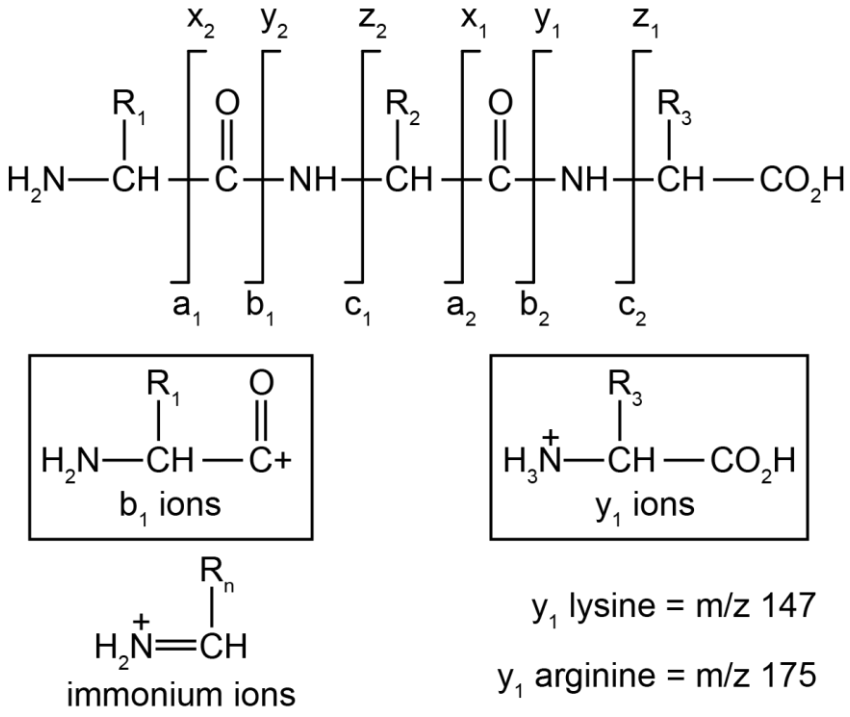


Figure 1.23 The peptide bond is the most common site cleaved following collision induced dissociation of peptides leading to formation of b- and *y*- ions as primary products. Adapted from[114].

The measured *m/z* of the peaks depends on many factors such as isotopic distribution, number of charges, and the accuracy of the measured *m/z* can be affected by calibration and internal error of the instrument. In addition, a fragment can lose small molecules such as ammonium and water, or carry supplementary molecules such as post-translational modifications.

1.4.4 Analytical modes

In a triple quadrupole instrument (Figure 1.24), (where Q1 is quadrupole 1, Q2 quadrupole 2 and Q3, quadrupole 3) there are many tandem MS scanning modes and *y*-ions are more prominent due to increased stability over b- ions during transmission. The four main modes are described below.

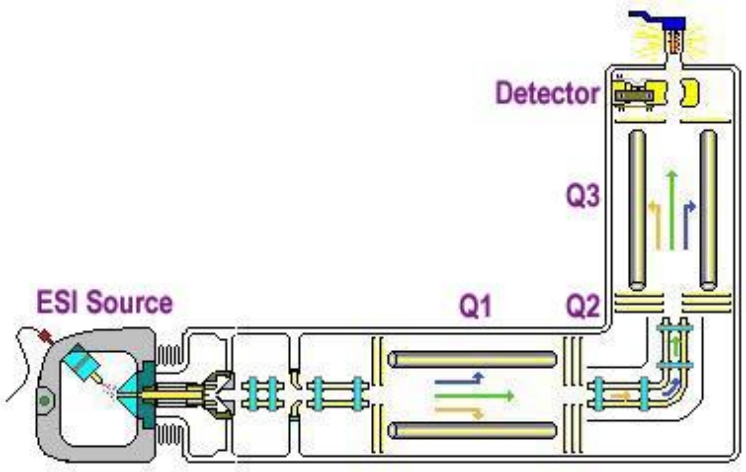


Figure 1.24 A typical triple quadruple mass spectrometer schematic. Reproduced from http://www.epa.gov/esd/chemistry/ice/asms04/fig1.JPG Accessed 21/08/09.

1.4.4.1 Product Ion Scan

During a product ion scan Q1 is set to pass ions of a single *m/z* ratio. These precursor ions are then passed to Q2 where collisions with argon gas cause fragmentation and result in product ions which are then scanned in Q3 to obtain a mass spectrum of the selected precursor ion.

1.4.4.2 Precursor Ion Scan

Q1 is scanned to transmit precursor ions sequentially into the collision cell, Q2, where product ions are formed by CID. These ions then pass to the Q3, which transmits a selected product ion. The precursor ion scan mode spectrum shows all the precursor ions that fragment to produce the selected product ion. Data for the *m/z* ratio axis is obtained from Q1 and data for the intensity axis is obtained from Q3.

1.4.4.3 Neutral Loss Scan

For a neutral loss scan Q1 and Q3 are scanned together at the same rate over the same width of mass range. However there is a constant mass offset between the two so that the product mass analyser scans a selected number of mass units below the precursor mass analyser. Q1 selects precursor ions by *m/z* ratio and these are introduced sequentially into the collision cell where they are fragmented. Then Q3

separates the product ions by their *m/z* ratio. For an ion to be detected, it must lose a neutral mass equal to the difference in mass being scanned by Q1 and Q3 e.g. for loss of water molecule Q3 will scan 18 amu lower than Q1. The spectrum will show all precursor ions that lose a neutral species of a selected mass.

1.4.4.4 Selected Reaction Monitoring

Selected reaction monitoring (SRM) relies on the selection of a precursor and product ion pair referred to as a transition. In a SRM assay, fragmentation of a peptide ion at a specific *m/z,* is used to generate a product ion diagnostic for a particular precursor ion. SRM is frequently used for ascertaining the presence and amount of specific peptides due to the increased specificity and therefore sensitivity that this targeted analysis offers.

Table 1.3 Summary of scan modes in a triple quadrupole mass spectrometer.

Scan Mode	**Q1 Quadrupole**	**Q2 Collision Cell**	**Q3 Quadrupole**
Product ion	Select ions of specific *m/z*	Product ions and then pass all products	Scan
Precursor ion	Scan		Select ions of specific *m/z*
Neutral Loss	Scan		Scan
Selected Reaction Monitoring	Select precursor ions of specific *m/z*		Select product ions of specific *m/z*

1.4.5 Fragmentation Methods

There are a variety of fragmentation methods used in MS however the most commonly used ones are described below.

1.4.5.1 CID

As mentioned above, collision-induced dissociation (CID), which is also known as collisionally activated dissociation (CAD), is a very common fragmentation method used in tandem MS. Molecular ions are collided with inert gas molecules such as

argon, causing the ions to fragment into smaller pieces. When peptides fragment at amide bonds, the resulting 'ladder' of ions can be used to deduce the amino acid sequence.

1.4.5.2 ECD

Electron capture dissociation uses low-energy electrons to fragment molecular ions. ECD results mainly in peptide backbone fragmentation to give c- and z- ions whilst leaving many post translational modifications (PTMs) intact. Therefore it is very good for elucidating PTMs and carried out mainly on Fourier Transform instruments e.g. Fourier-Transform Ion Cyclotron Resonance (FT-ICR). Also it is used for investigating whole proteins or large peptide fragments by 'top-down' mass spectrometry.

1.4.5.3 ETD

Electron transfer is an ion excitation method that uses free radical anions as a source of high energy electrons to fragment molecular ions. Like ECD, ETD results mainly in peptide backbone fragmentation, leaving many post-translational modifications intact, and as such is useful in 'top-down' approaches. Unlike ECD, ETD may be used on more common and cheaper instruments such as quadrupole ion trap instruments.

1.4.6 Data Independent Acquisitions, LC-MS^E

Normally Data Dependent Acquisition (DDA) modes are used during LC-MS/MS experiments where peptides are fragmented serially by MS/MS. The mass spectrometer cycles through a number of “survey scans” where peptide precursor ions are detected and usually the three most intense precursor ions are subject to CID. Unlike DDA, Data-Independent Acquisition (DIA) known as LC-MS^E, utilises parallel fragmentation of precursor ions such that all precursors, regardless of intensity are fragmented simultaneously by acquiring data from alternating scans of low-energy for detecting precursor ions and high- energy for detecting product ions of all precursors.[115,116] The resulting multiplex product ion data undergoes sophisticated processing based on exact chromatographic co-elution of each precursor and its product ions to give “reconstructed” product ion spectra for each precursor.

1.5 Protein Identification and Quantification

Protein identification relies on the presence of DNA and protein sequence databases. The majority of information within these databases has been accumulated within the last 10 years and has therefore perpetuated the growth of proteomics.[117] There are two main strategies by which proteins are identified using MS. The classic proteomics approach involves the separation of the proteins in a mixture by two dimensional gel electrophoresis (2DE), followed by in-gel tryptic digestion to give a distinct set of peptide masses. Analysis of these peptide fragments by MALDI –TOF MS generates a peptide mass fingerprint (PMF) which is searched against a database. There are a number of drawbacks to this approach which can make protein identification difficult. These include limited dynamic range as only the most abundant proteins are observed,[118] the requirement for relatively pure samples since mixtures of proteins will create mixtures of PMFs.[119] For these reasons PMF is used for rapid identification of a single protein component and it is not feasible to apply this technique to complex mixtures without considerable upstream fractionation.

The second major approach for protein identification is by tandem MS. In MS/MS, a peptide ion is isolated in the mass analyser and subjected to dissociation to give product ions. The amino acid sequence of the original precursor ion can be inferred from the masses of these product ions and this forms the basis of *de novo* sequencing though this can be challenging. This technique can be further subdivided into top-down and bottom-up approaches. In the top-down approach, pioneered by McLafferty and his co-workers, an intact protein is presented to the mass spectrometer where it is cleaved in the gas phase rather than solution.[120] This approach relies on high resolution and mass accuracy measurements and has therefore been most frequently applied to FT-ICR instruments.

In the bottom-up approach, complex protein mixtures are enzymatically digested into complex peptide mixtures, fractionated and then analysed by tandem MS.[121] Trypsin is the most popular enzyme used during proteomic studies. The enzyme cleaves specifically at the c-terminal side of the lysine and arginine residues. These peptide fragments are usually of less than 30 amino acids in length making them optimal for MS/MS fragmentation with the current instrumentation. As all tryptic peptide

fragments contain a basic lysine or arginine residue at the C-terminal they readily ionise to form doubly charged molecular ions.

These products ions are then typically subjected to MS/MS, which dissociates peptides in a predictable manner. From the MS/MS spectra peptide sequences can be determined by available search engines e.g., MASCOT[122] and SEQUEST[123] to search against a database containing a theoretical digest and predicted product ion spectra of resulting cleaved peptides. This is made possible since peptides fragment in a predictable manner (as described earlier).

As the bottom-up approach increases sample complexity, the resulting peptides often require a step (or more) of separation before the mass spectrometric analysis to achieve the most sensitive results. These techniques are discussed below.

1.6 Fractionation

Trypsin digestion of proteins results in a complex mixture of many tens of thousands of peptides. To resolve and concentrate this mixture the sample is separated based on the peptides' physical properties of charge and hydrophobicity. Separation is often by liquid chromatography coupled directly to MS. The reduction in complexity of the samples allows for an increase in the sensitivity of peptide identification. In addition, the dynamic range of the sample must be considered as each protein is present at significantly different concentrations. For example a typical cell lysate contains a range of 10^8 protein concentrations and therefore the analysis of the proteome can be difficult with less abundant peptides undetected considering that a typical mass spectrometer has a dynamic range of 10^3.

1.6.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoresis separates proteins by moving them in an electric field. A particular type of electrophoresis used frequently in proteomics is sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE separates proteins by their size (primarily) and charge for fractionation resulting in their visualisation. Excision of the individual protein spots from the gel and in-gel digestion with a protease such as trypsin provides samples suitable for MS analysis. SDS is an anionic detergent that interacts with the non-charged regions of the protein disrupting its hydrophobic core therefore causing denaturation. In addition,

mercaptoethanol or dithiotretiol is used to reduce any disulphide bonds that are present. Therefore differences in electrophoretic behaviour due to tertiary structure are negated and separation is based mainly on their relative molecular weights. The addition of strongly acidic sulfonic acid groups of SDS confers the denatured protein with a negative charge.

Polyacrylamide is a hydrophobic gel prepared by a mixture of N,N'-Methylenebisacrylamide and acrylamide where the polymerisation reaction is initiated with ammonium persulfate and TEMED (N, N, N', N'-tetramethylethylenediamine) is used as an accelerator. These gels have pores where large proteins will move through with more difficulty than smaller proteins. When current is applied to the gel the negatively charged protein-SDS complexes migrate towards the anode and as they pass through the gel the proteins separate owing to the molecular sieving properties of the gel. The smaller the protein, the more easily it can pass through the pores of the gel as the rate of migration depends on the resistive frictional force. The distance travelled in fixed time period is a log function of the molecular weight.[124] Highly charged proteins will be less uniformly bound by SDS and therefore may run at an erroneous molecular weight. Acrylamide concentrations can be varied to achieve different pore sizes and to optimise protein separation.

Polyacrylamide gels may be cast to have either continuous or discontinuous composition, as defined by the types of buffer systems and acrylamide concentrations used. Continuous gels have a constant acyrlamide concentration and the same buffer in both the gel and the tank is used. Although they are easy to prepare the bands are broader and therefore resolution is consequently poorer. Alternatively, Laemmli discontinuous gels are composed of an upper wide pore "stacking gel" layered on top of a lower narrow pore "resolving gel".[125] The stacking gel allows samples to be concentrated before separation in the resolving gel and results in improved band sharpness and resolution. In the Laemmli system different buffers with respect to both pH and ionic strength are used in the gel, and often two different buffers within the gel, as well as a third reservoir buffer. The upper gel of low acrylamide percentage (large pore size) has a pH of 6.8 and the resolving gel of higher acrylamide percentage (smaller pore size) a pH of 8.8. Both gels contain chloride ion as the mobile anion. The reservoir buffer contains glycine as its anion at

a pH of 8.8. When a current is passed the chloride ion migrates faster than the glycine ion and as these two current carrying species separate, a region of low conductivity, with a high voltage gradient is formed between them. This region is known as the Kohlrausch boundary and results in stacking the proteins into a very thin and highly concentrated narrow band.[126] All of the proteins will thus start their separation from approximately the same point. When the Kohlrausch boundary enters the resolving gel, unstacking (resolution) of the proteins occur. The abrupt increase in pH allows the glycine ions to run faster and the smaller pore size allows for the protein molecules. Therefore the boundary is dissipated and a relatively uniform voltage gradient is established and electrophoretic separation of the sample takes place.

SDS-PAGE is a one-dimensional (1D) gel separation whereby proteins are separated based on their molecular mass and then extracted for analysis by tandem mass spectrometry to identify proteins in moderately complex mixtures. To achieve a higher level of resolution and higher loading capacity onto the mass spectrometer then multidimensional separation may be required which uses two or more independent physical properties of the proteins. 2D gel electrophoresis separates proteins by isoelectric point in the first dimension and by molecular weight in the second dimension. There are a number of disadvantages when using gel-based separations. For example they are labour intensive due to the requirement for silver, Coomassie or fluorescent staining to enable visualisation of the protein band. Gel-based separations have a limited dynamic range and under standard conditions 2D separations are unable to detect membrane, highly basic or highly acidic proteins.[127] Another problem is that quantification may be unreliable as the gel spots can contain more than one protein.[128]

1.6.2 Liquid Chromatographic (LC) techniques

To overcome disadvantages outlined above with gel-based separations liquid chromatography (LC)-based separation techniques directly coupled to mass spectrometry were developed. LC-based separations have the advantages of speed, reproducibility and ease of coupling to the mass spectrometer compared to gel based separations. LC uses columns which are packed with support particles, onto which

the stationary phase is coated or chemically bonded or alternatively the stationary phase is located on the inner surface of the tube. The mixture to be analysed is dissolved in the mobile phase and is flowed through the stationary phase. Separation is based on differences in migration rates among the mobile phase components. As the mobile phase exits the column and passes through a detector (or series of detectors) that produce electronic signals which are plotted as function of time, distance or volume. The resulting display is a chromatogram, where eluting solutes are displayed graphically as series of peaks. The retention time is the time when a solute exits the column and passes through the detector.

Single dimensional LC separation of peptides based on their physiochemical properties is performed commonly using reversed phase (RP) or strong cation exchange (SCX). The combination of two different liquid chromatography techniques allows further separation of peptides and is known as multidimensional liquid chromatography.

1.6.2.1 Reversed Phase Chromatography

Reversed phase (RP) chromatography separates molecules based on differences in hydrophobicity. The stationary phase is non-polar and the mobile phase is relatively polar. Commonly alkylsilane groups such as C4, C8, C18 are chemically attached to the surface of silica beads. This creates a hydrophobic (non-polar) stationary phase which can bind polar molecules such as charged peptides. The mobile phase contains polar organic solvents such as methanol, isopropanol and acetonitrile and by increasing their concentration, polar compounds are eluted first while non-polar are retained. The alkyl chains of the column interact more strongly with more hydrophobic peptides and these are thus retained longer. Therefore salt is removed from the charged peptides and these are eluted directly into the ion source of an MS using an increasing gradient of non-polar organic solvent. Alternatively fractions can be collected “off line” and then subsequently analyzed.

1.6.2.2 Ion Exchange

Ion exchange separates analytes in solution based on differences in net charge. An analyte will have a net charge if it is dissolved in a buffer that has a pH either above or below its isolectric point. If the pH is below the pI of the analyte then it will become positively charged and cation exchange chromatography can be used as the analytes are attracted to a negatively charged solid support. Conversely, if the pH is above the pI of the analyte then it will be negatively charged and anion exchange chromatography will be used as the analytes are attracted to a positively charged solid support. Examples of cation exchange particles include sulfonate functional groups and for anion exchange quaternary amines.

The exchange of ions between the charged stationary surface and mobile phase of opposite charge ions is due to variation of mobile phase pH or ionic strength. In pH based ion exchange chromatography the pH of the starting buffer is maintained at a constant level to ensure analytes obtain the opposite charge of the stationary phase and bind to it. Analytes are eluted by altering the buffer pH so that the net charge on the analyte becomes the same as the stationary phase.

Alternatively, ionic strength can be varied using a salt such as sodium chloride to elute bound species. As the salt concentration increases there is competition between the buffer ions and proteins for the charged groups on the column. In the cation application, sodium ions are exchanged for bound protein which is then released. In the anion application, chloride ions are exchanged for bound analyte which is then eluted.

1.6.2.3 Size Exclusion Chromatography

Steric Exclusion Chromatography is also known as gel filtration, molecular-exclusion, or molecular–sieve chromatography, and separates species based on their molecular size. A range of materials are used for the stationary phase such as crosslinked dextran (Sephadex), agarose (Sepharose), polyacrylamide, porus glass and combinations of the above. Sephadex is found in HiTrap™ Desalting Columns which were used in this study to purify QconCAT proteins. The beads of size exclusion packing materials have tightly controlled pore sizes which exclude

molecules above a certain size. Therefore smaller molecules become trapped temporarily in the pores. Molecules larger than the pore size or those with shapes that prevent them from entering the pores remain in the mobile phase and are eluted more quickly from the column.

1.6.2.4 Multidimensional liquid chromatography

An example of a multidimensional protein identification technique is strong cation exchange (SCX) chromatography followed by reversed phase liquid chromatography (RPLC), and is commonly referred to as multidimensional protein identification technology (MudPIT).[129] Thus two orthogonal separations are applied prior to MS using an integrated multidimensional liquid chromatography column. The microcapillary column is packed with a C_{18} resin followed by a SCX resin. Serial separations are made on the basis of charge interactions between an acidified complex peptide mixture and the strong cation exchange column. A salt solution of increasing concentration (step-wise or gradient) elutes the peptides directly onto a RP (reversed phase) column, whereby salt is removed from the peptides and these are eluted directly into the ion source of an MS using an increasing gradient of non-polar organic solvent. If samples contain a high salt concentration they can desalted "off line" before loading onto the biphasic column or directly loaded onto a triphasic column which has a second C_{18} phase upstream from the SCX phase.

1.6.3 Nano-flow liquid chromatography (nano-flow LC)

The advances in nano-electrospray outlined above provided a major problem when coupling the technique to an analytical HPLC system. Analytical HPLC systems have much higher flow rates of the mobile phase, typically around 1 ml/min with the internal diameter of the column at 4.6 mm. To overcome these obstacles capillary high performance chromatography was developed. In comparison to analytical HPLC capillary LC columns are very narrow. Columns with internal diameters of 800 µm, 500 µm, 300 µm and 150 µm are known as microcapillary columns. Those with even smaller internal diameters, typically 100 µm, 75 µm, and 50 µm are termed nanoflow columns and separations are performed using very low flow rates

of between 100-300 nl/min.[130] Nano-LC columns are usually made from fused silica and are packed with a stationary phase beads of diameters between 5 µm or lower such as those used with ultra performance LC (UPLC) columns of 1.5-1.8 µm internal diameter.[131] Such columns require back pressures up to 15,000 p.s.i. to allow the mobile phase to pass through the tightly packed capillary. One such example of a commercially available system is the Waters ACQUITY UPLC system which can withstand pressure up to 15,000 psi and uses columns packed with 1.7-1.8 µm particles.[132] Trap columns are used to desalt the sample to avoid any effect of salts on the ionization of the analytes and to aid concentrating the peptide into a tight band. If peptides were loaded straight onto the analytical column then peak broadening would occur due to the low flow rates these columns can accept.

To obtain flow rates at the level of nanolitre per minute originally splitters were used with HPLC pumps which typically operate at low microlitre per minute flow rate. The splitter device divides the flow into two streams. However, the majority of the flow goes to waste. Splitless nano-LC instruments have now been introduced which use microfluidic flow control where electronic devices monitor parameters such as backpressure and flow rates and adjust them accordingly to maintain precise flow rates.[130]

1.6.4 OFF-GEL Fractionator System

The OFF-GEL fractionator system has recently been made available by Agilent Technologies.[133] Peptides or proteins are separated according to their isoelectric point (pI) using immobilised pH gradient (IPG) strips and a multiwell device. This allows the separated peptides to be recovered in solution at an appropriate volume for further analysis. OFF-GEL can be used as an alternative to SDS-PAGE or can be used as another level of pre-fractionation to observe lower abundance molecules in a complex mixture.

1.6.5 Bioinformatics

Peptide data directed MS/MS generates a huge amount of data, as many thousands of fragmentation spectra can be acquired during a single run. Manual verification of peptide assignments is time consuming, and not feasible for such large data sets. To identify these peptides their spectra are scanned against a protein sequence databases using one of a number of different algorithms.[134-136] The identification of a protein can be achieved if a single peptide uniquely matches to a protein sequence. The most commonly used are MASCOT and SEQUEST. MASCOT adopts a "probability based matching" approach; calculated fragment masses from peptide sequences in the database are compared to the experimentally observed peaks. From this a score is calculated which reflects the statistical significance of the match between the spectrum and the sequences contained in the database. SEQUEST adopts a cross correlation method, in which peptide sequences in the database are used to construct theoretical mass spectra and the degree of overlap or "cross correlation" between the predicted spectra and measured mass spectra determines the best match. For all these approaches, the identified peptides are compiled into a protein "hit-list", which is the output of a typical proteomic experiment.

For most database searching programmes there are parameters that can be adjusted which affect the significance of 'hits' returned. These factors include post-translational modifications, chemical modifications, mass tolerance, and approximations of the proteolytic enzyme. For a database search mass tolerances for precursor ions and fragment ions are inputted. The precursor or peptide mass tolerance setting is based on the mass accuracy of the analyser used to measure the precursor ion masses in the MS scans. Likewise, the fragment mass tolerance is based on the mass accuracy of the analyser to measure the products in the MS/MS scans. Therefore for hybrid instruments different accuracies for MS and MS/MS may be used.[137] The precursor mass tolerance acts a strict peptide filter and setting the parameter either too wide or too narrow compared to the actual mass tolerance of the analyser can have consequences to the database matches. If the precursor mass tolerance is too narrow when applied to the MS data then the correct peptide sequence may be omitted from the comparison to the tandem mass spectra. When the precursor mass tolerance is set to a value much larger than appropriate to the mass tolerance then longer search times can result as more candidate peptides are

analysed. Decreased search sensitivity and increased possibility for a false positive can result as there is a larger pool of candidate peptides that each putative peptide match must now compete against.[138]

For high mass accuracy instruments that are based on Fourier Transform mass spectrometry (FTMS) e.g. LTQ Orbitrap XL used in this study, high precursor accuracy can be achieved. Cottrell warns against the use of small database search when using such data.[137] This is because the combination of low ppm mass tolerance, tryptic cleavage search specificity, and a database limited to proteins from a single taxonomy can provide very specific matches limited to a single candidate peptide. Cottrell states that it is difficult to judge the reliability of such search output and therefore it would be better to search a larger database or use a wider mass tolerance to allow each spectrum to be tested against a range of candidate sequences. This approach was used in this study when searching the LTQ Orbitrap XL mass spectrum as although the reported mass accuracy of the instrument is 5 ppm,[139] during database searching the mass tolerance was opened up to 50 ppm. Post database searching peptide identification was accepted up to 10 ppm.

Modifications may be specified as either fixed or variable. A fixed modification is systematically applied to every occurrence of the specified residue and is equivalent to using a different mass for that modified residue. Variable modifications may or may not be present, such as oxidation which may be an unavoidable feature of sample preparation. For example the experimental data for a peptide that contains 3 methionine residues will be matched to all possible peptides containing 0, 1, 2, or 3 oxidised methionine residues (each + 15.99 amu). In addition selecting a type of instrument will also determine which product ion series is used for scoring. Lastly, proteolytic digestions are often imperfect and do not go to completion. A "missed cleavage" may account for such uncleaved sites and will extend the size of the database searched.

It is important to bear in mind that when inspecting the search output the algorithms used for are not infallible. False positive identifications can occur for several reasons such as the charge states are not always known, MS/MS spectra of poor quality or

mixtures of fragmentations of more than one peptide, scoring schemes in the current database are based on a simplified representation of the peptide ion fragmentation process. In an attempt to reduce the number of incorrect peptide identifications additional processing of the data can take place before database searching. Such processing usually involves removal of low quality spectra by the application of filtering criteria or the use of more advanced scoring systems. Other options include searching against a decoy (non-sense or reverse) database to estimate a false discovery rate.

1.6.6 Relative Quantification

Although MS provides sensitive detection and identification of proteins and peptides, the absolute signal intensity of a peptide ion is not directly quantitative. This is due to the ionization efficiency i.e. the proportion of peptide in solution converted to ions in the gas-phase. Therefore, the ion signals of two different peptides, even if they originate from the same protein compound and are therefore equimolar, cannot be compared to each other. Further, even the signal intensities of the same peptide ion obtained from two independent experiments will not produce exactly the same signal when analysed, owing to differences in co-eluting peptides, solvent composition, and the age and position of the electrospray needle. To circumvent this problem, it is essential to use an internal standard which normalises quantitative variations among different MS measurements. An ideal internal standard is one that is chemically and physically as similar as possible to the peptide being analysed. An established technique is the use of stable isotopes e.g. ^{13}C, ^{15}N, which differs in mass from naturally abundant isotopes e.g. ^{12}C, ^{14}N and are used to modify the peptide populations that are being compared.[140] These peptides are then pooled and analysed in a single MS run. Labelling with chemically identical tags results in a pair of peptides with the same chemical and physical features, therefore co eluting and with the same ionisation efficiency. Relative quantification and by extension absolute quantification is achieved by comparing the ratio of signal intensities of peptides with a known, isotope induced mass difference. An important aspect of all stable isotope labelling procedures is that the efficiency must be as close to 100 % as possible as under labelling leads to the generation of background noise. It is possible

to introduce labels at various levels including the metabolic, protein and peptide levels (Figure 1.25).

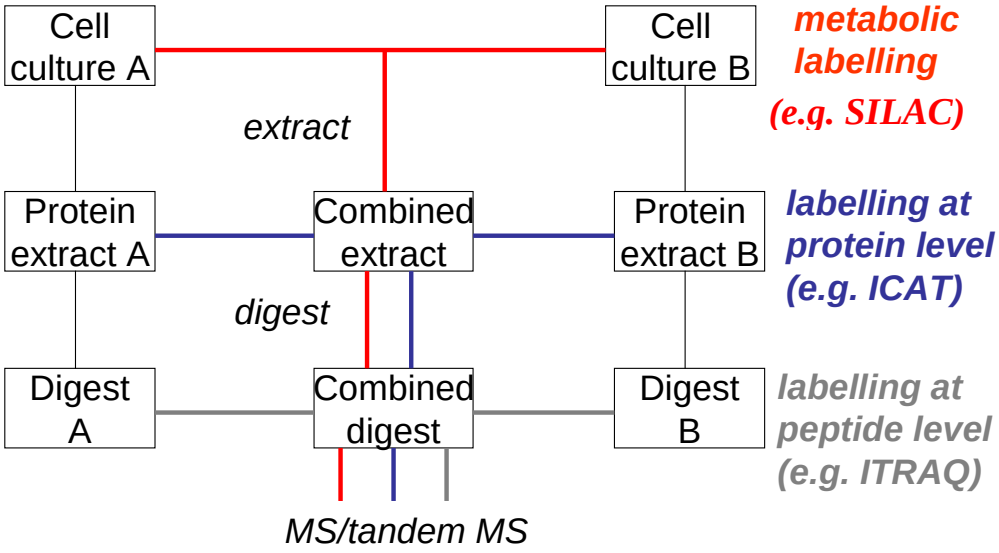


Figure 1.25 Workflow for relative quantification labelling using a variety of labelling methods.

1.6.7 Metabolic Labelling

Metabolic labelling involves incorporating stable isotopes *in vivo* into the proteins of cells grown in special medial containing these isotopes. Early efforts reported cell cultures of yeast and bacteria grown in media containing either ^{14}N (light) or ^{15}N (heavy) isotopes.[141-144] Following isotopic labelling, the two cell cultures were combined, processed and analysed by MS. There is a mass shift depending on the number of nitrogen atoms contained within the peptide. Since the method labels both the backbone and side-chain nitrogen atoms, the mass shift cannot be predicted for an unknown peptide and it is difficult to identify the labelled peptides in a complex spectrum unless very high accuracy mass spectrometer is used e.g. Fourier Transform Ion Cyclotron Resonance (FT -ICR).

An alternative approach is to use amino acids that contain a stable isotope, termed "stable isotope labelling with amino acids in cell culture (SILAC)". Proteins are labelled by growing cells in media containing isotopically labelled amino acids e.g. ^{13}C-lysine.[145] This method has the advantage of predictability of the mass shift which can be analysed by conventional mass spectrometers.

In general, stable isotope labelling *in vivo* has proved to be an effective method of quantitative proteomic analysis. As stable isotopes are incorporated into the early stages of sample preparation, reduction in variation between samples yields highly accurate quantification. There are number of disadvantages which include lack of application to biological samples which cannot be grown in cultures such as tissues and body fluids and the relatively long labelling incubation time in cell culture of five population doublings.[146] However, the recent approach of SuperSILAC circumvents this problem.[147]

1.6.8 Labelling at the Peptide and Protein Level

Labelling involves incorporation of stable isotopes *in vitro* at selective sites on peptides after harvesting from cell lines or tissues. These methods involve isotopic labelling of target peptides at their amino-(N-) or carboxyl- (C-) terminal or on specific amino acid residues, such as cysteine, lysine, tyrosine.

ICAT

Gygi and co-workers developed a method based on isotope-coded affinity tags (ICAT) and ESI MS.[148] The ICAT reagent consists of a biotin affinity tag for selective purification, a linker that incorporates a stable isotope like ^{13}C as the heavy isotope tag and a reactive iodoacetamide group that alkylates cysteine residues. Proteins from two samples are labelled with either light or heavy ICAT reagents, combined, digested with trypsin, and passed over an avidin affinity column. Isolation of only cysteine containing peptides results, and therefore the sample complexity is reduced allowing for increased selectivity. This method has been improved and made commercially available by Applied Biosystems. The heavy tag now contains ^{13}C instead of deuterium and each tag contains a cleavable linker which allows removal of the affinity tag before MS analysis thus simplifying the analysis and increasing the number of proteins identified and quantified. There are several additional advantages to the ICAT strategy. As the proteins are pooled, pre-fractionated and digested together, potential experimental variation by carrying out these steps separately is therefore reduced. A major disadvantage is that the procedure is limited to quantifying proteins that contain cysteines. However, only a small proportion of

proteins are cysteine free (8 % in yeast) and ICAT reagents with specificities to groups other than thiol could be synthesised.

ITRAQ

ITRAQ, an amine group based isotope labelling methodology, termed isobaric tags for relative and absolute quantification (iTRAQ) was developed and has since been made available commercially by Applied Biosystems (Figure 1.26).[149] This method can incorporate up to eight mass tags. Each tag consists of a reporter group, a balance group and a peptide reactive group. The reactive group labels the primary amine groups and free amines in lysine side chains of peptides. The reporter group is a tag with a mass of either 114, 115, 116 or 117 Da, and is balanced by a mass of from 28-31 Da to ensure that the combined mass of the reporter and balance groups remains constant at 145 Da for all reagents. As the tags are isobaric, when the four labelled samples are pooled the same peptides from each sample appear at the same mass in the MS spectrum. Thus the signal from each sample is summed in the MS mode and sensitivity is increased. On fragmentation each tag generates a specific distinct reporter ion at mass to charge ratio 114, 115, 116, or 117 and the relative intensity of these reporter ions provides the basis for relative quantification.

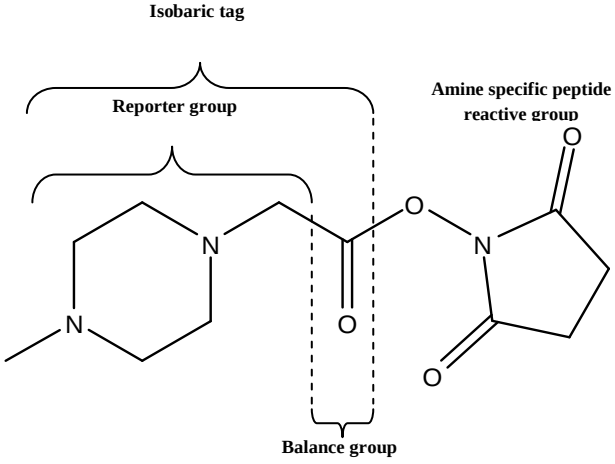


Figure 1.26 Structure of the iTRAQ reporter reagent. Key features are reporter group, mass range from 114 to 117 Da, a neutral balance group, mass range from 31 to 28 Da and an amine-specific peptide-reactive group.

The technology enables up to eight different samples within a single experiment to be combined, and is a useful strategy to quantify proteins of the same sample or

simultaneous comparison of normal and diseased states. Secondly, because all peptides are labelled, it is possible to obtain data on more peptides when compared to other methods, and this increases confidence in the quantification of each protein.[150]

1.6.9 Absolute Quantification

There are a few methods available for the absolute quantification of proteins and these include AQUA and QconCAT.

AQUA

AQUA, Absolute Quantification, uses a stable isotope labelled synthetic peptide in a similar approach to the stable isotope methods used for relative quantification.[151] This peptide is spiked into the analyte mixture at known concentration, and the ratio of unknown peptide to "heavy" labelled standard is used to calculate the concentration of the protein. There are several problems with this method, the first is that it is expensive and time consuming to chemically synthesise an internal standard for each peptide of interest. Secondly, the chemical synthesis of reference peptides are not always possible.[152] A significant advancement on this technique has been the development of QconCAT technology.

QconCAT

QconCAT uses a synthetic gene often commercially 'designed to order' (Polyquant, Germany) through a series of overlapping primers and cloned into an expression plasmid which when transformed into an appropriate host will encode a recombinant protein, the sequence of which is a concatamer of several tryptic peptides from proteins of interest (Figure 1.27).[153] Proteotypic peptides, peptides which are repeatedly and consistently identified from a protein in a complex mixture and used for quantitative analysis, specific for the target proteins are selected, and unless specifically required, peptides containing post-translational modifications are avoided. One or more Q-peptides per analyte protein, which is mass tagged and used for quantification, are chosen *in silico* based on various selection criteria. The cDNA for multiple Q-peptides are concatenated into a synthetic gene construct and

expressed in *E. coli* as an artificial QconCAT protein. Growth of the transfected *E. coli* in stable isotope containing media, either [^{15}N] or [$^{13}C_6$] lysine and [$^{13}C_6$] arginine, yields labelled QconCAT protein, which is then purified and quantified using standard assays. The labelled QconCAT is added to a complex biological sample in known amounts and co-digested with native proteins to give Q-peptide and native peptides. Absolute quantification of protein level is achieved by calculating the light to heavy ratio of native peptide to internal reference peptide. The QconCAT protein provides a relatively easy construction of a large set of mass tag peptides therefore enabling the absolute quantification of a large number of proteins within a biological system.

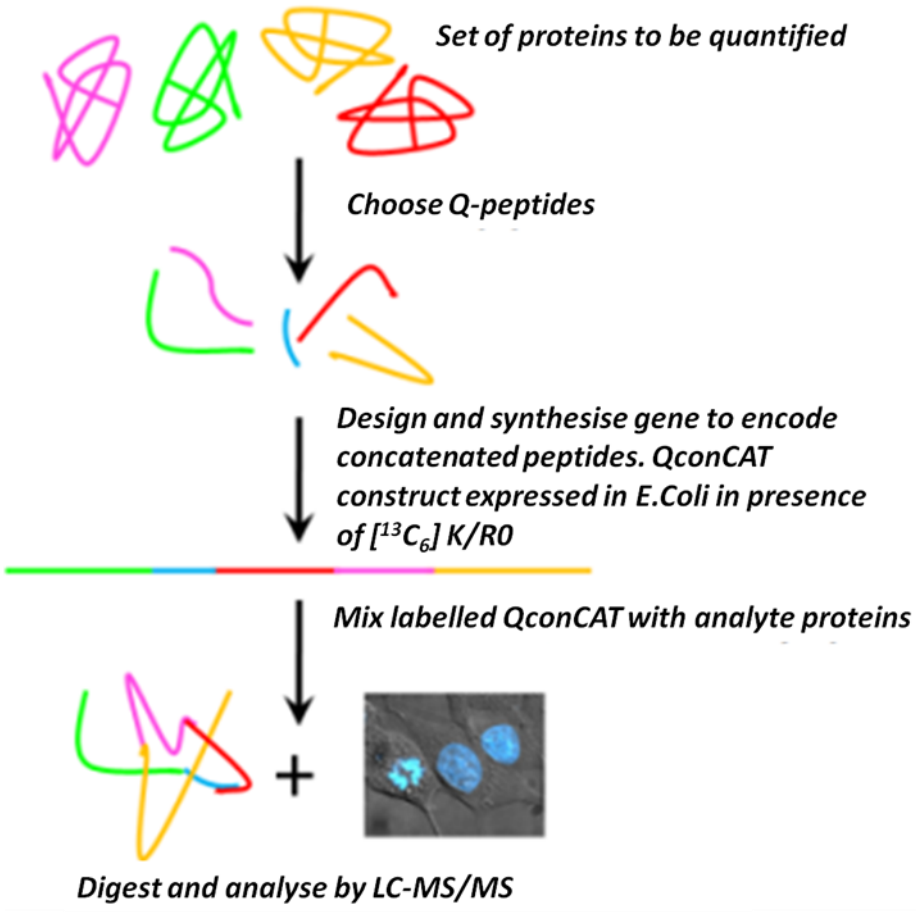


Figure 1.27 General principle of QconCAT quantification experiment. Adapted from[153].

1.6.10 Label-Free Quantification

Label-free quantification approaches are alternative methods to stable isotope labelling and are based on spectral counting. At first glance these methods look simple and cost effective, though such methods require validation. Pattern reproducibility within and between experimental runs becomes a major challenge as both methods rely heavily on the depth of MS/MS sampling.[154] As mentioned previously the MS^E method is based on the top three intensities of the most abundant peptides correlating to the amount of protein.[155] Protein ratios estimated from spectral counts are most significant for proteins with large numbers of spectra, and quantification is relative to an included protein standard of known concentration digested with the sample. Protein ratios estimated from peak intensities are limited by abundant proteins with high sequence coverage. It has been suggested that spectral counting and peak area intensity methods have lower sensitivity than stable isotope labelling.[156] In addition, a recent study by Carroll *et al* compared QconCAT quantitative data with that of label-free which suggests that there is general underestimation of protein abundance by label-free methods.[157]

1.7 Summary

The MAPK pathway plays a crucial role in regulating the cellular response to external stimuli. Signalling through this pathway is regulated by the coordinated function of specific protein kinases and protein phosphatases. As perturbation of this signalling pathway is often associated with diseases such as cancer, modelling is a useful means to help understand the outcomes that may result following changes in component levels. The aims of this study are to expand our current models of the MAPK pathway by providing quantitative information on total protein components, in particular DUSPs, scaffolds and substrates. QconCATs will be implemented to generate data from human colon cancer cell lines HT-29 (mutated Raf) and HCT 116 (mutated Ras) with human embryonic kidney cells, HEK-293, acting as a control. Although MAPK signal transduction has been subject to mathematical modelling for decades, these models have relied heavily on relative quantification data obtained using Western-blotting methods. Instead, absolute quantification data can be used to build mathematical models with improved accuracy, therefore providing a more realistic basis to represent and investigate biological behaviour.

2 Materials and Methods

2.1 Materials

All chemicals and HPLC grade solvents were acquired from Sigma Aldrich (Poole, UK) apart from the following: [$^{13}C_6$] arginine (R) and [$^{13}C_6$] lysine (K) 98 % purity, were supplied from Cambridge Isotope Laboratories (Andover, MA, USA). QconCAT constructs were supplied from PolyQuant (previously Entelechon, Regensburg, Germany).

2.2 QconCAT Construction

The chosen Q-peptide sequences were concatenated *in silico* and used to direct the design of a gene, codon-optimized for expression in *E. coli*. Additional sequences were added to provide a His_6 sequence (ALVALVHHHHHH) for affinity purification and a sacrificial peptide at the N-terminus and control peptide, GVNDNEEGFFSAR for QconCAT quantification. The gene was synthesized and cloned into the expression vector pET21a by PolyQuant GmbH (http://www.polyquant.com/).

2.3 Transformation

7 µg of QconCAT plasmids LM1 and LM2 were each dissolved in 100 µl HPLC grade water, and transformation was carried out according to manufacturer's standard protocol (Stratagene, Stockport, UK), with water control. Briefly, frozen competent cells were thawed on ice and gently mixed. For each transformation 20 µl of competent cells were aliquoted into pre chilled tubes. 1µl of plasmid DNA was added directly to the cells which were then incubated on ice for 5 min. The cells were then subjected to heat shock for 30 sec at 42 ^{o}C and placed on ice for 2 min. 80 µl of SOC medium (supplied by Stratagene) was added to each tube. 20 µl – 80 µl of cells were then spread onto LB agar plates containing the appropriate antibiotic for the plasmid and host strain. The plates were then incubated overnight at 37 ^{o}C.

2.4 Expression of QconCAT Constructs

This was carried out according to the protocol by Beynon and co workers.[153] A single colony of BL21-pQconCAT was used to inoculate 10 ml of LB containing ampicillin (50 µg/ml) to select the plasmid which carries the Amp^R gene, and incubated overnight at 37 °C with shaking at 200 rpm. The following morning 500 µl of the overnight culture was transferred to 50 ml of pre-warmed fresh LB medium

containing ampicillin (50 µg/ml) and incubated at 37 °C with shaking at 200 rpm. 1 ml samples were taken at hourly intervals and the absorbance at 600 nm (A_{600}) was monitored using a spectrophotometer using LB-Amp as a blank. When an A_{600} of 0.6-0.8 was reached 50 µl of 1 M IPTG was added to induce expression of the QconCAT protein. 1 ml sample was removed and A_{600} was measured immediately at time 0 and then every hour (up to 4 h). The samples were centrifuged at 8000 x *g* at 4 °C for 10 min. The supernatant was then removed and discarded and the cell pellet stored at -80 °C for later analysis by, sodium dodecyl sulfate polyacrylamide gel electrophoresis, SDS-PAGE.

After 4 hours the culture was transferred into a 50 ml centrifuge tube and centrifuged at 8000 x *g* for 10 min at 4 °C. The supernatant was decanted and cell pellet frozen at -20 °C until required for purification after confirmation of QconCAT expression by SDS-PAGE.

For isotopically labeled QconCAT the *E. coli* strain BL21(Δ)DE3 and BL21(Δ)pLysS cells were transformed with LM2 QconCAT pET21a and LM1 QconCAT pET21a respectively. These were cultured in minimal media (1 x M9 salts, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.00005 % (w/v) thiamine, 0.2 % (w/v) glucose, and unlabelled amino acids at 0.1 mg/ml), supplemented with labelled [$^{13}C_6$] arginine and [$^{13}C_6$] lysine at 0.1 mg/ml. Cells were grown to mid log phase (OD_{600}=0.6-0.8) at which point expression was induced by IPTG at 1 mM for 4 h. Cells were lysed with BugBuster Protein Extraction Reagent (Merck Chemicals, Nottingham, UK) and the artificial proteins were purified.

2.5 Purification

LM2 QconCAT was purified using AKTA explorer purification system (GE Healthcare, Little Chalfont, UK). SDS-PAGE confirmed that the protein of interest was present in inclusion bodies, which were solubilised in 6 M guanidinium chloride. 10 ml of soluble fraction in binding buffer (20 mM phosphate buffer, pH 7.4, 20 mM imidazole, 0.5 M NaCl, 6 M Guanidinium chloride) was uploaded on a 1 ml HisTrap™ Fast Flow column (GE Healthcare, Little Chalfont, UK) at flow rate of 0.2 ml/min. Unbound proteins were removed by extensive wash with 20 ml of binding buffer at flow rate of 1 ml/min. His tagged QconCAT eluted as a single peak in 2 ml fraction with elution buffer (50 mM phosphate buffer, pH 7.4, 500 mM

sodium chloride, 500 mM imidazole and 6 M guanidinium chloride). QconCAT was then desalted using two 5 ml HiTrap^{TM} Desalting Columns (GE Healthcare, Little Chalfont, UK) and desalting buffer (10mM ammonium bicarbonate, pH 8.5, 1 mM DTT). At each step of purification small fractions of sample were taken for SDS-PAGE analysis. Purified protein was stored at -80 °C.

The cell pellet of LM1 was resuspended in 2 ml of denaturing wash buffer, 50 mM sodium phosphate, 300 mM NaCl, pH 7.0 and agitated until it became translucent. The sample was centrifuged at 12000 x g for 20 min at 4 °C to pellet any insoluble material and the supernatant transferred to a new tube without disturbing the cell pellet. The supernatant is known as the clarified sample and a 20 µl sample was retained for SDS-PAGE analysis. TALON Resin (Clontech, Saint-Germain-en-Laye, France) was thoroughly resuspended in 20 % ethanol and 5 ml, equivalent to 10 times the required resin bed volume, was transferred to a Generon (Maidenhead, UK) Proteus "1 step batch" spin column. To pellet the resin the column was centrifuged at 700 x g for 2 min at 4 °C. The resin was washed by addition of 5 ml of 50 mM sodium phosphate, 300 mM NaCl, pH 7.0 and centrifuged at 700 x *g* for 2 min at 4 °C. The resin was washed again and the clarified sample added. The resin was resuspended and left on a shaker platform for 20 min at 4 °C with occasional inversion. The column was then centrifuged at 700 x *g* for 2 min at 4 °C and 50 µl of non-adsorbed material collected for analysis. 5 ml of 50 mM sodium phosphate, 300 mM NaCl, pH 7.0 was added and resin resuspended before centrifugation at 700 x *g* for 2 min at 4 °C. This wash step was repeated a twice more and 50 µl of each wash retained for analysis. For improved protein purity the resin was then washed with 5 ml of 50 mM sodium phosphate, 300 mM NaCl, 7.5 mM imidazole, pH 7.0 and resin resuspended before centrifugation at 700 x *g* for 2 min at 4 °C. This wash step was repeated twice more and 50 µl of each wash retained for analysis. The protein was eluted from the column by addition of 1 ml of 50 mM sodium phosphate, 300 mM NaCl, 150 mM imidazole, pH 7.0 followed by inversion for 2 min and centrifugation at 700 x *g* for 2 min at 4 °C. This was repeated a further four times and 50 µl of each retained for analysis.

2.6 Removal of Guanidinium Chloride

To Starting Material (SM), and Elution Bound Material (EBM) 10 ul of StrataClean Resin (Stratagene, Stockport, UK) bead suspension was added to each tube. StrataClean Resin removes the guanidinium chloride as this interferes with SDS-PAGE. Each sample was then vortexed for 1 min, centrifuged for 2 min at 230 x *g* at RT, and then supernatant was removed and discarded. Pellets were resuspended in 1 ml of distilled H_2O, vortexed and centrifuged at 230 x *g* for 2 min, and then supernatant was removed and discarded. 10 µl of 2 x SDS loading buffer containing DTT was added and the samples boiled at 96 °C for 4 min. Sample were resolved by SDS-PAGE under reduced conditions using 4-20 % precast gels (Expedeon, Cambridge, UK) or hand cast gels.

2.7 SDS-PAGE and Coomassie Staining

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using either precast mini gels (Expedeon, Cambridge UK) or those prepared between two glass plates. Firstly, a separating gel was poured, 10 % (w/v) acrylamide/ 0.4% (w/v) N, N'-methylene bisacrylamide, 400 mM Tris-HCl (pH 8.8), 0.1 % (w/v) SDS, 0.1 % (v/v) ammonium persulfate (w/v), 0.04 % (w/v) N,N,N,N-tetramethyl-ethylenediamine (TEMED). Next isopropanol (IPA) was layered over the gel solution to get an even surface and polymerisation proceeded for at least 15 min. The IPA was removed and the stacking gel, 5 % (w/v) acrylamide/ 0.4 % (w/v) N, N'-methylene bisacrylamide, 400 mM Tris-HCl (pH 6.8), 0.1 (w/v) % SDS, 0.1 (w/v) % ammonium persulfate, was poured over the separating gel. A comb was added to generate well for sample loading and gel left to set for at least 10 min.

Protein samples were mixed with loading buffer and denatured by heating for 5 min at 95 ºC. Proteins were resolved (running buffer: 25 mM Tris, 192 mM glycine (pH 8.3) and 0.1 % (w/v) SDS) at 120 V until the dye front reached the end of the gel. Prestained molecular weight markers were run on every gel (PageRuler, Fermentas). Following electrophoresis, gels were stained with Expedeon Blue (Expedeon, Cambridge, UK) and washed in water to visualise the protein bands.

2.8 In-gel Tryptic Digestion

The gel was placed on a clean glass plate and using a clean scalpel the gel band corresponding to the QconCAT protein was excised from the gel and transferred to a 1.5 ml microcentrifuge tube. To the sample 25 µl of 25 mM ammonium bicarbonate was added, after which the liquid was discarded. Next 25 µl of 25 mM ammonium bicarbonate- MeCN (2:1) was added, the sample was incubated at 37 °C for 15 min, and liquid discarded. These steps were repeated until the gel band was fully destained. As cysteine residues were present the protein was reduced and alkylated prior to trypsin digest. To the sample 25 µl of 10 mM dithiothreitol (DTT) in 25mM ammonium bicarbonate was added, then incubated at 56 °C for 60 min and liquid discarded. Next, 25 µl 55 mM iodoacetamide (IAM) in 25 mM ammonium bicarbonate was added, then sample was incubated in the dark at 37 °C for 45 min and liquid discarded. To the sample 25 µl of 25 mM ammonium bicarbonate was added, then incubated at 37 °C for 15 min, and liquid discarded. Next 25 µl of 25 mM ammonium bicarbonate- MeCN (2:1) was added, the sample was incubated at 37 °C for 15 min, and then liquid discarded. 10 µl of 25 mM ammonium bicarbonate containing 0.1 mg/ ml trypsin was added and samples incubated at 37 °C overnight.

2.9 In -Solution Digest

To a digest volume of 50 µl, 2 µl of 100 mM DTT was added to give a final DTT concentration of 4 mM. The sample was then incubated at 60 °C for at least 15 min to reduce oxidized cysteines (disulphide bonded cysteines). The sample was then cooled and to akylate the free sulfhydryl groups of the reduced cysteine residues 7.3 µl of 100 mM iodoacetamide was added to give a final concentration of 14 mM. The sample was incubated at room temperature in the dark for at least 40 min and then 2 µl of 100 mM DTT was added to quench the alkylation reaction. Trypsin was added at a ratio of 1:50 enzyme:protein and incubated at 37 °C overnight. The digest was quenched with 0.1 % (v/v) formic acid (FA) and stored at -20 °C prior to mass spectrometric analysis.

2.10 Lysis of Cell Pellets

Cell pellets in phosphate-buffered Saline (PBS) were received from the laboratory of our collaborator in Berlin, Dr. Nils Blüthgen and used for the in-solution RapiGest protocols (both non modified and modified for final experiments). For clarity all cell line names are as described in the American Tissue Culture Collection (Virginia,

US) catalogue. These were HCT 116 (K-Ras, mutant G13D) biological replicate 1, 11.5 x 10^8 cells; HCT 116 biological replicate 2, 18 x 10^8 cells; HT-29 biological replicate 1, 14 x 10^8 cells; HT-29 (B-Raf, mutant V600E) biological replicate 2, 13 x 10^8 cells; HEK-293 biological replicate 1, 17.5 x 10^7 cells; HEK-293 biological replicate 2, 9.5 x 10^7 cells. These were resuspended in 25 mM ammonium bicarbonate and then sonicated on ice for three 10 s pulses at 30 % amplitude at minute intervals. The lysates were then centrifuged at 17 000 x g for 1 h and the supernatant retained. The protein concentration of the supernatant was assessed by Bradford Assay.

2.11 Bradford Assay

Bradford reagent (Pierce, Rockford, USA) 990 µl was added to 10 µl of the protein sample (diluted as required) and the mixture vortexed briefly prior to leaving at room temperature for 5 min. The absorbance was measured in a plastic cuvette at 595 nm using a vis-spectrophotometer (Jenway, Stone, UK), against a blank of protein dilution buffer. The protein concentration was calculated with reference to a standard curve constructed by diluting Bovine Serum Albumin (BSA) in water.

2.12 In-solution digest with RapiGest

100 µg of protein was made up to a volume of 160 µl with 25 mM ammonium bicarbonate in low bind tubes. 10 µl of 1 % (w/v) RapiGest (Waters, Elstree, UK) was added and the sample heated at 80 °C for 10 min. After 5 min the sample was briefly vortexed and then pulse spun to return liquid to bottom of tube. 10 µl of DTT (9.2 mg/ml in 25 mM ammonium bicarbonate) was added, vortex mixed and heated at 60 °C for 10 min. The sample was then cooled to RT and pulse spun to return liquid to bottom of tube. 10 µl of iodoacetamide (33 mg/ml in 25 mM ammonium bicarbonate) was then added, vortexed and incubated at RT in the dark for 30 min. 10 µl of trypsin (0.2 µg/µl in 50 mM acetic acid) was added (protein:enzyme, 50:1) at and placed in heating block t at 37 °C. A second spike of trypsin was added 4 h later and the sample left overnight at 37 °C. The final digest volume was 200 µl. The following day, 1 µl of trifluoroacetic acid (TFA) was added, mixed and 1 µl of digest spotted onto pH paper to ensure the digest was acidic at approx pH 6. The digest was then incubated at 37 °C for 45 min. The digest was centrifuged for 15 min at 14 000 x *g* and white pellet was obtained.

2.13 In-solution digest with RapiGest (Modified)

To 100 µg of cell lysate and either 1.59 µg QconCAT LM1 or 0.159 µg QconCAT LM1 or 1.83 µg QconCAT LM2 or 0.183 µg QconCAT LM2 was added and the volume of each sample made up to 341 µl with 25 mM ambic/250 mM NaCl. Each tube was then vortexed and liquid spun to the bottom of the tube. To each sample 10 µl of 1 % (w/v) RapiGest was added and vortexed. The liquid was then spun down to the bottom of the tube and incubated at 80 °C with shaking at 400 rpm for 10 min. The liquid was again spun to the bottom of the tube. 12 µl of 100 mM DTT (15.4 mg/ ml) was added and each tube was vortexed and liquid spun to the bottom of the tubes prior to incubation at 60 °C with shaking at 400 rpm for 10 min. The samples were then cooled to room temperature and 12 µl of 300 mM IAM was added. The samples were then vortexed and liquid spun down to the bottom of the tube before incubating at room temperature in the dark for 30 min. 10 µl of 0.2 µg/µl trypsin was added to each sample and tubes were vortexed, liquid spun to the bottom and then incubated at 37 °C with shaking at 400 rpm for 4.5 h. The condensate was spun down and 5 µl of 0.1 M HCl was added. The samples were vortexed and liquid spun to the bottom of the tube. A second addition of 10 µl of 0.2 µg/µl trypsin was added and vortexed. The liquid was spun down to the bottom of the tube and incubated at 37 °C with shaking at 400 rpm overnight. The condensate was spun down and 9 µl of acetonitrile (2 % final concentration) was added. The samples were vortexed, liquid spun to the bottom of the tubes and then 1 µl of TFA (0.4 % final concentration) was added. The tubes were vortexed, liquid spun to the bottom of the tubes and incubated at 37 °C with shaking at 400 rpm for 2 h. Afterwards the condensate was spun down and incubated at 4 °C for at least 2 h which results in a white precipitate. The samples were then centrifuged at 16 000 x *g* at 4 °C for 15 min. The supernatant was removed using a gel loading pipette tip to avoid removal of the precipitant. Finally the supernatant was removed to a fresh tube.

2.14 Preparation of Cell Lysates for Filter-Aided Sample Preparation (FASP) Method

Cell lysates were received from our collaborators, Nils Blüthgen's laboratory in Berlin. Samples received were 15. 23 mg/ml of HT-29 (B-Raf, mutant V600E) colon cancer cells, 15.83 mg/ml of HCT 116 (K-Ras, mutant G13D) colon cancer cells and 8.17 mg/ml of U-2 OS osteosarcoma cells to act as a control. They were prepared by

the method detailed by Hanke and co workers.[158] Briefly, cell suspensions were aliquoted to microcentrifuge tubes and centrifuged at 400 x *g*. The cell pellet was then washed twice with ice-cold PBS and either stored at −80 °C or lysed directly in 1 % (w/v) Igepal (NP-40), 150 mM NaCl, 50 mM Tris, pH 7.5, and protease inhibitor cocktail Roche complete was added. After repeated vortexing and incubation on ice for 15 min, several freeze–thaw cycles were conducted until no intact cells could be observed by light microscope. Following centrifugation at 10 000 x *g* for 15 min at 4 °C, the supernatant was removed and the pellet washed three times with cold PBS.

2.15 Filter-Aided Sample Preparation (FASP) Method.

The method reported by Mann and colleagues was adapted to suit our requirements.[159] Aliquots of cell lysate and purified QconCAT protein corresponding to 100 µg and 3 µg respectively were added to 200 µl of 8 M urea in 0.1 M Tris/HCl, pH 8.5, in an Amicon Ultra 10 K membrane centrifugal unit (Millipore, Cork, Ireland). The unit was centrifuged at 14 000 x *g* for 40 min at room temperature and the flow through was discarded. 100 µl of 0.05 M iodoacetamide in 8M urea in 0.1 M Tris/HCl, pH 8.5, was added and the solution was mixed at 600 rpm in a thermo mixer for 1 min and then left to incubate in the dark for 20 min. The filter units were then centrifuged at 14 000 x *g* for 30 min. 100 µl of 8 M urea in 0.1 M Tris/HCl, pH 8.5, was added and the unit centrifuged at 14 000 x *g* for 40 min and this step was then repeated. 100 µl of 0.05 M ammonium bicarbonate in water was added and the unit centrifuged at 14 000 x *g* for 30 min. This step was repeated. An aliquot of trypsin corresponding to an enzyme to protein ratio of 1:100 was added and the solution mixed at 600 rpm in a thermo mixer for 1 min. Next the sample was incubated at 37 °C for 14 h. The filter units were then transferred to new collection tubes and the units centrifuged at 14 000 x *g* for 30 min. The units were then rinsed with 50 µl of 0.5 M NaCl and centrifuged again at 14 000 x *g* for 30 min. Lastly, the sample was acidified with 1 µl of 0.1 % FA and desalted.

2.16 Peptide Desalting

Samples containing high concentrations of salt require desalting prior to MS analysis to decrease contamination aid ionisation. Samples were desalted using C_{18} ZipTips^{TM} (Millipore, Cork, Ireland) prior to Matrix-Assisted Laser Desorption/Ionisation Time of Flight (MALDI-TOF) analysis. Initially the new tip was wetted by making 4

passes of 10 µl of methanol, where one pass is defined as a complete aspiration and dispensing. Then 4 passes of 10 µl of 1.0 % (v/v) TFA, the binding solution, was made. Next the sample was bound to the resin by making 20 passes of 10 ul sample. The sample was then washed with 4 passes of 10 µl of 0.1 % (v/v) formic acid to wash adsorbed salts. Lastly, the peptides were eluted by making 10 passes of 5 µl of 60 % (v/v) acetonitrile 0.1 % (v/v) FA.

Alternatively, to desalt large volumes a C_{18} macro trap (capacity 200 µg) was used. The trap was wetted with 300 µl of 80 % (v/v) MeCN, then cleaned with 300 µl of 0.1 % (v/v) TFA, 80 % (v/v) MeCN and then equilibrated with 500 µl of 0.1 % (v/v) TFA prior to loading the sample. A solution of 50 µl of 0.1 % (v/v) TFA and 90 µl of sample was added to the trap. The salt was washed from the column using 300 µl of 0.1 % (v/v) TFA. Lastly, 100 µl of 80 % (v/v) MeCN, 0.1 % (v/v) FA was added to elute the peptides.

2.17 LC-ESI-MS/MS analysis using a Q-TOF Ultima Global instrument

The digested peptide mixtures were diluted in 97 % (v/v) H_2O, 3 % (v/v) MeCN / 0.1 % (v/v) FA and injected into a Dionex (Surrey, UK) Ultimate 3000 capillary LC system via a FAMOS autosampler. At a flow rate of 200 nl/min, column pressure was 90 bar, master pressure was 140 bar, loading pressure was 58 bar. The eluent was sprayed into from a distal-coated fused silica PicoTip Emitter (New Objective, MA, USA) using a capillary voltage of 2.2 to 2.8 kV into the Z-Spray nano-electrospray source of the Q-TOF Ultima Global (Waters, Elstree, UK). Cone voltage was set to 100 V. Buffer A consisted of 2 % (v/v) MeCN , 0.1 % (v/v) FA, 98 % (v/v) H_2O and buffer B consisted of 100 % (v/v) MeCN and 0.1 % (v/v) FA. The gradient conditions were as follows: 3.5 min buffer B 2 %, 7.0 min / 10 % B, 41 min 40 % B, 47 min 95 %, 51.5 min/ 2 %.

Full scan MS spectra were acquired over the range of *m/z* 200-2000 in a data dependent mode with the top 3 most intense ions directing collision induced dissociation (CID).The three most abundant precursor were selected to fragment for 3 s. Collision energy offset values were entered for ions based on their *m/z* and charge and were typically between 25 and 40V. The MS/MS spectra were acquired over the range of *m/z* 50 to 2000.

2.18 LC-ESI-MS/MS analysis using a triple quadrupole Xevo TQ MS instrument

The digested peptide mixtures were diluted in 97 % (v/v) MeCN/0.1 % (v/v) FA (mobile phase A) and resolved by LCMS using a nanoACQUITY UPLC™ (Waters, Elstree, UK) coupled to a Xevo TQ MS (Waters, Elstree, UK). 1 µl of diluted sample was injected onto a trapping column (Waters C_{18} 180 µm X 20 mm), at a flow rate of 15 µl/min 99 % (v/v) A/1 % (v/v) B (where B consisted of 100 % (v/v) MeCN/0.1 % (v/v) FA), then after loading an analytical column (nanoACQUITY UPLC™ BEH C_{18} 75 µm x 150 mm, 1.7 µm column). The column temperature was set to 35 ºC, and a flow rate of 300 nl/min was maintained. The gradient conditions were as follows: 0.10 min 1 % B, 60 min, 40 % B, 62 min, 95 % B, 65 min, 1 % B, 90 min, 1 % B. The instrument was calibrated using Sodium-Caesium iodide (NaCsI) [NaI at 2 ug/µl, CsI at 50 ng/µl in water:isopropanol, 50:50 v/v]. The solution was infused into the ESI ion source using the integrated fluidics system. Calibration was performed between *m/z* 20-1974 using the automated IntelliStart procedure in the nanoACQUITY UPLC console (Instrument Driver Version 1.41.2872). Unless otherwise stated the quadrupoles were calibrated to operate at unit mass resolution.

2.19 LC-MS^E analysis using a Q-TOF Synapt HDMS instrument

The digested peptide mixtures were diluted in 97 (v/v) % MeCN / 0.1 (v/v) % formic FA (mobile phase A) and resolved by LCMS using a nanoACQUITY UPLC™ System (Waters, Elstree UK) coupled to a Q-TOF Synapt HDMS (Waters, Elstree UK). 1 µl of diluted sample was injected onto a trapping column (Waters C_{18} 180 µm X 20 mm), at a flow rate of 15 µl/min 99 % A/1 % B (where B consisted of 100 % (v/v) MeCN/0.1 % (v/v) FA), then after loading an analytical column (nanoACQUITY UPLC™ BEH C_{18} 75 µm x 150 mm, 1.7 µm column). The column temperature was set to 35 ºC, and a flow rate of 300 nl/min was maintained. The gradient conditions were as follows: 30 min 60 % A/40 % B, 32 min, 15 % A, 85 % B, 35.5 min 97 % A, 3 % B. A low energy survey scan was collected at 4 V collision energy offset between 200 and 2000 *m/z*. High energy fragmentation scans were collected as the collision energy was ramped from 15 to 50 V. The instrument was calibrated using a 500 fmol/µl solution of glu-fibrinopeptide B infused into the nano-ESI ion source at 500 nl/min. The product ion spectrum was acquired between *m/z* 50-1600 with the transfer cell collision energy set to 33 V for 5 min using a scan time of 5 s. Calibration was performed using the *y*-ion series and the b_2 product ion.

2.20 LC-ESI-MS/MS analysis using and LTQ-Orbitrap XL

The digested peptide mixtures were diluted in 97 (v/v) % MeCN / 0.1 (v/v) % FA (mobile phase A) and resolved by LCMS using a nanoACQUITY UPLC™ System (Waters, Elstree UK) coupled to an LTQ Orbitrap XL (ThermoFisher, Bremen, Germany). The eluent was injected directly into the nano-electrospray source fitted with a PicoTip Emitter silica tip 20 – 10 µm (New Objective, MA, USA). The sample temperature was maintained at 10 ºC, and 4 µl of each sample was injected initially onto a trapping column (Waters C_{18} 180 µm X 20 mm), at a flow rate of 18 µl/min 99 % A/ 1 % B (where B is 100 % MeCN/0.1 % (v/v) FA). The analytical column (nanoACQUITY UPLC™ BEH C_{18} 75 µm x 150 mm 1.7 µm column) was maintained at 35 ºC, and was developed at 300 nl/min by incrementing mobile phase B from 1 % (v/v) to 50 % (v/v) mobile phase B over 30 min, followed by a rapid ramp to 85 % mobile phase B over 1 min, and then a return to the starting mobile phase conditions for re-equilibration prior to the next injection. Data were acquired using Xcalibur version 2.0.7 SP1/Tuneplus version 2.2.0 Eng2/configured with Waters Acquity driver (build 1.0). Full scan MS (*m/z* range 300-1600 spectra were acquired in an LTQ-Orbitrap XL with the Orbitrap operating at a resolution of 30,000 (as defined @ *m/z* 400). In data-dependent mode with the top 5 most intense ions from the MS1 scan (full MS) were selected for tandem MS by collision-induced dissociation directing collision induced dissociation (CID) with helium gas in the LTQ, at a normalized collision energy of 30 %, and an activation q of 0.25. Dynamic exclusion was enabled for 30 s with a repeat count of two and an exclude duration width of 180 sec, and all product ion spectra were acquired in the LTQ. The Automatic Gain Control (AGC) feature was used to control the number of ions in the linear trap, and was set to 1 x 106 charges for a full MS scan, 6 X 104 SIM, 1 X 104 for the LTQ (MS^n), with ‘max’ injection times of 50 msec and 500 msec applied for the LTQ and Orbitrap respectively. All Orbitrap scans consisted of 1 microscan. The instrument was calibrated using MSCAL5 ProteoMass™ LTQ/FT-Hybrid ESI Pos.Mode CalMix (Sigma Aldrich, Poole, UK).

2.21 LC-ESI-MS/MS analysis using a triple quadrupole TSQ Vantage

The digested peptide mixtures were diluted in 97 (v/v) % MeCN / 0.1 (v/v) % FA (mobile phase A) and resolved by LCMS using a nanoACQUITY UPLC™ System (Waters, Elstree UK) coupled to an TSQ Vantage (ThermoFisher Scientific, Bremen,

Germany). The eluent was injected directly into the nano-electrospray source fitted with a PicoTip Emitter silica tip 20 – 10 µm (New Objective, MA, USA). The sample temperature was maintained at 10 ºC, and 4 µl of each sample was injected initially onto a trapping column (Waters C_{18} 180 µm X 20mm), at a flow rate of 18 µl/min 99 % A/ 1 % B (where B is 100 % MeCN/0.1 % (v/v) FA). The analytical column (nanoACQUITY UPLC™ BEH C_{18} 75 µm x 150 mm, 1.7µm column) was maintained at 35 ºC, and was developed at 300 nl/min by incrementing mobile phase B from 1 % (v/v) to 50 % (v/v) mobile phase B over 30 min, followed by a rapid ramp to 85 % mobile phase B over 1 min, and then a return to the starting mobile phase conditions for re-equilibration prior to the next injection. Data was acquired using Xcalibur version 2.0.7 SP1/Tuneplus version 2.2.0 Eng2/configured with Waters Acquity driver (build 1.0). Where possible, transitions were selected based on experimental tandem MS data obtained in the LTQ Orbitrap XL. The *y* series ions were selected as product ions as these are preferentially observed in the triple quadrupole and also because the isotopic variants of the tryptic peptides labelled with [$^{13}C_6$] Arg and [$^{13}C_6$] Lys retain their label at the C-terminus and thus the mass difference with native peptide. The vendor supplied software Pinpoint (v 1.1.12.0) was used in parallel to predict/confirm appropriate transitions (providing theoretical accurate *m/z* for product ions which is not possible experimentally in the LTQ) and for *in silico* calculation of collision energies by the solution of y= mx+c, where m=0.034 and 0.044 for +2 and +3 charge-states respectively, c=3.314 in both cases and x corresponds to mass *m/z*. The resolution of both Q1 and Q3 was set to 0.7 FWHM, and for high resolution analysis (H-SRM), Q1 was set to 0.2 FWHM. The scan time was set to 0.005 s, and the *m/z* transmission width 0.005. Lysates were applied (150 ng to 300 ng on column) with a range of QconCAT concentrations on column. The instrument was calibrated with polytyrosine – 1, 3, 6 Tuning and Calibration Solution (ThermoFisher Scientific, Bremen, Germany).

2.22 MALDI-TOF Analysis

MALDI-TOF MS analysis was performed using an Ultraflex™ II TOF/TOF (Bruker, Coventry, UK). Desalted tryptic digests (0.5 µl) were co-crystallised onto a MTP 34 target with a saturated solution of α-cyano-4-hydroxycinnamic acid in 50 %

(v/v) MeCN, 0.1 % (v/v) TFA. Detection was performed in reflector mode with delayed extraction and QCAL was used to calibrate the instrument.[160]

2.23 Peptide Identification and Quantification

Sequences corresponding to the recombinant QconCAT proteins LM1 and LM2 were added to a QconCAT reference database (FASTA), to assist in the identification of peptides and evaluation of miss cleavages, using the MASCOT (v 2.2.03, Matrix Science) and search engine. Tandem MS data acquired from the Obitrap was searched using the following parameters: up to 1 missed cleavage, peptides mass tolerance of 50 ppm, the fragment ion tolerance was set to 0.8 Da, carbamidomethylation of cysteine (57.02) was set as a fixed modifications and methionine oxidation (15.99), deamidation of asparagines (NG= known deamidation motif) and label ([$^{13}C_6$]Lys)/label ([$^{13}C_6$]Arg) were set as variable modifications.

For quantification, extracted ion chromatograms of the monoisotopic peaks were used to compare the ratios of analyte to standard following verification of tandem MS data from one or both L:H pairs of peptides per protein using Skyline v1.1 (MaCoss Laboratory, Seattle, US). The ratios were converted to copies per cell.

2.24 Computational

Pinpoint v.1.1.12.0 (ThermoFisher Scientific, Bremen, Germany) was used to predict SRM transition. Skyline v1.1 (MaCoss Laboratory, Seattle, US) was used to generate an SRM for the Xevo TQ MS.[161] Computational modelling and statistical analysis was carried out using Matlab v. R2010a (Mathworks, Massachusetts, US). The scaffold model detailed in Appendix 8 was written in conjunction with Nils Blüthgen's Laboratory, Humboldt University, Berlin. ODE 23 an inbuilt method of Runge-Kutta $2^{nd}/3^{rd}$ orders method for solving ordinary differential equations was used in combination with lsqnonlin, a nonlinear least squares method.

3 Results and Discussion I -LM1 and LM2 QconCAT Design, Expression, extraction and Purification

3.1 QconCAT Design

In this study QconCAT constructs were designed to generate Q-peptides for the quantification of 27 protein phosphatases, scaffold proteins and substrates of the MAPK pathway (Figure 3.1, Table 3.1 and 3.2). A consensus method, CONSeQuence using three different machine learning algorithms was employed to predict peptides likely to be observed based on their physicochemical properties.[162] The likelihood of these sequences undergoing complete tryptic proteolysis (generating limit peptides) was also predicted.[163] Q-peptides were then selected from a short list of potential candidates. Peptides with single nucleotide point mutations or internal sites of phosphorylation as defined in UniProt and Phosphosite databases were discarded as unsuitable.

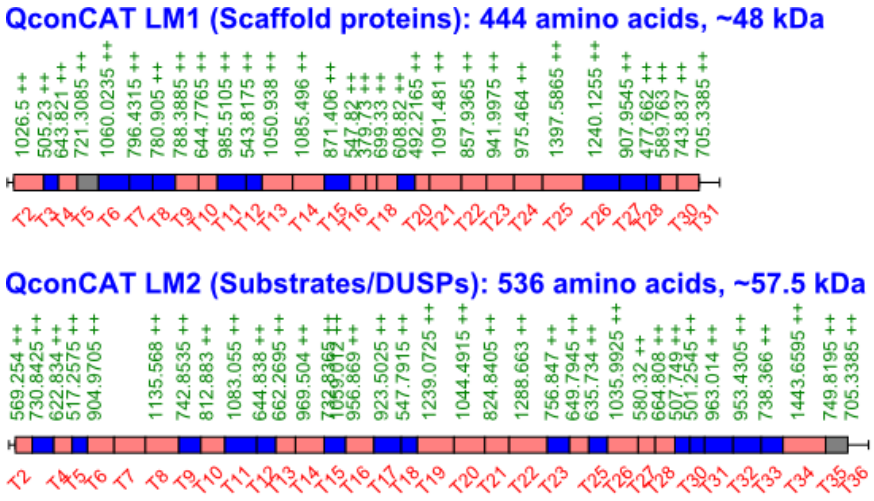


Figure 3.1 Schematic of the QconCAT constructs LM1 and LM2, designed to yield Q-peptides for quantification of the scaffold proteins and the substrates and DUSPs of the MAPK pathway respectively. Each T represents a tryptic peptide with their *m/z* ratio for double charged peptides shown. Blue and red boxes represent the Q peptides.

3.2 LM1 and LM2 QconCAT Expression Trial

QconCAT proteins LM1 and LM2 were first expressed and purified in *E. coli* in the absence of stable isotope label to confirm expression and test purification conditions. Following induction with 1 mM IPTG and incubation at 37 ºC with 200 rpm shaking, samples were taken at 0, 1, 2, 3 and 4 hours to assess protein expression by analysing total bacterial extract by SDS-PAGE. Following Coomassie staining a distinct band was present in both cases at ~65 kDa corresponding to a theoretical mass of 48 653 Da for LM1 and 58 036 Da for LM2 (Figure 3.2). These bands were subjected to in-gel digestion and subsequent analysis by MALDI-TOF (Figure 3.3) and ESI LTQ-Orbitrap XL (Figure 3.4) to confirm expression of LM1 and LM2.

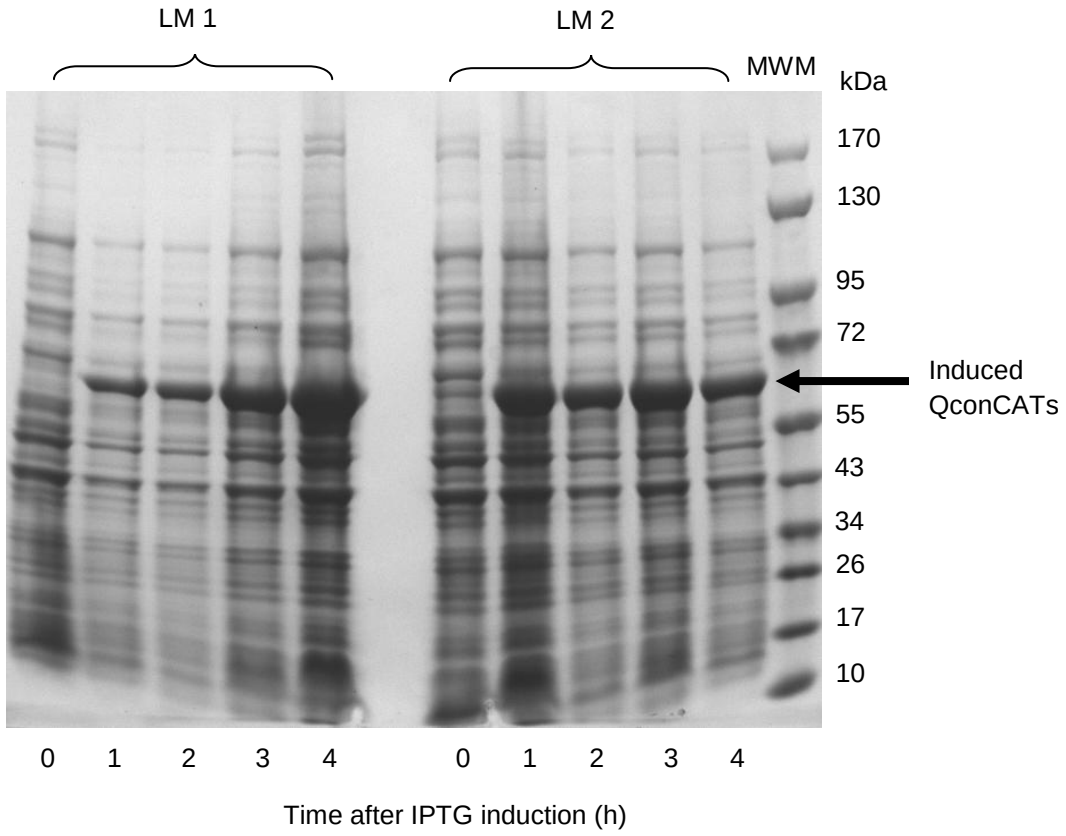


Figure 3.2 Induction of QconCAT LM1 and LM2 in *E. coli* BL21 cells. Cells were grown to an A_{600} of 0.8 and IPTG (1 mM) added to induce expression.

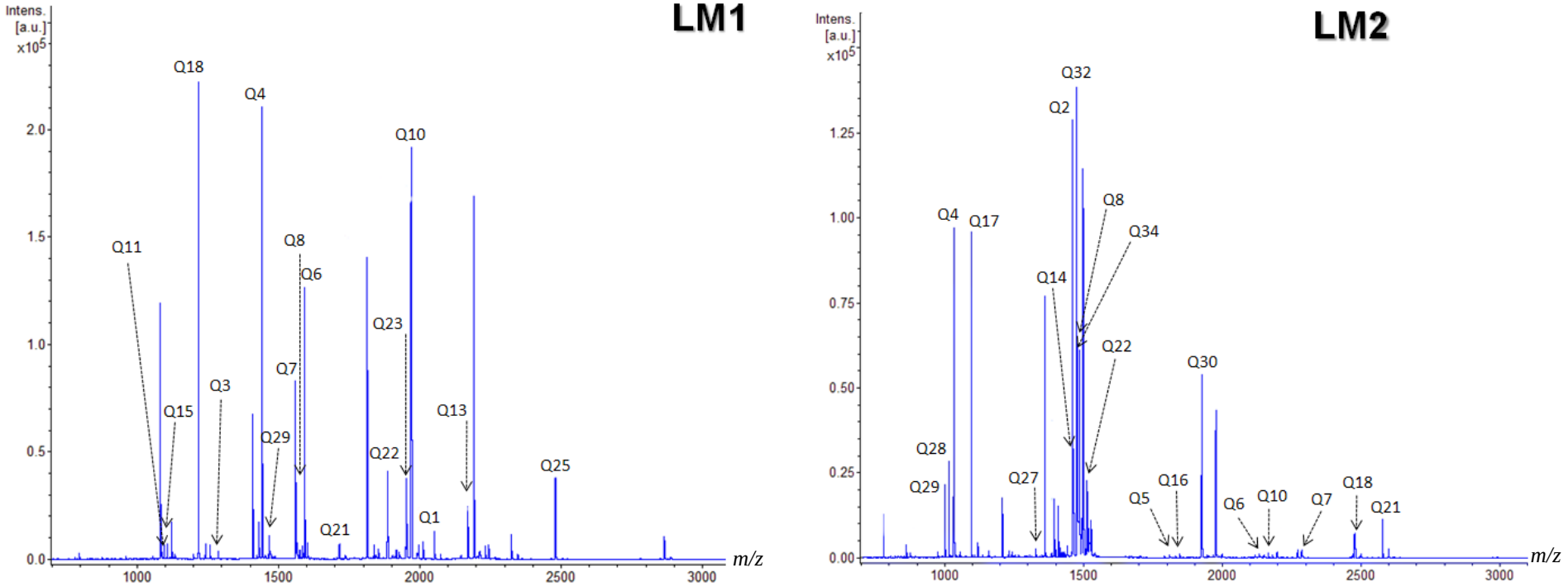


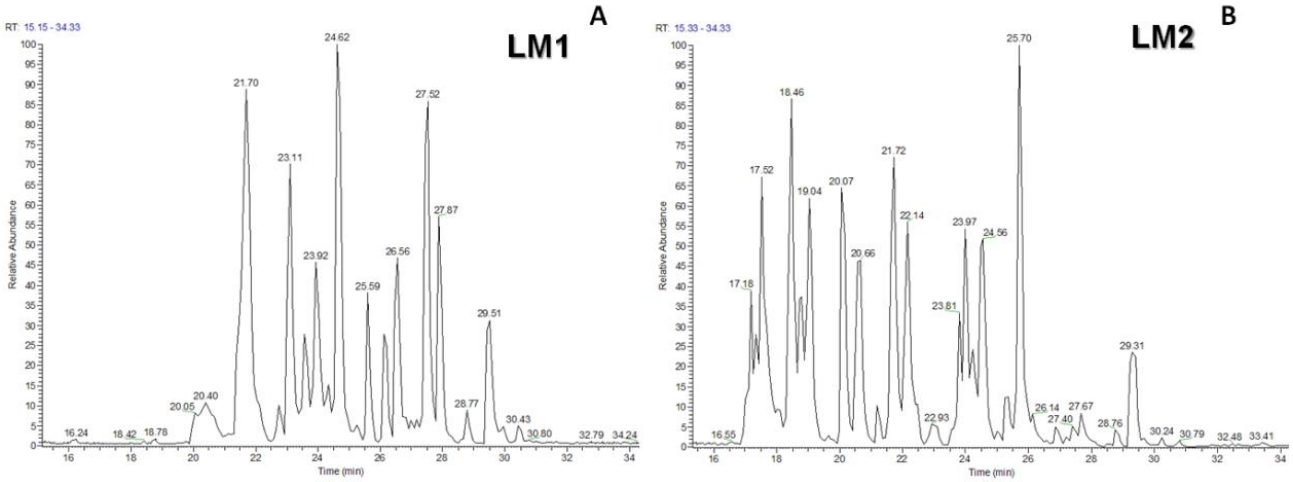
Figure 3.3 MALDI-TOF mass spectra for LM1 and LM2. The tryptic Q-peptides corresponding to QconCAT LM1 and LM2.

Table 3.1 Tryptic Q-peptides generated from the QconCAT LM1, a 49 kDa protein. Detailed are the names of the peptides selected for quantification of the scaffold proteins in the order that they appear in the construct, in addition to the sequence of the selected Q-peptide and their protonated monoisotopic mass. Bold indicates detection following MALDI-TOF MS analysis of the digested unpurified protein following induction in *E. coli* (Figure 3.3).

Pep.	Protein	Accession #	Q-Peptide Sequence	$[M+H]^+$	Detect
Q1	IQGAP1	P46940	NVIFEISPTEEVGDFEVK	2052.02	1
Q2	ARRB2 1/2/3	P32121	ACGVDFEIR	1009.48	0
Q3	KSR2 1	Q6VAB6	SEEQQPLSLQK	1286.66	1
Q4	Control	Control	GVNDNEEGFFSAR	1441.63	1
Q5	KSR2 1	Q6VAB6	LTVDAYPGLCPPPPLESGHR	2119.06	0
Q6	MPKS1	Q9UH	LPSVEGLHAIVVSDR	1591.88	1
Q7	PEBP1	P30086	LYTLVLTDPDAPSR	1560.83	1
Q8	KSR1	Q8IVT5	DLTLDALLEMNEAK	1575.79	1
Q9	PAXI 1/2/3	P49023	CYYCNGPILDK	1133.50	0
Q10	KSR2 1/2	Q6VAB6	QQFIFPDVVPVPETPTR	1970.04	1
Q11	KSR1	Q6VAB6	LIDISIGSLR	1086.65	1
Q12	ARRB1 1/2	P49407	CPVAMEEADDTVAPSSTFCK	2100.89	0
Q13	PEA_15	Q15121	SEEITTGSAWFSFLESHNK	2170.01	1
Q14	KSR2 2	Q6VAB6	IHSSVGSCENIPSQQR	1741.83	0
Q15	Sur8/Shoc2	Q9UQ13	SIHILPSSIK	1094.66	1
Q16	PAXI 1/2/3	P49023	LGVATVAK	758.48	0
Q17	MVP	Q14764	LFSVPDFVGDACK	1397.68	0
Q18	MPKS1	Q9UH	ELAPLFEELR	1216.66	1
Q19	ARRB1 1/2	P49407	ACGVDYEVK	983.45	0
Q20	ARRB2 1/2/3	P32121	CPVAQLEQDDQVSPSSTFCK	2181.98	0
Q21	Sur8/Shoc2	Q9UQ13	ELTQLTELYLYSNK	1714.89	1
Q22	IQGAP1	P46940	ILAIGLINEALDEGDAQK	1883.01	0
Q23	PEBP1	P30086	GNDISSGTVLSDYVGSGPPK	1949.95	1
Q24	MORG1	Q9BRX9	TYSGHGYEVLDAAGSFDNSSLCSGGGDK	2794.19	0
Q25	MORG1	Q9BRX9	VNTVQFNEEATVILSGSIDSSIR	2479.27	1
Q26	KSR1 1/2/3	Q8IVT5	LSHDWLCYLAPEIVR	1814.93	0
Q27	KSR1 1/2/3	Q8IVT5	CGASGDECGR	954.34	0
Q28	PEA_15	Q15121	ISEEDELDTK	1178.54	0
Q29	MVP	Q14764	ALQPLEEGEDEEK	1486.69	1

Table 3.2. Tryptic Q-peptides generated from the QconCAT LM2, 58 kDa protein containing peptides to quantify the substrates and the Dual Specificity Phosphatases (DUSPs). Detailed are the names of the Q-peptides selected and order in which they appear in the construct, in addition to the sequence of the selected Q-peptide and their protonated monoisotopic mass. Bold indicates detection following MALDI-TOF MS analysis of the digested unpurified protein following induction in *E. coli* (Figure 3.3).

Pep.	Protein	Accession #	Q-Peptide Sequence	$[M+H]^+$	Detect
Q1	DUSP16	Q9BY84	VPVNDSFCEK	1137.52	0
Q2	DUSP7	Q16829	SAEWLQEELEAR	1460.70	1
Q3	DUSP18	Q8NEJ0	QPSVSGLSQITK	1244.68	1
Q4	DUSP5	Q16690	YVLPDEAAR	1033.53	1
Q5	DUSP9	Q99956	YILNVTPNLPNFFEK	1808.96	1
Q6	DUSP18	Q8NEJ0	NTVHMVSSPVGMIPDIYEK	2117.04	1
Q7	DUSP14	O95147	VPLADMPHAPIGLYFDTVADK	2270.15	1
Q8	ELK1	P19419	EQGNGHIISWTSR	1484.72	1
Q9	ETS1 1/2	P14921	DWVMWAVNEFSLK	1624.78	0
Q10	DUSP 14	O95147	MISEGDIGGIAQITSSLFLGR	2165.13	0
Q11	DUSP 2	Q05923	ELECAALGTLLR	1288.69	0
Q12	DUSP 4	Q13115	FSSEYPEFCSK	1323.56	0
Q13	STAT3	P40763	SIVSELAGLLSAMEYVQK	1938.03	1
Q14	ETS2	P15036	NMDQVAPVANSYR	1464.69	0
Q15	DUSP 10	Q9Y6W6	SLNCGCSSASCCTVATYDK	1912.75	0
Q16	DUSP9	Q99956	ASFPVQILPNLYLGSAR	1846.02	1
Q17	DUSP2	Q05923	AGPTAVYFLR	1094.60	1
Q18	ETS1 1	P14921	VPSYDSFDSEDYPAALPNHKPK	2477.16	1
Q19	DUSP4	Q13115	ADISSWFMEAIEYIDAVK	2088.00	0
Q20	DUSP1	P28562	GGYEAFSASCPELCSK	1648.70	0
Q21	STAT3	P40763	LLQTAATAAQQGGQANHPTAAVVTEK	2576.34	1
Q22	DUS10	Q9Y6W6	AANLTYMPSSSGSAR	1512.71	1
Q23	ETS1 1/2	P14921	FCMNGAALCALGK	1298.61	0
Q24	DUSP7	P28562	ISSDCSDGESDR	1270.49	0
Q25	DUSP1	P28562	ADISSWFNEAIDFIDSIK	2071.00	0
Q26	DUSP16	Q9BY84	LVALLESGTEK	1159.66	1
Q27	MYC	P01106	DQIPELENNEK	1328.63	1
Q28	ELK1 1	P19419	GAGMAGPGGLAR	1014.51	1
Q29	ELK1	P19419	LVDAEEVAR	1001.53	1
Q30	DUSP5	Q16690	LLQEGGGGVAAVVVLDQGSR	1925.05	1
Q31	MYC	P01106	CHVSTHQHNYAAPPSTR	1905.88	0
Q32	ELK1 1	P19419	AEPEVPPQEGVPAR	1475.75	1
Q33	ETS2	P15036	GGLLDSMCPASTPSVLSSEQEFQMFPK	2886.34	0
Q34	Control	Control	GGVNDNEEGFFSAR	1498.66	1



C

	1	11	21	31	41	51	61	71	81	91
1	MGTKNVIFE	ISPTEEVGDF	EVKACGVDFE	IRSEEQQPLS	LQKGVNDNEE	GFFSARLTVD	AYPGLCPPPP	LESGHRLPSV	EGLHAIVVSD	RLYTLVLTDP
101	DAPSRDLTLD	ALLEMNEAKC	YYCNGPILDK	QQFIFPDVVP	VPETPTRLID	ISIGSLRCPV	AMEEADDTVA	PSSTFCKSEE	ITTGSAWFSF	LESHNKIHSS
201	VGSCENIPSQ	QRSIHILPSS	IKLGVATVAK	LFSVPDFVGD	ACKELAPLFE	ELRACGVDYE	VKCPVAQLEQ	DDQVSPSSTF	CKELTQLTEL	YLYSNKILAI
301	GLINEALDEG	DAQKGNDISS	GTVLSDYVGS	GPPKTYSGHG	YEVLDAAGSF	DNSSLCSGGG	DKVNTVQFNE	EATVILSGSI	DSSIRLSHDW	LCYLAPEIVR
401	CGASGDECGR	ISEEDELDTK	ALQPLEEGED	EEKLAAALEH	HHHHH					

	1	11	21	31	41	51	61	71	81	91
1	MGTKVPVNDS	FCEKSAEWLQ	EELEARQPSV	SGLSQITKYV	LPDEAARYIL	NVTPNLPNFF	EKNTVHMVSS	PVGMIPDIYE	KVPLADMPHA	PIGLYFDTVA
101	DKEQGNGHII	SWTSRDWVMW	AVNEFSLKMI	SEGDIGGIAQ	ITSSLFLGRE	LECAALGTLL	RFSSEYPEFC	SKSIVSELAG	LLSAMEYVQK	NMDQVAPVAN
201	SYRSLNCGCS	SASCCTVATY	DKASFPVQIL	PNLYLGSARA	GPTAVYFLRV	PSYDSFDSED	YPAALPNHKP	KADISSWFME	AIEYIDAVKG	GYEAFSASCP
301	ELCSKLLQTA	ATAAQQGGQA	NHPTAAVVTE	KAANLTYMPS	SSGSARFCMN	GAALCALGKI	SSDCSDGESD	RADISSWFNE	AIDFIDSIKL	VALLESGTEK
401	DQIPELENNE	KGAGMAGPGG	LARLVDAEEV	ARLLQEGGGG	VAAVVVLDQG	SRCHVSTHQH	NYAAPPSTRA	EPEVPPQEGV	PARGGLLDSM	CPASTPSVLS
501	SEQEFQMFPK	GGVNDNEEGF	FSARLAAALE	HHHHHH						

Figure 3.4 FT-MS base peak chromatograms for unlabelled LM1 (A) and unlabelled LM2 (B) tryptic digests separated by reverse-phase chromatography. (C) Sequence coverage for LM1 (top panel, 55.5 %) and LM2 (bottom panel, 73.7 %) when LC-ESI MS/MS data of the in-gel digested tryptic peptides were searched against a customised database using the SEQUEST algorithm. Sequence coverage is indicated by green and yellow shade which represents probability scores. Green indicates a high confidence level whilst yellow represents a modest confidence level.

3.3 LM2 Unlabelled QconCAT Extraction and Purification

Having confirmed expression of the two novel proteins from these expression constructs, protein was extracted from the bacterial cells as described in Materials and Methods. SDS-PAGE was used to confirm expressed LM2 was present in bacterial inclusion bodies (Figure 3.5). Initially, cells were lysed with BugBuster post-induction and preparations of the total fraction (TF), soluble fraction (SF), and starting material (SM) prior to purification were analysed by SDS-PAGE and Coomassie staining. The SDS-PAGE showed that there is a band below 73 kDa corresponding to the LM2 theoretical mass of 58 036 Da and that this band is present in the total fraction and starting material but not in the soluble fraction, indicating that LM2 is present in the inclusion bodies. LM2 from the inclusion bodies was therefore purified using a His_6 trap column; Eluted (E) fractions 1, 2, 3 correspond to the fractions obtained following purification (Figure 3.5). Less intense bands relative to LM2 in these fractions could indicate protein degradation or impurities. Following purification the pooled eluted material was desalted. The resulting chromatogram of LM2 factions and subsequent analysis by SDS-PAGE gel is shown in Figure 3.6.

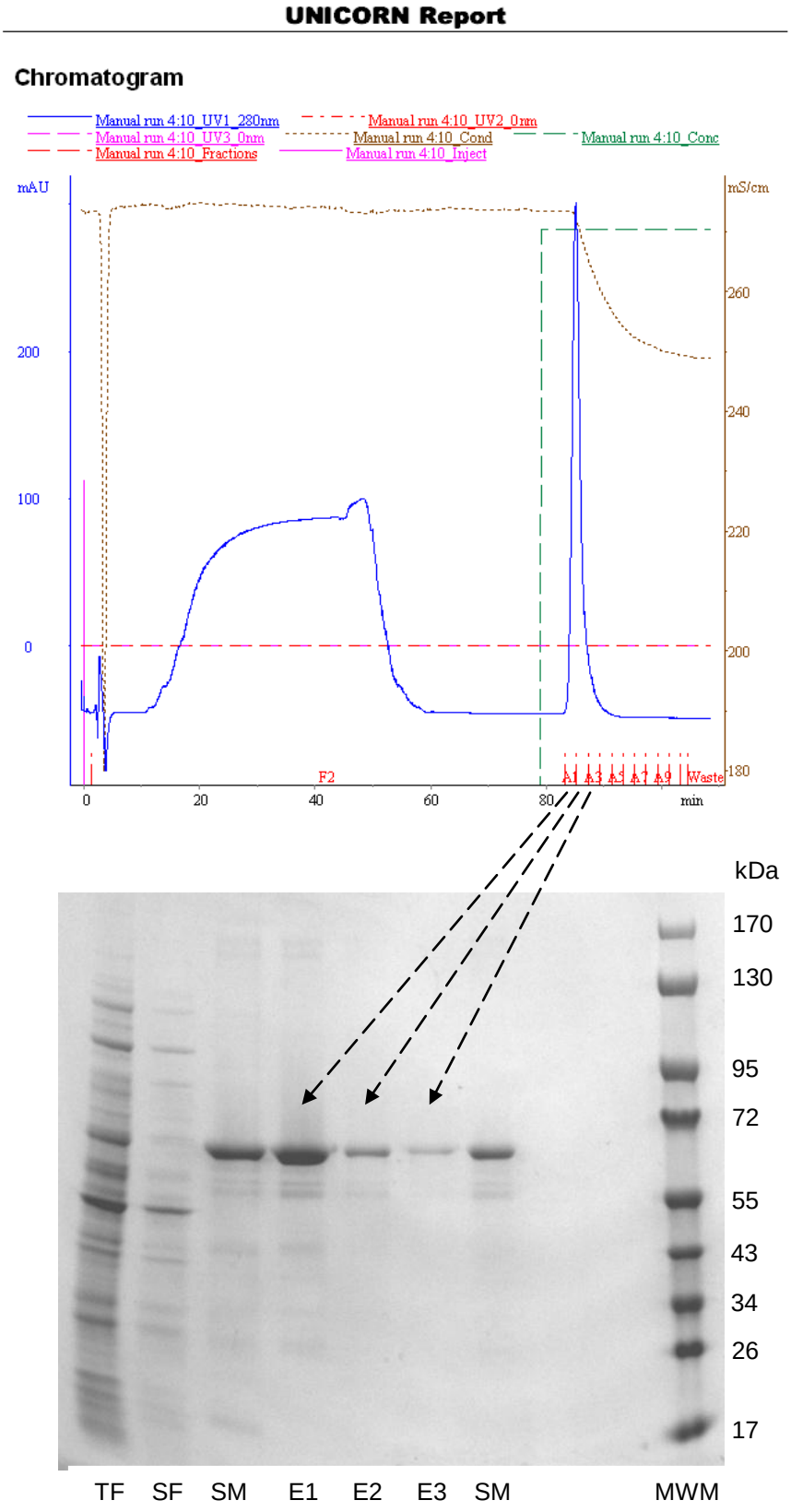


Figure 3.5. (A) UV chromatogram at 280 nm during purification of LM2 by FPLC using a His_6 trap column. (B) Coomassie stained SDS-PAGE gel showing protein in the total bacterial extract fraction (TF) and the soluble fraction (SF) prior to purification of LM2 from the TF with a His_6 trap column. SM: starting material prior to purification; E1, E2, E3 are eluted fractions following purification.

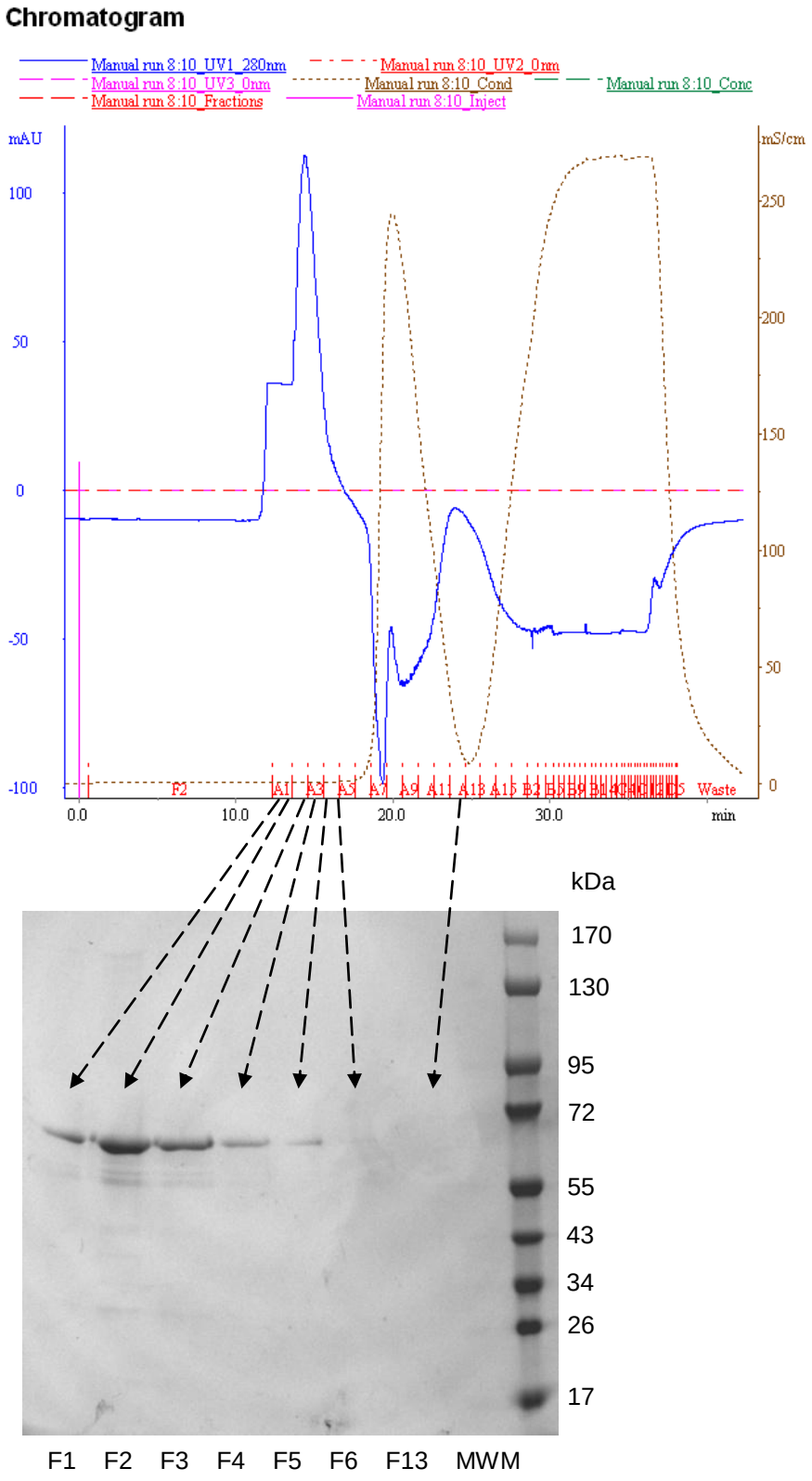


Figure 3.6 UV chromatogram at 280 nm of LM2 fractions (F) during elution of the bound material from a sephadex HiTrap™ desalting column and subsequent analysis of fractions by SDS-PAGE gel.

3.4 LM1 Unlabelled Expression, Extraction and Purification

Several repeat experiments were subsequently carried out to express LM1 unlabelled in *E. coli* BL21(Δ)DE3 cells. However, on each occasion expression of LM1 could not be detected based on appearance of a Coomassie stained band in SDS-PAGE analysis of total cell extract (in contrast to the data in Figures 3.5 and 3.6). Therefore new batches of competent cells were used to transform the plasmids from the original construct stocks and unfortunately LM1 could still not be detected. This problem was also experienced by other users of QconCATs. A possible explanation for the poor yields was expression of the QconCAT without presence of IPTG due to weak repression of the T7 promoter. Therefore the LM1 plasmids were transformed into freshly prepared competent cells of an alternative cell line, BL21(DE3) *pLysS*, the pLysS plasmid expresses T7 lysozyme, a natural T7 polymerase inhibitor, this alleviated the expression problem. LM1 was then expressed and purified in *E. coli* in the absence of stable isotope label. Following induction with 1 mM IPTG, samples were taken at 0, 1, 2, 3, 4, 5 and 6 hours to assess protein expression and later analysed by SDS-PAGE. Following coomassie staining a distinct band was detected corresponding to a theoretical mass of 48 kDa for LM1 (Figure 3.7).

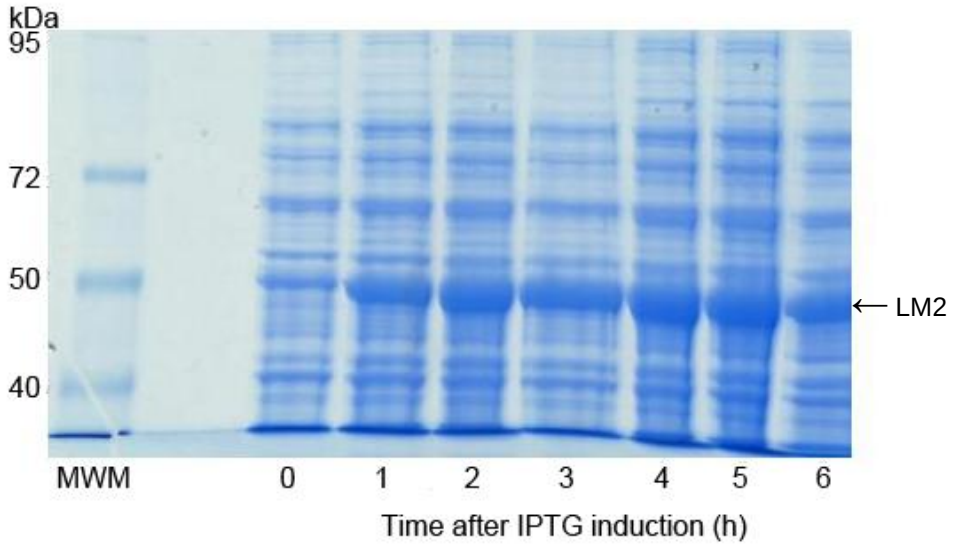


Figure 3.7 Induction of QconCAT LM1 in BL21(DE3) pLysS cells. Cells were grown to an A_{600} of 0.8 and then IPTG (1 mM) added to induce expression.

Cells were lysed with BugBuster post-induction and as in the case of LM2, expressed LM1 was present in the bacterial inclusion bodies. The inclusion bodies were resuspended in denaturing wash buffer to provide clarified sample as described in the methods section and this was purified using TALON beads. Five 1 ml

fractions were eluted from the beads where artificial protein was present in each of the fractions E1-3 (Figure 3.8). The eluents were pooled and desalted using dialysis.

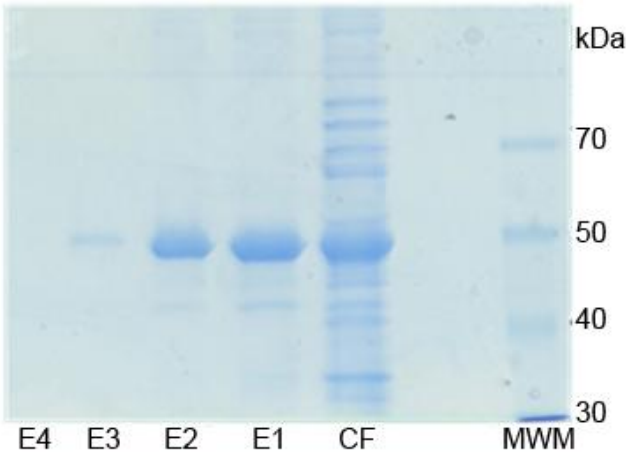


Figure 3.8 SDS-PAGE to show eluted fractions of unlabelled LM1 (E1-4) following purification of TALON beads, where CF represents clarified sample. MWM represents molecular weight marker.

3.5 LM2 Labelled QconCAT Expression, Extraction and Purification

LM2 plasmids were transformed into BL21(DE3) competent cells which were then grown in isotopically labelled media, [$^{13}C_6$]-R/K by Dr Kathleen Carroll. The SDS-PAGE gel showed a large band observed above 55 kDa corresponding to LM2 at 58 kDa (Figure 3.9). The recombinant protein was purified and desalted by dialysis.

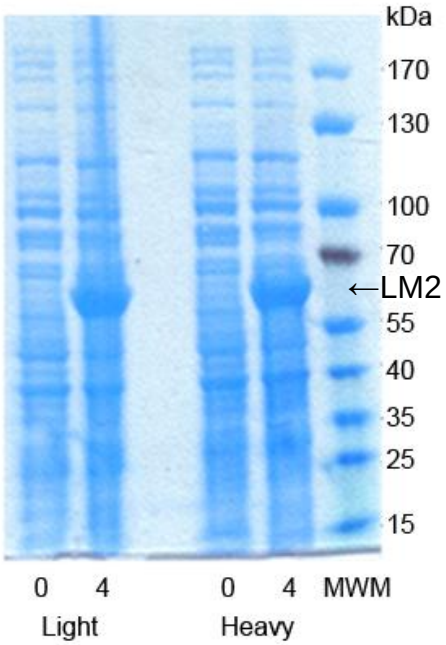


Figure 3.9 SDS-PAGE of expression of labelled and unlabelled QconCAT LM2 in minimal media before induction with IPTG 1 mM at T=0 and 4 h post induction at T=4 h.

MALDI-TOF analysis of in-gel digested bands was used to assess the extent of labelling of LM2. It was found that the unlabelled portion of the QconCAT is below

the limit of detection and is not taken into account for signal integration by the Bruker Daltonic flexAnalysis 3.0 software (Figure 3.10).

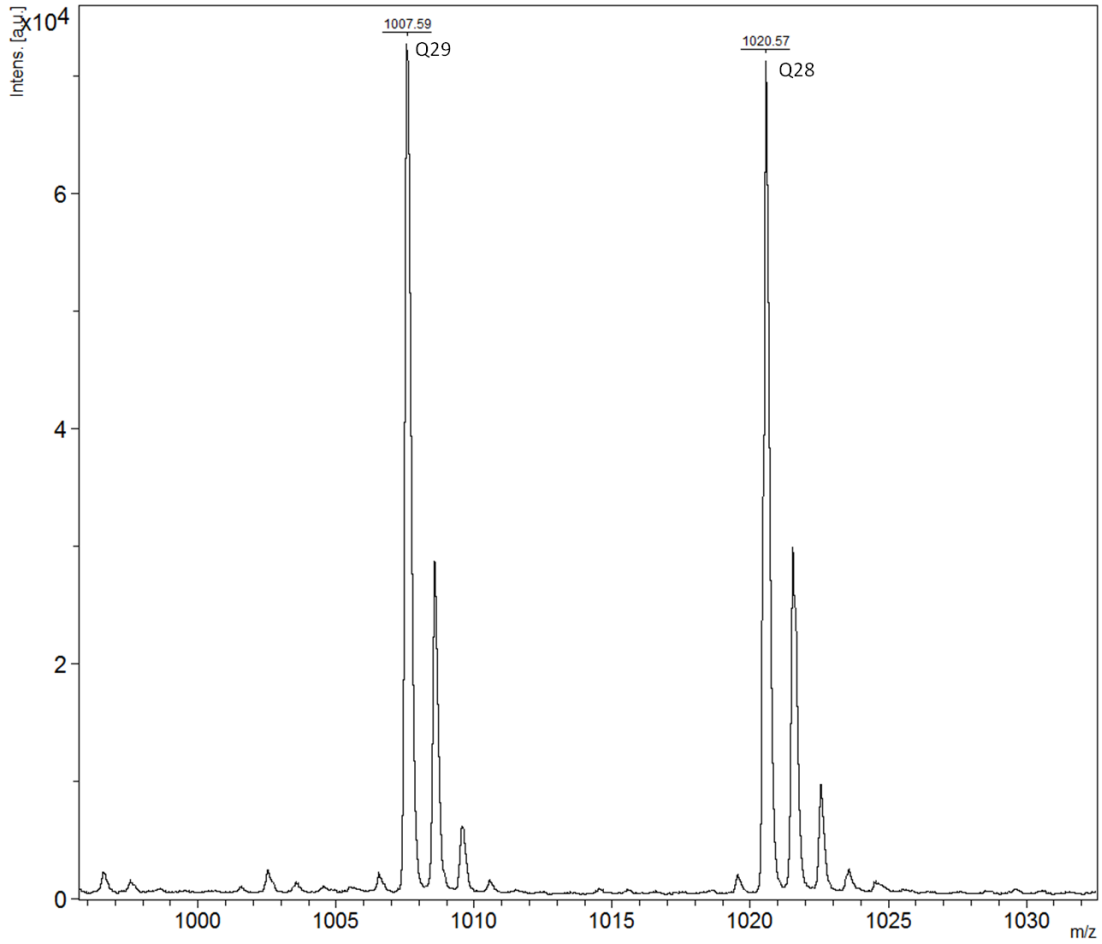


Figure 3.10 MALDI-TOF mass spectrum of trypsin digested labelled LM2, showing ions at *m/z* of 1007.59 and 1020.57, matching to isotope labelled Q29 and Q28 respectively. Ions at *m/*z 1001.59 and 1013.58, representative of unlabelled Q29 and Q28 are not observed.

3.6 LM1 Labelled QconCAT Expression, Extraction and Purification

The same general procedure as for the unlabelled QconCAT was used to express, extract and purify [$^{13}C_6$]-R/K labelled LM1. Following cell lysis with BugBuster the inclusion body cell pellet was resuspended in denaturing wash buffer and this clarified sample (CF) was purified using TALON beads and four 1 ml fractions were eluted from the beads (E1-4) (Figure 3.11). Several rounds of expression, extraction and purification were undertaken as protein degradation was a frequent problem. This issue was eventually overcome by undertaking extraction and purification in the cold room at 4 °C and the resulting SDS-PAGE shows a strong Coomassie staining band just below 50 kDa which corresponds to LM1 at 48 kDa and very little protein

degradation. The eluted fraction E3 was subsequently desalted using buffer exchange.

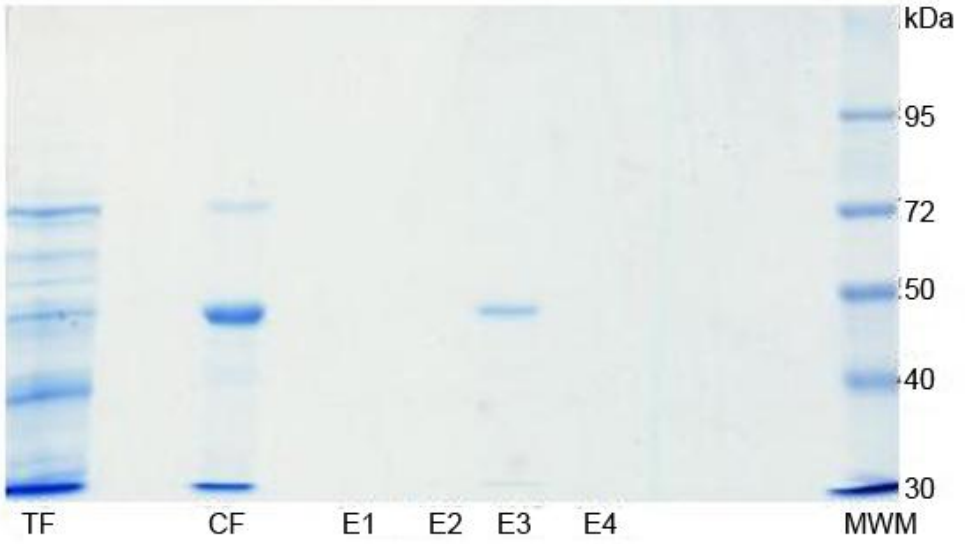


Figure 3.11 Coomassie stained SDS-PAGE gel to show expression of labelled LM1 in Total Fraction (TF), and Clarified Fraction (CF) prior to purification and eluted fractions E1-4 following purification using talon beads.

MALDI-TOF MS was used to assess the extent of labelling of LM1. The first round of protein expression in the presence of [$^{13}C_6$]-R/K resulted in incomplete protein labelling: comparison of peaks areas for the monoisotopic peaks of the labelled and unlabelled peptide ions indicated a ratio of approximately 1:0.75 labelled:unlabelled peptides (Figure 3.12). Ideally, > 95 % labelling should be achieved in order to simplify and optimise peptide based proteins quantification. It is thought that insufficient labelling in this case was due to using a starter culture that was grown in standard LB media and then transferred to labelled media without enough dilution to minimise the unlabelled portion. Unfortunately, the LM1 pLysS glycerol stocks would not grow in minimal media. The plasmid was therefore retransformed into fresh competent cells and new glycerol stocks were produced. These fresh glycerol stocks produced starter cultures of LM1 in minimal media and complete labelling of LM1 was achieved, as shown by the inability to quantify ions for the unlabelled peptides (Figure 3.12 and Figure 3.13).

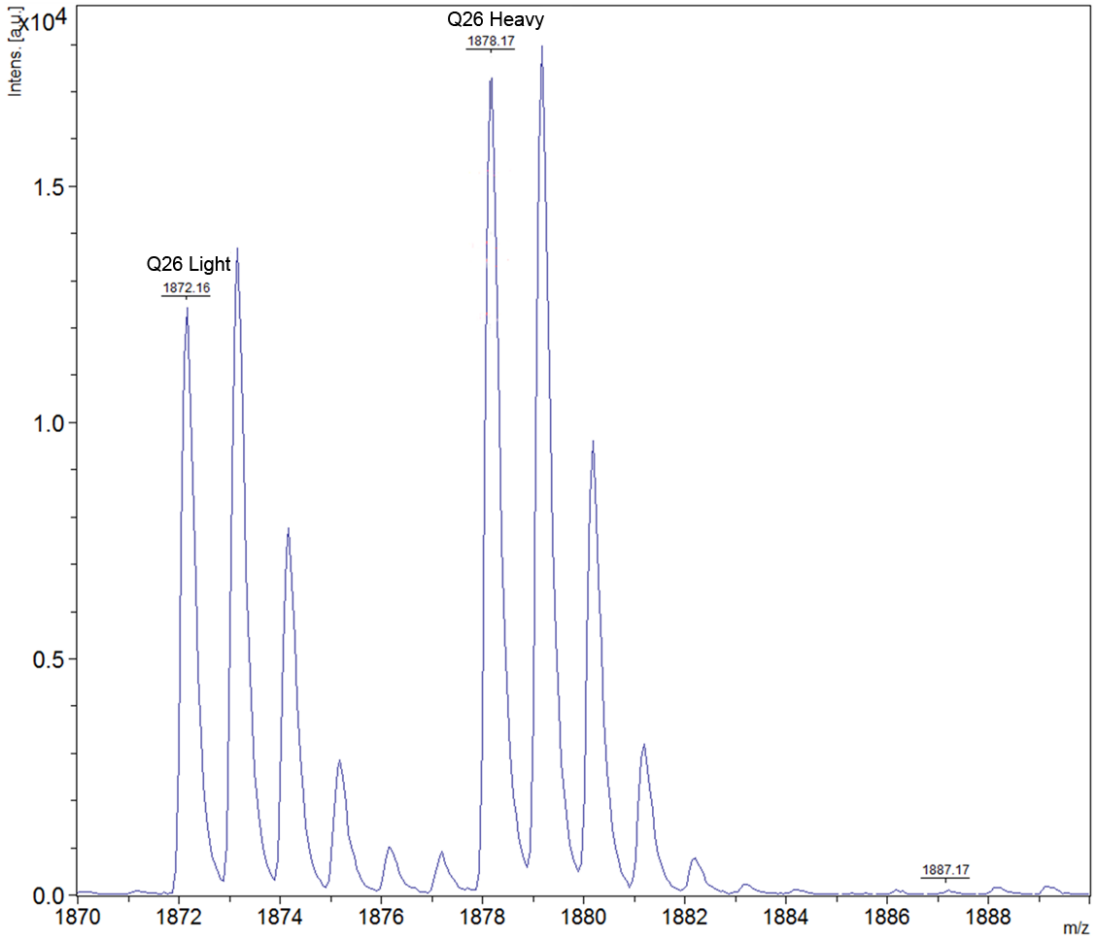


Figure 3.12 Selected region of a MALDI-TOF mass spectrum of LM1 tryptic peptides generated following expression in [$^{13}C_6$-K/R] containing medium. Shown are ions at *m/z* 1872.16 and 1878.17 matching to the light (unlabelled) and heavy (labelled) forms of Q26, demonstrating incomplete protein labelling. The ratio of labelled to unlabelled is 1:0.75.

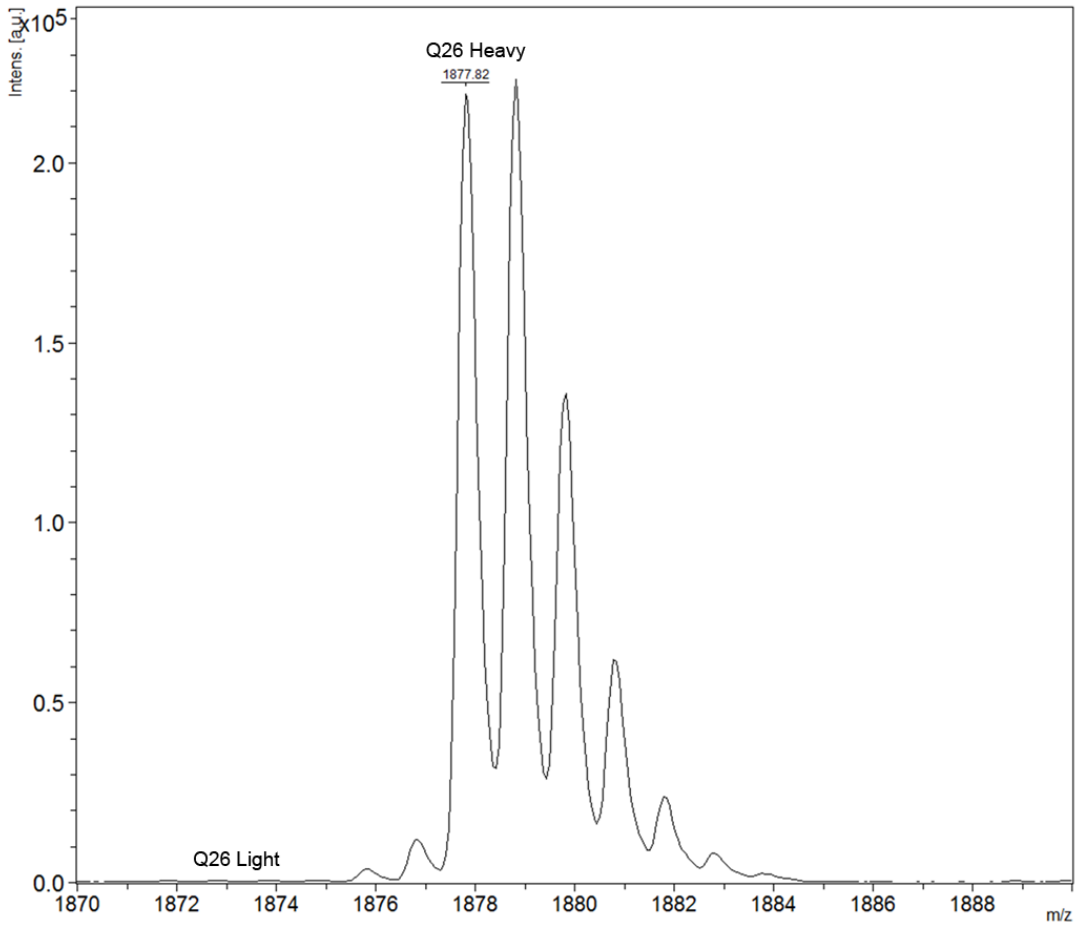


Figure 3.13 MALDI-TOF spectrum for Q26 of LM1 labelled QconCAT. Peptide ions representative of unlabelled Q26 were below the limit of quantification.

3.7 Conclusion

QconCAT proteins to generate peptide standards suitable for quantification of the scaffold proteins (LM1), substrates and DUSPs (LM2) of the MAPK pathway have been designed, synthesised, expressed, purified and analysed. Initial problems of expression for LM1 were overcome by retransforming into an alternative cell line BL21(DE3) *pLysS*. Labelled proteins have also been expressed, extracted and purified. The next chapter will describe experiments used to validate their behaviour in the mass spectrometer and to produce a list of transitions to enable peptide quantification by Selected Reaction Monitoring (SRM).

4 Results and Discussion II QconCAT LM2 SRM Method Development and Final Analysis

4.1 Selected Reaction Monitoring

Selected Reaction Monitoring (SRM), is a method of mass spectrometry-based quantification and is applied here for the quantification of selected peptides in a complex biological sample. A SRM experiment is carried out by specifying the precursor ion mass-to-charge ratio (*m/z*), for fragmentation and then monitoring specific product ions following collision-induced dissociation. The combination of each precursor and product ion monitored is known as a transition and in combination with a defined retention time for the peptide is a highly selective assay. SRM takes advantage of the triple quadrupole platform, as the first and third quadrupoles act as filters to select the mass-to-charge, values of the precursor and product ions respectively whilst the second quadrupole acts as the collision cell. Several transitions are monitored over an LC gradient resulting in a set of chromatographic traces. Unlike other MS proteomic methods no full scan mass spectra i.e. no product ion spectra are collected during SRM analysis. When many transitions are monitored during an LC gradient the dwell time (defined as the time spent by the instrument analysing each mass-to-charge ratio), is reduced. Poor quality data can result if too many transitions are monitored because the amount of time the analyser remains transmitting ions at a given *m/z* is reduced and the fraction of ions reaching the detector is small. Therefore, a fine balance must be struck between the number of transitions monitored (thereby maximising the capacity of the LC-MS experiment) without compromising the signal-to-noise ratio and causing excessive loss in sensitivity.

For SRM method development a workflow is described below (Figure 4.1) whereby a Q-TOF Ultima Global, was used to define the required transitions experimentally, which were then applied to perform a SRM experiment on apply the a triple quadrupole instrument, the Xevo TQ MS.

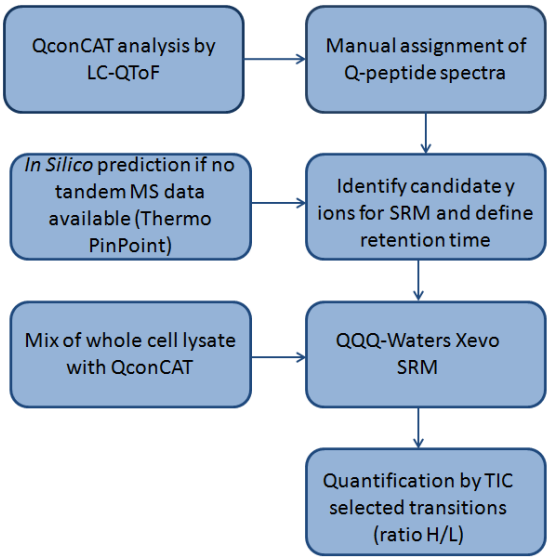


Figure 4.1 Schematic to show the workflow of SRM method development using the Q-TOF Ultima Global and Xevo TQ MS.

Initially, transitions for QconCATs LM2 were determined by analysing unlabelled QconCAT LM2 sample using a Q-TOF Global Ultima, coupled online to a LC system. A data-dependent acquisition was undertaken where the top 3 most abundant ions from a full scan MS spectrum were selected for CID and product ion spectra acquired in a serial manner (Figure 4.2). In addition, retention times for each Q-peptide were also defined. The predominant types of product ions generated by CID in a collision cell are *y*- series ions. These were selected amongst all the product ions for each Q-peptide precursor ion by locating the relevant tandem MS scan using the extracted ion chromatogram (XIC) of the precursor ion for each peptide. To provide transitions of high selectivity and sensitivity, *y*-ion products of high intensity were manually selected from the *m/z* region of the spectra greater than the *m/z* value for the (usually doubly-charged) precursor ion where there is less noise (Figure 4.2). Product ions with *m/z* values greater than that of the precursor ion were favoured in an attempt to increase the selectivity of the transitions by preventing interference from co-eluting singly-charged species that are also co-isolated in the first quadrupole. Due to the precursor ion only bearing a single charge, these contaminants can only generate product ions of lower *m/z* value. Therefore, by choosing to monitor product ions of higher *m/z* value than the precursor ion for multiply-charged peptides, these contaminants do not produce interfering ion current as their product ions have an unstable motion through the second quadrupole and

thus are not detected. For example, the full scan CID product ion spectrum of Q-peptide 12, FSSEYPEFCSK, in QconCAT LM2 shows that *y*6, *y*7 and *y*9 are the most intense products above $[M+2H]^{2+}$ and are therefore included in the transition list. Where peptides were not observed by the Q-TOF Ultima Global, a transition prediction program Pinpoint (v 1.1.12.0), was used to predict optimal transitions for SRM. Often Pinpoint only predicts one transition per peptide and ideally several are required per peptide for optimal selectivity of quantification in a complex mixture. (Table 4.1).

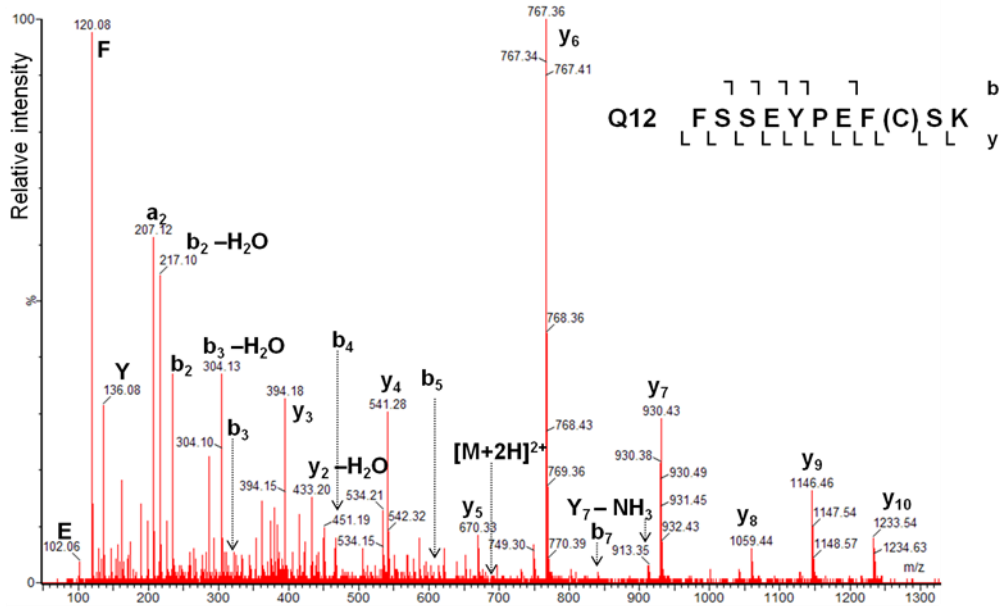


Figure 4.2 Example of full scan CID-MS/MS recorded on a Q-TOF Ultima Global instrument of light QconCAT LM2 Q12 peptide (*m/z* 690.79). (C) represents the fixed modification of carbamidomethylation of cysteine. F and Y represent phenylalanine and tyrosine immonium ions respectively**.** Product ions *y*6, *y*7 and *y*9 selected for the SRM assay.

Table 4.1 QconCAT LM2, designed to quantify substrates and DUSPs of the MAPK pathway. Q-peptides are listed along with their doubly- and triply- charged ions, retention times, and selected transitions. (C) represents cysteines modified by carbamidomethylation. For those peptides not detected by a Q-TOF Ultima Global, no retention time could be recorded and *tbc*, to be confirmed, appears in the retention time column.

Peptide	Protein	Accession #	Peptide Sequence	$[M+2H]^{2+}$	Rt, min	y-ion	*m/z*	y-ion	*m/z*	y-ion	*m/z*
Q1	DUSP16	Q9BY84	VPVNDSF(C)EK	597.78	18.89	y7	899.36	y8	998.42	y9	1095.48
Q2	DUSP7	Q16829	SAEWLQEELEAR	730.86	28.85	y5	617.32	y7	874.42	y8	987.51
Q3	DUSP18	Q8NEJ0	QPSVSGLSQITK	622.85	23.43	y8	833.47	y7	746.44	y10	1019.57
Q4	DUSP5	Q16690	YVLPDEAAR	517.27	22.88	y6	658.32	y7	771.40	y4	446.24
Q5	DUSP9	Q99956	YILNVTPNLPNFFEK	904.98	32.86	y9	1105.56	y10	1206.31	y11	1305.68
Q6	DUSP18	Q8NEJ0	NTVHMVSSPVGMIPDIYEK	1059.02	tbc	y6	764.38	y10	1164.60	y11	1261.65
Q7	DUSP14	O95147	VPLADMPHAPIGLYFDTVADK	1135.58	tbc	y12	1338.69	y7	795.39	y8	958.45
Q8	ELK1	P19419	EQGNGHIISWTSR	742.87	tbc	y7	862.48	-	-	-	-
Q9	ETS1 1/2	P14921	DWVMWAVNEFSLK	812.90	tbc	y8	907.49	-	-	-	-
Q10	DUSP 14	O95147	MISEGDIGGIAQITSSLFLGR	1083.07	tbc	y11	1192.67	-	-	-	-
Q11	DUSP 2	Q05923	ELE(C)AALGTLLR	673.36	32.34	y7	743.48	y8	814.51	y9	974.55
Q12	DUSP 4	Q13115	FSSEYPEF(C)SK	690.79	23.05	y6	767.34	y7	930.40	y9	1146.48
Q13	STAT3	P40763	SIVSELAGLLSAMEYVQK	969.52	tbc	y10	1181.62	-	-	-	-
Q14	ETS2	P15036	NMDQVAPVANSYR	732.85	22.43	y7	806.42	y8	877.45	y9	976.52
Q15	DUSP 10	Q9Y6W6	SLN(C)G(C)SSAS(C)(C)TVATYDK	1070.92	19.19	y10	1204.50	y12	1362.57	y11	1275.53
Q16	DUSP9	Q99956	ASFPVQILPNLYLGSAR	923.52	33.11	y9	990.53	y10	1103.62	y11	1216.70
Q17	DUSP2	Q05923	AGPTAVYFLR	547.80	30.39	y4	598.33	y5	697.40	y6	768.44
Q18	ETS1 1	P14921	VPSYDSFDSEDYPAALPNHKPK	1239.09	tbc	y13	1479.76	y10	1072.63	y6	720.42
Q19	DUSP4	Q13115	ADISSWFMEAIEYIDAVK	1044.50	tbc	y11	1281.60	-	-	-	-
Q20	DUSP1	P28562	GGYEAFSAS(C)PEL(C)SK	881.87	24.33	y8	980.42	y9	1051.45	y10	1138.49
Q21	STAT3	P40763	LLQTAATAAQQGGQANHPTAAVVTEK	1288.68	tbc	y9	915.51	y14	1422.73	-	-
Q22	DUS10	Q9Y6W6	AANLTYMPSSSGSAR	756.86	22.18	y9	879.39	y10	1042.46	y11	1143.50
Q23	ETS1 1/2	P14921	FCMNGAALCALGK	649.81	29.14	y7	675.38	y9	803.44	y10	917.48
Q24	DUSP7	P28562	ISSDCSDGESDR	635.75	tbc	y7	765.30	-	-	-	-
Q25	DUSP1	P28562	ADISSWFNEAIDFIDSIK	1036.01	tbc	y11	1281.64	-	-	-	-
Q26	DUSP16	Q9BY84	LVALLESGTEK	580.33	25.05	y6	650.30	y7	763.38	y9	947.50
Q27	MYC	P01106	DQIPELENNEK	664.82	20.72	y7	875.41	y8	972.46	y9	1085.55
Q28	ELK1 1	P19419	GAGMAGPGGLAR	507.76	21.46	y6	570.33	y7	627.35	y8	698.39
Q29	ELK1	P19419	LVDAEEVAR	501.27	21.18	y6	674.34	y7	789.37	y5	603.3097
Q30	DUSP5	Q16690	LLQEGGGGVAAVVVLDQGSR	963.03	28.18	y11	1114.58	y7	774.41	y9	1043.58
Q31	MYC	P01106	(C)HVSTHQHNYAAPPSTR	953.44	tbc	y4	460.25	y5	557.3	y10	1113.54
Q32	ELK1 1	P19419	AEPEVPPQEGVPAR	738.38	20.68	y9	950.5054	y10	1049.573	y8	853.4526
Q33	ETS2	P15036	GGLLDSM(C)PASTPSVLSSEQEFQMFPK	1443.67	tbc	y2	244.17	y14	1656.79	y15	1753.85
Q34	Control	Control	GGVNDNEEGFFSAR	749.83	31.49	y9	1056.47	y10	1171.5	y8	942.43

4.2 Data-Independent Acquisition by MS^E to Generate Transitions

An MS^E acquisition of QconCAT LM2 was carried out using a Synapt HDMS, a Q-TOF instrument, to compare manually selected transitions with those obtained by a data-independent acquisition (DIA). An MS^E experiment (also known as a DIA) is a dual scanning experiment. The first is a low-energy scan, whereby the intact peptides ions are detected. The second scan is a high-energy scan where CID is conducted in a non-selective manner i.e. all precursor ions are submitted to the collision cell. For the high-energy scan a collision energy ramp is preferred to allow efficient dissociation of all precursor ions given that larger ions require more collision energy for dissociation. The MS^E spectra were searched with variable modifications of carbamidomethylation and methionine oxidation against the UniProt *Homo sapiens* database in ProteinLynx Global Server (PLGS). The top five most intense product ions were reported for each peptide detected and 24 out of 34 peptides of QconCAT LM2 were detected. For the majority of cases there was an exact match with the *y*-ions manually selected for SRM. However, there was one case where MS^E disagreed with the manual selection; for peptide AEPEVPPQEGVPAR two out of three transitions selected did not match, (Table 4.2). On further inspection of the Q-TOF Ultima Global tandem MS data, it can be seen that although transitions involving *y*8 and *y*10 ions were not selected by the MS^E they are however the most intense ions above the *m/z* region of the spectra greater than the *m/z* value of the precursor ion, (Figure 4.3).These results added confidence to the manual *y*-ion transitions selected.

Table 4.2 ProteinLynx Global Server output for QconCAT LM2 Q32, Elk1 isoform, from an MS^E experiment carried out on a Synapt (Waters, UK).

Peptide	Rt, mins	Prec. Mass, *m/z*	Prod. Mass, *m/z*	y ion	Intensity
AEPEVPPQEGVPAR	20.04	738.38	950.50	y9	2431
AEPEVPPQEGVPAR	20.04	738.38	638.34	y12	132
AEPEVPPQEGVPAR	20.04	738.38	499.29	y5	90
AEPEVPPQEGVPAR	20.04	738.38	343.20	y3	65

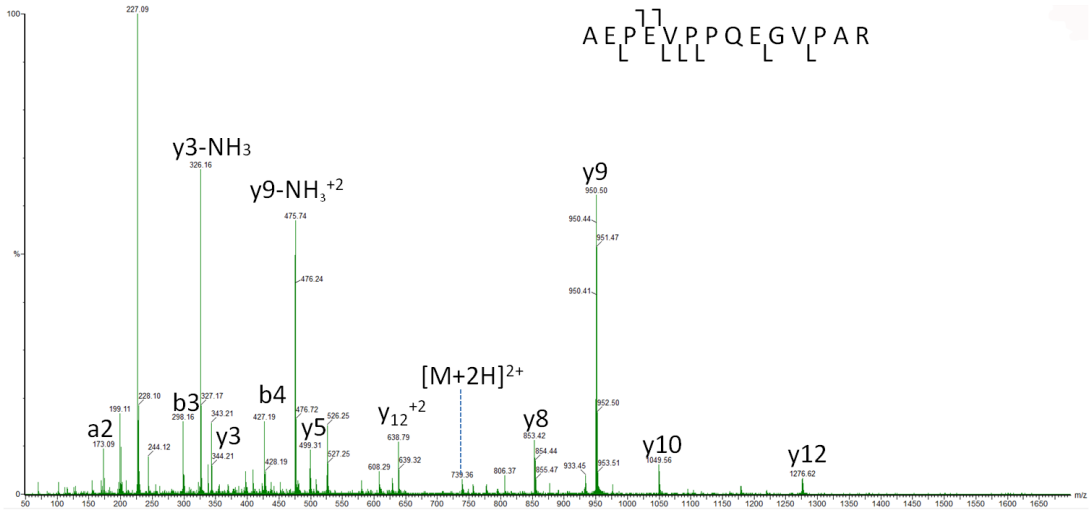


Figure 4.3 Full scan CID mass spectrum obtained using a Q-TOF Ultima Global for LM2 Q32, Elk1 isoform 1.

4.3 Cell Lysate Preparation

Cells lysate samples of HT-29 colon cancer cells (containing a mutated B-Raf) were prepared according to the paper by Hanke and co workers.[158] Lysates were digested to generate peptides for MS analysis using the Filter-Aided Sample Preparation (FASP) protocol[159] and prior to digestion were spiked with [$^{13}C_6$]-Lys/Arg labelled QconCAT LM2 to a final concentration of 3 %. Prior to analysis by mass spectrometry, samples were spiked with a 1:1 stoichiometric ratio, based upon Bradford assay of QconCAT, of internal standard GVNDNEEGFF*SAR, to QconCAT control peptide GGVNDEEGFFSAR for internal quantification of QconCAT. GVNDEEGFF*SAR is a synthetically labelled peptide where phenylalanine (F*) is labelled with [$^{13}C_9$] and [$^{15}N_7$]. The peptide sequence of the QconCAT control, GGVNDEEGFFSAR, has an additional glutamine at the start however the signal intensity is comparable.[160] Analysis of the digested cell lysate by mass spectrometry showed a large spike in signal at *m/z* 726.34 which was identified as the GVNDNEEGFF*SAR control peptide (Figure 4.4).

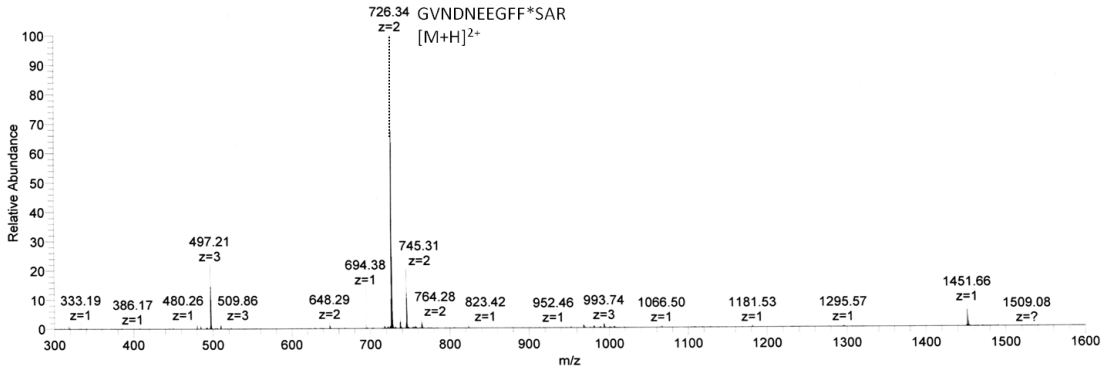


Figure 4.4 LTQ-Orbitrap XL full scan MS spectrum of a cell lysate, HT-29 with QconCAT LM2, processed by Filter-Aided Sample Preparation. GVNDNEEGFF*SAR where phenylalanine (F*) was labelled with [$^{13}C_9$] and [$^{15}N_7$] was spiked at a 1:1 stoichiometric ratio to QconCAT prior to processing by FASP.

Based on the relative signals for GVNDEEGFF*SAR and GGVNDEEGFFSAR Figure 4.4 suggests that processing cell samples by FASP results in extensive sample losses of up to 90 %. To investigate this further, 500 µl of 150 fmol/µl of trypsin digested bovine serum albumin (BSA) was mixed with 500 µl of 0.05 M ammonium bicarbonate and was added to the Amicon Ultra-0.5 centrifugal filter unit and was centrifuged at 14 000 x *g* for 2 hours. The optical density at 280 nm was measured before and after filtering through the membrane and a reduction from 0.020 a.u. to 0.004 a.u was observed. Therefore use of the Amicon Ultra-0.5 centrifugal filter unit suggests there are sample losses of up to 80 % observed by UV in the BSA experiment and up to 90 % observed by mass spectrometry.

To improve sample recovery from the filter unit, addition of an acetonitrile wash prior to eluting the desalted peptides was tested. The FASP protocol was followed (as described in the Materials and Methods section) and prior to the addition of trypsin, the sample mixture of QconCAT in whole cell lysate was then split equally between two new 3K filters. To split the mixture the 3K filter was inverted and centrifuged at 1 000 x *g* for 2 min to transfer the concentrated sample from the filter to the tube. The FASP protocol was then continued to be followed and trypsin was added for the overnight digestion. After desalting the peptides, a 500 µl wash of acetonitrile was added to one filter. An equivalent amount of the normal was solvent, 70 % acetonitrile in water was added to the other filter. Samples from both filters were then analysed in duplicate by SRM on the Xevo TQ MS and the mean Area

Under the Curve (AUC), for each transition was calculated and the percentage increase in sample recovery (Table 4.3). In all cases the addition of an acetonitrile improved sample recovery for the heavy peptides from 125.0 % to 147.0 % (Table 4.3). Therefore, use of the Amicon Ultra-0.5 centrifugal filter units results in sample losses, which is reduced by washing the filter with acetonitrile. It has since been reported by the Association of Biomolecular Resource Facilities' discussion forum (http://www.abrf.org/index.cfm/list.msg/abrf/79645 retrieved 30.11.2011) that the size and packing material of the filter had changed since the original protocol was published and this is potentially a source of the observed peptide losses. As it is extremely difficult to assess if these losses are uniform for a complex mixture, it was decided to use an alternative method of cell lysate preparation for quantitative MS analysis.

Table 4.3 Sample recovery with and without an acetonitrile (ACN) wash during the Filter-Aided Sample Preparation method for duplicate SRM analyses. Mean Area Under the Curve (AUC) was calculated and percentage recovery for each transition shown. Heavy peptides are denoted with R.

Peptide	**Prec. *m/z***	**Prod. *m/z***	**No ACN, Mean AUC**	**ACN, Mean AUC**	**Recovery, % incr.**
SAEWLQEELEAR	730.86	617.32	noise	Noise	
SAEWLQEELEAR	730.86	874.42	noise	Noise	
SAEWLQEELEAR	730.86	987.51	noise	Noise	
SAEWLQEELEA**R**	733.87	623.34	592	767	129.56
SAEWLQEELEA**R**	733.87	880.44	1799	2388	132.75
SAEWLQEELEA**R**	733.87	993.53	1498	1873	125.03
YVLPDEAAR	517.27	446.46	noise	Noise	
YVLPDEAAR	517.27	658.32	327	Noise	
YVLPDEAAR	517.27	771.40	226	Noise	
YVLPDEAA**R**	520.28	452.26	5618	8199	145.95
YVLPDEAA**R**	520.28	664.34	30505	43273	141.86
YVLPDEAA**R**	520.28	777.42	13703	20145	147.02
LVDAEEVAR	501.27	603.30	noise	Noise	
LVDAEEVAR	501.27	674.34	noise	Noise	
LVDAEEVAR	501.27	789.37	noise	23909	
LVDAEEVA**R**	504.28	609.32	noise	4488	129.09
LVDAEEVA**R**	504.28	680.36	noise	6921	144.35
LVDAEEVA**R**	504.28	795.39	noise	41307	134.31

4.4 In-solution digestion with RapiGest

An alternative method of preparing the cell lysates involved sonicating the cell pellets prior to an in-solution digest with RapiGest (Waters, UK). RapiGest is a surfactant that helps to solubilise proteins making them more susceptible to proteolytic cleavage; therefore the efficiency of digestion is increased. It was found that when using RapiGest with our cell lysates a white precipitate of RapiGest formed which is due to the hydrolysis of the RapiGest molecule under acidic conditions. At pH 2 the ester bond is hydrolysed releasing insoluble product is released which is normally removed by centrifugation, resulting in a detergent free sample that can be subjected to LC-MS analysis. The recommended protocol, by Waters, is to incubate at 37 °C for 45 minutes followed by centrifugation. However, this did not remove the RapiGest sufficiently to allow LC-MS analysis. If the samples were loaded onto a HPLC column then column blockages or loss of LC resolution may result. The protocol was therefore revised and the sample incubated at 37 °C for two hours, followed by further incubation at 4 °C for two hours and then centrifugation at 14000 x *g* at 4 °C for 15 minutes. The centrifugation step worked better and pelleted the insoluble RapiGest degradation products allowing collection of the supernatant for LC-MS analysis.

Samples processed by in-solution digestion with and without RapiGest were compared using an MS^E experiment carried out using the Synapt HDMS instrument. The MS^E spectra were searched with variable modifications of carbamidomethylation and methionine oxidation against the UniProt *Homo Sapiens* database in Protein Lynx Global Server PLGS. For the standard in-solution digestion, 5960 peptides had been identified. For the modified protocol using RapiGest there was increased peptide coverage as 9576 peptides had been identified and also the number of unalkylated cysteine residues was also decreased. This improved RapiGest protocol was therefore used to prepare all samples for final SRM analysis.

4.5 Initial Cell Lysate Experiments

Mammalian cell lysates are complex samples to analyse by mass spectrometry, and therefore it is necessary to carry out protein or peptide separation prior to the sample entering the mass spectrometer. By separating the analyte species, the signal interference can be reduced and thus an increased number of peptides can be detected. Generally reversed-phase separation of peptides based on their hydrophobicity is coupled on-line to a mass spectrometer. Here, the cell lysate of HT-29 colon cancer cells mixed with labelled QconCAT LM2 and underwent Filter Aided Sample Preparation (FASP, see methods section) prior to separation by LC and SRM analysis by Xevo TQ MS. The experiment was repeated (n=2) and the Area Under the Curve (AUC) calculated for each signal obtained by SRM. For each pair of light to heavy *y*-ion transitions an AUC light:heavy ratio was calculated. For each of these L:H ratios the mean, standard deviation and CV was calculated. Selected results are shown in Table 4.4 and data for the rest of the peptides are presented in Appendix 1.

Preliminary analyses showed that for several peptides light and heavy signals at the same retention time can be observed and therefore a L:H ratio for each transition can be calculated using the AUC. It was expected that the light: heavy (L:H) ratios of each *y*-ion transition for a peptide would be in agreement, for example, peptide YVLPDEAAR ratios *y6*, *y7* and *y4* are in agreement at 0.01. However this is not the case for other peptides. For example, in the case of peptide QPSVSGLSQITK, the *y8* ratio is 0.017 and the *y7* ratio is 1.88. For these situations it is most likely that the transition is not selective for the peptide of interest and another species is contributing to the signal which, given the complexity of the sample, is possible. For Q-peptides where only a heavy signal is observed, native proteins may be present at such low levels that their peptides are below the limit of detection. No light or heavy signals were observed in these SRM experiments for peptides previously not detected by Q-TOF or MALDI-TOF mass spectrometry. In addition, no light or heavy signals were observed for SLNCGCSSASCCTVATYDK, ASFPVQILPNLYLGSAR and FCMNGAALCALGK which had been previously been detected by Q-TOF-MS. However, these peptide signals were detected in pure QconCAT mixtures rather than in lysate which serves to highlight the importance of evaluating peptides in the true complex biological background. The mammalian cell

is very complex and it is likely that the loss of these heavy peptides is due to competitive ionisation between the analyte and any number of co-eluting matrix compounds and therefore the ionisation efficiency of the analyte will be decreased.[164,165] For highly complex samples, during the generation of positive ions by ESI for detection under positive ion mode it is likely that there is a limited amount of charge available on the ESI droplets. This causes competition for charge and in turn affects the amount of charged ions in the gas phase leading to suppression of the signal.

Table 4.4 shows that the chromatographic retention times for each *y*-ion for both heavy and light peptides are in agreement with each other. Co-elution of light with heavy peptides verifies the native peptide identity in complex mixtures and improves the selectivity of a SRM assay. In addition, by knowing the retention time scheduled SRMs can be used whereby each transition is monitored during a defined window of time around each retention time. Multiplexing SRMs in this way allows for an increase in dwell time for each ion and improvements in signal-to-noise and thus limit of quantification. However, it is still important to assess what dwell time is optimal for data collection of mammalian whole cell lysates using the Xevo TQ MS and this will be explained below (section 4.7).

Table 4.4 Results for selected peptides of SRM analysis using Xevo TQ MS of HT-29 colon cancer cells mixed with labelled QconCAT LM2 and sample preparation by Filter Aided Sample Preparation. AUC, Area under the Curve; L:H, light to heavy ratio; SD, standard deviation; CV, Coefficient of Variation; labelled amino acids are in bold.

Peptide	Prec. *m/z*	y ion	Prod. *m/z*	Rt, mins	AUC	L:H Ratio	Rt, mins	AUC	L:H Ratio	L:H Mean	SD	CV,%
YVLPDEAAR	517.27	y6	658.32	31.46	1488	-	31.6	1441				
YVLPDEAAR		y7	771.4	31.48	515	-	31.58	464				
YVLPDEAAR		y4	446.24	31.48	201	-	31.65	240				
YVLPDEAA**R**	520.28	y6	664.34	31.46	130780	0.011	31.6	100000	0.012	0.011	0.000	2.71
YVLPDEAA**R**		y7	777.42	31.48	55316	0.009	31.6	48380	0.01	0.0095	0.000	2.09
YVLPDEAA**R**		y4	452.26	31.48	19926	0.01	31.6	17750	0.014	0.0118	0.002	20.57
NMDQVAPVANSYR	732.85	y7	806.42	31.48	146	-	31.86	108				
NMDQVAPVANSYR		y8	877.45	31.46	123	-	31.79	50				
NMDQVAPVANSYR		y9	976.52	31.53	22	-	31.72	10				
NMDQVAPVANSY**R**	735.86	y7	812.44	31.5	15160	0.01	31.82	9633	0.011	0.01	0.001	10.73
NMDQVAPVANSY**R**		y8	883.47	31.5	11529	0.011	31.82	7864	0.006	0.009	0.003	35.8
NMDQVAPVANSY**R**		y9	982.54	31.5	3573	0.006	31.82	2368	0.004	0.005	0.001	26.35
AGPTAVYFLR	547.8	y4	598.33	40.96	342	-	41.2	174	0			
AGPTAVYFLR		y5	697.4	40.91	344	-	41.2	283	0			
AGPTAVYFLR		y6	768.44	40.93	236	-	41.17	267	0			
AGPTAVYFL**R**		y4	604.35	40.93	35209	0.01	41.2	23780	0.007	0.008	0.002	19.9
AGPTAVYFL**R**		y5	703.42	40.93	34360	0.01	41.2	23994	0.012	0.01	0.001	11.56
AGPTAVYFL**R**		y6	774.46	40.93	30835	0.008	41.2	18179	0.015	0.011	0.005	44.52
LVDAEEVAR	501.27	y6	674.34	28.12	440	-	28.22	530				
LVDAEEVAR		y7	789.37	28.14	1223	-	28.31	2006				
LVDAEEVAR		y5	603.31	28.14	0	-	28.34	267				
LVDAEEVA**R**	504.28	y6	680.36	28.12	23410	0.019	28.31	25122	0.021	0.019	0.002	8.16
LVDAEEVA**R**		y7	795.39	28.12	119382	0.01	28.31	100000	0.016	0.013	0.004	29.22
LVDAEEVA**R**		y5	609.33	28.12	16768	0	28.31	17492	0.015	0.007	0.010	141.42

4.6 Effect of Mass Resolution on Reproducibility of Data

The mass resolution of a mass spectrometer is a measure of its ability to separate ions of adjacent masses. Typically mass resolution is determined from a single peak and expressed as m/Δm, where m is the *m/z* centroid of the peak and Δm is the width of the peak at full-width-half-maximum (FWHM). As the mass resolution is increased the peak shape becomes narrower and within a set transmission window a larger proportion of ions contribute to the signal (Figure 4.5).

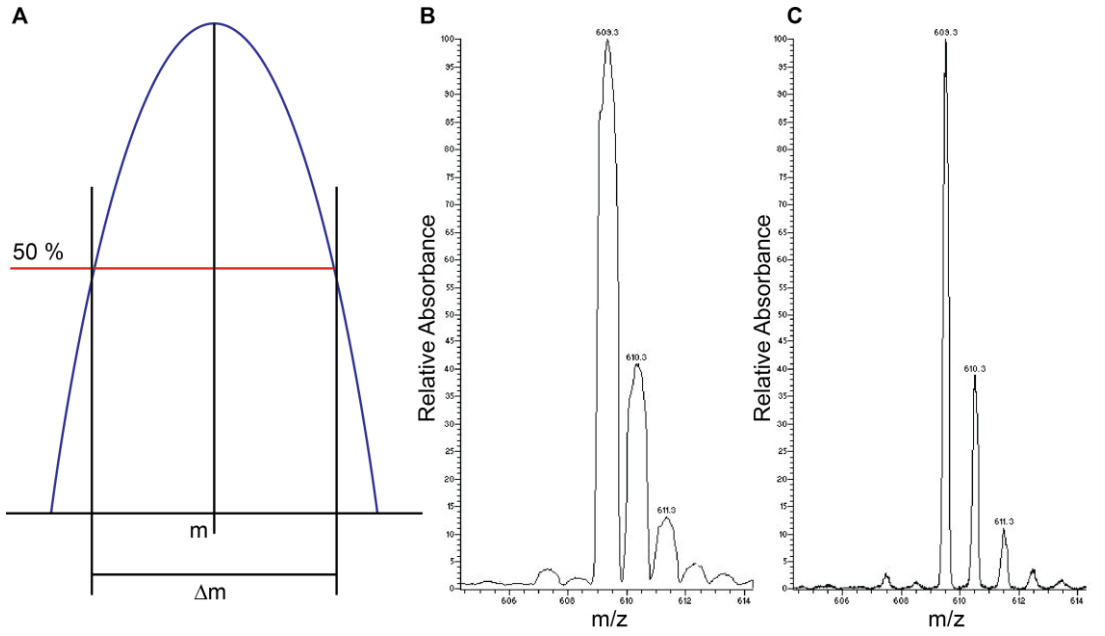


Figure 4.5 (A) mass resolution determined by m/Δm. (B) and (C) show the effect of increasing mass resolution on peak shape, where (B) has 0.70 *m/z* FWHM and (C) at 0.20 *m/z* FWHM.

Preliminary SRM analyses of whole cell lysates spiked with QconCAT were undertaken at unit mass resolution which is 0.70 *m/z*, for quadruple 1 and for quadruple 2. The data showed that for the same sample the ratio of L:H for the three *y*-ion transitions selected per peptide were not the same for each injection or between injections on the Xevo TQ MS. The CVs between three replicate injections for each *y*-ion L:H ratio were calculated and for the peptides shown these vary from 14.5 % to 173.2 %; ideally there should not be such large variation of the data. It is possible that the variation is due either unstable spray or insufficient dwell time.

Table 4.5 Results for selected peptides of SRM analysis using Xevo TQ MS of HT-29 colon cancer cells mixed with labelled QconCAT LM2 and sample preparation by Filter-Aided Sample Preparation. AUC, Area under the Curve; L:H, light to heavy ratio; SD, standard deviation; Coefficient of Variation; labelled amino acids are in bold.

Peptide	Prec. *m/z*	y ion	Prod. *m/z*	Rt, min	AUC	L:H Ratio	Rt, min	AUC	L:H Ratio	Rt, min	AUC	L:H Ratio	AUC, Mean	SD	CV, %
LVDAEEVAR	501.27	y6	674.34	28.12	440	-	28.22	530	-	28.41	278	-			
LVDAEEVAR		y7	789.37	28.14	1223	-	28.31	2006	-	28.39	1295	-			
LVDAEEVAR		y5	603.31	0	0	-	28.34	267	-	28.41	0	-			
LVDAEEVA**R**	504.28	y6	680.36	28.12	23410	0.019	28.31	25122	0.021	28.39	12093	0.023	0.013	0.012	87.0
LVDAEEVA**R**		y7	795.39	28.12	119382	0.010	28.31	128763	0.016	28.39	57482	0.023	0.009	0.008	92.0
LVDAEEVA**R**		y5	609.33	28.12	16768	0.000	28.31	17492	0.015	28.39	10061	0.000	0.005	0.009	173.2
NMDQVAPVANSR	732.85	y7	806.42	31.48	146	-	31.86	108	-	32.01	50	-			
NMDQVAPVANSR		y8	877.45	31.46	123	-	31.79	50	-	32.03	30	-			
NMDQVAPVANSR		y9	976.52	31.53	22	-	31.72	10	-	31.86	0	-			
NMDQVAPVANS**R**	735.86	y7	812.44	31.5	15160	0.010	31.82	9633	0.011	31.96	6901	0.007	0.009	0.002	21.3
NMDQVAPVANS**R**		y8	883.47	31.5	11529	0.011	31.82	7864	0.006	31.98	5963	0.005	0.007	0.003	40.1
NMDQVAPVANS**R**		y9	982.54	31.5	3573	0.006	31.82	2368	0.004	31.98	1712	0.000	0.003	0.001	39.5
QPSVSGLSQITK	622.85	y8	833.47	33.66	96	-	33.73	40	-	34.1	80	-			
QPSVSGLSQITK		y7	746.44	33.95	3628	-	34.09	3285	-	34.26	1627	-			
QPSVSGLSQITK		y10	1019.57	Noise		-	noise		-	noise	noise	-			
QPSVSGLSQIT**K**	625.86	y8	839.49	33.57	4597	0.021	33.8	2970	0.013	34.07	2034	0.039	0.025	0.013	54.2
QPSVSGLSQIT**K**		y7	752.46	33.59	2538	1.429	33.8	1747	1.880	34.1	735	2.210	1.841	0.394	21.4
QPSVSGLSQIT**K**		y10	1025.59	33.59	667	-	33.81	410	-	34.07	333	-			
AGPTAVYFLR	547.80	y4	598.33	40.96	342	-	41.2	174	-	41.39	88	-			
AGPTAVYFLR		y5	697.40	40.91	344	-	41.2	283	-	41.44	129	-			
AGPTAVYFLR		y6	768.44	40.93	236	-	41.17	267	-	41.43	59	-			
AGPTAVYFL**R**	550.81	y4	604.35	40.93	35209	0.010	41.2	23780	0.007	41.43	15007	0.006	0.008	0.002	25.5
AGPTAVYFL**R**		y5	703.42	40.93	34360	0.010	41.2	23994	0.012	41.43	14568	0.009	0.010	0.001	14.5
AGPTAVYFL**R**		y6	774.46	40.93	30835	0.008	41.2	18179	0.015	41.44	13325	0.004	0.009	0.005	58.8

Therefore to investigate this further, the effect of changing the mass accuracy of the quadrupoles 1 and 3 were considered. Peptides were monitored under a combination of different resolutions for each quadrupole (Table 4.6) and the experimental data was used to assess the precision of the peak area recorded over multiple injections of the same sample, therefore assessing reproducibility of the instrument.

Table 4.6 Mass resolution combinations for quadrupoles 1 and 3 of the Xevo TQ MS

Quadrupole 1, *m/z*	Quadrupole 3, *m/z*
0.70	0.70
0.70	0.50
0.50	0.70
0.50	0.50

Inspection of the data shows that as transmission window increases from 0.70 *m/z* to 0.50 *m/z* there is an increase in baseline noise and many ions are undetectable therefore these transitions are unacceptable for quantification, Table 4.7. There is no general improvement to CVs for replicate measurements and these remain inconsistent under different quadrupole resolutions. In addition there is little agreement of each individual light to heavy *y*-ion ratio calculated for the same peptide precursor across the different mass resolutions Tables 4.7 – 4.10.

Table 4.7 Quadrupoles set to 0.70 and 0.5 *m/z*. Results for selected peptides of SRM analysis using Xevo TQ MS of HT-29 colon cancer cells mixed with labelled QconCAT LM2 and sample preparation by Filter Aided Sample Preparation. AUC, Area under the Curve; L:H, light to heavy ratio; SD, standard deviation; Coefficient of Variation; labelled amino acids are in bold Grey indicates a signal on the edge of acceptability.

Peptide	Prec m/z	y ion	Prod m/z	AUC	Ratio	AUC	Ratio	AUC	Ratio	Mean	SD	CV,%
YVLPDEAAR	517.27	y6	658.32	1292		951		1007				
YVLPDEAAR		y7	771.40	469		285		452				
YVLPDEAAR		y4	446.24	49		69		126				
YVLPDEAA**R**	520.28	y6	664.34	102761	0.013	84301	0.011	80164	0.013	0.012	0.001	6.12
YVLPDEAA**R**		y7	777.42	48134	0.010	37830	0.008	36603	0.012	0.010	0.002	24.41
YVLPDEAA**R**		y4	452.26	20515	0.002	16535	0.004	15811	0.008	0.005	0.003	58.84
NMDQVAPVANSYR	732.85	y7	806.42	164		161		93				
NMDQVAPVANSYR		y8	877.45	125		87		72				
NMDQVAPVANSYR		y9	976.52	noise		noise		noise				
NMDQVAPVANSY**R**	735.86	y7	812.44	14668	0.011	12230	0.013	11012	0.008	0.011	0.002	21.68
NMDQVAPVANSY**R**		y8	883.47	13800	0.009	10131	0.009	9448	0.008	0.008	0.001	8.70
NMDQVAPVANSY**R**		y9	982.54	3884	0.000	2783	0.000	2690	0.000			
QPSVSGLSQITK	622.85	y8	833.47	noise		noise		noise				
QPSVSGLSQITK		y7	746.44	4825		3748		3882				
QPSVSGLSQITK		y10	1019.57	noise		noise		noise				
QPSVSGLSQIT**K**	625.86	y8	839.49	5006		2541		1231				
QPSVSGLSQIT**K**		y7	752.46	2573	1.88	1290	2.91	473	8.21	4.33	3.40	78.48
QPSVSGLSQIT**K**		y10	1025.59	659		341		112				

Table 4.8 Xevo TQ MS quadrupoles set to 0.70 and 0.70 *m/z*. Results for selected peptides of SRM analysis using Xevo TQ MS of HT-29 cells mixed with labelled QconCAT LM2 and sample preparation by Filter Aided Sample Preparation. AUC, Area under the Curve; L:H, light to heavy ratio; SD, standard deviation; Coefficient of Variation; labelled amino acids are in bold.

Peptide	Prec m/z	y ion	Prod m/z	AUC	Ratio	AUC	Ratio	AUC	Ratio	Mean	SD	CV,%
YVLPDEAAR	517.27	y6	658.32	1202		1332		1314				
YVLPDEAAR		y7	771.40	565		542		468				
YVLPDEAAR		y4	446.24	116		143		183				
YVLPDEAA**R**	520.28	y6	664.34	107266	0.011	122000	0.011	103761	0.013	0.012	0	8.071
YVLPDEAA**R**		y7	777.42	48609	0.012	53944	0.010	47054	0.010	0.011	0	8.924
YVLPDEAA**R**		y4	452.26	18623	0.006	23125	0.006	19362	0.009	0.007	0	25.71
NMDQVAPVANSYR	732.85	y7	806.42	225		180		163				
NMDQVAPVANSYR		y8	877.45	184		114		62				
NMDQVAPVANSYR		y9	976.52	noise		noise		noise				
NMDQVAPVANSY**R**	735.86	y7	812.44	16024	0.014	19531	0.009	15020	0.011	0.011	0	21.58
NMDQVAPVANSY**R**		y8	883.47	14452	0.013	16663	0.007	12355	0.005	0.008	0	49.18
NMDQVAPVANSY**R**		y9	982.54	4139	0.000	4985	0.000	3600	0.000			
QPSVSGLSQITK	622.85	y8	833.47	noise		noise		noise				
QPSVSGLSQITK		y7	746.44	4487		5084		4273				
QPSVSGLSQITK		y10	1019.57	noise		noise		noise				
QPSVSGLSQIT**K**	625.86	y8	839.49	5012		4558		2284				
QPSVSGLSQIT**K**		y7	752.46	2383	1.88	2485	2.05	1056	4.05	2.66	1.2	45.32
QPSVSGLSQIT**K**		y10	1025.59	578		542		342				

Table 4.9 Quadrupoles set to 0.5 and 0.70 *m/z*. Results for selected peptides of SRM analysis using Xevo TQ MS of HT-29 colon cancer cells mixed with labelled QconCAT LM2 and sample preparation by Filter Aided Sample Preparation. AUC, Area under the Curve; L:H, light to heavy ratio; SD, standard deviation; Coefficient of Variation; labelled amino acids are in bold Grey indicates a signal on the edge of acceptability.

Peptide	Prec m/z	y ion	Prod m/z	AUC	Ratio	AUC	Ratio	AUC	Ratio	Mean	SD	CV,%
YVLPDEAAR	517.27	y6	658.32	1055		1051		904				
YVLPDEAAR		y7	771.40	531		469		425				
YVLPDEAAR		y4	446.24	93		100		58				
YVLPDEAA**R**	520.28	y6	664.34	93857	0.011	93881	0.011	79709	0.011	0.011	0.000	0.66
YVLPDEAA**R**		y7	777.42	40366	0.013	43359	0.011	37190	0.011	0.012	0.001	10.28
YVLPDEAA**R**		y4	452.26	18252	0.005	18468	0.005	15300	0.004	0.005	0.001	18.05
NMDQVAPVANSYR	732.85	y7	806.42	148		124		87				
NMDQVAPVANSYR		y8	877.45	116		73		73				
NMDQVAPVANSYR		y9	976.52	noise		noise		83				
NMDQVAPVANSY**R**	735.86	y7	812.44	14056	0.011	12982	0.010	9805	0.009	0.010	0.001	8.63
NMDQVAPVANSY**R**		y8	883.47	12431	0.009	11159	0.007	8671	0.008	0.008	0.001	17.57
NMDQVAPVANSY**R**		y9	982.54	3244	0.000	3124	0.000	2419	0.034			
QPSVSGLSQITK	622.85	y8	833.47	noise		noise		noise				
QPSVSGLSQITK		y7	746.44	3849		4072		3814				
QPSVSGLSQITK		y10	1019.57	noise		noise		noise				
QPSVSGLSQIT**K**	625.86	y8	839.49	3746		2602		1097				
QPSVSGLSQIT**K**		y7	752.46	1991	1.93	1305	3.12	404	9.44	4.83	4.04	83.53
QPSVSGLSQIT**K**		y10	1025.59	459		286		146				

Table 4.10 Quadrupoles set to 0.5 and 0.5 *m/z*. For each peptide a star indicated the labelled version. Results for selected peptides of SRM analysis using Xevo TQ MS of HT-29 colon cancer cells mixed with labelled QconCAT LM2 and sample preparation by Filter Aided Sample Preparation. AUC, Area under the Curve; L:H, light to heavy ratio; SD, standard deviation; Coefficient of Variation; labelled amino acids are in bold Grey indicates a signal on the edge of acceptability.

Peptide	m/z	y ion	m/z	AUC	Ratio	AUC	Ratio	AUC	Ratio	Mean	SD	CV,%
YVLPDEAAR	517.27	y6	658.32	1004		972		883				
YVLPDEAAR		y7	771.40	565		307		352				
YVLPDEAAR		y4	446.24	208		128		103				
YVLPDEAA**R**	520.28	y6	664.34	101409	0.010	98121	0.010	73668	0.012	0.011	0.001	11.35
YVLPDEAA**R**		y7	777.42	43103	0.013	42106	0.007	31758	0.011	0.010	0.003	28.14
YVLPDEAA**R**		y4	452.26	19044	0.011	18129	0.007	14857	0.007	0.008	0.002	27.30
NMDQVAPVANSYR	732.85	y7	806.42	115		118		66				
NMDQVAPVANSYR		y8	877.45	noise		81		noise				
NMDQVAPVANSYR		y9	976.52	noise		noise		noise				
NMDQVAPVANSY**R**	735.86	y7	812.44	14334	0.008	13873	0.009	8973	0.007	0.008	0.001	7.26
NMDQVAPVANSY**R**		y8	883.47	11626	0.000	11884	0.007	7184	0	0.000	0.000	0.00
NMDQVAPVANSY**R**		y9	982.54	3419	0.000	3003	0	1933	0			
QPSVSGLSQITK	622.85	y8	833.47	noise		noise		noise				
QPSVSGLSQITK		y7	746.44	3754		3579		3470				
QPSVSGLSQITK		y10	1019.57	noise		noise		noise				
QPSVSGLSQIT**K**	625.86	y8	839.49	4431		3395		1267				
QPSVSGLSQIT**K**		y7	752.46	2122	1.77	1689	2.12	411	8.44	4.11	3.76	91.38
QPSVSGLSQIT**K**		y10	1025.59	591		356		103				

Although it is expected that as mass resolution increases a good signal, i.e. Gaussian peak shape and signal-to-noise above 10 should be obtained for each transition, this is not the case. The original quadrupole settings of 0.70 *m/z* and 0.70 *m/z* for quadrupole 1 and 2 gave the best signal-to-noise ratios for each transition. The poorer performance with higher mass resolution is likely to be due to reduced ion current getting to the detector as a result of the narrower mass resolutions on the quadrupoles. As the CVs for each transition are very high, alternative instrument settings or sample preparation should be optimised to reduce the CVs to achieve the best data possible for accurate quantification. In the next section, the effect of dwell time on reproducibility of data will be explored. Alternatively, it may be the sample itself causing these inconsistencies and the complexity maybe too much to obtain good quality data. Therefore it may be necessary to perform another separation step prior to mass spectrometric analysis either at the protein or peptide level e.g. OFF-GEL to reduce sample complexity and extend the dynamic range of the ultimate LC-MS analysis.

4.7 Effect of Dwell Time on Reproducibility of Data

Conditions for SRM analysis were optimised for use on a Xevo TQ MS, coupled to a nanoACQUITY UPLC system. Initially, a dwell time of 0.028 seconds was used to monitor transitions of tryptic peptides derived from osteosarcoma U-2 OS cell lysate. The CVs for the L:H peptide ratios were calculated over three replicates and showed that the data was outside of acceptable limits for CVs (up to 49.2 %), Table 4.11.

To improve the signal-to-noise ratio of the peptide signal, transitions were monitored using an increased dwell time of 0.161 seconds (Table 4.12). As the dwell time increased the CV for each *y*-ion transition reduced. For example, L:H ratios for the peptide NMDQVAPVANSYR gave CVs of up to 49.2 % at 0.028 seconds and as the dwell time increased to 0.161 seconds the CV reduced to between 1.5 - 6.5 % for *y*-ion transitions. At 0.028 seconds the *y9* transition for the light version of this peptide was not detectable as the signal was noisy. However, at 0.161 seconds a signal was obtained which when the calculated the L:H ratio for this transition is similar to those of *y8* and *y7*, therefore signal-to-noise is increased thus improving the lower limit of detection.

Table 4.11 SRM results obtained using a Xevo TQ MS for selected peptides of QconCAT LM2 mixed with U-2 OS cells of selected light and heavy peptides used to calculate ratios for Area Under the Curve (AUC), where H = heavy and L = light, CVs calculated after 3 replicate injections using a dwell time of 0.028 s.

Peptide	Prec m/z	y ion	Prod m/z	AUC	L:H Ratio	AUC	L:H Ratio	AUC	L:H Ratio	AUC, Mean	SD	CV,%
YVLPDEAAR	517.27	y6	658.32	1202		1332		1314				
YVLPDEAAR		y7	771.40	565		542		468				
YVLPDEAAR		y4	446.24	116		143		183				
YVLPDEAA**R**	520.28	y6	664.34	107266	0.011	122000	0.011	103761	0.013	0.012	0.001	8.1
YVLPDEAA**R**		y7	777.42	48609	0.012	53944	0.010	47054	0.010	0.011	0.001	8.9
YVLPDEAA**R**		y4	452.26	18623	0.006	23125	0.006	19362	0.009	0.007	0.002	25.7
NMDQVAPVANSYR	732.85	y7	806.42	225		180		163				
NMDQVAPVANSYR		y8	877.45	184		114		62				
NMDQVAPVANSYR		y9	976.52	noise		noise		noise				
NMDQVAPVANSY**R**	735.86	y7	812.44	16024	0.014	19531	0.009	15020	0.011	0.011	0.002	21.6
NMDQVAPVANSY**R**		y8	883.47	14452	0.013	16663	0.007	12355	0.005	0.008	0.004	49.2
NMDQVAPVANSY**R**		y9	982.54	4139	0.000	4985	0.000	3600	0.000			
QPSVSGLSQITK	622.85	y8	833.47	noise		noise		noise				
QPSVSGLSQITK		y7	746.44	4487		5084		4273				
QPSVSGLSQITK		y10	1019.57	noise		noise		noise				
QPSVSGLSQIT**K**	625.86	y8	839.49	5012		4558		2284	-	-	-	-
QPSVSGLSQIT**K**		y7	752.46	2383	1.883	2485	2.046	1056	4.046	2.658	1.205	45.3
QPSVSGLSQIT**K**		y10	1025.59	578		542		342	-	-	-	-

Table 4.12 SRM results obtained using a Xevo TQ MS for selected peptides of QconCAT LM2 in U-2 OS cell lysate, pairs of light and heavy peptides used to calculate ratios for Area Under the Curve (AUC), where H = heavy and L = light, CVs calculated after 3 injections. Dwell time 0.161 s.

Peptide	Prec m/z	y ion	Prod m/z	AUC	L:H Ratio	AUC	L:H Ratio	AUC	L:H Ratio	AUC, Mean	SD	CV,%
YVLPDEAAR	517.27	y6	658.32	3989		3464		3559				
YVLPDEAAR		y7	771.40	1830		1609		1652				
YVLPDEAAR		y4	446.24	604		513		553				
YVLPDEAA**R**	520.28	y6	664.34	324154	0.012	272655	0.013	269908	0.013	0.013	0.000	3.5
YVLPDEAA**R**		y7	777.42	152102	0.012	128373	0.013	136860	0.012	0.012	0.000	2.3
YVLPDEAA**R**		y4	452.26	42321	0.014	36585	0.014	39195	0.014	0.014	0.000	0.9
NMDQVAPVANSYR	732.85	y7	806.42	779		730		733				
NMDQVAPVANSYR		y8	877.45	691		630		708				
NMDQVAPVANSYR		y9	976.52	187		161		187				
NMDQVAPVANSY**R**	735.86	y7	812.44	59413	0.013	56412	0.013	57563	0.013	0.013	0.000	1.5
NMDQVAPVANSY**R**		y8	883.47	55798	0.012	46523	0.014	54629	0.013	0.013	0.001	4.5
NMDQVAPVANSY**R**		y9	982.54	16334	0.011	15018	0.011	15319	0.012	0.011	0.001	6.5
QPSVSGLSQITK	622.85	y8	833.47	718		740		754				
QPSVSGLSQITK		y7	746.44	10768		10527		10289				
QPSVSGLSQITK		y10	1019.57	97		125		98				
QPSVSGLSQIT**K**	625.86	y8	839.49	17218	0.042	18522	0.040	17816	0.042	0.041	0.001	3.0
QPSVSGLSQIT**K**		y7	752.46	8363	1.288	8774	1.200	8447	1.218	1.235	0.046	3.7
QPSVSGLSQIT**K**		y10	1025.59	3983	0.024	3768	0.033	3732	0.026	0.028	0.005	16.6

Figure 4.6 shows the improvement in chromatogram traces of both light and heavy versions of peptide YVLPDEAAR when the dwell time is increased from 0.028 seconds to 0.161 seconds. Therefore for SRMs it is important to maximise the dwell time for each transition and by scheduling SRMs this can be optimised to give the best quality data.

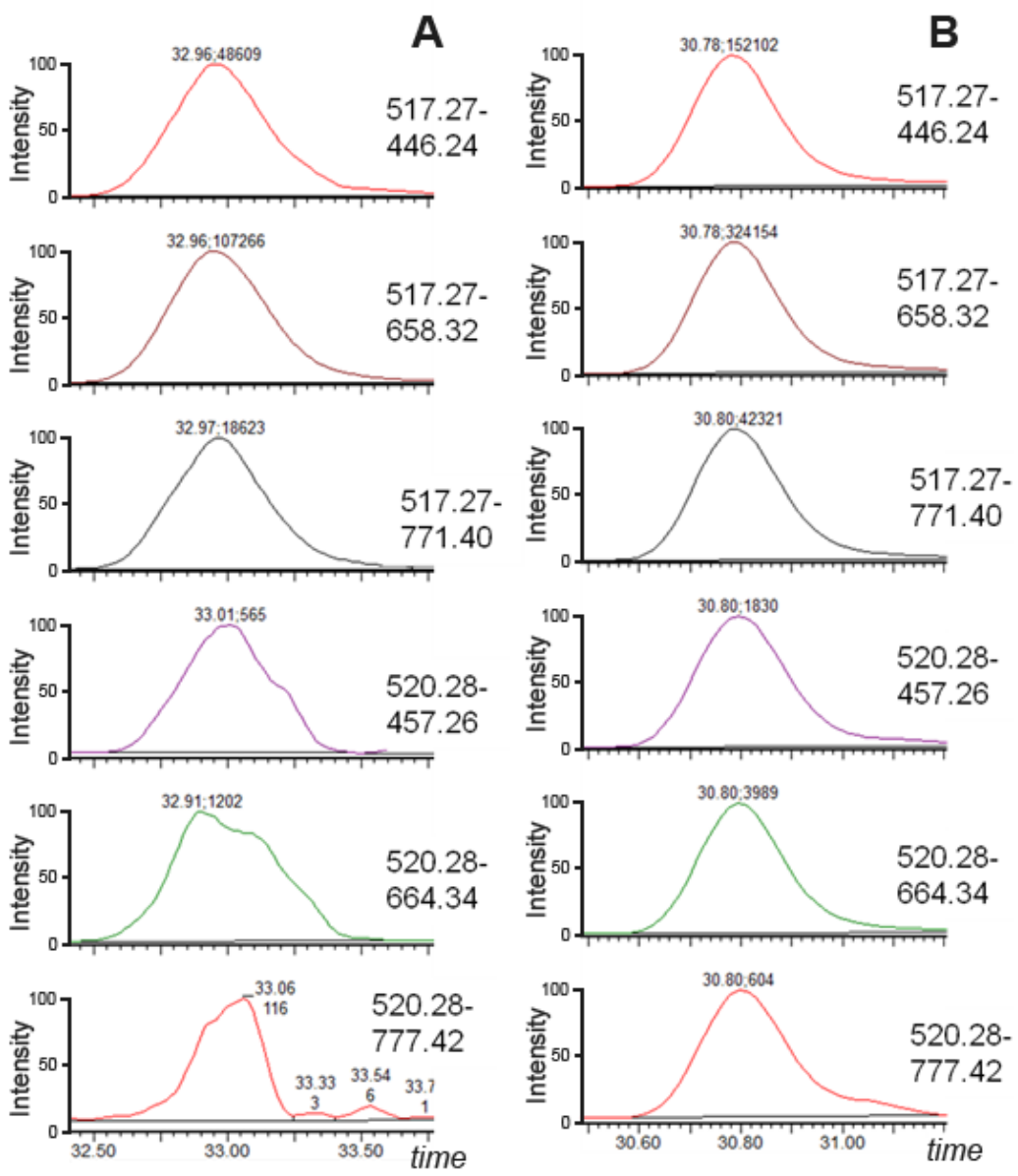


Figure 4.6 LC-SRM traces for light and heavy transitions at 0.028 s (A), and 0.161 s (B) dwell times, for of QconCAT LM2, Q4 peptide. Peaks are labelled with the retention time (min), of the eluted peptide, Area Under the Curve value and transition.

4.8 SRM Method Development for TSQ Vantage

There was limited time available on a Xevo TQ MS instrument and therefore it was necessary to transfer analyses to a TSQ Vantage based within the Manchester Centre for Integrative Systems Biology (MCISB). The established workflow within MCISB was followed to define transitions for the TSQ Vantage (Figure 4.7).

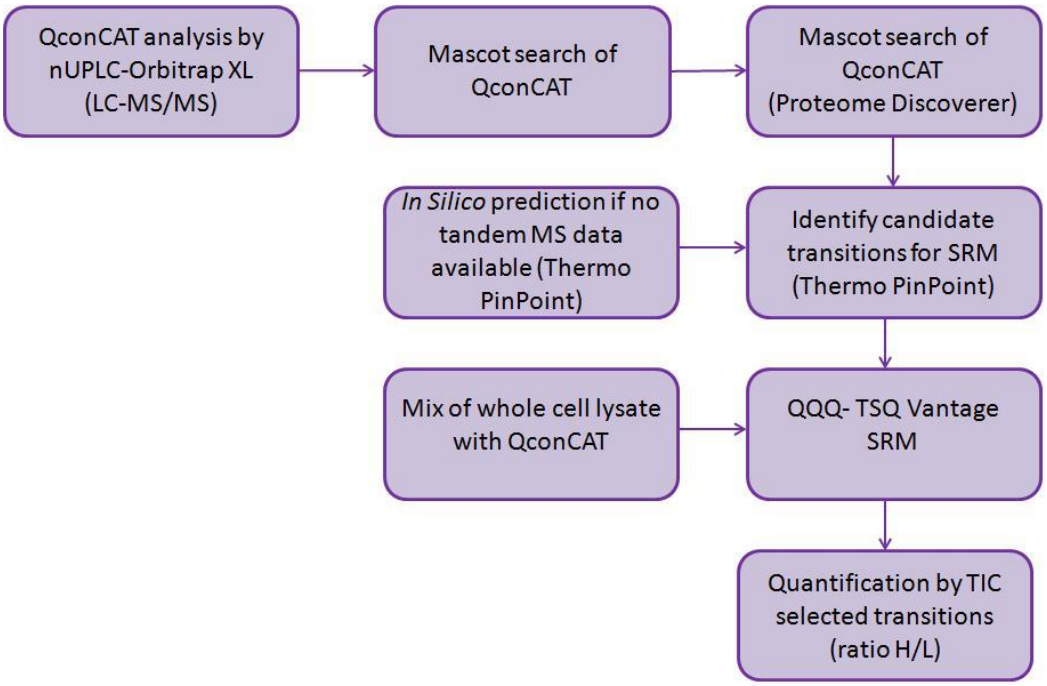
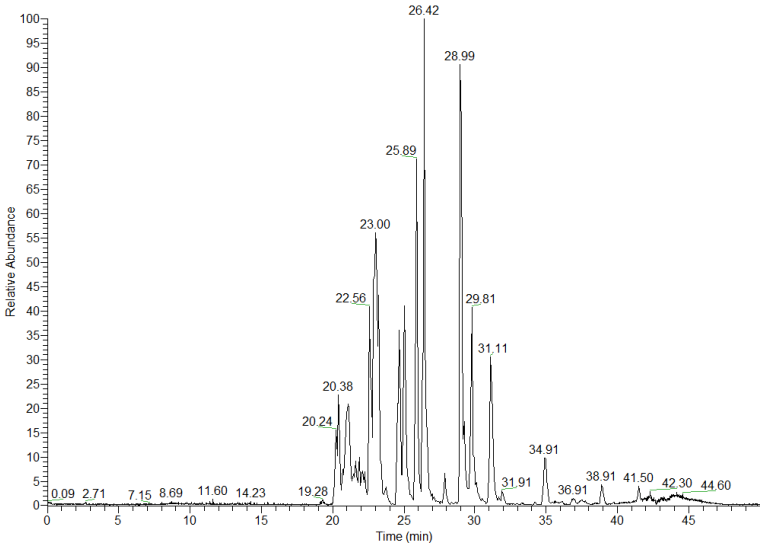


Figure 4.7 Workflow for Selected Reaction Monitoring (SRM) using the LTQ-Orbitrap XL and TSQ Vantage triple quadrupole instrument.

Initially, to define transitions for QconCAT LM2 for SRM analysis on the TSQ Vantage data-dependent acquisitions using a LTQ-Orbitrap XL were undertaken on a tryptic digest of the QconCAT proteins. Full scan tandem MS data were searched against the QconCAT database using MASCOT facilitated through the vendor supplied software Proteome Discoverer (version 1.0, build 43) using variable modifications of cysteine alkylation, methionine oxidations and one missed cleavage. For QconCAT LM2, 88.5 % sequence coverage was obtained (Figure 4.8). These MASCOT results were then imported into Pinpoint (v 1.1.12.0) to build an SRM method for the TSQ Vantage (Appendix 2).

A



B

	1	11	21	31	41	51	61	71	81	91
1	MGTKVPVNDS	FCEKSAEWLQ	EELEARQPSV	SGLSQITKYV	LPDEAARYIL	NVTPNLPNFF	EKNTVHMVSS	PVGMIPDIYE	KVPLADMPHA	PIGLYFDTVA
101	DKEQGNGHII	SWTSRDWVMW	AVNEFSLKMI	SEGDIGGIAQ	ITSSLFLGRE	LECAALGTLL	RFSSEYPEFC	SKSIVSELAG	LLSAMEYVQK	NMDQVAPVAN
201	SYRSLNCGCS	SASCCTVATY	DKASFPVQIL	PNLYLGSARA	GPTAVYFLRV	PSYDSFDSED	YPAALPNHKP	KADISSWFME	AIEYIDAVKG	GYEAFSASCP
301	ELCSKLLQTA	ATAAQQGGQA	NHPTAAVVTE	KAANLTYMPS	SSGSARFCMN	GAALCALGKI	SSDCSDGESD	RADISSWFNE	AIDFIDSIKL	VALLESGTEK
401	DQIPELENNE	KGAGMAGPGG	LARLVDAEEV	ARLLQEGGGG	VAAVVVLDQG	SRCHVSTHQH	NYAAPPSTRA	EPEVPPQEGV	PARGGLLDSM	CPASTPSVLS
501	SEQEFQMFPK	GGVNDNEEGF	FSARLAAALE	HHHHHH						

Figure 4.8 (A) Base peak chromatogram FT-MS for LM2 obtained using a LTQ-Orbitrap XL instrument, following a double trypsin digestion. (B) shows that LM2 QconCAT has a 88.5 % sequence coverage when ESI-MS data was searched against a customised database using the MASCOT algorithm

4.9 Lower Limit of Quantification Assessment

It was necessary to evaluate peptides in the complex biological background of the cell lysate and assess the lower limit of detection of the QconCAT. Ideally a 1:1 ratio of Q-peptide to endogenous peptide should be achieved to ensure accurate quantification; at ratios of approaching 100:1, the residual light peptide derived from the QconCAT will significantly contribute to the signal of the endogenous peptide, thus resulting in inaccurate quantification. Therefore cell lysate of HEK-293 was spiked with LM2 QconCAT, the column load ranged from 100 fmol to 0.005 fmol of QconCAT in a background of 150 ng of cell lysate, and was analysed by SRM on the TSQ Vantage. A good signal for each standard peptide derived from the QconCAT was obtained for the spikes at 100 fmol and 10 fmol, however below this the signal becomes noisy and would therefore be unsuitable for quantification (Figure 4.9).

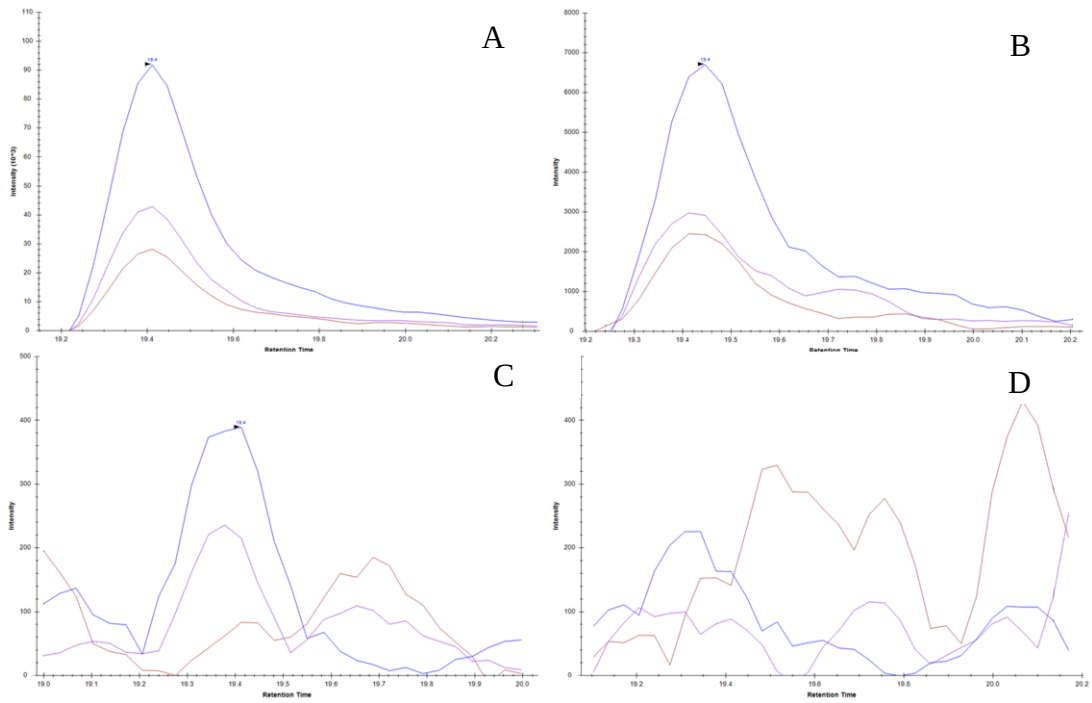


Figure 4.9 *y*-ion traces for heavy QPSVSGLSQITK peptide of DUSP 18 in 150 ng HEK-293 cell lysate, obtained on the TSQ Vantage . Transitions 625.89→582.35 (purple), 752.45 (blue), 839.47 (red). Trace A shows the standard peptide at 100 fmol on column (B) 10 fmol on column, (C) 1 fmol on column, (D) 0.1 fmol on column.

4.10 SRM Analysis of Pure LM2 QconCAT

Analysis of pure QconCAT was carried out using a Waters Xevo TQ MS to ascertain the heavy to light ratio for each QconCAT peptide and therefore assess the completeness of the isotopic labelling of the standard peptide. For all quantification data obtained within cell lines, the ratios between endogenous and heavy QconCAT protein were corrected according to Table 4.13 by subtracting any contributions made to the unlabelled signal from the light derived part of the QconCAT protein.

Table 4.13 Light to heavy ratio (L:H) for each peptide of pure LM2 QconCAT following SRM analysis by Waters Xevo TQ.

Q Peptide	Protein	Accession #	Q-Peptide Sequence	L:H
Q1	DUSP16	Q9BY84	VPVNDSFCEK	0.02
Q2	DUSP7	Q16829	SAEWLQEELEAR	0.01
Q3	DUSP18	Q8NEJ0	QPSVSGLSQITK	0.03
Q4	DUSP5	Q16690	YVLPDEAAR	0.01
Q5	DUSP9	Q99956	YILNVTPNLPNFFEK	0.03
Q11	DUSP 2	Q05923	ELECAALGTLLR	0.01
Q12	DUSP 4	Q13115	FSSEYPEFCSK	0.02
Q14	ETS2	P15036	NMDQVAPVANSYR	0.02
Q15	DUSP 10	Q9Y6W6	SLNCGCSSASCCTVATYDK	0.05
Q16	DUSP9	Q99956	ASFPVQILPNLYLGSAR	0.01
Q17	DUSP2	Q05923	AGPTAVYFLR	0.01
Q20	DUSP1	P28562	GGYEAFSASCPELCSK	0.03
Q22	DUS10	Q9Y6W6	AANLTYMPSSSGSAR	0.01
Q23	ETS1 1/2	P14921	F(C)MNGAAL(C)ALGK	0.02
Q23	ETS1 1/2	P14921	FCM(ox)NGAALCALGK	0.03
Q24	DUSP7	P28562	ISSDCSDGESDR	0.01
Q26	DUSP16	Q9BY84	LVALLESGTEK	0.03
Q27	MYC	P01106	DQIPELENNEK	0.03
Q28	ELK1 1	P19419	GAGMAGPGGLAR	0.02
Q29	ELK1	P19419	LVDAEEVAR	0.01
Q32	ELK1 1	P19419	AEPEVPPQEGVPAR	0.01
Q34	Control	Control	GGVNDNEEGFFSAR	0.01

4.11 Final Quantification Experiments

Based on the evidence provided for method development given earlier in this chapter, samples were prepared for final quantification experiments using the modified RapiGest protocol. Two biological and three technical replicates were analysed for each cell line (HCT 116, HT-29 and HEK-293). Whole cell lysates of these samples were mixed with various amounts of QconCAT LM1 and prepared for MS analysis using the modified RapiGest protocol (Materials and Methods). On-column loadings during SRM analysis were 300 ng of digested cell lysate with either

100 fmol or 10 fmol of co-digested QconCAT LM2. Cell lysates with no QconCAT addition, referred to as background samples, were also analysed with 300 ng loaded on the column. By analysing these background samples it was possible to check if there are any isobaric peptides contributing to the heavy peptide signal which may give an erroneous L:H ratio. No isobaric peptides were observed. Initially, SRM analysis proceeded with the TSQ Vantage. However due to a bug in the software, EZ tune component of Tuneplus v2.2.0Eng2, some of the scan settings were altered and therefore data collection did not proceed properly. As there was no further time available on the TSQ Vantage samples were alternatively analysed on the Xevo TQ MS. The loading amounts and ratio between lysate and QconCAT was kept constant between these two instruments.

4.11.1 Skyline generated SRM method for Xevo TQ MS

For the final analyses using the Xevo TQ MS, a SRM method for QconCAT LM2 was generated using Skyline v1.1 which is an open source software used to predict and optimally schedule SRM transitions.[161] Initially, SRM transitions (designed from Q-TOF data) were monitored over the entire LC gradient to obtain a retention time. Once LC retention times were known, a scheduled SRM assay for QconCAT LM2 was generated (Appendix 3). Skyline assigned a retention time to the peptide LLQTAATAAQQGGQANHPTAAVVTEK of STAT3. However, on manual interrogation of the data it was likely that this was a spurious assignment because the signal-to-noise ratio was poor and the retention time was assigned to the latter part of the re-equilibration phase. Furthermore, a lack of signal for this peptide was also observed by MALDI-TOF, Q-TOF and previous SRM analysis on the ThermoFisher TSQ Vantage instrument. Loss of signal for this peptide can be due to a missed cleavage owing to the glutamic acid residue next to the lysine. Siepen *et al*. proposed that an acidic sidechain adjacent to the cleavage site can form an intramolecular salt bridge to lysine, thus inhibiting its interaction with the P1 pocket of trypsin.[163] Therefore this peptide was omitted from the final analysis to increase the dwell time per transition for the remaining peptides and therefore the quality of data for quantification is optimised.

4.11.2 Final Results

The results of absolute quantification experiments such as QconCAT methodology used with SRM analyses it is possible to apply a classification to the peptides quantification.[157] For the data provided a "Type A" quantification is where both native peptide and Q-peptide are detected. "Type B" quantification is where Q-peptide can be observed but native peptide is absent. This allows for an upper limit to be placed on the quantification of that peptide and therefore protein. "Type C" quantification is where neither native peptide nor Q-peptide could be detected. Absence of both native and Q-peptide can be due to selection of a peptide with poor chromatographic properties, poor ionisation despite *in silico* predictions or unsuitable product ions to monitor either in terms of signal intensities or selectivity. The classification strategy described above can also be applied to protein quantification. Thus, when two Q-peptides are used for each protein which is known as the QconCAT replication level (QRL), the highest confidence level for quantification would be when both peptides are classified as Type A.[166]

The majority of peptides monitored were classified at Type B, where only standard peptide derived from the QconCAT could be observed (Figure 4.10 - 4.11 and Appendix 6). However, this does allow us to define an upper limit of copies per cell for proteins where the heavy peptide was detected. Figure 4.10 panel (A) shows that a calculation of the ratio between light and heavy might be possible and thus quantification could be obtained for this protein. However inspection of the *y*-ion transitions (Figure 4.10 panel (C)) for the light peptide shows that although the shape and retention time of these transitions might be indicative of a light signal it is not suitable for quantification due to poor signal-to-noise ratio of less than 10. It might be suggested that for such signals quantification may be achieved but at signal-to-noise ratio below the threshold. For these signals the spectra is of lower quality and care should be taken to interpret the results properly by additionally reporting the S:N ratio.[166] However, in this study and likewise with previous QconCAT studies all spectra used for quantification must have a good signal-to-noise ratio, specifically above a threshold of 10.[157] It is important to note that if the L:H ratio is calculated using a light signal of poor S:N, and additionally the peptide data is found to be in agreement with L:H ratio for pure QconCAT, this would indicate that the light signal is derived from the QconCAT and not endogenous protein.

Further evidence to support that any light signal observed is derived from QconCAT is indicated when the L:H ratio is maintained across each level (10 fmol and 100 fmol) of QconCAT spike, taking into account the light contamination by the unlabelled QconCAT. If the light signal was due endogenous peptide rather than QconCAT then as the level of spiked QconCAT is reduced from 100 fmol to 10 fmol and QconCAT is diluted out of the cell matrix, then the L:H ratio would increase. For these reasons all ratios obtained for QconCAT LM2 were classified as Type B quantifications, an upper limit of quantification was calculated at 10 fmol on column load, and this corresponds to the order of approximately 10^5 copies per cell for each of the cell lines (Table 4.14).

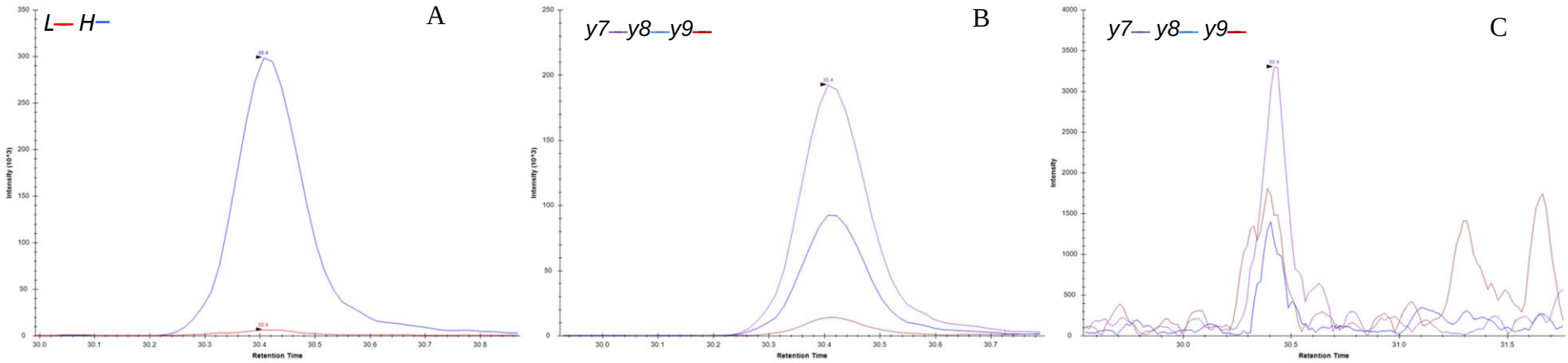


Figure 4.10 LC-SRM trace for NMDQVAPVANSYR of ETS2 obtained with a with a Xevo TQ MS. Standard peptide in 300 ng of cell lysate HEK-293 was applied on-column with 10 fmol equivalent of LM2. A Type B quantification, where (A) shows light (red) and heavy (blue) signals peptide. (B) shows the individual *y*-ions for heavy pepide and (C) shows individual *y*-ion transitions for light peptide.

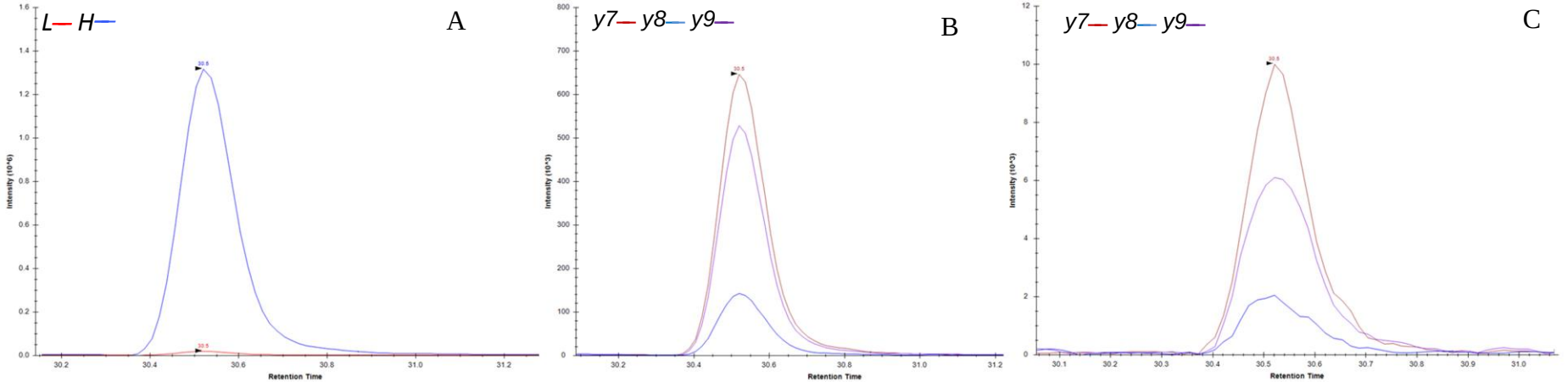


Figure 4.11 LC-SRM trace for NMDQVAPVANSYR of ETS2 obtained with a with a Xevo TQ MS. Standard peptide in 300 ng of cell lysate HT-29 was applied on-column with 100 fmol of QconCAT. An example of a Type B quantification. (A) shows light (red) and heavy (blue) signal. (B) shows heavy *y*-ion transitions, (C) pane shows light *y* -ion transitions.

There were also a number of Type C peptides recorded where neither standard nor analyte gave acceptable SRM data. However, the standard peptide maybe observed at a higher spike level therefore at increased on-column amounts these peptides are promoted to Type B. Up to 17 Q-peptides of QconCAT LM2 were detectable by final SRM analysis and classified as Type B quantifications. For all of these peptides it was possible to set an upper bound on the abundance of the protein. In a number of cases e.g. DUSP 18 data was obtained for one peptide only and although other quantification studies use only one peptide it would be more reliable to have quantification data available for both peptides.

Table 4.14 List of proteins and surrogate peptides derived from LM2 which have an upper limit of quantification (copies per cell, CPC) for each of the cell lines.

Pep	Protein	Q-Peptide Sequence	HEK-293 (2), CPC	HT-29 (1), CPC	HCT 116 (2), CPC
Q1	DUSP 16	VPVNDSFCEK	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q26	DUSP 16	LVALLESGTEK	Noise	Noise	Noise
Q2	DUSP 7	SAEWLQEELEAR	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q3	DUSP 18	QPSVSGLSQITK	Noise	< 2.6 x 10^5	<2.9 x 10^5
Q4	DUSP 5	YVLPDEAAR	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q5	DUSP 9	YILNVTPNLPNFFEK	<6.6 x 10^5	< 2.6 x 10^5	<2.9 x 10^5
Q16	DUSP 9	ASFPVQILPNLYLGSAR	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q12	DUSP 4	FSSEYPEFCSK	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q15	DUSP 10	SLNCGCSSASCCTVATYDK	Noise	Noise	Noise
Q22	DUSP 10	AANLTYMPSSSGSAR	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q17	DUSP 2	AGPTAVYFLR	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q20	DUSP 1	GGYEAFSASCPELCSK	<6.6 x 10^5	< 2.6 x 10^5	<2.9 x 10^5
Q23	ETS1 1/2	F(C)MNGAAL(C)ALGK	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q14	ETS2	NMDQVAPVANSYR	<6.7 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q28	ELK1 1	GAGMAGPGGLAR	<8.3 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q29	ELK 1	LVDAEEVAR	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5
Q32	ELK1 1	AEPEVPPQEGVPAR	<6.8 x 10^5	<2.7 x 10^5	<3.0 x 10^5

4.12 Conclusion

Various cell lysate sample preparation methods have been investigated including FASP, and in-solution digests with and without RapiGest. FASP resulted in significant samples losses and it could not be ruled out that those losses were sample specific. To retain as much sample material as possible and overcome potential complications with differential loss of native protein and QconCAT, in-solution digest was favoured. It was found that the use of RapiGest during an in-solution digest gave almost twice as many peptide identifications and minimal unalkylated cysteines residues. Method development of Selected Reaction Monitoring for each protein of interest was initiated using product ion analysis data acquired for each peptide from a Q-TOF Ultima Global. SRMs for each protein were manually selected and a method set up on the Xevo TQ MS instrument. SRMs were optimised to maximise the dwell time and thus enhance the signal obtained. This was later superseded by an *in-silico* method, Skyline, to generate an SRM assay. Unfortunately due to problems with available instrument time the project was moved between two instrument platforms, the second being the TSQ Vantage. Therefore a new cycle of SRM method development was initiated. Data-dependent acquisitions of QconCATs LM2 were obtained with the LTQ-Orbitrap XL and then used to generate a SRM method. Final SRM analyses of samples were undertaken using a Xevo TQ MS whereby 300 ng of cell lysate:10 fmol of standard were applied to the column. For the peptides detected an upper limit of quantification of the corresponding proteins could be calculated at approximately 10^5 copies per cell.

5 Results and Discussion III QconCAT LM1 SRM Method Development and Final Analysis

5.1 SRM determination

Initially, transitions for QconCAT LM1 were defined experimentally by analysing unlabelled LM1 using a Q-TOF Ultima Global coupled online to an LC system. A data-dependent acquisition was undertaken where the top 3 ions from a full scan MS spectrum were selected for CID and product ions spectra acquired in a serial manner during the LC gradient and retention times for each Q-peptide were also defined. As described in the previous chapter, *y*-ions were manually selected for each Q-peptide (Table 5.1). For example, the full scan CID product ion spectrum of Q-peptide 9 of QconCAT LM1 shows that *y7*, *y8* and *y9* are the most intense products above $[M+2H]^{2+}$ and are therefore included in the transition list (Figure 5.1). For histidine containing peptides higher charge states above 2+ were looked for. Where peptides were not observed by the Q-TOF Ultima Global, a transition prediction program Pinpoint (v 1.1.12.0) was used to predict optimal transitions for a SRM (Table 5.1).

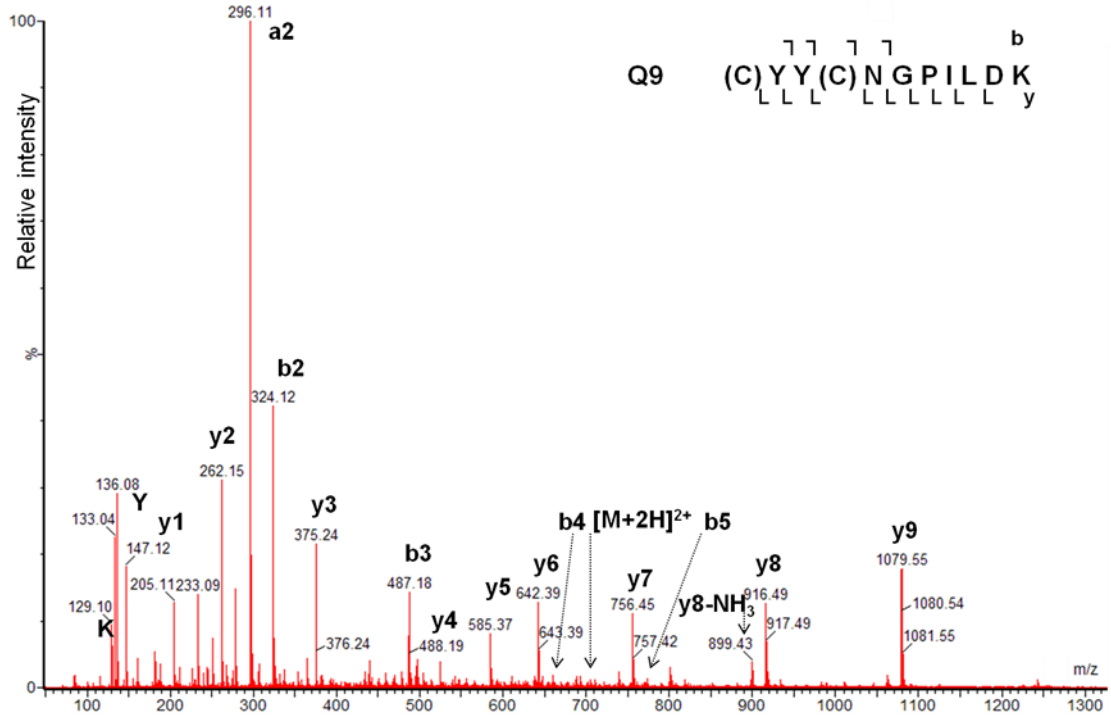


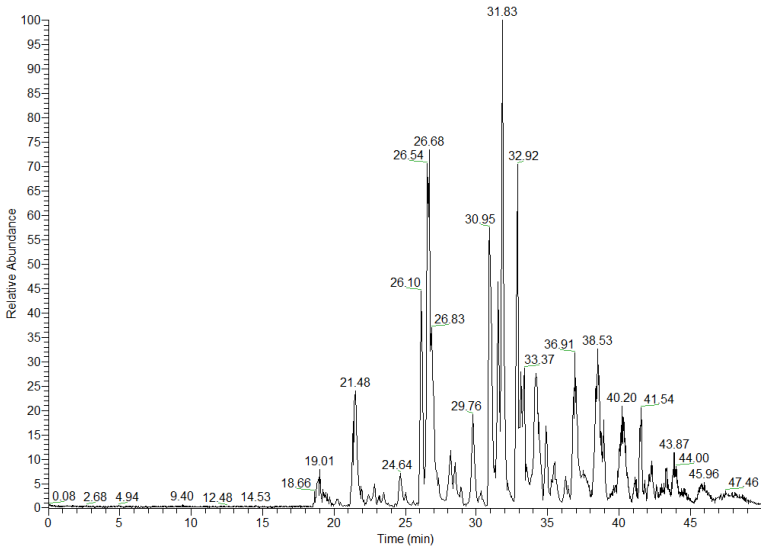
Figure 5.1 Example of a full scan CID-MS/MS recorded on a Q-TOF Ultima Global instrument of light QconCAT LM1 Q9 (*m/z* 701.81). (C) represents the fixed modification of carbamidomethylation of cysteine. K and Y represent lysine and tyrosine immonium ions respectively. Product ions *y*7, *y*8 and *y*9 selected for the SRM assay.

Table 5.1 QconCAT LM1 protein designed to quantify scaffold proteins of the MAPK pathway. Q-peptides are listed along with their doubly- and triply- charged ions, retention times and selected transitions. (C) represents cysteines modified by carbamidomethylation, starred transitions correspond to triply-charged precursors. For those peptides not detected by a Q-TOF Ultima Global no retention time could be recorded and tbc, to be confirmed, appears in the retention time column.

Peptide	Protein	Accession #	Peptide Sequence	$[M+2H]^{2+}$	Rt, min	y ion	*m/z*	y ion	*m/z*	y ion	*m/z*
Q1	IQGAP1	P46940	NVIFEISPTEEVGDFEVK	1026.51	28.08	y10	1152.54	y11	1249.59	y9	1051.49
Q2	ARRB2 1/2/3	P32121	A(C)GVDFEIR	533.75	22.21	y7	835.43	y6	778.41	y5	564.31
Q3	KSR2 1	Q6VAB6	SEEQQPLSLQK	643.83	18.25	y6	685.42	y8	941.54	y9	1070.58
Q4	Control	Control	GVNDNEEGFFSAR	721.32	23.07	y7	813.38	y8	942.43	y9	1056.47
Q5	KSR2 1	Q6VAB6	LTVDAYPGL(C)PPPPLESGHR	726.03*	32.86	y7	795.41	y8	892.46	y9	989.52
Q6	MPKS1	Q9UH	LPSVEGLHAIVVSDR	531.3*	24.79	y6	688.40	y7	759.40	y5	575.31
Q7	PEBP1	P30086	LYTLVLTDPDAPSR	780.92	25.23	y8	858.39	y9	971.47	y10	1070.55
Q8	KSR1	Q8IVT5	DLTLDALLEMNEAK	788.40	29.67	y7	834.40	y8	858.40	y9	971.48
Q9	PAXI 1/2/3	P49023	(C)YY(C)NGPILDK	701.81	20.85	y9	1079.52	y8	916.46	y7	756.43
Q10	KSR2 1/2	Q6VAB6	QQFIFPDVVPVPETPTR	985.52	28.21	y12	1306.70	y13	1453.76	y10	1094.62
Q11	KSR1	Q6VAB6	LIDISIGSLR	543.83	26.54	y8	860.48	y6	632.37	y7	745.46
Q12	ARRB1 1/2	P49407	(C)PVAMEEADDTVAPSSTF(C)K	1107.97	22.25	y12	1327.58	y13	1398.62	y14	1527.66
Q13	PEA_15	Q15121	SEEITTGSAWFSFLESHNK	1085.51	tbc	y10	1294.62	y8	961.47	-	-
Q14	KSR2 2	Q6VAB6	IHSSVGS(C)ENIPSQQR	899.93	16.73	y11	1275.57	y14	1548.71	y13	1461.68
Q15	Sur8/Shoc2	Q9UQ13	SIHILPSSIK	547.83	23.28	y6	644.40	y7	757.48	y8	894.54
Q16	PAXI 1/2/3	P49023	LGVATVAK	379.74	18.04	y7	645.39	y5	489.30	y6	588.37
Q17	MVP	Q14764	LFSVPDFVGDA(C)K	727.85	27.12	y9	1008.45	y7	796.37	y11	1194.55
Q18	MPKS1	Q9UH	ELAPLFEELR	608.83	28.18	y7	903.49	y5	693.35	y6	806.44
Q19	ARRB1 1/2	P49407	A(C)GVDYEVK	520.74	17.88	y7	809.40	y5	653.31	y4	538.29
Q20	ARRB2 1/2/3	P32121	(C)PVAQLEQDDQVSPSSTF(C)K	1148.51	21.96	y11	1255.56	y12	1370.59	y13	1498.65
Q21	Sur8/Shoc2	Q9UQ13	ELTQLTELYLYSNK	857.95	27.71	y9	1130.57	y10	1243.65	y11	1371.71
Q22	IQGAP1	P46940	ILAIGLINEALDEGDAQK	942.01	30.14	y14	1472.72	y11	1189.53	y12	1302.61
Q23	PEBP1	P30086	GNDISSGTVLSDYVGSGPPK	975.48	23.86	y10	1006.48	y11	1119.48	y12	1218.63
Q24	MORG1	Q9BRX9	TYSGHGYEVLDAAGSFDNSSLCSGGGDK	1397.60	tbc	y16	1501.62	-	-	-	-
Q25	MORG1	Q9BRX9	VNTVQFNEEATVILSGSIDSSIR	1240.14	tbc	y13	1347.75	-	-	-	-
Q26	KSR1 1/2/3	Q8IVT5	LSHDWLCYLAPEIVR	907.97	tbc	y5	613.37	y9	1063.56	-	-
Q27	KSR1 1/2/3	Q8IVT5	(C)GASGDE(C)GR	534.70	tbc	y6	693.26	y7	780.29	y8	851.33
Q28	PEA_15	Q15121	ISEEDELDTK	589.78	17.44	y9	1065.45	y8	978.42	y7	849.38
Q29	MVP	Q14764	ALQPLEEGEDEEK	743.85	19.19	y8	964.37	y10	1174.51	y8	964.37

5.2 SRM Method Development for the TSQ Vantage

To experimentally define the transitions for the TSQ Vantage, a tryptic digest of QconCAT LM1 was analysed using top 5 DDA on the LTQ Orbitrap XL. Full scan tandem MS data were searched against the QconCAT database using MASCOT facilitated through the vendor supplied software Proteome Discoverer (version 1.0, build 43), using variable modification of cysteine alkylation (to ensure complete alkylation), methionine oxidation and one missed cleavage. For QconCAT LM1, 91.0 % sequence coverage was obtained (Figure 5.2). These MASCOT results were then imported into Pinpoint (version 1.1.12.0) to build an SRM method to be used on the TSQ Vantage (Appendix 4).



	1	11	21	31	41	51	61	71	81	91
1	MGTKNVIFEI	SPTEEVGDFE	VKACGVDFEI	RSEEQQPLSL	QKGVNDNEEG	FFSARLTVDA	YPGLCPPPPL	ESGHRLPSVE	GLHAIVVSDR	LYTLVLTDPD
101	APSRDLTLDA	LLEMNEAKCY	YCNGPILDKQ	QFIFPDVVPV	PETPTRLIDI	SIGSLRCPVA	MEEADDTVAP	SSTFCKSEEI	TTGSAWFSFL	ESHNKIHSSV
201	GSCENIPSQQ	RSIHILPSSI	KLGVATVAKL	FSVPDFVGDA	CKELAPLFEE	LRACGVDYEV	KCPVAQLEQD	DQVSPSSTFC	KELTQLTELY	LYSNKILAIG
301	LINEALDEGD	AQKGNDISSG	TVLSDYVGSG	PPKTYSGHGY	EVLDAAGSFD	NSSLCSGGGD	KVNTVQFNEE	ATVILSGSID	SSIRLSHDWL	CYLAPEIVRC
401	GASGDECGRI	SEEDELDTKA	LQPLEEGEDE	EKLAAALEHH	HHHH					

Figure 5.2 (A) The base peak chromatogram nanoESI-FT-MS for LM1 obtained using an Orbitrap XL instrument, following a double trypsin digestion. (B) LM1 QconCAT has a 91 % sequence coverage when nanoESI-MS/MS data was searched against a customised database using the MASCOT algorithm.

5.3 SRM Analysis of Pure LM1 QconCAT

Analysis of pure QconCAT was carried out using a Waters Xevo TQ MS to ascertain the heavy to light ratio for each peptide. For all quantification data obtained within cell lines, the ratios between endogenous and heavy QconCAT protein were corrected according to Table 5.2 by subtracting any contributions made to the unlabelled signal from the light derived part of the QconCAT protein.

Table 5.2 Light to heavy ration (L:H) for each peptide of pure LM1 QconCAT following SRM analysis by Waters Xevo TQ.

Q Peptide	Protein	Accession #	Q-Peptide Sequence	L:H
Q1	IQGAP1	P46940	NVIFEISPTEEVGDFEVK	0.02
Q2	ARRB2 1/2/3	P32121	ACGVDFEIR	0.01
Q3	KSR2 1	Q6VAB6	SEEQQPLSLQK	0.01
Q4	Control	Control	GVNDNEEGFFSAR	0.01
Q6	MPKS1	Q9UH	LPSVEGLHAIVVSDR	0.01
Q7	PEBP1	P30086	LYTLVLTDPDAPSR	0.01
Q9	PAXI 1/2/3	P49023	CYYCNGPILDK	0.02
Q10	KSR2 1/2	Q6VAB6	QQFIFPDVVPVPETPTR	0.01
Q14	KSR2 2	Q6VAB6	IHSSVGSCENIPSQQR	0.01
Q15	Sur8/Shoc2	Q9UQ13	SIHILPSSIK	0.01
Q16	PAXI 1/2/3	P49023	LGVATVAK	0.01
Q18	MPKS1	Q9UH	ELAPLFEELR	0.01
Q19	ARRB1 1/2	P49407	ACGVDYEVK	0.02
Q20	ARRB2 1/2/3	P32121	CPVAQLEQDDQVSPSSTFCK	0.03
Q21	Sur8/Shoc2	Q9UQ13	ELTQLTELYLYSNK	0.02
Q22	IQGAP1	P46940	ILAIGLINEALDEGDAQK	0.02
Q23	PEBP1	P30086	GNDISSGTVLSDYVGSGPPK	0.01
Q26	KSR1 1/2/3	Q8IVT5	LSHDWLCYLAPEIVR	0.04
Q28	PEA_15	Q15121	ISEEDELDTK	0.02
Q29	MVP	Q14764	ALQPLEEGEDEEK	0.02

5.4 Final Quantification Experiments Using the TSQ Vantage

Final SRM quantification experiments were carried out using TSQ Vantage. Two biological and three technical replicates were analysed for each cell line (HCT 116, HT-29 and HEK-293). Whole cell lysates of these samples were mixed with various amounts of QconCAT LM1 and prepared for MS analysis using the modified RapiGest protocol (Materials and Methods). On-column quantities during SRM analysis were 300 ng of digested cell lysate with either 100 fmol or 10 fmol of co-digested QconCAT LM1. Cell lysates with no QconCAT addition, referred to as background samples, were also analysed with 300 ng applied to the column. As described in the previous chapter, the results of QconCAT methodology used with SRM analysis can be classified as either Type A, Type B or Type C. Data for all three analytical replicates across three cell lines is shown in Appendix 6.

There was a number of Type A quantifications, where endogenous and standard QconCAT peptide could be observed, across all cell lines. All data was manually verified by examining the contributions to the light and heavy signal made by each of the *y*-ion transitions. The elution peak profiles and retention times for each of contributing *y*-ion transitions were in agreement with each other per light or heavy signal. More importantly these were also in agreement between light and heavy signals per peptide. For example Figure 5.3 shows *y*-ion contributions for the light and heavy signals for the protein of KSR2, peptide SEEQQPLSLQK, in HT-29 cell lysate. Another example of a Type A quantification is Sur8/Shoc2, peptide SIHLPSSIK, (Figure 5.4).

Type A quantifications were achieved with either the 100 fmol or 10 fmol on-column load, due the differences in signal intensity for the different Q-peptides. These differences can be due to either reduced ionisation efficiency or abundance of the native peptide. For example, a signal for peptide NVIFEISPTEEVGDFEVK of IQGAP1, suitable for quantification was observed in all samples at 100 fmol but not for Q-peptide at 10 fmol on column. The signal at 10 fmol applied on-column may be improved if the dwell time is optimised in future analyses.

For some Type B quantifications it was possible to observe peptides at 10 fmol applied on column: for example, LIDISIGSLR provided data suitable for quantification at this lower amount (Figure 5.5). Although there is a slight shoulder on the tail end of the *y8* transition which would indicate some interference, there is sufficient evidence e.g. equal retention times for all *y*- ions to indicate that the heavy signal is derived from the Q-peptide. Therefore, for Type B quantifications where Q-peptide was observed an upper limit can be placed on the copies per cell values derived from either 100 fmol or 10 fmol on-column loads depending on signal quality.

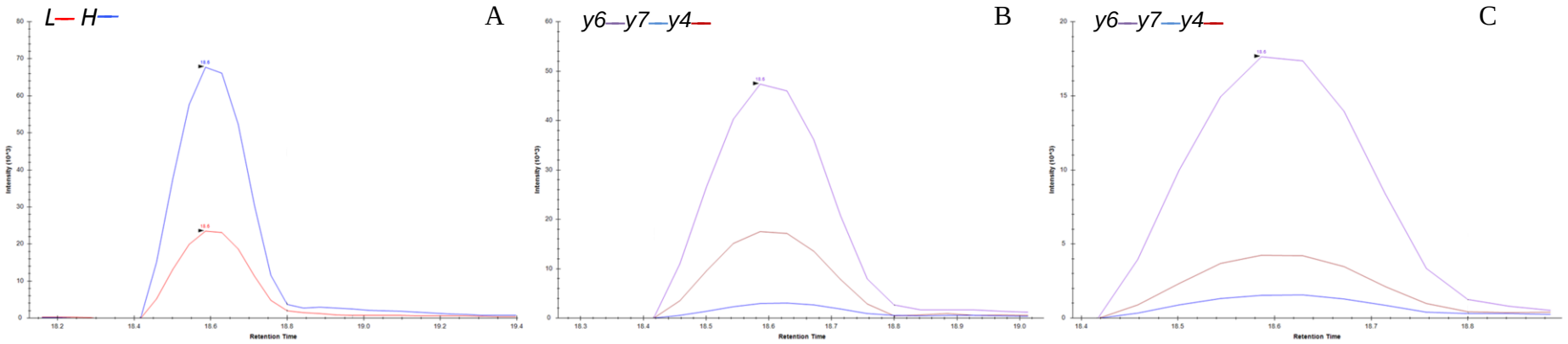


Figure 5.3 (A) shows light and heavy trace for SEEQPLSLQK, KSR2 isoform 1. (B) shows light *y*-ion transitions and (C) shows heavy *y*-ion transitions.300 ng of digested HT-29 and 100 fmol of co-digested LM1 QconCAT on-column load.

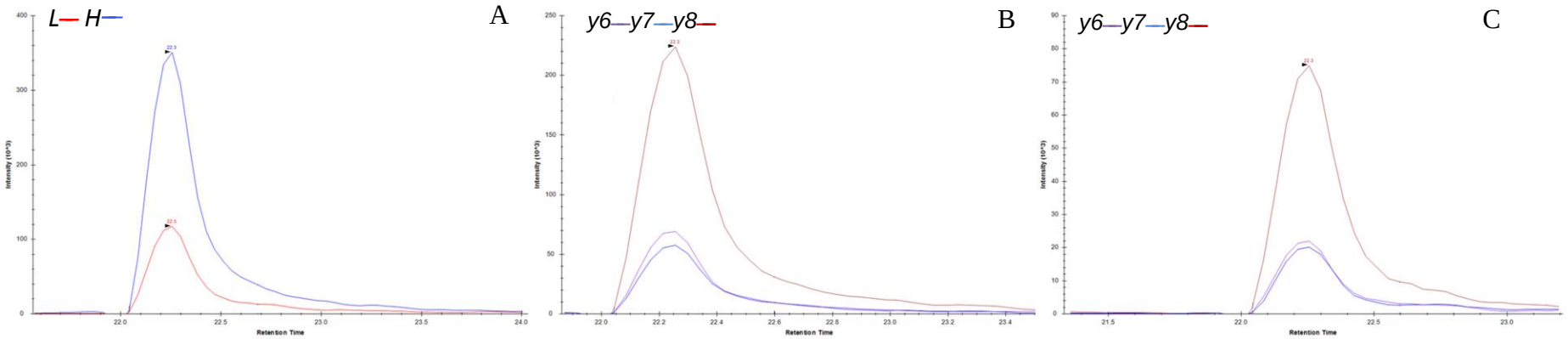


Figure 5.4 (A) shows light and heavy transitions for SIHILPSSIK Sur8/Shoc2. (B) shows light *y*-ion transitions and (C) shows heavy *y*-ion transitions. 300 ng of digested HCT 116 and 100 fmol of co-digested QconCAT LM1 on-column load.

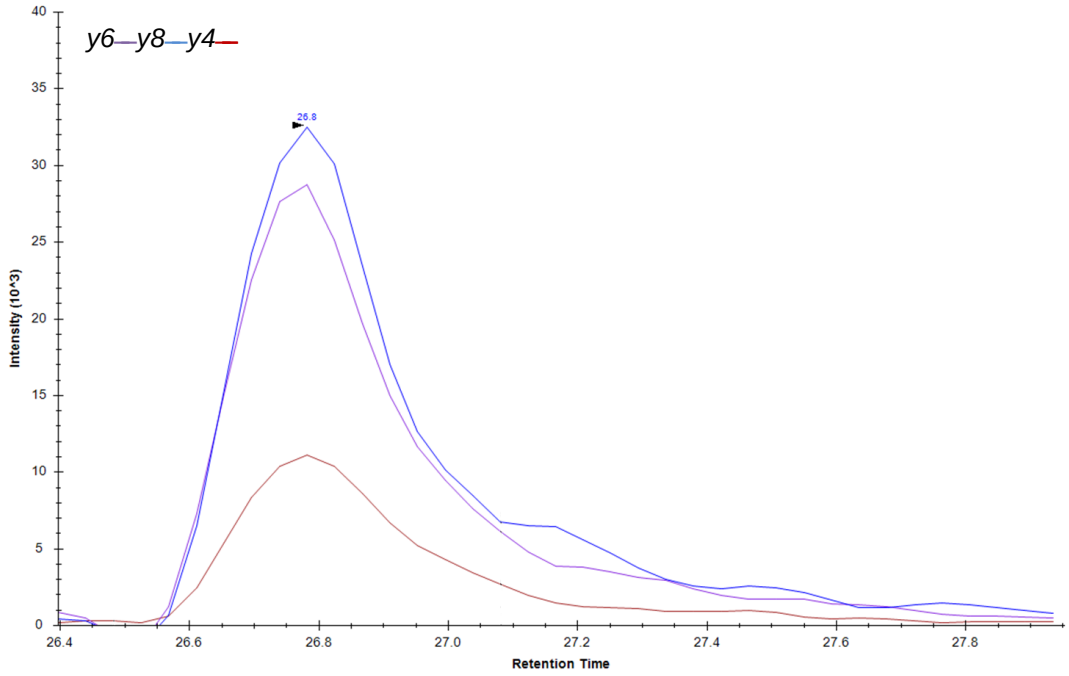


Figure 5.5 LC-SRM trace showing *y*-ion transitions for heavy peptide LIDISIGSLR of KSR1 obtained with a TSQ Vantage. 300 ng of digested HT-29 and 10 fmol of co-digested QconCAT LM1 on-column load.

Surprisingly, the peptide LSHDWL(C)YLAPEIVR of KSR1 was detected by SRM at 100-fmol applied on-column. Previously, a signal for this peptide was elusive and was not detectable either by MALDI-TOF or Q-TOF. A signal may not have been observed due to the different ionisation mechanism of MALDI. As this peptide was observed by highly selective SRM assay following ionisation by electrospray it possible that the signal may have been too low to be observed during a top three DDA by Q-TOF.

There were also a number of peptides which were either Type B or Type C quantifications depending on the cell line analysed such as ILAIGLINEALDEGDAQK of protein IQGAP1. This peptide was classified as a Type B for HT-29 or Type C for HEK-293 and HCT 116 whereby neither native or Q-peptide could be detected. The Q-peptide had been previously detected by MALDI-TOF and Q-TOF. A reason for its loss in most cell lysates could be because these analyses by MALDI-TOF and Q-TOF were undertaken on the pure digested QconCAT protein -rather than the biological sample spiked with the standard. The

lack of detection of a peptide in a complex biological matrix is likely due to ion suppression effects.[167]

Another peptide that is infrequently observed is (C)PVAMEEADDTVAPSSTF(C)K of beta-arrestin 1 protein, isoforms 1 and 2. A reason for this is that there is a methionine residue which may become oxidised during sample handling and therefore the amount of (C)PVAMEEADDTVAPSSTF(C)K detected would be reduced. Another modification of this peptide could be cyclisation of the N-terminal carbamidomethyl-cysteine.[168]

Following classification of the data, the upper limit of quantification was calculated which corresponds to approximately 10^6 copies per cell for each of the cell lines (Tables 5.4 to 5.6). The data shows that there are approximately 10 fold differences in the upper limits for quantification of PEB1 (peptides GNDISSGTVLSDYVGSGPPK and LYTLVLTDPDAPSR) across all cell lines. For example for HCT 116 biological replicate 1 peptide GNDISSGTVLSDYVGSGPPK measured 4.3 x 10^6 copies per cell whereas LYTLVLTDPDAPSR measured 4.9 x 10^5 copies per cell. A possible reason for this could be that there were missed cleavages due to differences in digestion efficiency between Q-peptide and analyte. Missed cleavage can cause either an overestimate in protein expression depending on whether the endogenous peptide is excised to completeness and the Q-peptide is not or an underestimate in protein expression if the reverse occurs. As described earlier (section 4.11.1), an acidic sidechain next to a cleavage site may be responsible for a missed cleavage. This would explain the relatively low CPC value for LYTLVLTDPDAPSR of PEB1, as both cleavage sites either side of this peptide contain an aspartic acid adjacent to the arginine or lysine (LPSVEGLHAIVVSDR-LYTLVLTDPDAPSR-DLTLDALLEMNEAK). Unfortunately, no data was collected for HCT 116 biological replicate 2 during the analytical run and it is likely that this was due to a spray issue or a problem with sample preparation (Table 5.5).

Table 5.3 Copies per cell calculated for biological replicates 1 and 2 for colon cancer cells HEK293 when 300 ng of cell lysate and either 10 fmol or 100 fmol LM1 QconCAT was applied on column. Type A and Type B quantifications are shown.

			10fmol LM1		100fmol LM1	
Peptide	**Protein**	**Q-Peptide Sequence**	**HEK-293 (1),**	**HEK-293 (2), CPC**	**HEK-293 (1), CPC**	**HEK-293 (2), CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	-	-	2.3 x 10^6	2.4 x10^6
Q22	IQGAP1	ILAIGLINEALDEGDAQK	-	-	-	-
Q2	ARRB2 1/2/3	ACGVDFEIR	-	<6.7 x 10^5	-	<6.7 x 10^6
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	-	-	-	-
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	-	-	-	-
Q19	ARRB1 1/2	ACGVDYEVK	-	-	-	-
Q3	KSR2 1	SEEQQPLSLQK	-	-	2.6 x 10^6	2.3 x 10^6
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	-	-	-	-
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	<8.3 x 10^5	-	<8.3 x 10^5	<6.7 x 10^6
Q14	KSR2 2	IHSSVGSCENIPSQQR	-	-	-	-
Q6	MPKS1	LPSVEGLHAIVVSDR	-	-	-	<6.7 x 10^6
Q18	MPKS1	ELAPLFEELR	<8.3 x 10^5	-	<8.3 x 10^5	<6.7 x 10^6
Q8	KSR1	DLTLDALLEMNEAK	-	-	-	-
Q11	KSR1	LIDISIGSLR	-	<6.7 x 10^5	<8.3 x 10^5	<6.7 x 10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	-	-	-	-
Q9	PAXI 1/2/3	CYYCNGPILDK	-	-	-	<6.7 x 10^6
Q13	PEA_15	SEEITTGSAWFSFLESHNK	-	-	-	2.3 x 10^6
Q28	PEA_15	ISEEDELDTK	-	-	-	-
Q17	MVP	LFSVPDFVGDACK	<8.3 x 10^5	<6.7 x 10^5	-	2.4 x 10^6
Q29	MVP	ALQPLEEGEDEEK	-	-	-	<6.7 x 10^6
Q15	Sur8/Shoc2	SIHILPSSIK	<8.3 x 10^5	-	<8.3 x 10^5	2.2 x 10^6
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	-	-	2.7 x 10^6	-
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	-	-	3.8 x 10^6	5 x 10^6
Q7	PEBP1	LYTLVLTDPDAPSR	-	<6.7 x 10^5	<8.3 x 10^5	<6.7 x 10^6

Table 5.4 Copies per cell calculated for biological replicates 1 and 2 for colon cancer cells HCT 116 when 300 ng of cell lysate and either 10 fmol or 100 fmol LM1 QconCAT was applied on column. Type A and Type B quantifications are shown, where Type B quantifications provide an upper bound.

			10fmol LM1		100fmol LM1	
Peptide	**Protein**	**Q-Peptide Sequence**	**HCT 116(1), CPC**	**HCT 116(2), CPC**	**HCT 116 (1), CPC**	**HCT 116 (2), CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	-	-	4.1x 10^6	1 x 10^6
Q22	IQGAP1	ILAIGLINEALDEGDAQK	-	-	-	-
Q2	ARRB2 1/2/3	ACGVDFEIR	<4.1 x 10^6	-	<4.1x 10^6	<3 x 10^6
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	-	-	<4.1 x 10^6	-
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	-	-	<4.1 x 10^6	-
Q19	ARRB1 1/2	ACGVDYEVK	-	-	<4.1 x 10^6	9.3 x 10^5
Q3	KSR2 1	SEEQQPLSLQK	<4.1 x 10^6	-	1 x10^6	1 x 10^6
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	-	-	-	-
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	<4.1 x 10^6	-	<4.1 x 10^6	<3 x 10^6
Q14	KSR2 2	IHSSVGSCENIPSQQR	-	-	-	-
Q6	MPKS1	LPSVEGLHAIVVSDR	-	-	-	<3 x 10^6
Q18	MPKS1	ELAPLFEELR	<4.1 x 10^6	-	<4.1 x 10^6	< 3 x 10^6
Q8	KSR1	DLTLDALLEMNEAK	-	-	-	<3 x 10^6
Q11	KSR1	LIDISIGSLR	<4.1 x 10^6	-	<4.1 x 10^6	<3 x 10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	-	-	-	-
Q9	PAXI 1/2/3	CYYCNGPILDK	-	-	1 x10^6	<3 x 10^6
Q13	PEA_15	SEEITTGSAWFSFLESHNK	-	-	-	-
Q28	PEA_15	ISEEDELDTK	-	-	-	-
Q17	MVP	LFSVPDFVGDACK	<4.1 x 10^6	-	<4.1 x 10^6	1 x 10^6
Q29	MVP	ALQPLEEGEDEEK	-	-	<4.1 x 10^6	-
Q15	Sur8/Shoc2	SIHILPSSIK	<4.1 x 10^6	-	<4.1 x 10^6	-
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	-	-	1 x10^6	1 x 10^6
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	-	-	4.3 x 10^6	4 x 10^6
Q7	PEBP1	LYTLVLTDPDAPSR	7 x 10^5	-	4.9 x 10^5	4.7 x 10^5

Table 5.5 Copies per cell (CPC) calculated for biological replicates 1 and 2 for colon cancer cells HT-29 when 300 ng of cell lysate and either 10 fmol or 100 fmol LM1 QconCAT was applied on column. Type A and Type B quantifications are shown, where Type B quantifications provide an upper bound.

			10fmol LM1		100fmol LM1	
Peptide	**Protein**	**Q-Peptide Sequence**	**HT-29 (1), CPC**	**HT-29 (2), CPC**	**HT-29(1), CPC**	**HT-29 (2), CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	-	-	1 x 10^6	1 x 10^6
Q22	IQGAP1	ILAIGLINEALDEGDAQK	-	-	<2.6 x 10^6	<2.5 x 10^6
Q2	ARRB2 1/2/3	ACGVDFEIR	<2.5 x 10^5	-	<2.6 x 10^6	1.6 x 10^6
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	-	-	<2.6 x 10^6	-
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	-	-	-	<2.5 x 10^6
Q19	ARRB1 1/2	ACGVDYEVK	-	-	<2.6 x 10^6	-
Q3	KSR2 1	SEEQQPLSLQK	1 x 10^6	-	9.4 x 10^5	8.5 x 10^5
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	-	-	-	-
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	<2.5 x 10^5	<2.6 x 10^5	<2.6 x 10^6	<2.5 x 10^6
Q14	KSR2 2	IHSSVGSCENIPSQQR	-	-	-	-
Q6	MPKS1	LPSVEGLHAIVVSDR	-	-	<2.6 x 10^6	<2.5 x 10^6
Q18	MPKS1	ELAPLFEELR	<2.5 x 10^5	<2.6 x 10^5	<2.6 x 10^6	<2.5 x 10^6
Q8	KSR1	DLTLDALLEMNEAK	-	-	-	<2.5 x 10^6
Q11	KSR1	LIDISIGSLR	-	<2.6 x 10^5	<2.6 x 10^6	<2.5 x 10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	-	-	-	-
Q9	PAXI 1/2/3	CYYCNGPILDK	-	-	-	<2.5 x 10^6
Q13	PEA_15	SEEITTGSAWFSFLESHNK	-	-	<2.6 x 10^6	-
Q28	PEA_15	ISEEDELDTK	-	-	-	-
Q17	MVP	LFSVPDFVGDACK	-	-	9.9 x 10^5	8.1 x 10^5
Q29	MVP	ALQPLEEGEDEEK	-	-	-	-
Q15	Sur8/Shoc2	SIHILPSSIK	-	-	1 x 10^6	8.8 x 10^5
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	-	-	9.2 x 10^5	9.3 x 10^5
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	-	-	2.8 x 10^6	3 x 10^6
Q7	PEBP1	LYTLVLTDPDAPSR	4.9 x 10^6	4.2 x 10^5	2.8 x 10^5	3.8 x 10^5

The quantifications may be compared to values obtained by other MS-based methods for quantifying proteins in human tissue culture cell lines. Two such studies are by Aebersold[169] and Mann.[170] In the Aebersold study AQUA peptides were generated to obtain absolute quantification of proteins in U-2 OS (human osteosarcoma) cell line. Mann and co workers recently published a study based on Protein Epitope Signature Tag (PrEST) and SILAC to determine protein copy numbers for HeLa cells. Their values for proteins measured in this study are shown in Table 5.6 along with a summary of values obtained by QconCAT LM1 and SRM. Where there are two peptides measured per protein by LM1 QconCAT the larger number is included in this table. This is because a loss of signal can be rationalised due to a number of scenarios including missed cleavage of the peptide and peptide modifications which may not be equal between standard and endogenous peptide. The same principle has been applied to values included in the table when comparing two Type A quantifications between two biological replicates for the same peptide. Type B quantifications, where an upper bound can be placed on the copy number per cell have also been included for reference.

Table 5.6 Copies per cell (CPC) measured in this study (HEK-293, HT-29, HCT 116) compared to those measured in two alternative studies by Mann (HeLa)[170] and Aebersold (U-2 OS).[169]

Protein	Accession Number	U-2 OS, CPC	HeLa, CPC	HEK-293, CPC	HT-29, CPC	HCT 116, CPC
IQGAP1	P46940	1.7 x 10^5	1.29 x 10^6	2.4 x10^6	1 x 10^6	1 x 10^6
ARRB2 1/2/3	P32121	2.4 x 10^3	-	<6.7 x 10^6	1.6 x 10^6	<4.1x 10^6
KSR2 1	Q6VAB6	-	-	2.6 x 10^6	9.4 x 10^5	1 x 10^6
MPKS1	Q9UHA4	1.63 x 10^4	1.41 x 10^5	<8.3 x 10^5	<2.6 x 10^6	<4.1x 10^5
PEBP1	P30086	8.4 x 10^5	-	5 x 10^6	3 x 10^6	4x10^6
KSR1	Q8IVT5	-	-	<6.7 x 10^5	<2.6 x 10^5	<4.1 x 10^6
PAXI 1/2/3	P49023	5.4 x 10^4	-	<6.7 x 10^6	<2.5 x 10^6	1 x10^6
ARRB1 1/2	P49407	-	-	<6.7 x 10^6	1.6 x 10^6	<4.1 x 10^5
PEA_15	Q15121	1.0 x 10^6	-	2.3 x 10^6	<2.6 x 10^6	1 x10^6
Sur8/Shoc2	Q9UQ13	1.9 x 10^3	-	2.2 x 10^6	1 x 10^6	<4.1 x 10^5
MVP	Q14764	1.5 x 10^5	-	2.4 x 10^6	9.9 x 10^5	<4.1 x 10^5
MORG1	Q9BRX9	<5 x 10^2	-	-	-	-

By comparing the data obtained in this study to previous studies it is clear that the copy number per cell are generally within the same range as previously measured. The values obtained for each of the scaffold proteins are all under 10^7 copies per cell and range across one order of magnitude from $<4.1 \times 10^5$ to 4.14×10^6 copies per cell. The three component scaffold protein, KSR2 isoform 1, provides absolute quantifications in the range 0.94×10^6 to 2.6×10^6 copies per cell depending on cell line and ß-arrestin-2 gave an absolute quantification value of 1.6×10^6 copies per cell for HT-29. These values are within the range predicted by mathematical modelling in Chapter 6 and will be discussed further in that chapter.

Importantly, this study has provided copy number per cells across a number of different cell lines allowing a direct comparison between these. It is interesting to note that absolute quantification for the scaffold proteins KSR1 and KSR2 has not been measured previously and copy numbers per cell for this scaffold protein are provided across all three cell lines measured in this study.

This study shows that there are subtle differences in scaffold quantities (i.e. 2 to 3 fold) when comparing cell lines HCT 116 and HT-29 to control cells HEK293. These changes might be due to the cell compensating for changes in the MAPK pathway due to either B-Raf (HT-29) or K-Ras (HCT 116) mutations.

5.5 Failure to Detect Endogenous Peptides for LM1 and LM2

For a number of Q-peptides for both LM1 and LM2 it was not possible to detect endogenous peptides. The main limitation of quantitative analysis of MS by 1D reversed phase chromatography column is the amount of cell material that can be loaded. Typically, the maximum load for a Waters Acquity column, 75 micron ID, is 500 ng of cell material. However, in the laboratory to extend the life and maintain the performance of the column this is limited to 300 ng and equates to a maximum of 7200 cells applied on column, depending on cell line. At a limit of quantification 10 fmol, this sets a limit of detection at approximately 260,000 copies per cell.

As mentioned previously, a recently published study by Carroll *et al* uses QconCAT to perform absolute quantification of the glycolytic pathway in yeast.[157] They studied abundant proteins of this pathway and 150 ng of protein applied to the column was representative of approximately 30,000 cells (at 5 pg/cell). The lowest number of copies per cell was approximately 20,000 where the analyte signal was equivalent to 1 fmol. For mammalian cells the protein content is much higher and cells occupy a larger volume. Therefore a 150 ng protein load of mammalian cells would account for only 7000 cells based on an average value in Table A, Appendix 5. At a limit of quantification of 1 fmol the limit of detection is set at approximately 100,000 copies per cell. These theoretical calculations demonstrate the effect of sample complexity and cell volume on the sensitivity copies per cell for mammalian cells compared to yeast.

As described earlier, Mann and co workers recently published a study based on SILAC for absolute quantification of HeLa cells.[170] They employed pre-fractionation using SAX prior to MS analysis. Their work provided a copy number of 6,643 for the transcription factor c-Fos. Although this specific protein is not included in LM2 other transcription factors such as ELK and c-MYC are measured but unfortunately endogenous protein was not detected. If these transcription factors are present in approximately the same amount to c-Fos then at a limit of quantification of 1 fmol, approximately 90,000 cells (2 µg protein) would need to be loaded onto the column which far exceeds its capacity.

Therefore to improve sensitivity when using mammalian cells sample pre-fractionation such as isoelectric focussing can be used prior to analysis. However,

these techniques require increased number of LC-MS/MS runs which can be costly and time consuming. Another limiting factor is the sensitivities of current instruments as to achieve copies per cell values in the order of 10^2 for mammalian cells (based on a 500 ng protein load). Without extensive pre-fractionation an exceptional performance of good signal-to-noise at a limit of quantification of 1 amol would be required.

5.6 Future Experimental Work

If time permitted to continue these studies, then to effectively increase the amount of cell material that is analysed by mass spectrometry and improve sensitivity a pre-fractionation strategy would be employed. High pH (pH 10) RP-LC has been shown to be favourable over traditional methods of SCX due to increased number of peptides detected and elimination of desalting step prior to analysis by mass spectrometry which reduces samples loss and speeds up sample processing time.[171] Alternatively OFF-GEL isoelectric focussing might also be employed as a pre-fractionation strategy. As described above, Aebersold has used this method to successfully detect and quantify standard and endogenous peptide pairs as low as 4500 protein copies per cell in human cell line U-2 OS. To overcome poorly performing peptides such as Type C the best course of action would be to select new peptides for inclusion in the next generation of QconCAT.

Further refinement of these experiments may include the optimisation of collision energy. It has been shown in the literature that the collision energy (CE) used per transition i.e. optimised for each precursor and product rather than just per precursor could enhance the signal intensity, rather than relying on the generic equation $CE=a(precursor\ m/z)+b$ where a and b are determined for the given instrument and predict CE for a precursor and not product ion.[172] Thus a CE optimised per transition would thus improve the sensitivity for each *y*-ion transition.

Ideally, it would also be useful to quantify the proteins involved in the central MAPK cascade such as Raf, MEK and ERK to thereby provide a complete set of parameters for the cascade for modelling. Previous to this study QconCATs were

designed and expressed for some of these proteins. Future work could include the method development of an SRM assay for these QconCATs and then determine copy values within normal and mutant cell lines.

5.7 Conclusion

Method development of the SRM assays for each protein of interest was initiated using product ion analysis data acquired for each peptide from a Q-TOF Ultima Global. SRMs for each protein were manually selected and a method set up on the Xevo TQ MS instrument. Unfortunately due to problems with available instrument time the project was moved between two instrument platforms, the second being the TSQ Vantage. Therefore a new cycle of SRM method development was initiated (as described in the previous chapter). Data-dependent acquisitions of QconCATs LM1 were obtained with the LTQ-Orbitrap XL and then used to generate a SRM method. Final SRM analyses of samples were undertaken using a Vantage TSQ with 300 ng cell lysate with different amounts of QconCAT LM1 at 100 fmol and 10 fmol applied on column, due to peptide dependent signal intensities. Type A quantifications were achieved for up to ten different proteins measured in HEK-293, HCT 116 and HT-29.

The boundaries of the current implementation of QconCAT methodology for the quantification of copies per cell for DUSPs, transcription factor substrates and some scaffold proteins was reached as indicated by the failure to detect endogenous peptides for these targets. The main reason for this is due to the limits of current MS hardware which were discussed and possible solutions presented including pre-fractionation and collision energy optimisation.

6 Results and Discussion IV Mathematical Modelling

6.1 Scaffold Modelling

Over the past twenty years it has been accepted that signalling specificity for a certain pathway must be achieved through spatial organisation.[173] Scaffold proteins organise members of the signalling cascade into specific complexes, localise signalling proteins at the membrane and exert positive and negative feedback on the signalling cascade.[174] Typically, in a system where a scaffold molecule binds to two or more signalling proteins (Raf, MEK, and ERK) then as the concentration of the scaffold increases the amount of complex formed will initially increase but then decrease.[175] A similar effect is also seen between antibody and antigen binding and is known as the prozone effect.[176] The first scaffold model was proposed by Levchenko and explores in depth an Ordinary Differential Equation (ODE) model based on a two signalling proteins binding to a scaffold protein.[52] The case of three signalling proteins binding to a scaffold protein has only been briefly explored in the literature. Therefore it is important to understand in more depth the optimal scaffold concentration and behaviour of a three signalling proteins-scaffold complex.

An ODE model, describing all possible dynamics of up to three signalling proteins bound to the scaffold, was constructed in MATLAB. The model assumes that binding of each signalling protein is independent of the presence of the other proteins. As such, cooperativity and binding order is not considered. The model variables are the differently occupied scaffold complexes. Considering all possible binding states of three proteins to a scaffold protein leads to a seven component ODE system. Each process is modelled using mass-action kinetics, whereby the rate of binding is assumed to be proportional to the concentrations of each reactant. The parameters of the model are E_T total ERK; R_T total Raf; M_T total MEK; k_r Raf rate of association; k_m MEK rate of association; k_e ERK rate of association; d_r Raf rate of dissociation; d_m MEK rate of dissociation; d_e ERK rate of dissociation.

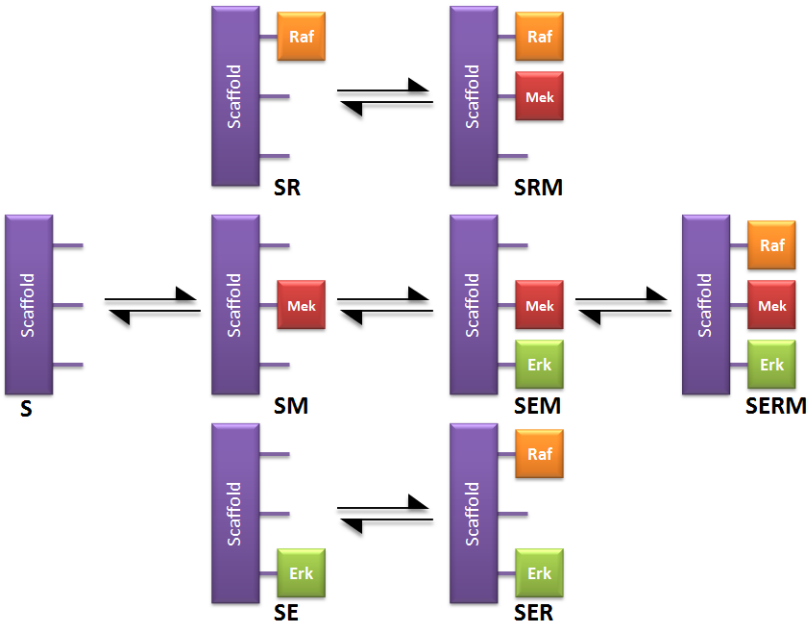


Figure 6.1 Schematic of the scaffold model, which assumes no cooperativity and random order binding. The variables of the model are differently occupied scaffold complexes. *E*, *M*, *R* and *S* denote concentrations of ERK, MEK, Raf and scaffold, respectively.

$$\frac{dSR}{dt} = R.S.k_r - d_r.SR + SMR.d_m + SRE.d_e - SR.M.k_m - SR.E.k_e \quad (1)$$

$$\frac{dSE}{dt} = E.S.k_e - d_e.SE + SEM.d_m + SRE.d_r - SE.M.k_m - SE.R.k_r \quad (2)$$

$$\frac{dSM}{dt} = M.S.k_m - d_m.SM + SMR.d_r + SEM.d_e - SM.R.k_r - SM.E.k_e \quad (3)$$

$$\frac{dSRE}{dt} = SR.E.k_e + SE.R.k_r - SRE.(d_r + d_e) - SRE.M.k_m + SERM.d_m \quad (4)$$

$$\frac{dSEM}{dt} = SE.M.k_m + SM.E.k_e - SEM.(d_e + d_m) - SEM.R.k_r + SERM.d_r \quad (5)$$

$$\frac{dSMR}{dt} = SM.R.k_r + SR.M.k_m - SMR.(d_m + d_r) - SMR.E.k_e + SERM.d_e \quad (6)$$

$$\frac{dSERM}{dt} = SMR.E.k_e + SEM.R.k_r + SRE.M.k_m - (d_m + d_r + d_e).SERM \quad (7)$$

Considering the laws of conservation the total amount of each signalling parameter and scaffold parameter are described below.

$$E_T = E + SE + SRE + SEM + SERM \quad (8)$$

$$M_T = M + SM + SEM + SMR + SERM \quad (9)$$

$$R_T = R + SR + SRE + SMR + SERM \quad (10)$$

$$S_T = S + SR + SE + SM + SRE + SEM + SMR + SERM \quad (11)$$

The conservation laws (8-11) were replaced into the system (1-7) so, for example:

$$\begin{aligned}\frac{dSERM}{dt} = {} & SMR.(E_T - SE - SRE - SEM - SERM).k_e \\ & + SEM.(R_T - SR - SRE - SMR - SERM).k_r \\ & + SRE.(M_T - SM - SEM - SMR - SERM).k_m \\ & - (d_m + d_r + d_e).SERM\end{aligned}$$

We simulated the model under different scaffold concentrations and dissociation constants. Two normalised parameter sets of Raf/MEK/ERK were used (Table 6.1 and 6.2).[55,67]

Table 6.1 The parameters used in the scaffold model as depicted in Figure 6.1. Values for signalling proteins of Raf, MEK and ERK were obtained from Schoeberl *et al* and normalised with respect to Raf.[55] The results of the simulation are shown in Figure 6.2.

Parameter	Number of copies per cell	Value
R_T	4×10^4	1
M_T	2.20×10^7	1000
E_T	2.10×10^7	1000
S_T	-	0-3000
k_r	-	2
k_m	-	0.1-1000
k_e	-	2
d_r	-	1
d_m	-	1
d_e	-	1

Table 6.2 The parameters used in the scaffold model as depicted in Figure 6.1. Values for signalling parameters of Raf, MEK and ERK were obtained from Legewie *et al* and normalised with respect to Raf.[67] The results of the simulation are shown in Figure 6.3.

Parameter	Total cellular concentration, µM	Value
R_T	1	1
M_T	1	1
E_T	10	10
S_T	-	0-20
k_r	-	2
k_m	-	0.1-1000
k_e	-	2
d_r	-	1
d_m	-	1
d_e	-	1

The model shows that the optimal scaffold concentration is between the lowest and second lowest signalling concentration (Figure 6.2 and 6.3). This outcome defines the range of optimal scaffold concentration. The effect of dissociation constants for MEK on optimal scaffold concentration and maximum SERM complex are also shown. As the dissociation constant increases, the maximum SERM complex concentration decreases which is expected when a signalling protein is bound weakly to the scaffold.

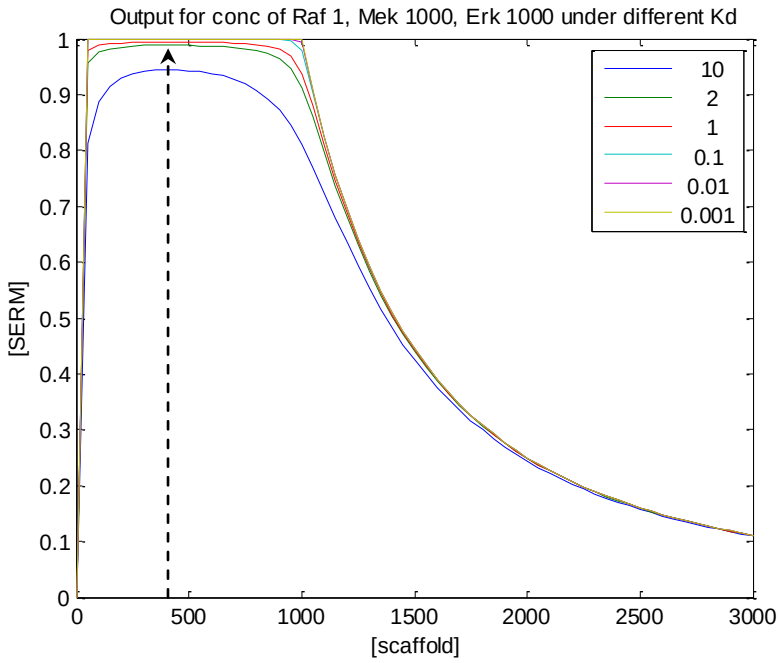


Figure 6.2 ODE simulations under varying scaffold concentration and arbitrary dissociation constants reveal that the optimal scaffold concentration for three component scaffold complex (SERM). The graph depicts scaffold concentration under normalised Raf/MEK/ERK parameter sets published by Schoeberl *et al.*[55] Concentrations are presented in normalised units. Optimal scaffold concentration is 400. Arbitrary values of K_d under which the model was tested under are shown in the legend.

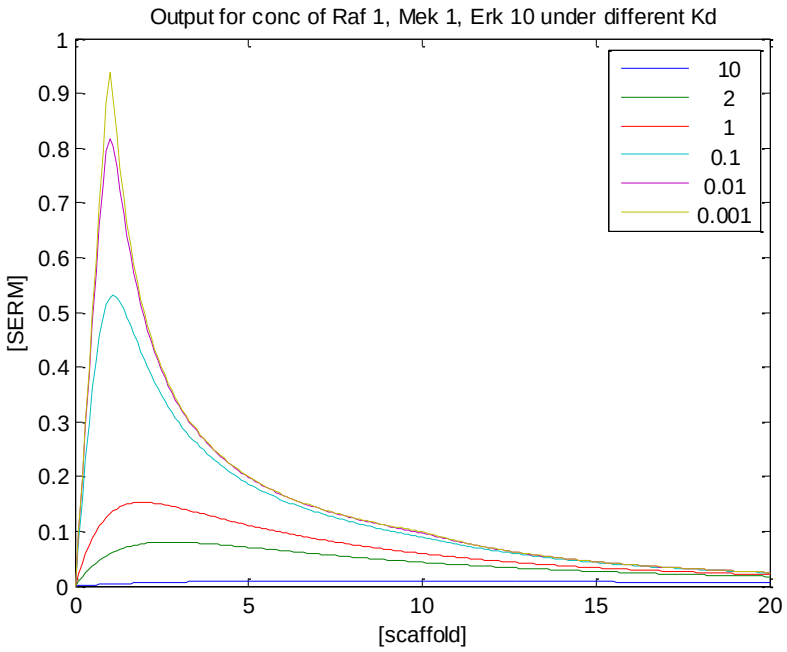


Figure 6.3 ODE simulations under normalised Raf/MEK/ERK value sets published by Legewie *et al*.[67] The concentration of scaffold is plotted against the concentration of the SERM complex, where Raf, MEK and ERK are bound under a range of dissociation constants. Concentrations are presented in normalised units. Arbitrary values of K_d under which the model was tested under are shown in the legend.

To understand why, when there is full binding of Raf, MEK and ERK to the scaffold to give SERM complex, optimal scaffold concentration is dependent on the lowest concentration of the signalling parameter, we consider model behaviour under differing numbers of scaffold molecules.

6.1.1 Model behaviour under 3 scaffold molecules

For example, if the system contains 3 scaffold molecules and 1 Raf molecule, 5 MEK molecules and 100 ERK molecules then, as there is only 1 Raf molecule available, only 1 full scaffold complex can be formed. Therefore optimal scaffold concentration is when 1 full complex is formed. In this example system as there is so much more ERK than scaffold protein, the likelihood that all 3 scaffolds will be bound to ERK is very high. If MEK binds strongly, then the 3 scaffolds will also be occupied with MEK, leaving 2 molecules in solution. As all 3 scaffolds feature both MEK and ERK, no matter which scaffold Raf binds to, one full scaffold complex will be formed.

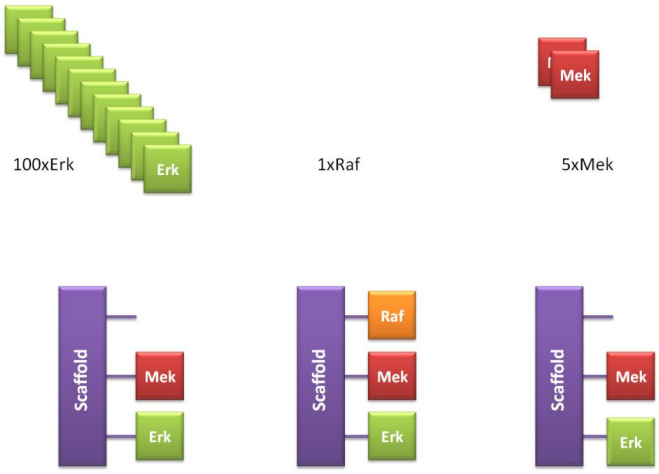


Figure 6.4 A schematic of model behaviour under three scaffold molecules.

6.1.2 Model behaviour under 5 scaffold molecules

When the number of scaffold molecules available increase to 5 then again ERK will have bound to all scaffolds as it is in excess. The scaffold concentration now equals the concentration of MEK, which is the second lowest signalling parameter. If the dissociation constant for MEK is very small, then MEK will bind strongly to all 5 scaffold molecules. So when Raf binds, one full scaffold complex is formed which is the optimal situation.

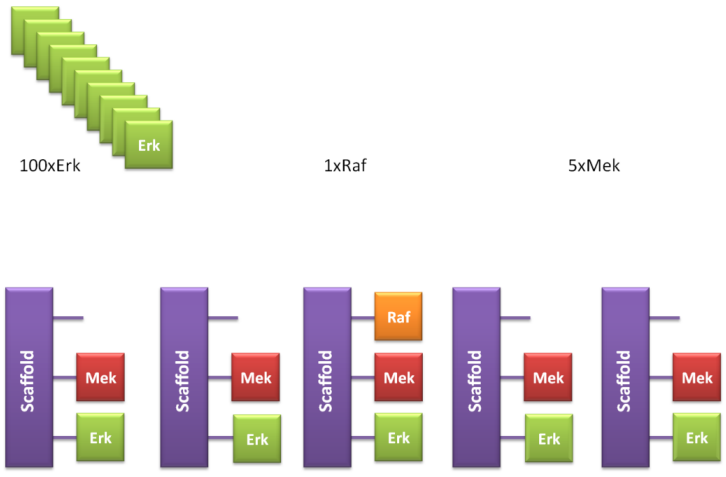


Figure 6.5 Schematic of model behaviour with five scaffold molecules present

6.1.3 Model behaviour under 6 scaffold molecules

The number of scaffold molecules increases to 6 and now there are more scaffold molecules than MEK molecules. Therefore even when MEK binds strongly there will be one scaffold molecule without MEK. Again as ERK is in excess, it is likely that all scaffolds will be bound to it. Raf may now bind to a scaffold molecule that contains ERK and MEK or it may bind to a scaffold which only contains ERK. There is a 5/6 probability that Raf will bind to the scaffold containing both ERK and MEK, but a 1/6 probability that no full scaffold will form and the complex SER will accumulate. Therefore scaffold concentrations greater than the second lowest signalling parameter are suboptimal.

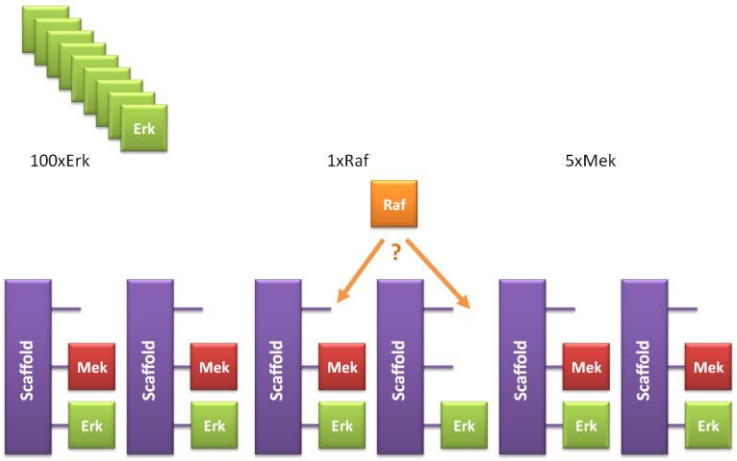


Figure 6.6 Schematic of model behaviour when there are six scaffolds present and less molecules of MEK.

6.1.4 Simplified Scaffold Model

The model described above can be simplified and instead it is possible to consider the variables of the binding sites of the scaffold rather than the differently occupied scaffold complexes. The model can now be set up with only 3 variables where each binding site can be either occupied or empty. It is assumed that each binding site interacts independently with each signalling protein and that the protein binds monovalently. The ratio of a signalling protein bound to the scaffold versus being free in solution is only a function of the dissociation constant.

The occupied binding site for Raf, MEK and ERK is termed SR^+, SM^+, SE^+ while the corresponding empty binding sites are called SR^-, SM^-, SE^-. The following ODEs can be constructed.

$$\frac{dSR^+}{dt} = k.SR^-.R - d.SR^+ \quad (12)$$

$$\frac{dSM^+}{dt} = k.SM^-.M - d.SM^+ \quad (13)$$

$$\frac{dSE^+}{dt} = k.SE^-.E - d.SE^+ \quad (14)$$

Considering the law of conservation, all occupied and unoccupied binding sites must sum up to the total concentration of scaffold S_T which is the total concentration of a binding site.

$$S_T = SR^+ + SR^- = SM^+ + SM^- = SE^+ + SE^- \quad (15)$$

The same law is applied to every signalling parameter:

$$R_T = R + SR^+ \quad (16)$$

$$M_T = M + SM^+ \quad (17)$$

$$E_T = E + SE^+ \quad (18)$$

By replacing the conservation laws (15) and (16-14) in the system (12-14), we obtain the following ODEs.

$$\frac{dSR^+}{dt} = k.(S_T - SR^+).(R_T - SR^+) - d.SR^+ \quad (19)$$

$$\frac{dSM^+}{dt} = k.(S_T - SM^+).(M_T - SM^+) - d.SM^+ \quad (20)$$

$$\frac{dSE^+}{dt} = k.(S_T - SE^+).(E_T - SE^+) - d.SE^+ \quad (21)$$

As each differential equation is independent it can be solved on its own and as each has a similar structure it is possible to write a generic ODE:

$$\frac{dC}{dt} = k.(S_T - C).(X_T - C) - d.C \quad (22)$$

Where C is any occupied binding site and X_T is the total concentration of any signalling component. By setting (22) to zero, a quadratic equation is obtained and the general solution applied. The solution then reads, due to the physical constraint that the minimum values of X_T and S_T are larger than the bound site, C:

$$C = \frac{X_T+S_T+K_D}{2} - \sqrt{\left(\frac{(X_T+S_T+K_D}{2}\right)^2 - S_T.X_T} \quad (23)$$

Where $K_D = d/k$ is the dissociation constant

The solution shows that the occupied binding site depends on the total concentration of scaffold and the respective signalling component that binds to this site, and on the respective dissociation constant of that binding process.

6.2 Conclusion

The models presented show that the amount of an occupied binding site depends on the total concentration of scaffold or scaffold binding site, the signalling protein that binds to this site and also the dissociation constant of the binding process. The values for the parameters used were based on experimental data previously reported in the literature by Schoeberl and Legewie as either number of copies per cell or total cellular concentration respectively. For simplicity the molecule per cell values were rescaled and arbitrary dissociation and association constants used. The model predicts that when all three signalling molecules are bound to scaffold then the upper limit determined is optimal at a scaffold concentration of 400, which when converted back to copies per cell (based on Schoeberl's values for Raf, MEK and ERK) is 8.8 x 10^6 copies per cell. In this study the Type A quantification values for the three component scaffold, KSR2 isoform 1 are calculated between 0.94 x 10^6 and 2.6 x 10^6 copies per cell across the cell lines (Table 5.6, Chapter 5). For ß-arrestin-2, another 3 component scaffold, a Type A quantification was calculated for cell line HT-29 at 1.6 x 10^6 copies per cell. Therefore the predicted values are close to the experimentally derived values and deviate at most by a factor of 10. Ideally, the model would be further validated if Type A quantifications for a range of three component scaffold proteins had been available. However, Type B quantifications where an upper limit on the abundance of the three component scaffold protein KSR1 was reported to be in the range of less than 0.26 x 10^6 to less than 4.1 x 10^6 copies per cell depending on the cell line. Given that the optimal scaffold concentration must be between the smallest and second smallest signalling component, these values and those obtained for KSR2 isoform 1 and ß-arrestin-2 fit with those quoted by Schoeberl where by Raf is 4 x 10^4 copies per cell and ERK is 2.1 x 10^7 copies per cell. However it is important to note that these values for Raf and ERK have been arrived at by an alternative method of biochemical quantification to QconCAT. There are inherent difficulties in reconciling data from different analyses due to differences in cell lines, growth conditions and analytical

workflows used. Instead it would be better to apply QconCAT methodology and SRM to the central MAPK cascade to obtain copies per cell for the central components of Raf, MEK and ERK. The comparison of data generated by the same method and then incorporated into a new model would be more predictive than trying to compare and use data for modelling which has been produced by different analytical methods. Overall, to fully validate this model then measurement by QconCAT of the concentration of all parameters including scaffold proteins and Raf, MEK, ERK, provided by QconCAT methodology would be extremely useful. Alongside determining absolute quantification data for the parameters of the model a further improvement would be to use experimentally derived rate constants rather than using arbitrary values.

7 Summary and Conclusion

The aim of the work presented in this thesis was to quantify the DUSPs, scaffold proteins and substrates of the MAPK pathway using the QconCAT technology and to undertake mathematical modelling to investigate this pathway. QconCATs were designed and synthesised to encode two analyte peptides per protein and, after stable isotope labelling of the standard *in vivo*, protein levels were determined by LC-MS, using ultra high performance liquid chromatography (UHPLC) coupled mass spectrometry. Absolute protein concentrations for some scaffold proteins were determined in the colon cancer cell lines HCT 116 and HT-29, and control cell line HEK-293. Where no native signal was observed an upper limit of copies per cell was possible for substrates, DUSPS and transcription factors. ODE modelling suggested that the optimal scaffold concentration was between the lowest and second lowest concentration of signalling components. The copies per cell value for scaffold protein KSR2, ß-arrestin-2 and upper limit calculated for KSR1 agrees with the range of values predicted by this model.

The development and production of QconCAT standards are discussed in Chapter 3. The constructs allowed the application of this methodology to proteins of the MAPK pathway which have largely been ignored by large scale models such as those by Schoeberl. Absolute quantification data for a subset of DUSPs, substrates and scaffold proteins were obtained.

Chapters 4 and 5 consider the development of a SRM assay for these QconCAT derived Q-peptides. Experimental and *in silico* predictions for *y*-ion transitions were used across two different platforms, the Xevo TQ MS and the TSQ Vantage. Prior to final experiments, sample preparation was optimised using RapiGest, peptide signals were assessed in a complex biological background and dwell times optimised to give the best quality of data for quantification. Final experiments assessed the upper limit of copies per cell for DUSPs, substrates and scaffold proteins for those peptides where a good signal (signal-to-noise ratio > 10) could be observed in whole cell lysates of colon cancer (HCT 116, HT-29) and control cell line (HEK-293). The lower limit of quantification was 10 fmol on column and although sensitivities of 0.1 fmol on column can be achieved for some peptides, this was often with pure peptides

rather than a complex biological sample due to ion suppression effects in the complex matrix. Native protein signals were observed and thus quantification was therefore achieved for some of the scaffold proteins. Where no native signal was observed it was possible to assign a lower limit of quantification when the QconCAT peptide was detected. It is likely that future advancements in instrumentation will enhance sensitivity and reduce the lower limit of quantification.

To better understand the contributions of the scaffold proteins to the MAPK pathway a mass action kinetics model was produced. The model, using normalised parameter sets and simulations, predicted that the optimal scaffold protein is between the lowest and second lowest concentration of signalling component. There is good agreement between this predicted value and that obtained for scaffold protein, KSR1, by absolute quantification. Ideally, data generated in the same way, i.e. using the QconCAT technology, would be used for quantifying the signalling components Raf, MEK and ERK, to develop models faithful to the biology.

Further developments to uncover more details would be to obtain quantification data for central cascade of the MAPK pathway including Raf, MEK and ERK proteins. This would produce a cohesive data set in which to produce a large scale model for the MAPK pathway to include previously ignored components including the scaffolds, DUSPs and substrates. To further understand the biological behaviour quantification of kinase signalling could be undertaken.[177] A QconCAT can be employed to determine the phosphorylation stoichiometry of the proteins. A further extension would be to apply the QconCAT approach to other signalling pathways such as PI3K/PTEN/Akt/mTOR which share common components and to investigate crosstalk between these pathways, as well as to define further the role of the scaffold proteins. The expansion of models using absolute quantification data is essential to gain a greater understanding of the normal and disease biology at the system level.

References

1. Schaeffer, H.J. and Weber, M.J., Mitogen-Activated Protein Kinases: Specific Messages from Ubiquitous Messengers. *Molecular and Cellular Biology*, 1999, **19**, 2435-2444.
2. Cargnello, M. and Roux, P.P., Activation and Function of the MAPKs and their Substrates, the MAPK-Activated Protein Kinases. *Microbiology and Molecular Biology Review*, 2011, **75**, 50-83.
3. Huang, C.Y.F. and Ferrell, J.E., Ultrasensitivity in the Mitogen-Activated Protein Kinase Cascade. *Proceedings of the National Academy of Sciences of the United States of America,* 1996, **93**, 10078-10083.
4. Goldbeter, A. and Koshland, D.E., An Amplified Sensitivity Arrising from Covalent Modification in Biological Systems. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 1981, **78**, 6840-6844.
5. Wellbrock, C., Karasarides, M., and Marais, R., The RAF Proteins Take Centre Stage. *Nature Reviews. Molecular and Cellular Biology,* 2004, **5**, 875-885.
6. Kolch, W., Meaningful Relationships: the Regulation of the Ras/Raf/MEK/ERK Pathway by Protein Interactions. *Biochemical Journal*, 2000, **351**, 289-305.
7. Yoon, S. and Seger, R., The Extracellular Signal-Regulated Kinase: Multiple Substrates Regulate Diverse Cellular Functions. *Growth Factors*, 2006, **24**, 21 - 44.
8. Lewis, T.S., Shapiro, P.S., and Ahn, N.G., Signal Transduction Through MAP Kinase Cascades. *Advances in Cancer Research*, 1998, **74,** 49-139.
9. Dhanasekaran, D.N., *et al.*, Scaffold Proteins of MAP-Kinase Modules. *Oncogene*, 2007, **26**, 3185-3202.
10. Schaeffer, H.J., Catling, A. D., *et al*., MP1: A MEK Binding Partner That Enhances Enzymatic Activation of the MAP Kinase Cascade. *Science*, 1998, **281**, 1668-1671.
11. Raman, M., Chen, W., and Cobb, M.H., Differential Regulation and Properties of MAPKs. *Oncogene*, 2007, **26**, 3100-3112.
12. DeFea, K.A., *et al*., The Proliferative and Antiapoptotic Effects of Substance P are Facilitated by Formation of a Beta-Arrestin Dependent Scaffolding Complex. *Proceedings of the National Academy of Sciences of the United States of America*, 2000, **97**, 11086-11091.
13. DeFea, K.A., *et al*., Beta-Arrestin Dependent Endocytosis of Proteinase-Activated Receptor 2 is Required for Intracellular Targeting of Activated ERK1/2. *The Journal of Cell Biology*, 2000, **148**, 1267-1281.
14. Tohgo, A., *et al*., Beta -Arrestin Scaffolding of the ERK Cascade Enhances Cytosolic ERK Activity but Inhibits ERK-Mediated Transcription Following Angiotensin AT1a Receptor Stimulation. *The Journal of Biological Chemistry*, 2002, **277**, 9429-9436.
15. Brown, M.C. and Turner, C.E., Paxillin: Adapting to Change. *Physiological Reviews*, 2004, **84**, 1315-1339.
16. Turner, C.E., Paxillin interactions. *Journal of Cell Science*, 2000, **113,** 4139-4140.

17. Wunderlich, W., *et al*., A Novel 14-Kilodalton Protein Interacts with the Mitogen-Activated Protein Kinase Scaffold MP1 on a Late Endosomal/Lysosomal Compartment. *The Journal of Cell Biology*, 2001, **152**, 765-776.
18. Teis, D., Wunderlich, W., and Huber, L.A., Localization of the MP1-MAPK Scaffold Complex to Endosomes is Mediated by p14 and Required for Signal Transduction. *Developmental Cell*, 2002, **3**, 803-814.
19. Pullikuth, A., *et al*., The MEK1 Scaffolding Protein MP1 Regulates Cell Spreading by Integrating PAK1 and Rho Signals. *Molecular and Cellular Biology*, 2005, **25**, 5119-5133.
20. Li, W.Q., Han, M., and Guan, K.L., The Leucine-Rich Repeat Protein SUR-8 Enhances MAP Kinase Activation and Forms a Complex with Ras and Raf. *Genes & Development*, 2000, **14**, 895-900.
21. Rodriguez-Viciana, P., *et al*., A Phosphatase Holoenzyme Comprised of Shoc2/Sur8 and the Catalytic Subunit of PP1 Functions as an M-Ras Effector to Modulate Raf Activity. *Molecular Cell*, 2006, **22**, 217-230.
22. Hill, J.M., *et al*., Recognition of ERK MAP Kinase by PEA-15 Reveals a Common Docking Site within the Death Domain and Death Effector Domain. *EMBO Journal*, 2002, **21,** 6494-6504.
23. Callaway, K., *et al.*, The Anti-Apoptotic Protein PEA-15 is a Tight Binding Inhibitor of ERK1 and ERK2, which Blocks Docking Interactions at the D-Recruitment Site. *Biochemistry*, 2007, **46**, 9187-9198.
24. Whitehurst, A.W., *et al*., The Death Effector Domain Protein PEA-15 Prevents Nuclear Entry of ERK2 by Inhibiting Required Interactions. *The Journal of Biological Chemistry*, 2004, **279**, 12840-12847.
25. Formstecher, E., *et al*., PEA-15 Mediates Cytoplasmic Sequestration of ERK MAP Kinase. *Developmental Cell*, 2001, **1**, 239-250.
26. Vaidyanathan, H., *et al*., ERK MAP Kinase is Targeted to RSK2 by the Phosphoprotein PEA-15. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 19837-19842.
27. Roy, M., Li, Z.G., and Sacks, D.B., IQGAP1 Binds ERK2 and Modulates its Activity. *The Journal of Biological Chemistry*, 2004, **279**, 17329-17337.
28. Jadeski, L., *et al*., IQGAP1 Stimulates Proliferation and Enhances Tumorigenesis of Human Breast Epithelial Cells. *The Journal of Biological Chemistry*, 2008, **283**, 1008-1017.
29. Rome, L., Kedersha, N., and Chugani, D., Unlocking Vaults: Organelles in Search of a Function. *Trends in Cell Biology*, 1991, **1**, 47-50.
30. Kolli, S., *et al.*, The Major Vault Protein Is a Novel Substrate for the Tyrosine Phosphatase SHP-2 and Scaffold Protein in Epidermal Growth Factor Signaling. *The Journal of Biological Chemistry*, 2004, **279**, 29374-29385.
31. Yeung, K., *et al*., Suppression of Raf-1 Kinase Activity and MAP Kinase Signalling by RKIP. *Nature*, 1999, **401,** 173-177.
32. Park, S., *et al*., RKIP Downregulates B-Raf Kinase Activity in Melanoma Cancer Cells. *Oncogene*, 2005, **24**, 3535-3540.
33. Chatterjee, D., *et al*., RKIP Sensitizes Prostate and Breast Cancer Cells to Drug-Induced Apoptosis. *The Journal of Biological Chemistry*, 2004, **279**, 17515-17523.

34. Fu, Z., *et al*., Effects of Raf Kinase Inhibitor Protein Expression on Suppression of Prostate Cancer Metastasis. *Journal of the National Cancer Institute,* 2003, 878-889.
35. Schuierer, M.M., *et al*., Reduction in Raf Kinase Inhibitor Protein Expression is Associated with Increased Ras-Extracellular Signal-Regulated Kinase Signaling in Melanoma Cell Lines. *Cancer Research*, 2004, **64**, 5186-5192.
36. Keller, E.T., Metastasis Suppressor Genes: a Role for Raf Kinase Inhibitor Protein (RKIP). *Anti-Cancer Drugs*, 2004, **15**, 663-669.
37. Theodosiou, A. and Ashworth, A., MAP Kinase Phosphatases. *Genome Biology*, 2002, **3**, 3009.1-3009.10.
38. Kondoh, K. and Nishida, E., Regulation of MAP Kinases by MAP Kinase Phosphatases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 2007, **1773**, 1227-1237.
39. Karlsson, M., *et al*., Both Nuclear-Cytoplasmic Shuttling of the Dual Specificity Phosphatase MKP-3 and Its Ability to Anchor MAP Kinase in the Cytoplasm Are Mediated by a Conserved Nuclear Export Signal. *The Journal of Biolological Chemistry*, 2004, **279**, 41882-41891.
40. Caunt, C.J., *et al*., Epidermal Growth Factor Receptor and Protein Kinase C Signaling to ERK2: Spatiotemporal Regulation of ERK2 by DUSPs. *The Journal of Biological Chemistry,* 2008. **283**, 6241-6252.
41. Chen, R.E. and Thorner, J., Systems Biology Approaches in Cell Signaling Research. *Genome Biology*, 2005, **6**, 235.1-235.5.
42. Press, W.H., *et al*., *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing.* 2nd edition, 1992, Cambridge University Press, 718-725.
43. Orton, R.J., *et al*., Computational Modelling of the Receptor-Tyrosine-Kinase-Activated MAPK Pathway. *Biochemical Journal*, 2005, **392**, 249-261.
44. Blüthgen, N., *et al*., Effects of Sequestration on Signal Transduction Cascades. *FEBS Journal*, 2006, **273**, 895-906.
45. Blüthgen, N. and Legewie, S., Systems Analysis of MAPK Signal Transduction. *Essays in Biochemistry: Systems Biology,* 1st edition, 2008, Portland Press Ltd, 95-107.
46. Fell, D.A. and Sauro, H.M., Metabolic Control and its Analysis- Additional Relationships Between Elastacities and Control Coefficients. *European Journal of Biochemistry*, 1985, **148**, 555-561.
47. Ferrell, J.E. and Bhatt, R.R., Mechanistic Studies of the Dual Phosphorylation of Mitogen-Activated Protein Kinase. *Journal of Biological Chemistry*, 1997, **272**, 19008-19016.
48. Burack, W.R. and Sturgill, T.W., The Activating Dual Phosphorylation of MAPK by MEK is Nonprocessive. *Biochemistry*, 1997, **36**,5929-5933.
49. Ferrell, J.E., and Machleder, E.M., The Biochemical Basis of an All-or-None Cell Fate Switch in Xenopus Oocytes. *Science*, 1998, **280**, 895-898.
50. Kholodenko, B.N., Negative Feedback and Ultrasensitivity can Bring About Oscillations in the Mitogen-Activated Protein Kinase Cascades. *European Journal of Biochemistry*, 2000, **267**, 1583-1588.
51. Kolch, W., Calder, M., and Gilbert, D., When Kinases Meet Mathematics: the Systems Biology of MAPK Signalling. *FEBS Letters*, 2005, **579**, 1891-1895.

52. Levchenko, A., Bruck, J., and Sternberg, P.W., Scaffold Proteins may Biphasically Affect the Levels of Mitogen-Activated Protein Kinase Signaling and Reduce its Threshold Properties. *Proceedings of the National Academy of Sciences of the United States of America,* 2000, **97**, 5818-5823.
53. Somsen, O.J.G., *et al*., Selectivity in Overlapping MAP Kinase Cascades. *Journal of Theoretical Biology,* 2002, **218**,343-354.
54. Brightman, F.A. and Fell, D.A., Differential Feedback Regulation of the MAPK Cascade Underlies the Quantitative Differences in EGF and NGF Signalling in PC12 cells. *FEBS Letters,* 2000, **482**, 169-174.
55. Schoeberl, B., *et al*., Computational Modeling of the Dynamics of the MAP Kinase Cascade Activated by Surface and Internalized EGF Receptors. *Nature Biotechnology*, 2002, **20**, 370-375.
56. Hatakeyama, M., *et al.,* A Computational Model on the Modulation of Mitogen-Activated Protein Kinase (MAPK) and Akt Pathways in Heregulin-Induced ErbB Signalling. *Biochemical Journal*, 2003, **373**, 451-463.
57. Gong, Y.C. and Zhao, X., Shc-Dependent Pathway is Redundant but Dominant in MAPK Cascade Activation by EGF Receptors: a Modeling Inference. *FEBS Letters*, 2003, **554**, 467-472.
58. Hornberg, J.J., *et al*., Control of MAPK Signalling: from Complexity to What Really Matters. *Oncogene*, 2005, **24**, 5533-5542.
59. Swain, P.S. and Siggia, E.D., The Role of Proofreading in Signal Transduction Specificity. *Biophysical Journal*, 2002, **82**, 2928-2933.
60. Markevich, N.I., Hoek, J.B., and Kholodenko, B.N., Signaling Switches and Bistability Arising from Multisite Phosphorylation in Protein Kinase Cascades. *The Journal of Cell Biology*, 2004, **164**, 353-359.
61. Chapman, S. and Asthagiri, A.R., Resistance to Signal Activation Governs Design Features of the MAP Kinase Signaling Module. *Biotechnology and Bioengineering*, 2004, **85**,311-322.
62. Ferrell, J.E., Tripping the Switch Fantastic: How a Protein Kinase Cascade can Convert Graded Inputs into Switch-Like Outputs. *Trends in Biochemical Sciences*, 1996, **21**, 460-466.
63. Goldbeter, A., A Minimal Cascade Model for the Mitotic Oscillator Involving Cyclin and CDC2 Kinase. *Proceedings of the National Academy of Sciences of the United States of America,* 1991, **88**, 9107-9111.
64. Blüthgen, N., Sequestration Shapes the Response of Signal Transduction Cascades. *IUBMB Life*, 2006, **58**, 659-663.
65. Ferrell, J.E., Self-Perpetuating States in Signal Transduction: Positive Feedback, Double-Negative Feedback and Bistability. *Current Opinion in Cell Biology,* 2002, **14**, 140-148.
66. Bhalla, U.S. and Iyengar, R., Emergent Properties of Networks of Biological Signaling Pathways. *Science*, 1999, **283**, 381-387.
67. Legewie, S., *et al*., Competing Docking Interactions can Bring About Bistability in the MAPK Cascade. *Biophysical Journal*, 2007, **93**, 2279-2288.
68. Borman S., Russell, H., Siuzdak, G., A Mass Spectrometery Timeline. *Today's Chemist*, 2003, 47-49.
69. Dole, M., Mack, L.L., and Hines, R.L., Molecular Beams of Macroions. *Journal of Chemical Physics,* 1968, **49**, 2240-2249.

70. Fenn, J.B., Ion Formation from Charged Droplets- Roles of Geometry, Energy, and Time. *Journal of the American Society for Mass Spectrometry*, 1993, **4**, 524-535.
71. Gaskell, S.J., Electrospray: Principles and Practice. *Journal of Mass Spectrometry*, 1997, **32**, 677-688.
72. Barber, M., *et al*., Fast Atom Bombardment of Solids (F.A.B.): a New Ion Source for Mass Spectrometry. *Journal of the Chemical Society, Chemical Communications*, 1981, 325-327.
73. Mack, L.L., *et al*., Molecular Beams of Macroions. II. *The Journal of Chemical Physics*, 1970, **52**, 4977-4986.
74. Iribarne, J.V. and Thomson, B.A., Evaporation of Small Ions from Charged Droplets. *Journal of Chemical Physics*, 1976, **64**, 2287-2294.
75. Wilm, M., Principles of Electrospray Ionization. *Molecular & Cellular Proteomics*, 2011, **10**, 9407.1-9407.8.
76. Fernandez de la Mora, J., Electrospray Ionization of Large Multiply Charged Species Proceeds via Dole's Charged Residue Mechanism. *Analytica Chimica Acta*, 2000, **406**, 93-104.
77. Wang, G. and Cole, R.B., Charged Residue Versus Ion Evaporation for Formation of Alkali Metal Halide Cluster Ions in ESI. *Analytica Chimica Acta*, 2000, **406**, 53-65.
78. Gamero-Castaño, M. and Mora, J.F., Kinetics of Small Ion Evaporation from the Charge and Mass Distribution of Multiply Charged Clusters in Electrosprays. *Journal of Mass Spectrometry*, 2000, **35**, 790-803.
79. Mamyrim, B.A, Time-of-Flight Mass Spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry*, 2001, **206**, 251-266.
80. Lee, M.J., *et al*., A Preliminary Study Using Fast Gradient Liquid Chromatography Coupled to a Quadrupole Orthogonal Time-of-Flight Mass Spectrometer. *Rapid Communications in Mass Spectrometry*, 1999, **13**, 216-221.
81. Glish, G.L. and Goeringer, D.E., Tandem Quadrupole-Time-of-Flight Instrument for Mass-Spectrometry. *Analytical Chemistry*, 1984, **56**, 2291-2295.
82. Glish, G.L. and Burinsky, D.J., Hybrid Mass Spectrometers for Tandem Mass Spectrometry. *Journal of the American Society for Mass Spectrometry*, 2008, **19**, 161-172.
83. Dawson, J.H.J. and Guilhaus, M., Orthogonal-Acceleration Time-of-Flight Mass Spectrometer. *Rapid Communications in Mass Spectrometry*, 1989, **3**, 155-159.
84. Guilhaus, M., Spontaneous and Deflected Drift-Trajectories in Orthogonal Acceleration Time-of-Flight Mass Spectrometry. *Journal of the American Society for Mass Spectrometry*, 1994, **5**, 588-595.
85. Guilhaus, M., Essential Elements of Time-of-Flight Mass Spectrometry in Combination with the Inductively Coupled Plasma Ion Source. *Spectrochimica Acta Part B-Atomic Spectroscopy*, 2000, **55**, 1511-1525.
86. Paul, W. and Steinwedel, H., Ein neues Massenspektrometer Ohne Magnetfeld. *A Journal of Physical Sceinces (Zeitschrift für Naturforschung A)*, 1953, **8**, 448–450.

87. Stafford, G.C., *et al.*, Recent Improvements in and Analytical Applications of Advanced Ion Trap Technology. *International Journal of Mass Spectrometry and Ion Processes,* 1984, **60**, 85-98.
88. Paul, W., Electromagnetic Traps for Charged and Neutral Particles (Nobel Lecture). *Angewandte Chemie International Edition in English*, 1990, **29**, 739-748.
89. Douglas, D.J., Frank, A.J., and Mao D., Linear Ion Traps in Mass Spectrometry. *Mass Spectrometry Reviews*, 2005, **24**, 1-29.
90. Ahmed, F.E., Utility of Mass Spectrometry for Proteome and Lysis: Part I. Conceptual and Experimental Approaches. *Expert Review of Proteomics*, 2008, **5**, 841-864.
91. Schwartz, J.C., Senko, M.W., and Syka, J.E.P., A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer. *Journal of the American Society for Mass Spectrometry*, 2002, **13**, 659-669.
92. Yates, J.R., Ruse, C.I., and Nakorchevsky, A., Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annual Review of Biomedical Engineering*, 2009, **11**, 49-79.
93. Olsen, J.V., *et al*., A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular & Cellular Proteomics*, 2009, **8**,2759-2769.
94. Hager, J.W., A New Linear Ion Trap Mass Spectrometer. *Rapid Communications in Mass Spectrometry*, 2002, **16**, 512-526.
95. Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry*, 2000, **72**, 1156-1162.
96. Hu, Q., *et al*., The Orbitrap: a New Mass Spectrometer. *Journal of Mass Spectrometry*, 2005, **40**, 430-443.
97. Creaser, C.S., *et al*., Ion Mobility Spectrometry: a Review. Part 1. Structural Analysis by Mobility Measurement. *Analyst*, 2004, **129**, 984-994.
98. Hogg, A.M. and Kebarle, P., Mass-Spectrometric Study of Ions at Near-Atmospheric Pressure. II. Ammonium Ions Produced by the Alpha Radiolysis of Ammonia and Their Solvation in the Gas Phase by Ammonia and Water Molecules. *The Journal of Chemical Physics*, 1965, **43**, 449-456.
99. Kebarle, P. and Hogg, A.M., Mass-spectrometric Study of Ions at Near Atmospheric Pressures. I. The Ionic Polymerization of Ethylene. *The Journal of Chemical Physics*, 1965, **42**, 668-674.
100. Cohen, M.J. and Karasek, F.W., Plasma Chromatography – a New Dimension for Gas Chromatography and Mass Spectrometry. Journal *of Chromatographic Science*, 1970, **8**, 330-337
101. Purves, R.W., *et al*., Mass Spectrometric Characterization of a High-Field Asymmetric Waveform Ion Mobility Spectrometer. *Review of Scientific Instruments,* 1998, **69**, 4094-4105.
102. Giles, K., *et al*., Applications of a Travelling Wave-Based Radio-Frequency-Only Stacked Ring Ion Guide. *Rapid Communications in Mass Spectrometry*, 2004, **18**, 2401-2414.
103. Hoaglund, C.S., *et al*., Three-Dimensional Ion Mobility/TOFMS Analysis of Electrosprayed Biomolecules. *Analytical Chemistry*, 1998, **70**, 2236-2242.

104. Henderson, S.C., *et al*., ESI/Ion Trap/Ion Mobility/Time-of-Flight Mass Spectrometry for Rapid and Sensitive Analysis of Biomolecular Mixtures. *Analytical Chemistry*, 1999, **71**, 91-301.
105. Dugourd, P., *et al*., High-Resolution Ion Mobility Measurements. *Review of Scientific Instruments*, 1997, 1122-1129.
106. Bluhm, B.K., Gillig, K.J., and Russell, D.H., Development of a Fourier-Transform Ion Cyclotron Resonance Mass Spectrometer-Ion Mobility Spectrometer. *Review of Scientific Instruments*, 2000, **71**, 4078-4086.
107. Harvey, S.R., MacPhee, C.E. and Barran, P.E., Ion Mobility Mass Spectrometry for Peptide Analysis. *Methods*, 2011, **54**, 454-461.
108. Pringle, S.D., *et al*., An Investigation of the Mobility Separation of some Peptide and Protein Ions Using a New Hybrid Quadrupole/Travelling Wave IMS/oa-ToF Instrument. *International Journal of Mass Spectrometry*, 2007, **261**, 1-12.
109. Houel, S, *et al.*, Quantifying the Impact of Chimera MS/MS Spectra on Peptide Identificaiton in Large-Scale Proteomic Studies. *Journal of Proteome Research*, 2010, **9**, 4152-4160.
110. Giles, K., Williams, J.P., and Campuzano, I., Enhancements in Travelling Wave Ion Mobility Resolution. *Rapid Communications in Mass Spectrometry*, 2011, **25,** 1559-1566.
111. Steen, H. and Mann, M., The abc's (and xyz's) of Peptide Sequencing. *Nature Reviews. Molecular Cellular Biology*, 2004, 699-711.
112. Biemann, K., Contributions Of Mass-Spectrometry To Peptide And Protein-Structure. *Biomedical And Environmental Mass Spectrometry*, 1988, **16**, 99-111.
113. Summerfield, S.G. and Gaskell, S.J., Fragmentation Efficiencies of Peptide Ions Following Low Energy Collisional Activation. *International Journal of Mass Spectrometry and Ion Processes*, 1997, **165,** 509-521.
114. Ashcroft, A.E., Protein and Peptide Identification: The Role of Mass Spectrometry in Proteomics. *Natural Product Report*, 2003, **20**, 202 - 215.
115. Silva, J.C., *et al*., Absolute Quantification of Proteins by LCMSE. *Molecular & Cellular Proteomics*, 2006, **5**, 144-156.
116. Silva, J.C., *et al*., Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs. *Analytical Chemistry*, 2005, **77**, 2187-2200.
117. Graves, P.R. and Haystead, T.A.J., Molecular Biologist's Guide to Proteomics. *Microbiology and Molecular Biology Review*, 2002, **66**, 39-63.
118. Gygi, S.P., *et al*., Evaluation of Two-Dimensional Gel Electrophoresis-Based Proteome Analysis Technology. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 2000, **97,** 9390-9395.
119. Clauser, K.R., Baker, P., and Burlingame, A.L., Role of Accurate Mass Measurement (+/- 10 ppm) in Protein Identification Strategies Employing MS or MS MS and Database Searching. *Analytical Chemistry*, 1999, **71**, 2871-2882.
120. Kelleher, N.L., *et al*., Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *Journal of the American Chemical Society*, 1999, **121**, 806-812.
121. Yates, J.R., *et al*., Method To Correlate Tandem Mass-Spectra Of Modified Peptides To Amino-Acid-Sequences In The Protein Database. *Analytical Chemistry*, 1995, **67**, 1426-1436.

122. David, N.P., *et al*., Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data, *Electrophoresis*, 1999, **20**, 3551-3567.
123. Eng, J.K., McCormack, A.L., and Yates, J.R., An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *Journal of the American Society for Mass Spectrometry*, 1994, **5**, 976-989.
124. Weber, K and Osborn, M., The Reliability of Molecular Weight Determinations by Dodecyl Sulfate-Polyacrylamide Gel Electophoresis. *The Journal of Biological Chemistry*, 1969, **244**, 4406-4412.
125. Laemmli, U.K., Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 1970, **227**, 680-685.
126. Kohlrausch, F., Annalen der Physik und Chemie, 1897, **62**, 209-211.
127. Delahunty, C. and Yates, J.R., Protein Identification Using 2D-LC-MS/MS. *Methods*, 2005, **35**, 248-255.
128. Cutillas, P., Burlingame A., and Unwin, R., Proteomic Strategies and Their Application in Studies of Renal Function. *Physiology*, 2004, **19**, 114-119.
129. Link, A.J., *et al*., Direct Analysis of Protein Complexes Using Mass Spectrometry. *Nature Biotechnology*, 1999, **17**, 676-682.
130. Noga, M., *et al*., A Practical Guide to Nano-LC Troubleshooting. *Journal of Separation Science*, 2007, **30**, 2179-2189.
131. Hernandez-Borges, J., *et al*., Recent Applications in Nanoliquid Chromatography. *Journal of Separation Science*, 2007, **30**, 1589-1610.
132. Plumb, R.S., *et al*., A Rapid Screening Approach to Metabonomics Using UPLC and oa-TOF Mass Spectrometry: Application to Age, Gender and Diurnal Variation in Normal/Zucker Obese Rats and Black, White and Nude Mice. *Analyst*, 2005, **130**, 844-849.
133. Philippe, E.M., *et al*., Protein Fractionation in a Multicompartment Device Using OFF-GEL Isoelectric Focusing. *Electrophoresis*, 2003, **24**, 3-11.
134. Eng, J.K., McCormack, A.L., and Yates, J.R., An Approach To Correlate Tandem Mass-Spectral Data Of Peptides With Amino-Acid-Sequences In A Protein Database. *Journal Of The American Society For Mass Spectrometry*, 1994, **5**, 976-989.
135. Perkins, D.N., *et al*., Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis*, 1999, **20**, 3551-3567.
136. Mann, M. and Wilm, M., Error Tolerant Identification Of Peptides In Sequence Databases By Peptide Sequence Tags. *Analytical Chemistry*, 1994, **66**, 4390-4399.
137. Cottrell, J.S., Protein Identification Using MS/MS Data. *Journal of Proteomics*, 2011, **74**,1842-1851.
138. Eng, J.K., *et al*., A Face in the Crowd: Recognizing Peptides Through Database Search. *Molecular & Cellular Proteomics*, 2011, **10**, R111.009522–1- R111.009522–9.
139. Makarov, A., *et al*., Dynamic Range of Mass Accuracy in LTQ Orbitrap Hybrid Mass Spectrometer. *Journal of the American Society for Mass Spectrometry*, 2006, **17**, 977-982.
140. Barr, J.R., *et al*., Isotope Dilution-Mass Spectrometric Quantification of Specific Proteins: Model Application with Apolipoprotein A-I. *Clinical Chemistry*, 1996, **42**, 1676-1682.

141. Oda, Y., *et al*., Accurate Quantitation of Protein Expression and Site-Specific Phosphorylation. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 1999, **96**, 6591-6596.
142. Lipton, M.S., *et al*., Global Analysis of the Deinococcus Radiodurans Proteome by Using Accurate Mass Tags. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 2002, **99**, 11049-11054.
143. Smith, R.D., *et al*., An Accurate Mass Tag Strategy for Quantitative and High-Throughput Proteome Measurements. *Proteomics*, 2002, **2**, 513-523.
144. Smith, R.D., *et al*., Rapid Quantitative Measurements of Proteomes by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Electrophoresis*, 2001, **22**, 1652-1668.
145. Everley, P.A., *et al*., Quantitative Cancer Proteomics: Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) as a Tool for Prostate Cancer Research. *Molecular & Cellular Proteomics*, 2004, **3**, 729-735.
146. Yan, W. and Chen, S.S., Mass Spectrometry-Based Quantitative Proteomic Profiling. *Briefings in Functional Genomics and Proteomics*, 2005, **4**, 27-38.
147. Neubert, T.A. and Tempst, P., Super-SILAC for Tumors and Tissues. *Nature Methods*, 2010, **7**, 361-362.
148. Gygi, S.P., *et al*., Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nature Biotechnology*, 1999, **17**, 994-999.
149. Ross, P.L., *et al*., Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics*, 2004, **3**, 1154-1169.
150. Unwin, R.D., *et al*., Quantitative Proteomic Analysis Using Isobaric Protein Tags Enables Rapid Vomparison of Changes in Transcript and Protein Levels in Transformed Cells. *Molecular & Cellular Proteomics*, 2005, **4**, 924-935.
151. Gerber, S.A., *et al*., Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 2003, **100**, 6940-6945.
152. Picotti, P., *et al*., Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. *Cell*, 2009, **138**, 795-806.
153. Pratt, J.M., *et al*., Multiplexed Absolute Quantification for Proteomics Using Concatenated Signature Peptides Encoded by QconCAT Genes. *Nature Protocols*, 2006, **1**, 1029-1043.
154. Mirzaei, H., *et al*., Comparative Evaluation of Current Peptide Production Platforms Used in Absolute Quantification in Proteomics. *Molecular & Cellular Proteomics*, 2008, **7**, 813-823.
155. Silva, J.C., *et al*., Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms Acquisition. *Molecular & Cellular Proteomics*, 2006, **5**, 144-156.
156. Old, W.M., *et al*., Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Molecular & Cellular Proteomics*, 2005, **4**, 1487-1502.
157. Carroll, K.M., *et al*., Absolute Quantification of the Glycolytic Pathway in Yeast. *Molecular & Cellular Proteomics*, 2011, **10**, M111.007633-1-M111.007633-15.

158. Hanke, S., *et al*., Absolute SILAC for Accurate Quantitation of Proteins in Complex Mixtures Down to the Attomole Level. *Journal of Proteome Research,* 2008, **7**, 1118-1130.
159. Wisniewski, J.R., *et al*., Universal Sample Preparation Method for Proteome Analysis. *Nature Methods,* 2009, **6**, 359-362.
160. Eyers, C.E., *et al*., QCAL a Novel Standard for Assessing Instrument Conditions for Proteome Analysis. *Journal of the American Society for Mass Spectrometry,* 2008, **19**, 1275-1280.
161. MacLean, B., *et al*., Skyline: an Open Source Document Editor for Creating and Analyzing Targeted Proteomics experiments. *Bioinformatics,* 2010. **26**, 966-968.
162. Eyers, C.E., *et al*., CONSeQuence: Prediction of Reference Peptides for Absolute Quantitative Proteomics Using Consensus Machine Learning Approaches. *Molecular & Cellular Proteomics,* 2011, **10**, M110.003384-1-M110.003384-12.
163. Siepen, J.A., *et al*., Prediction of Missed Cleavage Sites in Tryptic Peptides Aids Protein Identification in Proteomics. *Journal of Proteome Research,* 2006, **6**, 399-408.
164. Van De Steene, J.C., Mortier, K.A., and Lambert, W.E., Tackling Matrix Effects During Development of a Liquid Chromatographic–Electrospray Ionisation Tandem Mass Spectrometric Analysis of Nine Basic Pharmaceuticals in Aqueous Environmental Samples. *Journal of Chromatography A,* 2006, **1123**, 71-81.
165. Zhang, G. and Wujcik, C.E., Overcoming Ionization Effects Through Chromatography: A Case Study for the ESI-LC–MS/MS Quantitation of a Hydrophobic Therapeutic Agent in Human Serum Using a Stable-Label Internal Standard. *Journal of Chromatography B,* 2009, **877**, 2003-2010.
166. Brownridge, P., *et al.*, Global Absolute Quantification of a Proteome: Challenges in the Deployment of a QconCAT Strategy. *Proteomics,* 2011, **11**, 2957-2970
167. King, R., *et al.*, Mechanistic Investigation of Ionization Suppression in Electrospray Ionization. *Journal of the American Society for Mass Spectrometry,* 2000, **11**, 942-950.
168. Reimer, J., *et al*., Effect of Cyclization of N-Terminal Glutamine and Carbamidomethyl-Cysteine (Residues) on the Chromatographic Behavior of Peptides in Reversed-Phase Chromatography. *Journal of Chromatography A.* **1218,** 5101-5107.
169. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Rinner, O., Ellenberg, J. and Aebersold, R., The Quantitative Proteome of a Human Cell Line. *Molecular Systems Biology*, 2011, **7**, 549.1-549-8.
170. Zeiler, M., Straube, W.L., Lundberg, E., Uhlen, M. and Mann, M., A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines. *Molecular & Cellular Proteomics,* 2012, **11,** O111.009613-1-O111.009613-13.
171. Wang, Y., *et al*., Reversed-Phase Chromatography with Multiple Fraction Concatenation Strategy for Proteome Profiling of Human MCF10A Cells. *Proteomics*, 2011, **11**, 2019-2026.

172. Holstein, C.A., Gafken, P.R., and Martin, D.B., Collision Energy Optimization of b- and y-Ions for Multiple Reaction Monitoring Mass Spectrometry. *Journal of Proteome Research*, 2010, **10**, 231-240.
173. Good, M.C., Zalatan, J.G., and Lim, W.A., Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science*, 2011, **332**, 680-686.
174. Locasale, J.W., Shaw, A.S., and Chakraborty, A.K., Scaffold Proteins Confer Diverse Regulatory Properties to Protein Kinase Cascades. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 13307-13312.
175. Burack, W.R. and Shaw, A.S., Signal Transduction: Hanging on a Scaffold. *Current Opinion in Cell Biology*, 2000, **12**, 211-216.
176. Bray, D. and Lay, S., Computer-Based Analysis of the Binding Steps in Protein Complex Formation. *Proceedings of the National Academy of Sciences*, 1997, **94**, 13493-13498.
177. Cutillas, P.R. and Jorgensen, C., Biological Signalling Activity Measurements Using Mass Spectrometry. *Biochemical Journal*, 2011, **434**, 189-199.

Appendices

Appendix 1

Initial results from cell lysate, HT-29, and QconCAT LM2 experiment. (C) represents carbamidomethylation modification and bold represents labelled amino acid residue. Table continues over several pages.

Q	Sequence	Precursor m/z	y	y-ion m/z	Ret t, min	AUC	Ratio	Ret t, min	AUC	Ratio	Mean	Std dev	CV, %
1	VPVNDSF(C)EK	597.78	y7	899.36	28.51	0		0	0				
	VPVNDSF(C)EK		y8	998.42	0	0		0	0				
	VPVNDSF(C)EK		y9	1095.48	0	0		0	0				
	VPVNDSF(C)E**K**	600.79	y7	905.38	28.58	603		28.66	483				
	VPVNDSF(C)E**K**		y8	1004.44	28.59	292		28.71	268				
	VPVNDSF(C)E**K**		y9	1101.5	28.58	150		28.68	61				
2	SAEWLQEELEAR	730.86	y5	617.32	0	0		0	0				
	SAEWLQEELEAR		y7	874.42	0	0		0	0				
	SAEWLQEELEAR		y8	987.51	0	0		0	0				
	SAEWLQEELEA**RR**	733.87	y5	623.34	42.83	341		43.36	37				
	SAEWLQEELEA**RR**		y7	880.44	42.83	963		43.33	148				
	SAEWLQEELEA**RR**		y8	993.53	42.75	783		43.31	107				
3	QPSVSGLSQITK	622.85	y8	833.47	33.66	96		33.73	40				
	QPSVSGLSQITK		y7	746.44	33.95	3628		34.09	3285				
	QPSVSGLSQITK		y10	1019.57	0			noise					
	QPSVSGLSQIT**K**	625.86	y8	839.49	33.57	4597	0.021	33.8	2970	0.013	0.017	0.0052	30.53
	QPSVSGLSQIT**K**		y7	752.46	33.59	2538	1.429	33.8	1747	1.88	1.65	0.32	19.27
	QPSVSGLSQIT**K**		y10	1025.59	33.59	667		33.81	410				

4	YVLPDEAAR	517.27	y6	658.32	31.46	1488		31.6	1441				
	YVLPDEAAR		y7	771.4	31.48	515		31.58	464				
	YVLPDEAAR		y4	446.24	31.48	201		31.65	240				
	YVLPDEAA**R**	520.28	y6	664.34	31.46	130780	0.011	31.6	1E+05	0.012	0.011	0.00031	2.71
	YVLPDEAA**R**		y7	777.42	31.48	55316	0.009	31.6	48380	0.01	0.0095	0.0002	2.09
	YVLPDEAA**R**		y4	452.26	31.48	19926	0.01	31.6	17750	0.014	0.0118	0.0024	20.57
5	YILNVTPNLPNFFEK	904.98	y9	1105.56	0	0		0	0				
	YILNVTPNLPNFFEK		y10	1206.31	0	0		0	0				
	YILNVTPNLPNFFEK		y11	1305.68	0	0		0	0				
	YILNVTPNLPNFFE**K**	907.99	y9	1111.58	0	0		0	0				
	YILNVTPNLPNFFE**K**		y10	1212.33	0	0		0	0				
	YILNVTPNLPNFFE**K**		y11	1311.7	0	0		0	0				
6	NTVHMVSSPVGMIPDIYEK	(3+) 706.3	y7	877.467	31.05	21		0	0				
	NTVHMVSSPVGMIPDIYEK		y8	1008.51	0	0		0	0				
	NTVHMVSSPVGMIPDIYEK		y9	1065.53	0	0		0	0				
	NTVHMVSSPVGMIPDIYE**K**	(3+)708.3067	y7	883.487	0	0		0	0				
	NTVHMVSSPVGMIPDIYE**K**		y8	1014.53	0	0		0	0				
	NTVHMVSSPVGMIPDIYE**K**		y9	1071.55	30.67	13		0	0				

7	VPLADMPHAPIGLYFDTVADK	(3+)757.389	y7	795.388	0	0		0	0				
	VPLADMPHAPIGLYFDTVADK		y8	958.452	0	0		0	0				
	VPLADMPHAPIGLYFDTVADK		y10	1128.56	0	0		0	0				
	VPLADMPHAPIGLYFDTVAD**K**	(3+) 759.3957	y7	801.408	0	0		0	0				
	VPLADMPHAPIGLYFDTVAD**K**		y8	964.472	0	0		0	0				
	VPLADMPHAPIGLYFDTVAD**K**		y10	1134.58	0	0		0	0				
8	EQGNGHIISWTSR	742.87	y7	862.48	0	0		0	0				
	EQGNGHIISWTS**R**	745.88	y7	868.5	0	0		0	0				
	EQGNG(deam)HIISWTSR	495.908	y5	636.31	0	0		0	0				
	EQGNG(deam)HIISWTSR		y4	549.278	0	0		0	0				
	EQGNG(deam)HIISWTSR		y3	363.199	0	0		0	0				
	EQGNG(deam)HIISWTS**R**	497.914	y5	642.33	0	0		0	0				
	EQGNG(deam)HIISWTS**R**		y4	555.298	0	0		0	0				
	EQGNG(deam)HIISWTS**R**		y3	369.219	0	0		0	0				
9	DWVMWAVNEFSLK	812.9	y8	907.49	0	0		0	0				
	DWVMWAVNEFSL**K**	815.91	y8	913.51	0	0		0	0				
10	MISEGDIGGIAQITSSLFLGR	1083.07	y11	1192.67	0	0		0	0				
	MISEGDIGGIAQITSSLFLG**R**	1086.08	y11	1198.69	0	0		0	0				
11	ELE(C)AALGTLLR	673.36	y7	743.48	0	0		0	0				
	ELE(C)AALGTLLR		y8	814.51	0	0		0	0				
	ELE(C)AALGTLLR		y9	974.55	0	0		0	0				
	ELE(C)AALGTLL**R**	676.37	y7	749.5	45.23	255		45.55	39				
	ELE(C)AALGTLL**R**		y8	820.53	45.28	186		45.55	5				
	ELE(C)AALGTLL**R**		y9	980.57	45.24	142		45.55	7				

12	FSSEYPEF(C)SK	690.79	y6	767.34	34.42	37		34.73	71				
	FSSEYPEF(C)SK		y7	930.4	0	0		0	0				
	FSSEYPEF(C)SK		y9	1146.48	0	0		34.66	8				
	FSSEYPEF(C)S**K**	693.8	y6	773.36	34.48	3043	0.012	34.73	1886	0.038	0.024	0.018	72.37
	FSSEYPEF(C)S**K**		y7	936.42	34.49	534		34.73	458				
	FSSEYPEF(C)S**K**		y9	1152.5	34.49	257		34.72	138				
13	SIVSELAGLLSAMEYVQK	969.52	y10	1181.62	0	0		0	0				
	SIVSELAGLLSAMEYVQ**K**	972.53	y10	1187.64	0	0		0	0				
14	NMDQVAPVANSYR	732.85	y7	806.42	31.48	146		31.86	108				
	NMDQVAPVANSYR		y8	877.45	31.46	123		31.79	50				
	NMDQVAPVANSYR		y9	976.52	31.53	22		31.72	10				
	NMDQVAPVANSY**R**	735.86	y7	812.44	31.5	15160	0.01	31.82	9633	0.011	0.01	0.001	10.73
	NMDQVAPVANSY**R**		y8	883.47	31.5	11529	0.011	31.82	7864	0.006	0.009	0.0031	35.80
	NMDQVAPVANSY**R**		y9	982.54	31.5	3573	0.006	31.82	2368	0.004	0.005	0.0014	26.35
15	SLN(C)G(C)SSAS(C)(C)TVATYDK	1070.92	y10	1204.5	0	0		0	0			0	
	SLN(C)G(C)SSAS(C)(C)TVATYDK		y12	1362.57	0	0		0	0			0	
	SLN(C)G(C)SSAS(C)(C)TVATYDK		y11	1275.53	0	0		0	0			0	
	SLN(C)G(C)SSAS(C)(C)TVATYD**K**	1073.93	y10	1210.52	0	0		0	0			0	
	SLN(C)G(C)SSAS(C)(C)TVATYD**K**		y12	1368.59	0	0		0	0			0	
	SLN(C)G(C)SSAS(C)(C)TVATYD**K**		y11	1281.55	0	0		0	0			0	
16	ASFPVQILPNLYLGSAR	923.52	y9	990.53	0	0		0	0			0	
	ASFPVQILPNLYLGSAR		y10	1103.62	0	0		0	0			0	
	ASFPVQILPNLYLGSAR		y11	1216.7	0	0		0	0			0	
	ASFPVQILPNLYLGSA**R**	926.53	y9	996.55	0	0		0	0			0	
	ASFPVQILPNLYLGSA**R**		y10	1109.64	0	0		0	0			0	
	ASFPVQILPNLYLGSA**R**		y11	1222.72	0	0		0	0			0	

17	AGPTAVYFLR	547.8	y4	598.33	40.96	342	0	41.2	174	0			
	AGPTAVYFLR		y5	697.4	40.91	344	0	41.2	283	0			
	AGPTAVYFLR		y6	768.44	40.93	236	0	41.17	267	0			
	AGPTAVYFL**R**	550.81	y4	604.35	40.93	35209	0.01	41.2	23780	0.007	0.008	0.0017	19.90
	AGPTAVYFL**R**		y5	703.42	40.93	34360	0.01	41.2	23994	0.012	0.010	0.0013	11.56
	AGPTAVYFL**R**		y6	774.46	40.93	30835	0.008	41.2	18179	0.015	0.011	0.0049	44.52
18	VPSYDSFDSEDYPAALPNHKPK	1239.09	y13	1479.76	0	0		0	0				
	VPSYDSFDSEDYPAALPNHKPK		y10	1072.63	0	0		0	0				
	VPSYDSFDSEDYPAALPNHKPK		y6	720.42	0	0		0	0				
	VPSYDSFDSEDYPAALPNHKP**K**	1242.1	y13	1485.78	0	0		0	0				
	VPSYDSFDSEDYPAALPNHKP**K**		y10	1078.65	0	0		0	0				
	VPSYDSFDSEDYPAALPNHKP**K**		y6	726.44	0	0		0	0				
19	ADISSWFMEAIEYIDAVK	1044.5	y11	1281.6	0	0		0	0				
	ADISSWFMEAIEYIDAV**K**	1047.51	y11	1287.62	0	0		0	0				
20	GGYEAFSAS(C)PEL(C)SK	881.87	y8	980.42	0	0		0	0				
	GGYEAFSAS(C)PEL(C)SK		y9	1051.45	0	0		0	0				
	GGYEAFSAS(C)PEL(C)SK		y10	1138.49	0	0		0	0				
	GGYEAFSAS(C)PEL(C)S**K**	884.88	y8	986.44	36.13	41		0	0				
	GGYEAFSAS(C)PEL(C)S**K**		y9	1057.47	36.18	18		0	0				
	GGYEAFSAS(C)PEL(C)S**K**		y10	1144.51	36.18	27		36.25	0				
21	LLQTAATAAQQGGQANHPTAAVVTEK	1288.68	y9	915.51	0	0		0	0				
	LLQTAATAAQQGGQANHPTAAVVTEK		y14	1422.73	0	0		0	0				
	LLQTAATAAQQGGQANHPTAAVVTE**K**	1291.69	y9	921.53	0	0		0	0				
	LLQTAATAAQQGGQANHPTAAVVTE**K**		y14	1428.75	0	0		0	0				

22	AANLTYMPSSSGSAR	756.86	y9	879.39	31.15	28		31.37	21				
	AANLTYMPSSSGSAR		y10	1042.46	31.2	28		31.29	29				
	AANLTYMPSSSGSAR		y11	1143.5	31.17	15		31.37	0				
	AANLTYMPSSSGSA**R**	759.87	y9	885.41	31.15	5239	0.005	31.36	3274				
	AANLTYMPSSSGSA**R**		y10	1048.48	31.15	4210		31.34	2391				
	AANLTYMPSSSGSA**R**		y11	1149.52	31.15	2797		31.36	1875				
23	FCMNGAALCALGK	649.81	y7	675.38	noise			0	0				
	FCMNGAALCALGK		y9	803.44	noise			0	0				
	FCMNGAALCALGK		y10	917.48	noise			0	0				
	FCMNGAALCALG**K**	652.82	y7	681.4	noise			0	0				
	FCMNGAALCALG**K**		y9	809.46	noise			0	0				
	FCMNGAALCALG**K**		y10	923.5	noise			0	0				
24	ISSDCSDGESDR	635.75	y7	765.3	0	0		0	0				
	ISSDCSDGESD**R**	638.76	y7	771.32	0	0		0	0				
25	ADISSWFNEAIDFIDSIK	1036.01	y11	1281.64	0	0		0	0				
	ADISSWFNEAIDFIDSI**K**	1039.02	y11	1287.66	0	0		0	0				
26	LVALLESGTEK	580.33	y6	650.3	37.07	10	0	0	0	0			
	LVALLESGTEK		y9	947.5	0	0	0	37.04	21	0			
	LVALLESGTEK		y7	763.38	36.94	8	0	37.2	14	0			
	LVALLESGTE**K**	583.34	y6	656.32	36.95	858	0.012	37.1	657	0			
	LVALLESGTE**K**		y9	769.4	36.94	999	0	37.1	1089	0.019			
	LVALLESGTE**K**		y7	953.52	36.95	1638	0.005	37.12	1489	0.009			

27	DQIPELENNEK	664.82	y7	875.41	0	0	0	0	0	0			
	DQIPELENNEK		y8	972.46	31.21	44	22.7	31.36	29	37.55			
	DQIPELENNEK		y9	1085.55	0	0	0	0	0	0			
	DQIPELENNE**K**	667.83	y7	881.43	31.22	59	0	31.27	29	0			
	DQIPELENNE**K**		y8	978.48	31.16	2424	0.018	31.36	2224	0.013			
	DQIPELENNE**K**		y9	1091.57	31.21	90	0	31.34	88	0			
28	GAGMAGPGGLAR	507.76	y6	570.33	28.14	212		28.34	210				
	GAGMAGPGGLAR		y7	627.35	28.09	377		28.39	409				
	GAGMAGPGGLAR		y8	698.39	28.11	336		28.34	237				
	GAGMAGPGGLA**R**	510.77	y6	576.35	28.11	20026	0.011	28.38	19688	0.011			
	GAGMAGPGGLA**R**		y7	633.37	28.11	36884	0.01	28.38	35905	0.011	0.010	0.00083	7.66
	GAGMAGPGGLA**R**		y8	704.41	28.11	32463	0.01	28.38	32486	0.007	0.008	0.0022	24.48
29	LVDAEEVAR	501.27	y6	674.34	28.12	440	45.51	28.22	530				
	LVDAEEVAR		y7	789.37	28.14	1223	30.16	28.31	2006				
	LVDAEEVAR		y5	603.31	0	0	0	28.34	267				
	LVDAEEVA**R**	504.28	y6	680.36	28.12	23410	0.019	28.31	25122	0.021	0.019	0.0016	8.16
	LVDAEEVA**R**		y7	795.39	28.12	119382	0.01	28.31	1E+05	0.016	0.013	0.0038	29.22
	LVDAEEVA**R**		y5	609.33	28.12	16768	0	28.31	17492	0.015	0.007	0.010	141.42
30	LLQEGGGGVAAVVVLDQGSR	963.03	y11	1114.58	0	0		0	0				
	LLQEGGGGVAAVVVLDQGSR		y7	774.41	0	0		0	0				
	LLQEGGGGVAAVVVLDQGSR		y9	1043.58	0	0		0	0				
	LLQEGGGGVAAVVVLDQGS**R**	966.04	y11	1120.6	0	0		0	0				
	LLQEGGGGVAAVVVLDQGS**R**		y7	780.43	0	0		0	0				
	LLQEGGGGVAAVVVLDQGS**R**		y9	1049.6	0	0		0	0				

31	(C)HVSTHQHNYAAPPSTR	953.44	y4	460.25	0	0		0	0				
	(C)HVSTHQHNYAAPPSTR		y5	557.3	0	0		0	0				
	(C)HVSTHQHNYAAPPSTR		y10	1113.54	0	0		0	0				
	(C)HVSTHQHNYAAPPST**R**	956.45	y4	466.27	0	0		0	0				
	(C)HVSTHQHNYAAPPST**R**		y5	563.32	0	0		0	0				
	(C)HVSTHQHNYAAPPST**R**		y10	1119.56	0	0		0	0				
32	AEPEVPPQEGVPAR	738.38	y9	950.505	0	0		0	0				
	AEPEVPPQEGVPAR		y10	1049.57	0	0		0	0				
	AEPEVPPQEGVPAR		y8	853.453	0	0		0	0				
	AEPEVPPQEGVPA**R**	741.39	y9	956.526	0	0		0	0				
	AEPEVPPQEGVPA**R**		y10	1055.59	0	0		0	0				
	AEPEVPPQEGVPA**R**		y8	859.473	0	0		0	0				
33	GGLLDSM(C)PASTPSVLSSEQEFQMFPK	1443.67	y2	244.17	0	0		0	0				
	GGLLDSM(C)PASTPSVLSSEQEFQMFPK		y14	1656.79	0	0		0	0				
	GGLLDSM(C)PASTPSVLSSEQEFQMFPK		y15	1753.85	0	0		0	0				
	GGLLDSM(C)PASTPSVLSSEQEFQMFP**K**	1446.68	y2	250.19	0	0		0	0				
	GGLLDSM(C)PASTPSVLSSEQEFQMFP**K**		y14	1662.81	0	0		0	0				
	GGLLDSM(C)PASTPSVLSSEQEFQMFP**K**		y15	1759.87	0	0		0	0				
34	GGVNDNEEGFFSAR	749.83	y9	1056.47	0	0		0	0				
	GGVNDNEEGFFSAR		y10	1171.5	36.71	21		0	0				
	GGVNDNEEGFFSAR		y8	942.43	0	0		0	0				
	GGVNDNEEGFFSA**R**	752.84	y9	1062.49	36.76	1557		36.76	554				
	GGVNDNEEGFFSA**R**		y10	1177.52	36.73	933	0.023	36.78	252				
	GGVNDNEEGFFSA**R**		y8	948.45	36.78	1419		36.75	632				

Appendix 2

LM2 SRM assay, TSQ Vantage.

Peptide Sequence	**Prec. *m/z***	**Prod. *m/z***
AANLTYMPSSSGSAR	756.869	748.3578
AANLTYMPSSSGSAR	756.869	879.3983
AANLTYMPSSSGSAR	756.869	1042.462
AANLTYMPSSSGSA**R**	759.869	754.3779
AANLTYMPSSSGSA**R**	759.869	885.4183
AANLTYMPSSSGSA**R**	759.869	1048.482
ADISSWFMEAIEYIDAVK	1044.503	837.4347
ADISSWFMEAIEYIDAVK	1044.503	1428.707
ADISSWFMEAIEYIDAVK	1044.503	1281.639
ADISSWFMEAIEYIDAV**K**	1047.513	843.4547
ADISSWFMEAIEYIDAV**K**	1047.513	1434.727
ADISSWFMEAIEYIDAV**K**	1047.513	1287.659
ADISSWFNEAIDFIDSIK	1036.004	837.4347
ADISSWFNEAIDFIDSIK	1036.004	1411.71
ADISSWFNEAIDFIDSIK	1036.004	1264.641
ADISSWFNEAIDFIDSI**K**	1039.014	843.4547
ADISSWFNEAIDFIDSI**K**	1039.014	1417.73
ADISSWFNEAIDFIDSI**K**	1039.014	1270.661
AEPEVPPQEGVPAR	738.378	638.338
AEPEVPPQEGVPAR	738.378	950.5049
AEPEVPPQEGVPAR	738.378	525.2903
AEPEVPPQEGVPA**R**	741.388	641.348
AEPEVPPQEGVPA**R**	741.388	956.5248
AEPEVPPQEGVPA**R**	741.388	528.3003
AGPTAVYFLR	547.803	483.7737
AGPTAVYFLR	547.803	697.4026
AGPTAVYFLR	547.803	598.3342
AGPTAVYFL**R**	550.813	486.7837
AGPTAVYFL**R**	550.813	703.4226
AGPTAVYFL**R**	550.813	604.3542
ASFPVQILPNLYLGSAR	923.514	990.5361
ASFPVQILPNLYLGSAR	923.514	770.9457
ASFPVQILPNLYLGSAR	923.514	1103.62
ASFPVQILPNLYLGSA**R**	926.524	996.5562
ASFPVQILPNLYLGSA**R**	926.524	773.9557
ASFPVQILPNLYLGSA**R**	926.524	1109.64
DQIPELENNEK	664.82	972.4627
DQIPELENNEK	664.82	1085.547
DQIPELENNEK	664.82	486.735

DQIPELENNE**K**	667.83	978.4827
DQIPELENNE**K**	667.83	1091.567
DQIPELENNE**K**	667.83	489.745
DWVMWAVNEFSLK	812.895	907.4878
DWVMWAVNEFSLK	812.895	347.2283
DWVMWAVNEFSLK	812.895	737.3823
DWVMWAVNEFSL**K**	815.905	913.5078
DWVMWAVNEFSL**K**	815.905	353.2484
DWVMWAVNEFSL**K**	815.905	743.4023
ELECAALGTLLR	644.85	559.3557
ELECAALGTLLR	644.85	743.4769
ELECAALGTLLR	644.85	502.3342
ELECAALGTLL**R**	647.86	565.3757
ELECAALGTLL**R**	647.86	749.4968
ELECAALGTLL**R**	647.86	508.3542
ELE(C)AALGTLLR	673.36	559.3557
ELE(C)AALGTLLR	673.36	743.4769
ELE(C)AALGTLLR	673.36	502.3342
ELE(C)AALGTLL**R**	676.37	565.3757
ELE(C)AALGTLL**R**	676.37	749.4968
ELE(C)AALGTLL**R**	676.37	508.3542
EQGNGHIISWTSR	742.865	636.3094
EQGNGHIISWTSR	742.865	749.3935
EQGNGHIISWTSR	742.865	614.3148
EQGNGHIISWTS**R**	745.875	642.3295
EQGNGHIISWTS**R**	745.875	755.4135
EQGNGHIISWTS**R**	745.875	617.3248
FSSEYPEF(C)SK	690.792	670.2859
FSSEYPEF(C)SK	690.792	767.3387
FSSEYPEF(C)SK	690.792	617.258
FSSEYPEF(C)S**K**	693.802	676.306
FSSEYPEF(C)S**K**	693.802	773.3587
FSSEYPEF(C)S**K**	693.802	620.268
GAGMAGPGGLAR	507.761	627.3568
GAGMAGPGGLAR	507.761	698.3939
GAGMAGPGGLAR	507.761	570.3353
GAGMAGPGGLA**R**	510.771	633.3767
GAGMAGPGGLA**R**	510.771	704.4139
GAGMAGPGGLA**R**	510.771	576.3553
GGLLDSMCPASTPSVLSSEQEFQMFPK	962.783	1055.511
GGLLDSMCPASTPSVLSSEQEFQMFPK	962.783	927.9505
GGLLDSMCPASTPSVLSSEQEFQMFPK	962.783	971.4665

GGLLDSMCPASTPSVLSSEQEFQMFP**K** GGLLDSMCPASTPSVLSSEQEFQMFP**K** GGLLDSMCPASTPSVLSSEQEFQMFP**K**	964.79 964.79 964.79	1058.521 930.9604 974.4765
GGLLDSM(C)PASTPSVLSSEQEFQMFPK GGLLDSM(C)PASTPSVLSSEQEFQMFPK GGLLDSM(C)PASTPSVLSSEQEFQMFPK	981.79 981.79 981.79	1055.511 927.9505 971.4665
GGLLDSM(C)PASTPSVLSSEQEFQMFP**K** GGLLDSM(C)PASTPSVLSSEQEFQMFP**K** GGLLDSM(C)PASTPSVLSSEQEFQMFP**K**	983.797 983.797 983.797	1058.521 930.9604 974.4765
GGVNDNEEGFFSAR GGVNDNEEGFFSAR GGVNDNEEGFFSAR	749.831 749.831 749.831	480.256 684.3458 813.3884
GGVNDNEEGFFSA**R** GGVNDNEEGFFSA**R** GGVNDNEEGFFSA**R**	752.841 752.841 752.841	486.276 690.3658 819.4084
GGYEAFSAS(C)PEL(C)SK GGYEAFSAS(C)PEL(C)SK GGYEAFSAS(C)PEL(C)SK	881.874 881.874 881.874	636.3016 733.3544 507.259
GGYEAFSAS(C)PEL(C)S**K** GGYEAFSAS(C)PEL(C)S**K** GGYEAFSAS(C)PEL(C)S**K**	884.884 884.884 884.884	642.3216 739.3743 513.279
LLQEGGGGVAAVVVLDQGSR LLQEGGGGVAAVVVLDQGSR LLQEGGGGVAAVVVLDQGSR	963.026 963.026 963.026	873.4783 774.4099 675.3415
LLQEGGGGVAAVVVLDQGS**R** LLQEGGGGVAAVVVLDQGS**R** LLQEGGGGVAAVVVLDQGS**R**	966.036 966.036 966.036	879.4983 780.4299 681.3615
LLQTAATAAQQGGQANHPTAAVVTEK LLQTAATAAQQGGQANHPTAAVVTEK LLQTAATAAQQGGQANHPTAAVVTEK	1288.674 1288.674 1288.674	915.514 1479.755 1052.573
LLQTAATAAQQGGQANHPTAAVVTE**K** LLQTAATAAQQGGQANHPTAAVVTE**K** LLQTAATAAQQGGQANHPTAAVVTE**K**	1291.684 1291.684 1291.684	921.534 1485.775 1058.593
LVALLESGTEK LVALLESGTEK LVALLESGTEK	580.332 580.332 580.332	947.5038 763.3827 650.2986
LVALLESGTE**K** LVALLESGTE**K** LVALLESGTE**K**	583.342 583.342 583.342	953.5239 769.4026 656.3186
LVALLESGTEKDQIPELENNEK LVALLESGTEKDQIPELENNEK LVALLESGTEKDQIPELENNEK	1235.139 1235.139 1235.139	972.4627 1456.727 1328.632

LVALLESGTEKDQIPELENNE**K** LVALLESGTEKDQIPELENNE**K** LVALLESGTEKDQIPELENNE**K**	1241.159 1241.159 1241.159	978.4827 1468.767 1334.652
LVDAEEVAR LVDAEEVAR LVDAEEVAR	501.266 501.266 501.266	789.3732 603.3091 674.3463
LVDAEEVA**R** LVDAEEVA**R** LVDAEEVA**R**	504.276 504.276 504.276	795.3932 609.3291 680.3662
MISEGDIGGIAQITSSLFLGR MISEGDIGGIAQITSSLFLGR MISEGDIGGIAQITSSLFLGR	1083.067 1083.067 1083.067	1192.668 1419.795 961.0046
MISEGDIGGIAQITSSLFLG**R** MISEGDIGGIAQITSSLFLG**R** MISEGDIGGIAQITSSLFLG**R**	1086.077 1086.077 1086.077	1198.688 1425.815 964.0146
NMDQVAPVANSYR NMDQVAPVANSYR NMDQVAPVANSYR	732.848 732.848 732.848	806.415 877.4521 976.5205
NMDQVAPVANSY**R** NMDQVAPVANSY**R** NMDQVAPVANSY**R**	735.858 735.858 735.858	812.435 883.4721 982.5405
NTVHMVSSPVGMIPDIYEK NTVHMVSSPVGMIPDIYEK NTVHMVSSPVGMIPDIYEK	1059.023 1059.023 1059.023	764.382 1435.713 1261.649
NTVHMVSSPVGMIPDIYE**K** NTVHMVSSPVGMIPDIYE**K** NTVHMVSSPVGMIPDIYE**K**	1062.033 1062.033 1062.033	770.402 1441.733 1267.669
QPSVSGLSQITK QPSVSGLSQITK QPSVSGLSQITK	622.845 622.845 622.845	833.4722 576.3346 746.4401
QPSVSGLSQIT**K** QPSVSGLSQIT**K** QPSVSGLSQIT**K**	625.855 625.855 625.855	839.4922 582.3546 752.4601
SAEWLQEELEAR SAEWLQEELEAR SAEWLQEELEAR	730.854 730.854 730.854	874.426 987.51 746.3674
SAEWLQEELEA**R** SAEWLQEELEA**R** SAEWLQEELEA**R**	733.864 733.864 733.864	880.4459 993.53 752.3874
SIVSELAGLLSAMEYVQK SIVSELAGLLSAMEYVQK SIVSELAGLLSAMEYVQK	969.516 969.516 969.516	955.4548 537.3026 868.4228

SIVSELAGLLSAMEYVQ**K** SIVSELAGLLSAMEYVQ**K** SIVSELAGLLSAMEYVQ**K**	972.526 972.526 972.526	961.4748 543.3226 874.4427
SLN(C)G(C)SSAS(C)(C)TVATYDK SLN(C)G(C)SSAS(C)(C)TVATYDK SLN(C)G(C)SSAS(C)(C)TVATYDK	1070.923 1070.923 1070.923	597.2873 1027.407 526.2502
SLN(C)G(C)SSAS(C)(C)TVATYD**K** SLN(C)G(C)SSAS(C)(C)TVATYD**K** SLN(C)G(C)SSAS(C)(C)TVATYD**K**	1073.933 1073.933 1073.933	603.3073 1030.417 532.2702
VPSYDSFDSEDYPAALPNHKPK VPSYDSFDSEDYPAALPNHKPK VPSYDSFDSEDYPAALPNHKPK	826.392 826.392 826.392	793.3692 1141.024 720.4146
VPSYDSFDSEDYPAALPNHKP**K** VPSYDSFDSEDYPAALPNHKP**K** VPSYDSFDSEDYPAALPNHKP**K**	830.405 830.405 830.405	797.3825 1147.044 732.4546
VPVNDSF(C)EK VPVNDSF(C)EK VPVNDSF(C)EK	597.776 597.776 597.776	583.2539 670.2859 548.2421
VPVNDSF(C)E**K** VPVNDSF(C)E**K** VPVNDSF(C)E**K**	600.786 600.786 600.786	589.2739 676.306 551.2521
YILNVTPNLPNFFEK YILNVTPNLPNFFEK YILNVTPNLPNFFEK	904.982 904.982 904.982	1105.567 1206.615 781.3874
YILNVTPNLPNFFE**K** YILNVTPNLPNFFE**K** YILNVTPNLPNFFE**K**	907.992 907.992 907.992	1111.587 1212.635 787.4073
YVLPDEAAR YVLPDEAAR YVLPDEAAR	517.269 517.269 517.269	658.3149 771.399 386.2031
YVLPDEAA**R** YVLPDEAA**R** YVLPDEAA**R**	520.279 520.279 520.279	664.3349 777.419 389.2131
F(C)MNGAAL(C)ALGK F(C)MNGAAL(C)ALGK F(C)MNGAAL(C)ALGK	706.827 706.827 706.827	860.4653 388.2549 317.2178
F(C)MNGAAL(C)ALG**K** F(C)MNGAAL(C)ALG**K** F(C)MNGAAL(C)ALG**K**	709.837 709.837 709.837	866.4853 394.2749 323.2378

Appendix 3

LM2 QconCAT SRM assay generated using Skyline software and unscheduled SRM data generated from Waters Xevo TQ MS.

Peptide	Prec. *m/z*	Prod. *m/z*	Rt, min
LVDAEEVAR	501.2667	603.3097	26.71
LVDAEEVAR	501.2667	674.3468	26.71
LVDAEEVAR	501.2667	789.3737	26.71
LVDAEEVA**R**	504.2768	609.3298	26.71
LVDAEEVA**R**	504.2768	680.3669	26.71
LVDAEEVA**R**	504.2768	795.3939	26.71
GAGMAGPGGLAR	507.7611	570.3358	26.83
GAGMAGPGGLAR	507.7611	627.3573	26.83
GAGMAGPGGLAR	507.7611	698.3944	26.83
GAGMAGPGGLA**R**	510.7712	576.356	26.83
GAGMAGPGGLA**R**	510.7712	633.3774	26.83
GAGMAGPGGLA**R**	510.7712	704.4146	26.83
VPVNDSF(C)EK	597.7766	899.3564	27.33
VPVNDSF(C)EK.	597.7766	998.4247	27.33
VPVNDSF(C)EK.	597.7766	1095.478	27.33
VPVNDSF(C)E**K**	600.7867	905.3765	27.33
VPVNDSF(C)E**K**	600.7867	1004.445	27.33
VPVNDSF(C)E**K**	600.7867	1101.498	27.33
AEPEVPPQEGVPAR	738.3781	853.4526	27.38
AEPEVPPQEGVPAR	738.3781	950.5054	27.38
AEPEVPPQEGVPAR	738.3781	1049.574	27.38
AEPEVPPQEGVPA**R**	741.3881	859.4728	27.38
AEPEVPPQEGVPA**R**	741.3881	956.5255	27.38
AEPEVPPQEGVPA**R**	741.3881	1055.594	27.38
AANLTYMPSSSGSAR	756.8592	879.3989	30.12
AANLTYMPSSSGSAR	756.8592	1042.462	30.12
AANLTYMPSSSGSAR	756.8592	1143.51	30.12
AANLTYMPSSSGSA**R**	759.8693	885.419	30.12
AANLTYMPSSSGSA**R**	759.8693	1048.4821	30.12
AANLTYMPSSSGSA**R**	759.8693	1149.53	30.12
NMDQVAPVANSYR	732.8486	806.4155	30.35
NMDQVAPVANSYR	732.8486	877.4526	30.35
NMDQVAPVANSYR	732.8486	976.5211	30.35
NMDQVAPVANSY**R**	735.8587	812.4357	30.35
NMDQVAPVANSY**R**	735.8587	883.4728	30.35
NMDQVAPVANSY**R**	735.8587	982.5412	30.35

YVLPDEAAR YVLPDEAAR YVLPDEAAR	517.2693 517.2693 517.2693	446.2358 658.3155 771.3995	30.51 30.51 30.51
YVLPDEAA**R** YVLPDEAA**R** YVLPDEAA**R**	520.2794 520.2794 520.2794	452.2559 664.3356 777.4197	30.51 30.51 30.51
QPSVSGLSQITK QPSVSGLSQITK QPSVSGLSQITK	622.8459 622.8459 622.8459	746.4407 833.4727 1019.573	32.58 32.58 32.58
QPSVSGLSQIT**K** QPSVSGLSQIT**K** QPSVSGLSQIT**K**	625.856 625.856 625.856	752.4608 839.4929 1025.593	32.58 32.58 32.58
FSSEYPEF(C)SK FSSEYPEF(C)SK FSSEYPEF(C)SK	690.7925 690.7925 690.7925	767.3392 930.4026 1146.4771	33.57 33.57 33.57
FSSEYPEF(C)S**K** FSSEYPEF(C)S**K** FSSEYPEF(C)S**K**	693.8026 693.8026 693.8026	773.3594 936.4227 1152.4969	33.57 33.57 33.57
GGYEAFSAS(C)PEL(C)SK GGYEAFSAS(C)PEL(C)SK GGYEAFSAS(C)PEL(C)SK	881.8742 881.8742 881.8742	980.4176 1051.455 1138.4871	35.33 35.33 35.33
GGYEAFSAS(C)PEL(C)S**K** GGYEAFSAS(C)PEL(C)S**K** GGYEAFSAS(C)PEL(C)S**K**	884.8843 884.8843 884.8843	986.4377 1057.475 1144.507	35.33 35.33 35.33
GGVNDNEEGFFSAR GGVNDNEEGFFSAR GGVNDNEEGFFSAR GGVNDNEEGFFSAR	749.8315 749.8315 749.8315 749.8315	813.389 942.4316 1056.474 1171.501	35.81 35.81 35.81 35.81
GGVNDNEEGFFSA**R** GGVNDNEEGFFSA**R** GGVNDNEEGFFSA**R** GGVNDNEEGFFSA**R**	752.8416 752.8416 752.8416 752.8416	819.4091 948.4517 1062.495 1177.522	35.81 35.81 35.81 35.81
SLN(C)G(C)SSAS(C)(C)TVATYDK SLN(C)G(C)SSAS(C)(C)TVATYDK SLN(C)G(C)SSAS(C)(C)TVATYDK	1070.924 1070.924 1070.924	1204.4969 1275.5341 1362.566	36.22 36.22 36.22
SLN(C)G(C)SSAS(C)(C)TVATYD**K** SLN(C)G(C)SSAS(C)(C)TVATYD**K** SLN(C)G(C)SSAS(C)(C)TVATYD**K**	1073.934 1073.934 1073.934	1210.517 1281.5551 1368.587	36.22 36.22 36.22
LVALLESGTEK LVALLESGTEK LVALLESGTEK	580.3321 580.3321 580.3321	650.2991 763.3832 947.5044	39.19 39.19 39.19

LVALLESGTE**K** LVALLESGTE**K** LVALLESGTE**K**	583.3422 583.3422 583.3422	656.3193 769.4034 953.5245	39.19 39.19 39.19
AGPTAVYFLR AGPTAVYFLR AGPTAVYFLR	547.8033 547.8033 547.8033	598.3348 697.4032 768.4403	41.02 41.02 41.02
AGPTAVYFL**R** AGPTAVYFL**R** AGPTAVYFL**R**	550.8134 550.8134 550.8134	604.3549 703.4233 774.4604	41.02 41.02 41.02
F(C)MNGAAL(C)ALGK F(C)MNGAAL(C)ALGK F(C)MNGAAL(C)ALGK	706.8278 706.8278 706.8278	732.4073 803.4444 860.4658	41.31 41.31 41.31
F(C)MNGAAL(C)ALG**K** F(C)MNGAAL(C)ALG**K** F(C)MNGAAL(C)ALG**K**	709.8379 709.8379 709.8379	738.4274 809.4645 866.486	41.31 41.31 41.31
SAEWLQEELEA**R** SAEWLQEELEA**R** SAEWLQEELEA**R**	733.8645 733.8645 733.8645	623.3455 880.4466 993.5307	42.79 42.79 42.79
SAEWLQEELEAR SAEWLQEELEAR SAEWLQEELEAR	730.8544 730.8544 730.8544	617.3253 874.4265 987.5106	42.79 42.79 42.79
YILNVTPNLPNFFEK YILNVTPNLPNFFEK YILNVTPNLPNFFEK	904.9827 904.9827 904.9827	781.3879 1105.568 1206.615	53.83 53.83 53.83
YILNVTPNLPNFFE**K** YILNVTPNLPNFFE**K** YILNVTPNLPNFFE**K**	907.9928 907.9928 907.9928	787.408 1111.588 1212.635	53.83 53.83 53.83
ASFPVQILPNLYLGSAR ASFPVQILPNLYLGSAR ASFPVQILPNLYLGSAR	923.5147 923.5147 923.5147	990.5367 1103.621 1216.705	54.05 54.05 54.05
ASFPVQILPNLYLGSA**R** ASFPVQILPNLYLGSA**R** ASFPVQILPNLYLGSA**R**	926.5248 926.5248 926.5248	996.5568 1109.641 1222.725	54.05 54.05 54.05

Appendix 4

LM1 SRM Assay, TSQ Vantage.

Peptide Sequence	Prec. *m/z*	Prod. *m/z*
A(C)GVDYEVK	520.74	538.29
A(C)GVDYEVK	520.74	653.31
A(C)GVDYEVK	520.74	809.40
A(C)GVDYEV**K**	523.75	544.31
A(C)GVDYEV**K**	523.75	659.33
A(C)GVDYEV**K**	523.75	815.42
A(C)GVDFEIR	533.75	564.31
A(C)GVDFEIR	533.75	679.34
A(C)GVDFEIR	533.75	835.43
A(C)GVDFEI**R**	536.76	570.33
A(C)GVDFEI**R**	536.76	685.36
A(C)GVDFEI**R**	536.76	841.45
LIDISIGSLR	543.83	432.26
LIDISIGSLR	543.83	632.37
LIDISIGSLR	543.83	860.48
LIDISIGSL**R**	546.84	438.28
LIDISIGSL**R**	546.84	638.39
LIDISIGSL**R**	546.84	866.50
SIHILPSSIK	547.83	447.77
SIHILPSSIK	547.83	644.40
SIHILPSSIK	547.83	757.48
SIHILPSSI**K**	550.84	450.78
SIHILPSSI**K**	550.84	650.42
SIHILPSSI**K**	550.84	763.50
ISEEDELDTK	589.77	476.27
ISEEDELDTK	589.77	849.38
ISEEDELDTK	589.77	978.43
ISEEDELDT**K**	592.78	482.29
ISEEDELDT**K**	592.78	855.40
ISEEDELDT**K**	592.78	984.45
ELAPLFEELR	608.83	452.25
ELAPLFEELR	608.83	546.29
ELAPLFEELR	608.83	903.49
ELAPLFEEL**R**	611.84	455.26
ELAPLFEEL**R**	611.84	552.31
ELAPLFEEL**R**	611.84	909.51
SEEQQPLSLQK	643.83	475.29
SEEQQPLSLQK	643.83	685.42
SEEQQPLSLQK	643.83	813.48

SEEQQPLSLQ**K** SEEQQPLSLQ**K** SEEQQPLSLQ**K**	646.84 646.84 646.84	481.31 691.44 819.50
(C)YY(C)NGPILDK (C)YY(C)NGPILDK (C)YY(C)NGPILDK	701.81 701.81 701.81	375.22 642.38 756.43
(C)YY(C)NGPILD**K** (C)YY(C)NGPILD**K** (C)YY(C)NGPILD**K**	704.82 704.82 704.82	381.24 648.40 762.45
GVNDNEEGFFSAR GVNDNEEGFFSAR GVNDNEEGFFSAR	721.32 721.32 721.32	480.26 684.35 813.39
GVNDNEEGFFSA**R** GVNDNEEGFFSA**R** GVNDNEEGFFSA**R**	724.33 724.33 724.33	486.28 690.37 819.41
LFSVPDFVGDA(C)K LFSVPDFVGDA(C)K LFSVPDFVGDA(C)K	727.85 727.85 727.85	550.23 554.26 796.37
LFSVPDFVGDA(C)**K** LFSVPDFVGDA(C)**K** LFSVPDFVGDA(C)**K**	730.86 730.86 730.86	556.25 557.27 802.39
ALQPLEEGEDEEK ALQPLEEGEDEEK ALQPLEEGEDEEK	743.85 743.85 743.85	651.79 835.33 1174.51
ALQPLEEGEDEE**K** ALQPLEEGEDEE**K** ALQPLEEGEDEE**K**	746.86 746.86 746.86	654.80 841.35 1180.53
LYTLVLTDPDAPSR LYTLVLTDPDAPSR LYTLVLTDPDAPSR	780.92 780.92 780.92	642.32 858.40 971.48
LYTLVLTDPDAPS**R** LYTLVLTDPDAPS**R** LYTLVLTDPDAPS**R**	783.93 783.93 783.93	648.34 864.42 977.50
DLTLDALLEMNEAK DLTLDALLEMNEAK DLTLDALLEMNEAK	788.40 788.40 788.40	461.24 721.32 834.40
DLTLDALLEMNEA**K** DLTLDALLEMNEA**K** DLTLDALLEMNEA**K**	791.41 791.41 791.41	467.26 727.34 840.42
LPSVEGLHAIVVSDR LPSVEGLHAIVVSDR LPSVEGLHAIVVSDR	796.44 796.44 796.44	739.90 896.49 1066.60

LPSVEGLHAIVVSD**R** LPSVEGLHAIVVSD**R** LPSVEGLHAIVVSD**R**	799.45 799.45 799.45	742.91 902.51 1072.62
ELTQLTELYLYSNK ELTQLTELYLYSNK ELTQLTELYLYSNK	857.95 857.95 857.95	511.25 787.40 1130.57
ELTQLTELYLYSN**K** ELTQLTELYLYSN**K** ELTQLTELYLYSN**K**	860.96 860.96 860.96	517.27 793.42 1136.59
IHSSVGS(C)ENIPSQQR IHSSVGS(C)ENIPSQQR IHSSVGS(C)ENIPSQQR	899.93 899.93 899.93	615.32 842.45 1275.57
IHSSVGS(C)ENIPSQQ**R** IHSSVGS(C)ENIPSQQ**R** IHSSVGS(C)ENIPSQQ**R**	902.94 902.94 902.94	621.34 848.47 1281.59
LSHDWL(C)YLAPEIVR LSHDWL(C)YLAPEIVR LSHDWL(C)YLAPEIVR	936.48 936.48 936.48	613.37 684.40 797.49
LSHDWL(C)YLAPEIV**R** LSHDWL(C)YLAPEIV**R** LSHDWL(C)YLAPEIV**R**	939.49 939.49 939.49	619.39 690.42 803.51
ILAIGLINEALDEGDAQK ILAIGLINEALDEGDAQK ILAIGLINEALDEGDAQK	942.01 942.01 942.01	762.33 1189.53 1472.72
ILAIGLINEALDEGDAQ**K** ILAIGLINEALDEGDAQ**K** ILAIGLINEALDEGDAQ**K**	945.02 945.02 945.02	768.35 1195.55 1478.74
TYSGHGYEVLDAAGSFDNSSL(C)SGGGDK TYSGHGYEVLDAAGSFDNSSL(C)SGGGDK TYSGHGYEVLDAAGSFDNSSL(C)SGGGDK	951.08 951.08 951.08	433.20 520.24 793.35
TYSGHGYEVLDAAGSFDNSSL(C)SGGGD**K** TYSGHGYEVLDAAGSFDNSSL(C)SGGGD**K** TYSGHGYEVLDAAGSFDNSSL(C)SGGGD**K**	953.08 953.08 953.08	439.22 526.26 799.37
GNDISSGTVLSDYVGSGPPK GNDISSGTVLSDYVGSGPPK GNDISSGTVLSDYVGSGPPK	975.48 975.48 975.48	542.29 1006.48 1119.57
GNDISSGTVLSDYVGSGPP**K** GNDISSGTVLSDYVGSGPP**K** GNDISSGTVLSDYVGSGPP**K**	978.49 978.49 978.49	548.31 1012.50 1125.59
QQFIFPDVVPVPETPTR QQFIFPDVVPVPETPTR QQFIFPDVVPVPETPTR	985.52 985.52 985.52	700.36 896.48 1306.70

QQFIFPDVVPVPETPT**R** QQFIFPDVVPVPETPT**R** QQFIFPDVVPVPETPT**R**	988.53 988.53 988.53	706.38 902.50 1312.72
NVIFEISPTEEVGDFEVK NVIFEISPTEEVGDFEVK NVIFEISPTEEVGDFEVK	1026.51 1026.51 1026.51	694.34 1249.59 1336.63
NVIFEISPTEEVGDFEV**K** NVIFEISPTEEVGDFEV**K** NVIFEISPTEEVGDFEV**K**	1029.52 1029.52 1029.52	700.36 1255.61 1342.65
SEEITTGSAWFSFLESHNK SEEITTGSAWFSFLESHNK SEEITTGSAWFSFLESHNK	1085.51 1085.51 1085.51	961.47 1108.54 1294.62
SEEITTGSAWFSFLESHN**K** SEEITTGSAWFSFLESHN**K** SEEITTGSAWFSFLESHN**K**	1088.52 1088.52 1088.52	967.49 1114.56 1300.64
LTVDAYPGL(C)PPPPLESGHR LTVDAYPGL(C)PPPPLESGHR LTVDAYPGL(C)PPPPLESGHR	1088.55 1088.55 1088.55	456.23 892.46 989.52
LTVDAYPGL(C)PPPPLESGH**R** LTVDAYPGL(C)PPPPLESGH**R** LTVDAYPGL(C)PPPPLESGH**R**	1091.56 1091.56 1091.56	462.25 898.48 995.54
(C)PVAMEEADDTVAPSSTF(C)K (C)PVAMEEADDTVAPSSTF(C)K (C)PVAMEEADDTVAPSSTF(C)K	1107.97 1107.97 1107.97	828.86 1027.96 1327.58
(C)PVAMEEADDTVAPSSTF(C)**K** (C)PVAMEEADDTVAPSSTF(C)**K** (C)PVAMEEADDTVAPSSTF(C)**K**	1110.98 1110.98 1110.98	831.87 1030.97 1333.60
(C)PVAQLEQDDQVSPSSTF(C)K (C)PVAQLEQDDQVSPSSTF(C)K (C)PVAQLEQDDQVSPSSTF(C)K	1148.51 1148.51 1148.51	454.21 457.21 970.44
(C)PVAQLEQDDQVSPSSTF(C)**K** (C)PVAQLEQDDQVSPSSTF(C)**K** (C)PVAQLEQDDQVSPSSTF(C)**K**	1151.52 1151.52 1151.52	460.22 460.23 973.45

Appendix 5

Upper Limit for Copies per cell, example calculation

1. $Number\ of\ molecules\ of\ standard\ on\ coloumn = Number\ of\ moles\ loaded\ x\ Avogadro's\ constant.$

$$6.022E^{10} = 100E^{-15}\ x\ 6.022E^{23}$$

2. $Molecules\ per\ cell = \frac{Number\ of\ molecules\ of\ standard\ on\ coloumn}{Number\ of\ cells\ on\ coloumn}$

$$1E^{6} = \frac{6.022E^{10}}{5100}$$

Table A. Number of cells on column for a 300 ng load

Cell Line	Number of cells on column
HEK-293, Biological replicate 1	7147
HEK-293, Biological replicate 2	8799
HT-29, Biological replicate 1	22222
HT-29, Biological replicate 2	23063
HCT 116, Biological replicate 1	14295
HCT 116, Biological replicate 2	19580

Appendix 6

Table A. Light to heavy (L:H) ratios calculated for HEK-293 biological replicate 1 and either 10 fmol or 100 fmol LM2 QconCAT on column. For each technical replicate light to heavy ratios, the mean, standard deviation (SD) are provided. All L:H ratios have been adjusted by subtracting the light signal derived from QconCAT peptide. Mean is used to calculate copies per cell (CPC) values. B- type quantifications (B) and noise (N).

			HEK-293 (1) + 10 fmol LM2 QconCAT						HEK-293 (1) + 100 fmol LM2 QconCAT					
Pep	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**
Q1	DUSP16	VPVNDSFCEK	0.03	0.03	0.03	0.03	0.00	<6.8 x 10^5	0.04	0.03	0.06	0.04	0.02	<8.3 x 10^6
Q2	DUSP7	SAEWLQEELEAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	N	N	N	N		N
Q3	DUSP18	QPSVSGLSQITK	N	N	N	N		N	N	NSE	NSE	NSE		N
Q4	DUSP5	YVLPDEAAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.02	0.02	0.02	0.02	0.00	<8.3 x 10^6
Q5	DUSP9	YILNVTPNLPNFFEK	0.03	0.03	0.03	0.03	0.00	<6.6 x 10^5	0.02	0.02	0.02	0.02	0.00	<8.2 x10^6
Q12	DUSP 4	FSSEYPEFCSK	0.03	0.03	0.03	0.03	0.00	<6.8 x 10^5	0.03	0.04	0.03	N		N
Q14	ETS2	NMDQVAPVANSYR	0.01	0.01	0.01	0.01	0.00	<6.7 x 10^5	0.02	0.02	0.01	0.02	0.01	<8.3 x 10^6
Q15	DUSP 10	SLNCGCSSASCCTVATYDK	N	N	N	N		N	N	N	N	N		N
Q16	DUSP9	ASFPVQILPNLYLGSAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.02	0.03	0.01	0.02	0.01	<8.3 x 10^6
Q17	DUSP2	AGPTAVYFLR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.01	0.01	0.01	0.01	0.00	<8.3 x 10^6
Q20	DUSP1	GGYEAFSASCPELCSK	0.03	0.03	0.03	0.03	0.00	<6.6 x 10^5	0.03	0.04	0.03	0.03	0.01	<8.2 x10^6
Q22	DUS10	AANLTYMPSSSGSAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.02	0.02	0.02	0.02	0.00	<8.3 x 10^6
Q23	ETS1 1/2	F(C)MNGAAL(C)ALGK	0.03	0.03	0.02	0.03	0.00	<6.8 x 10^5	N	Noise	Noise	Noise		N
Q26	DUSP16	LVALLESGTEK	N	N	N	N		N	N	N	N	N		N
Q28	ELK1 1	GAGMAGPGGLAR	0.02	0.02	0.02	0.02	0.00	<8.3 x 10^5	0.02	0.01	0.03	0.02	0.01	<8.3 x 10^6
Q29	ELK1	LVDAEEVAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.05	0.02	0.02	0.03	0.02	<8.3 x 10^6
Q32	ELK1 1	AEPEVPPQEGVPAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.02	0.02	0.02	0.02	0.00	<8.3 x 10^6

Table B. Light to heavy (L:H) ratios calculated for HEK-293 biological replicate 2 and either 10 fmol or 100 fmol LM2 QconCAT on column. For each technical replicate light to heavy ratios, the mean, standard deviation (SD) are provided. All L:H ratios have been adjusted by subtracting the light signal derived from QconCAT peptide. Mean is used to calculate copies per cell (CPC) values. B- type quantifications (B) and noise (N).

			HEK-293 (2) + 10 fmol LM2 QconCAT						HEK-293 (2) + 100 fmol LM2 QconCAT					
Pep	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**
Q1	DUSP16	VPVNDSFCEK	0.03	0.03	0.03	0.03	0.00	<6.8 x 10^5	0.02	0.04	0.04	0.03	0.01	<6.7 x 10^5
Q2	DUSP7	SAEWLQEELEAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	N	N	N	N	-	-
Q3	DUSP18	QPSVSGLSQITK	N	N	N	N	-	-	N	N	N	N	-	-
Q4	DUSP5	YVLPDEAAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.03	0.02	0.02	0.02	0.01	<6.8 x 10^6
Q5	DUSP9	YILNVTPNLPNFFEK	0.03	0.03	0.03	0.03	0.00	<6.6 x 10^5	0.03	0.02	0.02	0.02	0.01	<6.6 x 10^6
Q12	DUSP 4	FSSEYPEFCSK	0.03	0.03	0.03	0.03	0.00	<6.8 x 10^5	0.03	0.03	0.03	0.03	0.00	<6.6 x 10^6
Q14	ETS2	NMDQVAPVANSYR	0.01	0.01	0.01	0.01	0.00	<6.7 x 10^5	0.02	0.02	0.02	0.02	0.00	<6.7 x 10^5
Q15	DUSP 10	SLNCGCSSASCCTVATYDK	N	N	N	N	-	-	N	N	N	N	-	_
Q16	DUSP9	ASFPVQILPNLYLGSAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.02	0.02	0.02	0.02	0.00	<6.8 x 10^6
Q17	DUSP2	AGPTAVYFLR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^6
Q20	DUSP1	GGYEAFSASCPELCSK	0.03	0.03	0.03	0.03	0.00	<6.6 x 10^5	0.03	0.05	0.03	0.04	0.01	<6.8 x 10^6
Q22	DUS10	AANLTYMPSSSGSAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^6
Q23	ETS1 1/2	F(C)MNGAAL(C)ALGK	0.03	0.03	0.03	0.03	0.00	<6.8 x 10^5	0.03	0.03	0.03	0.03	0.00	<6.7 x 10^5
Q26	DUSP16	LVALLESGTEK	N	N	N	N	-	-	N	N	N	N	-	-
Q28	ELK1 1	GAGMAGPGGLAR	0.01	0.01	0.01	0.01	0.00	<8.3 x 10^5	0.05	0.02	0.02	0.03	0.02	<6.7 x 10^5
Q29	ELK1	LVDAEEVAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.05	0.05	0.05	0.05	0.00	<6.7 x 10^5
Q32	ELK1 1	AEPEVPPQEGVPAR	0.01	0.01	0.01	0.01	0.00	<6.8 x 10^5	0.02	0.02	0.02	0.02	0.00	<6.8 x 10^6

Table C. Light to heavy (L:H) ratios calculated for HT-29 biological replicate 1 and either 10 fmol or 100 fmol LM2 QconCAT on column. For each technical replicate light to heavy ratios, the mean, standard deviation (SD) are provided. All L:H ratios have been adjusted by subtracting the light signal derived from QconCAT peptide. Mean is used to calculate copies per cell (CPC) values. B- type quantifications (B) and noise (N).

			HT-29 (1)+ 10 fmol LM2 QconCAT						HT-29 (1) + 100 fmol LM2 QconCAT					
Pep	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**
Q1	DUSP16	VPVNDSFCEK	0.03	0.03	0.03	0.03	0.00	<2.7 x 10^5	0.04	0.03	0.04	0.04	0.01	<2.7 x 10^6
Q2	DUSP7	SAEWLQEELEAR	0.01	0.01	0.01	0.01	0.00	<2.7 x 10^5	B	B	B	B		<2.7 x 10^6
Q3	DUSP18	QPSVSGLSQITK	B	B	B	B		< 2.6 x 10^5	B	B	B	B		< 2.6 x 10^6
Q4	DUSP5	YVLPDEAAR	0.01	0.01	0.01	0.01	0.00	<2.7 x 10^5	B	B	B	B		<2.7 x 10^6
Q5	DUSP9	YILNVTPNLPNFFEK	0.03	0.04	0.03	0.03	0.01	< 2.6 x 10^5	B	B	B	B		< 2.6 x 10^6
Q12	DUSP 4	FSSEYPEFCSK	0.03	0.03	0.03	0.03	0.00	<2.7 x 10^5	0.03	0.03	0.03	0.03	0.00	<2.7 x 10^6
Q14	ETS2	NMDQVAPVANSYR	0.02	0.01	0.01	0.01	0.01	<2.7 x 10^5	B	B	B	B		<2.7 x 10^6
Q15	DUSP 10	SLNCGCSSASCCTVATYDK	N	B	N	N		N	N	N	N	N		N
Q16	DUSP9	ASFPVQILPNLYLGSAR	0.01	0.01	0.01	0.01	0.00	<2.7 x 10^5	B	B	B	B		<2.7 x 10^6
Q17	DUSP2	AGPTAVYFLR	B	NSE	B	B		<2.7 x 10^5	B	B	B	B		<2.7 x 10^6
Q20	DUSP1	GGYEAFSASCPELCSK	0.03	0.03	0.03	0.03	0.00	< 2.6 x 10^5	B	0.04	B	B		< 2.6 x 10^6
Q22	DUS10	AANLTYMPSSSGSAR	0.01	0.01	0.01	0.01	0.00	<2.7 x 10^5	B	0.01	B	B		<2.7 x 10^6
Q23	ETS1 1/2	F(C)MNGAAL(C)ALGK	B	B	B	B		<2.7 x 10^5	B	B	B	B		<2.7 x 10^6
Q26	DUSP16	LVALLESGTEK	N	N	N	N		N	N	N	N	N		N
Q28	ELK1 1	GAGMAGPGGLAR	0.01	0.01	0.01	0.01	0.00	<2.7 x 10^5	B	B	B	B		<2.7 x 10^6
Q29	ELK1	LVDAEEVAR	B	B	B	B		<2.7 x 10^5	B	B	B	B		<2.7 x 10^6
Q32	ELK1 1	AEPEVPPQEGVPAR	B	B	B	B		<2.7 x 10^5	B	B	B	B		<2.7 x 10^6

Table D. Light to heavy (L:H) ratios calculated for HT-29 biological replicate 2 and either 10 fmol or 100 fmol LM2 QconCAT on column. For each technical replicate light to heavy ratios, the mean, standard deviation (SD) are provided. All L:H ratios have been adjusted by subtracting the light signal derived from QconCAT peptide. Mean is used to calculate copies per cell (CPC) values. B- type quantifications (B) and noise (N).

			HT-29 (2) + 10 fmol LM2 QconCAT						HT (29) + 100 fmol LM2 QconCAT					
Pep	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**
Q1	DUSP16	VPVNDSFCEK	0.03	0.04	0.03	0.03	0.01	<2.6 x 10^6	B	B	B	B		<2.6 x 10^6
Q2	DUSP7	SAEWLQEELEAR	0.01	0.01	0.01	0.01	0.00	<2.6 x 10^5	B	B	B	B		<2.6 x 10^6
Q3	DUSP18	QPSVSGLSQITK	B	B	B	B		<2.5 x 10^5	B	B	B	B		<2.5 x 10^6
Q4	DUSP5	YVLPDEAAR	0.01	0.01	0.01	0.01	0.00	<2.6 x 10^5	B	B	B	B		<2.6 x 10^6
Q5	DUSP9	YILNVTPNLPNFFEK	0.03	0.03	0.03	0.03	0.00	<2.5 x 10^5	B	B	B	B		N
Q12	DUSP 4	FSSEYPEFCSK	0.03	0.03	0.03	0.03	0.00	<2.6 x 10^6	0.04	0.04	0.04	0.04	0.00	<2.6 x 10^6
Q14	ETS2	NMDQVAPVANSYR	0.01	0.01	0.02	0.01	0.01	<2.6 x 10^5	0.02	0.02	0.02	0.02	0.00	<2.6 x 10^6
Q15	DUSP 10	SLNCGCSSASCCTVATYDK	N	N	N	N		N	N	N	N	N		N
Q16	DUSP9	ASFPVQILPNLYLGSAR	0.01	0.01	0.01	0.01	0.00	<2.6 x 10^5	B	B	0.02	0.02		<2.6 x 10^6
Q17	DUSP2	AGPTAVYFLR	B	B	B	B		<2.6 x 10^5	0.02	0.01	B	0.02	0.01	<2.6 x 10^6
Q20	DUSP1	GGYEAFSASCPELCSK	0.03	0.03	0.03	0.03	0.00	<2.5 x 10^5	0.02	B	B	B		<2.6 x 10^6
Q22	DUS10	AANLTYMPSSSGSAR	NSE	0.01	0.01	0.01	0.00	<2.6 x 10^5	B	B	0.01	B		<2.6 x 10^6
Q23	ETS1 1/2	F(C)MNGAAL(C)ALGK	B	B	B	B		<2.6 x 10^5	B	B	B	B		<2.6 x 10^6
Q26	DUSP16	LVALLESGTEK	N	N	N	N		N	N	N	N	N		N
Q28	ELK1 1	GAGMAGPGGLAR	0.01	0.01	0.01	0.01	0.00	<2.6 x 10^5	B	B	B	B		<2.6 x 10^6
Q29	ELK1	LVDAEEVAR	B	B	0.01	B		<2.6 x 10^5	B	B	B	B		<2.6 x 10^6
Q32	ELK1 1	AEPEVPPQEGVPAR	NSE	0.01	0.01	0.01	0.00	<2.6 x 10^5	0.02	0.02	B	0.02	0.00	<2.6 x 10^6

Table E. Light to heavy (L:H) ratios calculated for HCT 116 biological replicate 1 and either 10 fmol or 100 fmol LM2 QconCAT on column. For each technical replicate light to heavy ratios, the mean, standard deviation (SD) are provided. All L:H ratios have been adjusted by subtracting the light signal derived from QconCAT peptide. Mean is used to calculate copies per cell (CPC) values. B-type quantifications (B) and noise (N).

			HCT 116 (1) + 10 fmol LM2 QconCAT						HCT 116 (1) + 100 fmol LM2 QconCAT					
Pep	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**
Q1	DUSP16	VPVNDSFCEK	0.03	0.03	0.03	0.03	0.00	<4.1 x 10^5	0.03	0.03	0.03	0.03	0.00	<4.1 x 10^6
Q2	DUSP7	SAEWLQEELEAR	B	B	B	B		<4.2 x 10^5	B	B	B	B		<4.2 x 10^6
Q3	DUSP18	QPSVSGLSQITK	B	B	B	B		<4.1 x 10^5	B	B	B	B		<4.1 x 10^5
Q4	DUSP5	YVLPDEAAR	0.01	0.01	0.01	0.01	0.00	<4.2 x 10^5	B	0.03	0.02	0.03	0.01	<4.2 x 10^6
Q5	DUSP9	YILNVTPNLPNFFEK	0.03	0.03	0.04	0.03	0.01	<4.1 x 10^5	N	N	N	N		N
Q12	DUSP 4	FSSEYPEFCSK	0.03	0.03	0.03	0.03	0.00	<4.1 x 10^5	B	B	B	B		<4.1 x 10^6
Q14	ETS2	NMDQVAPVANSYR	0.01	0.01	0.01	0.01	0.00	<4.1 x 10^5	0.02	0.02	0.02	0.02	0.00	<4.1 x 10^6
Q15	DUSP 10	SLNCGCSSASCCTVATYDK	N	N	N	N		N	N	N	N	N		N
Q16	DUSP9	ASFPVQILPNLYLGSAR	0.01	0.01	0.01	0.01	0.00	<4.2 x 10^5	B	B	B	B		<4.2 x 10^6
Q17	DUSP2	AGPTAVYFLR	0.01	B	B	B		<4.2 x 10^5	B	0.01	B	0.01	0.00	<4.2 x 10^6
Q20	DUSP1	GGYEAFSASCPELCSK	B	0.03	B	B		<4.1 x 10^5	B	0.04	B	0.04	0.00	<4.1 x 10^5
Q22	DUS10	AANLTYMPSSSGSAR	0.01	0.01	0.01	0.01	0.00	<4.2 x 10^5	B	0.02	0.01	0.02	0.01	<4.2 x 10^6
Q23	ETS1 1/2	F(C)MNGAAL(C)ALGK	B	B	B	B		<4.1 x 10^5	B	B	B	B		<4.1 x 10^6
Q26	DUSP16	LVALLESGTEK	N	N	N	N		N	N	N	N	N		N
Q28	ELK1 1	GAGMAGPGGLAR	0.01	0.01	B	0.01	0.00	<4.1 x 10^5	B	B	B	B		<4.1 x 10^6
Q29	ELK1	LVDAEEVAR	B	0.01	0.01	0.01	0.00	<4.2 x 10^5	B	B	B	B		<4.2 x 10^6
Q32	ELK1 1	AEPEVPPQEGVPAR	B	B	B	B		<4.2 x 10^5	N	B	B	B		<4.2 x 10^6

Table F. Light to heavy (L:H) ratios calculated for HCT 116 biological replicate 2 and either 10 fmol or 100 fmol LM2 QconCAT on column. For each technical replicate light to heavy ratios, the mean, standard deviation (SD) are provided. All L:H ratios have been adjusted by subtracting the light signal derived from QconCAT peptide. Mean is used to calculate copies per cell (CPC) values. B-type quantifications (B) and noise (N).

			HCT 116 (2) + 10 fmol LM2 QconCAT						HCT 116 (2) + 100 fmol LM2 QconCAT					
Pep	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CPC**
Q1	DUSP16	VPVNDSFCEK	0.03	0.03	0.03	0.03	0.00	<3.0 x 10^5	B	B	0.03	0.03	0.03	<3.0 x 10^6
Q2	DUSP7	SAEWLQEELEAR	B	B	B	B		<3.0 x 10^5	N	N	N	N		N
Q3	DUSP18	QPSVSGLSQITK	B	B	B	B		<2.9 x 10^5	B	B	B	B		<2.9 x 10^6
Q4	DUSP5	YVLPDEAAR	0.01	0.01	0.01	0.01	0.00	<3.0 x 10^5	B	B	B	B		<3.0 x 10^6
Q5	DUSP9	YILNVTPNLPNFFEK	0.03	0.03	0.03	0.03	0.00	<2.9 x 10^5	N	N	N	N		N
Q12	DUSP 4	FSSEYPEFCSK	0.03	0.03	0.03	0.03	0.00	<3.0 x 10^5	0.05	0.05	0.05	0.05	0.00	<3.0 x 10^6
Q14	ETS2	NMDQVAPVANSYR	0.01	0.01	0.01	0.01	0.00	<3.0 x 10^5	0.02	0.02	0.03	0.02	0.01	<3.0 x 10^6
Q15	DUSP 10	SLNCGCSSASCCTVATYDK	N	N	N	N		N	N	N	N	N		N
Q16	DUSP9	ASFPVQILPNLYLGSAR	0.01	0.01	0.01	0.01	0.00	<3.0 x 10^5	N	N	N	N		N
Q17	DUSP2	AGPTAVYFLR	0.01	0.01	B	0.01	0.00	<3.0 x 10^5	B	B	B	B		<3.0 x 10^6
Q20	DUSP1	GGYEAFSASCPELCSK	0.03	0.03	0.03	0.03	0.00	<2.9 x 10^5	B	B	B	B		<2.9 x 10^6
Q22	DUS10	AANLTYMPSSSGSAR	0.01	0.01	0.01	0.01	0.00	<3.0 x 10^5	B	B	B	B		<3.0 x 10^6
Q23	ETS1 1/2	F(C)MNGAAL(C)ALGK	B	B	B	B		<3.0 x 10^5	N	N	N	B		<3.0 x 10^6
Q26	DUSP16	LVALLESGTEK	N	N	N	N		N	N	N	N	N		N
Q28	ELK1 1	GAGMAGPGGLAR	0.01	0.01	0.01	0.01	0.00	<3.0 x 10^5	B	B	B	B		<3.0 x 10^6
Q29	ELK1	LVDAEEVAR	B	0.01	0.01	0.01	0.00	<3.0 x 10^5	B	B	B	B		<3.0 x 10^6
Q32	ELK1 1	AEPEVPPQEGVPAR	B	B	B	B		<3.0 x 10^5	N	N	B	B		<3.0 x 10^6

Appendix 7

Table A. HEK-293 biological replicate 1 and either 100 fmol or 10 fmol ofLM1 QconCAT. For each technical replicate light to heavy (L:H) ratios, the mean, standard deviation (SD), and coefficient of variation (CV) are provided. All L:H ratios have light signal derived from QconCAT peptide subtracted. Mean is used to calculate copies per cell (CPC). B-type quantificaitons (B) and noise (N).

			HEK293 (1) + 10 fmol LM1 QconCAT					HEK293 (1) +100 fmol LM1 QconCAT						
Pep.	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**CPC**	**L:H_1**	**L:H_2**	**L:H_2**	**Mean**	**SD**	**CV, %**	**CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	N	N	N	N		0.28	0.28	0.26	0.27	0.01	4.22	2.30x10^6
Q2	ARRB2 1/2/3	ACGVDFEIR	N	N	N	N		N	N	N	N			
Q3	KSR2 1	SEEQQPLSLQK	N	N	N	N		0.25	0.32	0.36	0.31	0.06	17.96	2.61x10^6
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	N	N	N	N		N	N	N	N			
Q6	MPKS1	LPSVEGLHAIVVSDR	N	N	N	N		N	N	N	N			
Q7	PEBP1	LYTLVLTDPDAPSR	N	N	N	N		B	B	B	B			<8.3 x 10^6
Q8	KSR1	DLTLDALLEMNEAK	N	N	N	N		N	N	N	N			
Q9	PAXI 1/2/3	CYYCNGPILDK	N	N	N	N		N	N	N	N			
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	B	N	B	B	<8.3x10^5	B	B	B	B			<8.3x10^6
Q11	KSR1	LIDISIGSLR	N	N	N	N		B	B	B	B			<8.3x10^6
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	N	N	N	N		N	N	N	N			
Q13	PEA_15	SEEITTGSAWFSFLESHNK	N	N	N	N		N	N	N	N			
Q14	KSR2 2	IHSSVGSCENIPSQQR	N	N	N	N		N	N	N	N			
Q15	Sur8/Shoc2	SIHILPSSIK	B	B	B	B	<8.3x10^5	B	B	B	B			<8.3x10^6
Q16	PAXI 1/2/3	LGVATVAK												
Q17	MVP	LFSVPDFVGDACK	B	N	B	B	<8.3 x10^5	N	N	N	N			
Q18	PEBP1	ELAPLFEELR	B	N	B	B	<8.3 x10^5	B	B	B	B			<8.3x10^6
Q19	ARRB1 1/2	ACGVDYEVK	N	N	N	N		N	N	N	N			
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	N	N	N	N		N	N	N	N			
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	N	N	N	N		0.33	0.36	0.30	0.33	0.03	9.09	2.78x10^6
Q22	IQGAP1	ILAIGLINEALDEGDAQK	N	N	N	N		N	N	N	N			
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	N	N	N	N		0.50	0.45	0.43	0.46	0.04	7.84	3.88x10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	N	N	N	N		N	N	N	N			
Q28	PEA_15	ISEEDELDTK	N	N	N	N		N	N	N	N			
Q29	MVP	ALQPLEEGEDEEK	N	N	N	N		N	N	0.43	N			

Table B. HEK-293 biological replicate 2 and either 100 fmol or 10 fmol of LM1 QconCAT. For each technical replicate light to heavy (L:H) ratios, the mean, standard deviation (SD), and coefficient of variation (CV) are provided. All L:H ratios have light signal derived from QconCAT peptide subtracted. Mean is used to calculate copies per cell (CPC). B-type quantifications (B) and noise (N).

			HEK-293 (2) + 10 fmol LM1 QconCAT					HEK-293 (2) + 100 fmol LM1 QconCAT						
Pep	**Protein**	**Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CV, %**	**CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	N	N	N	N		0.34	N	0.39	0.37	0.04	9.69	2.5 x 10^6
Q2	ARRB2 1/2/3	ACGVDFEIR	B	B	B	B	<6.7 x 10^5	B	B	B	B			<6.7 x 10^6
Q3	KSR2 1	SEEQQPLSLQK	N	N	N	N		0.33	0.34	0.36	0.34	0.02	4.45	2.3 x 10^6
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	N	N	N	N		N	N	N	N			
Q6	MPKS1	LPSVEGLHAIVVSDR	N	N	N	N		N	N	B	B			<6.7 x 10^6
Q7	PEBP1	LYTLVLTDPDAPSR	0.63	B	B	B	<6.7 x 10^5	B	B	B	B			<6.7 x 10^6
Q8	KSR1	DLTLDALLEMNEAK	N	N	N	N		N	N	N	N			
Q9	PAXI 1/2/3	CYYCNGPILDK	N	N	N	N		B	B	N	B			<6.7 x 10^6
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	N	N	N	N		B	B	B	B			<6.7 x 10^6
Q11	KSR1	LIDISIGSLR	B	B	B	B	<6.7 x 10^5	B	B	B	B			<6.7 x 10^6
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	N	N	N	N		N	N	N	N			
Q13	PEA_15	SEEITTGSAWFSFLESHNK	N	N	N	N		N	N	N	N			
Q14	KSR2 2	IHSSVGSCENIPSQQR	N	N	N	N		N	N	N	N			
Q15	Sur8/Shoc2	SIHILPSSIK	N	N	N	N		0.29	0.35	0.33	0.32	0.03	9.45	2.2 x10^6
Q17	MVP	LFSVPDFVGDACK	B	B	B	B	<6.7 x 10^5	0.40	0.36	0.33	0.36	0.04	9.67	2.5 x 10^6
Q18	MPKS1	ELAPLFEELR	N	N	N	N		0.01	B	B	B	0.00		<6.7 x 10^6
Q19	ARRB1 1/2	ACGVDYEVK	N	N	N	N		B	N	N	N			
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	N	N	N	N		N	N	N	N			
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	N	N	N	N		N	N	N	N			
Q22	IQGAP1	ILAIGLINEALDEGDAQK	N	N	N	N		N	N	N	N			
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	N	N	N	N		0.64	0.72	0.85	0.74	0.11	14.39	5.0 x 10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	N	N	N	N		N	N	N	N			
Q28	PEA_15	ISEEDELDTK	N	N	N	N		N	N	N	N			
Q29	MVP	ALQPLEEGEDEEK	N	N	N	N		N	B	B	B			<6.7 x 10^6

Table C. HT-29 biological replicate 1 and either 100 fmol or 10 fmol of LM1 QconCAT. For each technical replicate light to heavy (L:H) ratios, the mean, standard deviation (SD), and coefficient of variation (CV) are provided. All L:H ratios have light signal derived from QconCAT peptide subtracted. Mean is used to calculate copies per cell (CPC). B-type quantifications (B) and noise (N).

			HT-29 (1) + 10 fmol LM1 QconCAT							HT-29 (1) + 100 fmol LM1 QconCAT						
Pep	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CV, %**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CV, %**	**CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	N	N	N	N				0.35	0.40	0.42	0.39	0.04	9.25	1.06x10^6
Q2	ARRB2 1/2/3	ACGVDFEIR	B	B	B	B			<2.68x10^5	B	B	B	B			<2.68x10^6
Q3	KSR2 1	SEEQQPLSLQK	0.37	0.38	N	0.38	0.01	1.89	1.02x10^5	0.37	0.38	0.30	0.35	0.01	2.02	9.48x10^5
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	N	N	N	N				N	N	N	N			
Q6	MPKS1	LPSVEGLHAIVVSDR	N	N	N	N				B	B	B	B			<2.68x10^6
Q7	PEBP1	LYTLVLTDPDAPSR	1.62	1.83	1.99	1.81	0.15	8.19	4.91 x10^6	0.10	0.12	0.10	0.11	0.01	13.26	2.89E+05
Q8	KSR1	DLTLDALLEMNEAK	N	N	N	N				N	N	N	N			
Q9	PAXI 1/2/3	CYYCNGPILDK	N	N	N	N				N	N	N	N			
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	B	B	B	B			<2.68x10^5	B	B	B	B			<2.68x10^6
Q11	KSR1	LIDISIGSLR	N	N	N	N				B	B	B	B			<2.68x10^6
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	N	N	N	N				N	N	N	N			
Q13	PEA_15	SEEITTGSAWFSFLESHNK	N	N	N	N				B	B	B	B			<2.68x10^6
Q14	KSR2 2	IHSSVGSCENIPSQQR	N	N	N	N				N	N	N	N			
Q15	Sur8/Shoc2	SIHILPSSIK	N	N	N	N				N	0.35	0.39	0.37	0.03	7.64	1.00x10^6
Q17	MVP	LFSVPDFVGDACK	N	N	N	N				0.35	0.31	0.44	0.37	0.03	7.71	9.94x10^5
Q18	PEBP1	ELAPLFEELR	B	B	B	B			<2.68x10^5	B	B	B	B			<2.68x10^6
Q19	ARRB1 1/2	ACGVDYEVK	N	N	N	N				B	N	B	B			<2.68x10^6
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	N	N	N	N				N	B	B	B			<2.68x10^6
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	N	N	N	N				0.34	0.34	0.34	0.34	0.00	0.00	9.21x10^5
Q22	IQGAP1	ILAIGLINEALDEGDAQK	N	N	N	N				B	B	B	B			<2.68x10^6
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	N	N	N	N				N	1.03	1.07	1.05	0.03	2.69	2.85x10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	N	N	N	N				N	N	N	N			
Q28	PEA_15	ISEEDELDTK	N	N	N	N				N	N	N	N			
Q29	MVP	ALQPLEEGEDEEK	N	N	N	N				N	N	N	N			

Table D. HT-29 biological replicate 2 and either 100 fmol or 10 fmol of LM1 QconCAT. For each technical replicate light to heavy (L:H) ratios, the mean, standard deviation (SD), and coefficient of variation (CV) are provided. All L:H ratios have light signal derived from QconCAT peptide subtracted. Mean is used to calculate copies per cell (CPC). B-type quantifications (B) and noise (N).

			HT-29 (2) + 10 fmol LM1 QconCAT							HT-29 (2) + 100 fmol LM1 QconCAT						
Pep	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CV,%**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CV, %**	**CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	N	N	N	N				0.37	0.42	0.41	0.40	0.03	0.66	1.04x10^6
Q2	ARRB2 1/2/3	ACGVDFEIR	B	N	N	N				B	B	B	B			<2.58x10^6
Q3	KSR2 1	SEEQQPLSLQK	N	N	N	N				0.33	0.32	0.33	0.33	0.01	0.14	8.53x10^5
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	N	N	N	N				N	N	N	N			
Q6	MPKS1	LPSVEGLHAIVVSDR	N	N	N	N				B	B	B	B			<2.58x10^6
Q7	PEBP1	LYTLVLTDPDAPSR	1.72	1.61	1.60	1.64	0.07	4.05	4.29x10^5	0.14	0.15	0.15	0.15	0.01	0.14	3.83x10^5
Q8	KSR1	DLTLDALLEMNEAK	N	N	N	N				N	B	B	B			<2.58x10^6
Q9	PAXI 1/2/3	CYYCNGPILDK	N	N	N	N				B	B	B	B			<2.58x10^6
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	B	N	B	B			<2.58x10^5	B	B	B	B			<2.58x10^6
Q11	KSR1	LIDISIGSLR	B	B	B	B			<2.58x10^5	B	B	B	B			<2.58x10^6
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	N	N	N	N				B	B	N	B			<2.58x10^6
Q13	PEA_15	SEEITTGSAWFSFLESHNK	N	N	N	N				N	N	N	N			
Q14	KSR2 2	IHSSVGSCENIPSQQR	N	N	N	N				N	N	N	N			
Q15	Sur8/Shoc2	SIHILPSSIK	N	N	N	N				0.37	0.32	0.33	0.34	0.03	0.66	8.88x10^5
Q17	MVP	LFSVPDFVGDACK	N	N	N	N				0.36	0.30	0.31	0.31	0.03	0.80	8.09x10^5
Q18	PEBP1	ELAPLFEELR	B	B	N	B			<2.58x10^5	B	B	B	B			<2.58x10^6
Q19	ARRB1 1/2	ACGVDYEVK	N	N	N	N				N	N	N	N			
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	N	N	N	N				N	N	N	N			
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	N	N	N	N				0.32	0.36	0.40	0.36	0.04	1.00	9.40x10^5
Q22	IQGAP1	ILAIGLINEALDEGDAQK	N	N	N	N				0.62	0.67	0.61	0.63	0.03	0.80	1.65x10^6
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	N	N	N	N				1.13	1.13	1.21	1.16	0.05	1.15	3.02x10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	N	N	N	N				N	N	N	N			
Q28	PEA_15	ISEEDELDTK	N	N	N	N				N	N	N	N			
Q29	MVP	ALQPLEEGEDEEK	N	N	N	N				N	N	N	N			

Table E. HCT 116 biological replicate 1 and either 100 fmol or 10 fmol of LM1 QconCAT. For each technical replicate light to heavy (L:H) ratios, the mean, standard deviation (SD), and coefficient of variation (CV) are provided. All L:H ratios have light signal derived from QconCAT peptide subtracted. Mean is used to calculate copies per cell (CPC). B-type quantifications (B) and noise (N).

			HCT 116 (1) + LM1 10 fmol							HCT 116 (1) + LM1 100 fmol						
Peptide	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CV, %**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CV, %**	**CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	N	N	N	N	N			N	B	B	B			<4.1 x 10^6
Q2	ARRB2 1/2/3	ACGVDFEIR	B	B	B	B	B		<4.1 x 10^5	B	B	B	B			<4.1 x 10^6
Q3	KSR2 1	SEEQQPLSLQK	B	B	B	B			<4.1 x 10^5	0.32	0.32	0.31	0.32	0.01	1.82	1.33x10^6
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	N	N	N	N				N	N	N	N			
Q6	MPKS1	LPSVEGLHAIVVSDR	N	N	N	N				N	N	N	N			
Q7	PEBP1	LYTLVLTDPDAPSR	1.77	1.66	1.59	1.67	0.09	5.42	7.04x10^5	0.11	0.12	0.12	0.12	0.01	4.95	4.91x10^5
Q8	KSR1	DLTLDALLEMNEAK	N	N	N	N				N	N	B	N			
Q9	PAXI 1/2/3	CYYCNGPILDK	N	N	N	N				0.39	0.28	0.35	0.34	0.06	16.38	1.43x10^6
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	N	B	B	B			<4.1 x 10^5	B	B	N	B			<4.1 x 10^6
Q11	KSR1	LIDISIGSLR	N	B	B	B			<4.1 x 10^5	N	B	B	B			<4.1 x 10^6
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	N	N	N	N				N	B	B	B			<4.1 x 10^6
Q13	PEA_15	SEEITTGSAWFSFLESHNK	N	N	N	N				N	N	N	N			
Q14	KSR2 2	IHSSVGSCENIPSQQR	N	N	N	N				N	N	N	N			
Q15	Sur8/Shoc2	SIHILPSSIK	B	B	N	B			<4.1 x 10^5	0.30	0.32	0.33	0.32	0.02	4.82	<4.1 x 10^6
Q17	MVP	LFSVPDFVGDACK	B	B	B	B			<4.1 x 10^5	0.28	0.34	0.38	0.33	0.05	15.10	<4.1 x 10^6
Q18	PEBP1	ELAPLFEELR	B	B	B	B			<4.1 x 10^5	B	B	B	B			<4.1 x 10^6
Q19	ARRB1 1/2	ACGVDYEVK	N	N	N	N				N	B	B	B			<4.1 x 10^6
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	N	N	N	N				N	N	N	N			<4.1 x 10^6
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	N	N	N	N				0.28	0.34	0.36	0.33	0.04	12.99	1.38 x 10^6
Q22	IQGAP1	ILAIGLINEALDEGDAQK	N	N	N	N				N	N	N	N			
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	N	N	B	N				1.12	0.98	0.98	1.03	0.08	7.87	4.32 x 10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	N	N	N	N				N	N	N	N			
Q28	PEA_15	ISEEDELDTK	N	N	N	N				N	N	N	N			
Q29	MVP	ALQPLEEGEDEEK	N	N	N	N				B	B	B	B	B		<4.1 x 10^6

Table F. HCT 116 biological replicate 2 and either 100 fmol or 10 fmol of LM1 QconCAT. For each technical replicate light to heavy (L:H) ratios, the mean, standard deviation (SD), and coefficient of variation (CV) are provided. All L:H ratios have light signal derived from QconCAT peptide subtracted. Mean is used to calculate copies per cell (CPC). B-type quantifications (B) and noise (N).

			HCT 116 (2) +LM1 10 fmol						HCT 116 (2) +LM1 100 fmol					
Peptide	**Protein**	**Q-Peptide Sequence**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**CPC**	**L:H_1**	**L:H_2**	**L:H_3**	**Mean**	**SD**	**CV, %**	**CPC**
Q1	IQGAP1	NVIFEISPTEEVGDFEVK	N	N	N	N	-	0.33	0.35	0.36	0.35	0.02	4.4	1.07x 10^6
Q2	ARRB2 1/2/3	ACGVDFEIR	N	N	N	N	-	B	B	B	B			<3.0 x 10^6
Q3	KSR2 1	SEEQQPLSLQK	N	N	N	N	-	0.34	0.37	0.34	0.35	0.02	4.9	1.08 x 10^6
Q5	KSR2 1	LTVDAYPGLCPPPPLESGHR	N	N	N	N	-	N	N	N	N			
Q6	MPKS1	LPSVEGLHAIVVSDR	N	N	N	N	-	B	B	B	B			<3.0 x 10^6
Q7	PEBP1	LYTLVLTDPDAPSR	N	N	N	N	-	0.15	0.15	0.16	0.15	0.01	3.8	4.72 x 10^5
Q8	KSR1	DLTLDALLEMNEAK	N	N	N	N	-	B	N	B	B			<3.0 x 10^6
Q9	PAXI 1/2/3	CYYCNGPILDK	N	N	N	N	-	B	B	B	B			<3.0 x 10^6
Q10	KSR2 1/2	QQFIFPDVVPVPETPTR	N	N	N	N	-	B	B	B	B			<3.0 x 10^6
Q11	KSR1	LIDISIGSLR	N	N	N	N	-	B	B	B	B			<3.0 x 10^6
Q12	ARRB1 1/2	CPVAMEEADDTVAPSSTFCK	N	N	N	N	-	N	N	N	N			
Q13	PEA_15	SEEITTGSAWFSFLESHNK	N	N	N	N	-	N	N	N	N			
Q14	KSR2 2	IHSSVGSCENIPSQQR	N	N	N	N	-	N	N	N	N			
Q15	Sur8/Shoc2	SIHILPSSIK	N	N	N	N	-	0.32	0.36	0.33	0.34	0.02	6.1	1.04 x 10^6
Q17	MVP	LFSVPDFVGDACK	N	N	N	N	-	0.35	0.37	0.36	0.36	0.01	2.8	1.11 x 10^6
Q18	PEBP1	ELAPLFEELR	N	N	N	N	-	0.00	0.00	0.00	0.00	0.00	0.00	<3.0 x 10^6
Q19	ARRB1 1/2	ACGVDYEVK	N	N	N	N	-	0.29	0.27	0.35	0.30	0.04	1.4	9.33 x 10^5
Q20	ARRB2 1/2/3	CPVAQLEQDDQVSPSSTFCK	N	N	N	N	-	N	N	N	N			
Q21	Sur8/Shoc2	ELTQLTELYLYSNK	N	N	N	N	-	0.34	0.32	0.37	0.34	0.03	7.3	1.06 x 10^6
Q22	IQGAP1	ILAIGLINEALDEGDAQK	N	N	N	N	-	N	N	N	N			
Q23	PEBP1	GNDISSGTVLSDYVGSGPPK	N	N	N	N	-	1.29	1.48	1.51	1.43	0.12	8.3	4.39 x 10^6
Q26	KSR1 1/2/3	LSHDWLCYLAPEIVR	N	N	N	N	-	N	N	N	N			
Q28	PEA_15	ISEEDELDTK	N	N	N	N	-	0.33	N	N	N			
Q29	MVP	ALQPLEEGEDEEK	N	N	N	N	-	N	N	B	N			

Appendix 8

```
%scaffold model

function dydt=scaffold_mod2(y,ke,kr,km,de,dr,dm,St,Mt,Et,Rt)

dydt=zeros(7,1);

% dSR/dt=R*S*kr-dr*SR+SMR*dm+SRE*de-SR*M*km-SR*E*ke;              y(5)
% dSE/dt=E*S*ke-de*SE+SEM*dm+SRE*dr-SE*M*km-SE*R*kr;              y(7)
% dSM/dt=M*S*km-dm*SM+SMR*dr+SEM*de-SM*R*kr-SM*E*ke;              y(6)
%
% dSRE/dt=SR*E*ke+SE*R*kr-SRE*(dr+de)-SRE*M*km+SMRE*dm;           y(2)
% dSEM/dt=SE*M*km+SM*E*ke-SEM*(de+dm)-SEM*R*kr-SMRE*dr;           y(3)
% dSMR/dt=SM*R*kr+SR*M*km-SMR*(dm+dr)-SMR*E*ke-SMRE*de;           y(4)
%
% dSMRE/dt=SMR*E*ke+SEM*R*kr+SRE*M*km-(dm+de+dr)*SMRE;            y(1)

%E=(Et-y(7)-y(2)-y(3)-y(1));
%M=(Mt-y(6)-y(3)-y(4)-y(1));
%R=(Rt-y(5)-y(2)-y(4)-y(1));
%S=(St-y(5)-y(6)-y(7)-y(2)-y(3)-y(4)-y(1));


dydt(1)=y(4)*(Et-y(7)-y(2)-y(3)-y(1))*ke+y(3)*(Rt-y(5)-y(2)-y(4)-
y(1))...
        *kr+y(2)*(Mt-y(6)-y(3)-y(4)-y(1))*km-(dm+de+dr)*y(1);

dydt(2)=y(5)*(Et-y(7)-y(2)-y(3)-y(1))*ke+y(7)*(Rt-y(5)-y(2)-y(4)-
y(1))....
        *kr-y(2)*(dr+de)-y(2)*(Mt-y(6)-y(3)-y(4)-y(1)*km+y(1))*dm;

dydt(3)=y(7)*(Mt-y(6)-y(3)-y(4)-y(1))*km+y(6)*(Et-y(7)-y(2)-y(3)-
y(1))....
        *ke-y(3)*(de+dm)-y(3)*(Rt-y(5)-y(2)-y(4)-y(1))*kr-y(1)*dr;

dydt(4)=y(6)*(Rt-y(5)-y(2)-y(4)-y(1))*kr+y(5)*(Mt-y(6)-y(3)-y(4)-
y(1))....
        *km-y(4)*(dm+dr)-y(4)*(Et-y(7)-y(2)-y(3)-y(1))*ke-y(1)*de;

dydt(5)=(Rt-y(5)-y(2)-y(4)-y(1))*(St-y(5)-y(6)-y(7)-y(2)-y(3)-y(4)-
y(1))...
        *kr-dr*y(5)+y(4)*dm+y(2)*de-y(5)*(Mt-y(6)-y(3)-y(4)-
y(1))*km-y(5)...
        *(Et-y(7)-y(2)-y(3)-y(1))*ke;

dydt(6)=(Mt-y(6)-y(3)-y(4)-y(1))*(St-y(5)-y(6)-y(7)-y(2)-y(3)-y(4)-
y(1))...
        *km-dm*y(6)+y(4)*dr+y(3)*de-y(6)*(Rt-y(5)-y(2)-y(4)-y(1))...
        *kr-y(6)*(Et-y(7)-y(2)-y(3)-y(1))*ke;

dydt(7)= (Et-y(7)-y(2)-y(3)-y(1))*(St-y(5)-y(6)-y(7)-y(2)-y(3)-y(4)-
y(1))...
        *ke-de*y(7)+y(3)*dm+y(2)*dr-y(7)*(Mt-y(6)-y(3)-y(4)-y(1))...
        *km-y(7)*(Rt-y(5)-y(2)-y(4)-y(1))*kr;



clear;
```

```
close all;


ke=1;
kr=1;
km=1;

de=1;
dr=1;
dm=1;

St=0:200;###
Mt=10;
Rt=1;
Et=100;

time=0:0.001:10;
initials=zeros(1,7);


optionshere = optimset('TolFun',1e-15,'TolX',1e-15,...
              'LevenbergMarquardt','on','MaxFunEvals',1e8,...
              'MaxIter',1e8);

stationary_level=zeros(length(St),7);

for i=1:length(St)


[T,X] = ode23s(@(t,y)
scaffold_mod2(y,ke,kr,km,de,dm,dr,St(i),Mt,Rt,Et),time...
        ,initials);

stationary=lsqnonlin(@(y)
scaffold_mod2(y,ke,kr,km,de,dm,dr,St(i),Mt,Rt,Et),X(end,:),[],[],opt
ionshere);

stationary_level(i,:)=stationary;

end

total=sum(stationary_level');
total_E=stationary_level(:,1)+stationary_level(:,2)+stationary_level
(:,3)+stationary_level(:,7);
total_M=stationary_level(:,1)+stationary_level(:,3)+stationary_level
(:,4)+stationary_level(:,6);
total_R=stationary_level(:,1)+stationary_level(:,2)+stationary_level
(:,4)+stationary_level(:,5);
```