# Investigating the heat shock response of the yeast proteome via quantitative proteomics

A thesis submitted to the University of Manchester for the degree of PhD in the Faculty of Biology, Medicine & Health

2016

Rebecca Mackenzie

School of Biological Sciences

University of Manchester

## Table of Contents

## Table of Figures

## Table of Tables

## List of Abbreviations

2D-DIGE: two-dimensional difference in-gel electrophoresis.
2D-PAGE: two-dimensional polyacyrlamide gel electrophoresis.
AAA+: ATPases Associated with diverse cellular Activities.
AQUA: peptides for absolute quantification.
CCD: ChapCAT containing digest.
CCT: Chaperonin-Containing TCP-1 complex.
ChapCATs: chaperone QconCATs.
CHIP: the C-terminus of Hsp70-interacting protein.
CID: collision induced dissociation.
CLIPs: Chaperones Linked to Protein Synthesis.
CoPY: Census of the Proteome of Yeast.
CopyCATs: chaperone QconCATs using 2 Q-peptides per chaperone.
cpc: copies per cell.
DC: direct current.
DDA: data dependent acquisition.
DIA: Data independent acquisition.
DTT: dithiolthreitol.
emPAI: exponentially modified Protein Abundance Index.
ER: Endoplasmic Reticulum.
ERAD: Endoplasmic Reticulum Associated Degradation.
ESI-MS: electrospray ionisation mass spectrometry.
ESR: Environmental Stress Response.
ETD: electron transfer dissociation.
FDR: false discovery rate.
HF: high field.
HPD: histidine-proline-aspartic acid.
HS: heat shock.
HSE: Heat Shock Element.
HSF: Heat Shock transcription Factor.
Hsps: Heat Shock Proteins.
HSR: Heat Shock Response.
IB: Inclusion Bodies.
iBAQ: intensity-based absolute quantification.
IPTG: Isopropyl β-D-1-thiogalactopyranoside.
iTRAQ: isobaric tag for relative and absolute quantification.
LA: Luria Agar.
LB: luria broth.
LC-ESI-MS: liquid chromatography – electrospray ionisation – mass spectrometry.
LC-MS/MS: liquid chromatography tandem mass spectrometry.
m/z: mass to charge ratio. ,
MaxLFQ: MaxQuant's Label Free Quantification.
mod-cpc: 'modelled' copies per cell.
MS: Mass spectrometry.
$MS^{E}$: $MS^{Excess}$.
NBD: Nucleotide Binding Domain.
NEF: nucleotide exchange factor.
NeuCode: neutron encoding.
NG: normal growth.

PFD: Prefoldin complex.
PLGS: Protein Lynx Global Server.
PPI: protein-protein interaction.
ppm: parts per million.
PQC: Protein Quality Control.
PRM: parallel reaction monitoring.
PSMs: Peptide Spectrum Matches.
QconCAT: Q-peptide concatemer.
Q-peptides: quantification peptides.
RAC: ribosome-associated complex.
rCV: robust coefficient of variance.
RF: radio frequency.
ROS: Reactive Oxygen Species.
rpm: revolutions per minute.
SBD: Substrate Binding Domain.
SDS: sodium dodecyl sulphate.
SF: Soluble Fraction.
sHsps: Small Heat Shock Proteins.
SILAC: stable-isotope labelling of amino acids in cell culture.
SM: Starting Material.
SRM: selected reaction monitoring.
STRE: stress response element.
TAP: tandem affinity purification.
TFA: trifluoroacetic acid.
TIM: translocase of the inner membrane.
TMTs: tandem mass tags.
ToF: Time of Flight.
TPR: tetratricopeptide repeat.
TQ: tandem quadrupole.
UPR: Unfolded Protein Response.
XIC: extracted ion chromatogram.
YOD: Yeast only digest.

## Abstract

Institution: The University of Manchester

Name: Rebecca Mackenzie

Degree: PhD Biochemistry

Thesis Title: Investigating the heat shock response of the yeast proteome via quantitative proteomics

Date: September 2016

Proteostasis, the regulation of protein abundance and function in cellular systems, underpins the ability of organisms to deal with environmental challenge such as heat shock stress. In this thesis, a quantitative study of this process at the molecular level has been undertaken using the model eukaryote *S. cerevisiae* to characterise the protein level response to heat shock. Although well studied, much of the previous work has focussed on changes at the mRNA levels or relative changes in protein expression. To address this, an absolute quantification strategy was developed utilising the QconCAT approach.

A total of 10 recombinant QconCAT proteins were designed to target the 63 chaperones in *S. cerevisiae*, with up to 5 Q-peptides selected per chaperone where possible. Subsequently, absolute copy per cell values were determined for 49 of the 63 chaperones in *S. cerevisiae* under conditions of normal growth and heat shock (42 °C, 30 minutes). Chaperones that are known targets of the heat shock response activating transcription factor HSF1 are significantly upregulated in response to heat shock.

Furthermore, this dataset has been extended towards proteome-wide quantification, for which SRM-normalised label free quantification values for 1644 proteins in both conditions were determined. Using these values and a high quality chaperone-client interaction dataset, progress has been made towards modelling the change in the protein volume and workload of each chaperone in response to heat shock. Interestingly, for the chaperone Ssb2, both its workload and absolute abundance were significantly upregulated in response to heat shock. However, across all chaperones, the relationship between protein volume, workload fold change and abundance fold change is minimal; further work is required to investigate this.

## Declaration

No portion of the work referred to in this 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

## Acknowledgements

I would like to thank my main supervisors, Professor Simon Hubbard and Professor Claire Eyers for their endless support during my studies and for allowing me to undertake my studies in Bioinformatics and Mass Spectrometry.

I would specifically like to thank those who went out of their way to teach me new techniques, and mentor me: Craig Lawless, Stephen Holman, Philip Brownridge, Karin Lanthaler, Rachel Watkins, Paraskevi Kritsiligkou, Manuel Garcia, Victoria Harman and Professor Rob Beynon.

I would like to acknowledge all members of the Hubbard, Eyers, Grant and Schwartz laboratories, as well as the members of the Centre for Proteome Research in Liverpool; your feedback to my work has always been valued.

Finally, I wish to thank my best friend Rosanna, and my partner Leon, for their endless kindness, support and motivation during the last few years.

# Chapter 1 - Introduction

## 1.1 Protein Folding

Within living systems, there exist a wide variety of highly specific protein structures resulting from protein folding processes that bring into close proximity key functional groups, enabling organisms to develop both diversity and selectivity in their underlying chemical processes. In addition, protein folding is coupled to a range of other biological activities, including the trafficking of molecules to particular cellular locations, regulation of cellular growth and differentiation and communication between biological pathways, all crucial for the viability of the cell. Subsequently, a protein must be able to fold to its active native state repeatedly – first folding upon emerging from the ribosome as a newly synthesised polypeptide chain, and then refolding after various unfolding events such as misfolding and translocation across a membrane as part of a cellular pathway (Englander and Mayne, 2014).

A protein requires only its amino acid sequence to fold to the native state, as demonstrated by the spontaneous folding of ribonuclease and the five-helix bundle protein $\lambda_{6\text{-}85}$ (Anfinsen, 1973, Yang and Gruebele, 2003). Folding *in vivo* may be co-translational, initiated before completion of synthesis whilst the polypeptide chain is still attached to the ribosome (Cabrita et al., 2010). However, protein folding beyond the α helical secondary structure is prohibited due to the narrow polypeptide channel on the ribosome (Hartl and Hayer-Hartl, 2002). Protein structures including multiple domains and beta sheets can only form after release of the entire sequence from the ribosome. This post-translational protein folding can occur in the cytosol or other cellular compartments such as the mitochondria or endoplasmic reticulum (ER) after translocation through membranes according to an internal targeting sequence (Bukau and Horwich, 1998, Hartl and Hayer-Hartl, 2002). Such proteins are less efficient at folding, and are particularly prone to misfolding and aggregation after becoming trapped in kinetic intermediates due to the crowded cellular environment (Ellis, 2001a, Kerner et al., 2005). Protein aggregation is also observed under conditions of high temperatures (Acampora and Hermans, 1967). The protein aggregates can differ in size, render the protein non-functional, reduce the efficiency of protein folding within all cells, and can eventually lead to the formation of toxic aggregates (Ellis and Minton, 2006, Ellis, 2001a). In bacteria, protein aggregates become sequestered within inclusion bodies (Singh and Panda, 2005). In humans, protein aggregates can become structured fibrillar amyloids that are characteristic of a number of disease states, including Alzheimer's (aggregation of β-amyloid) and Parkinson's diseases

(aggregation of α-synuclein) (Dobson, 2001). Although amyloid fibrils are one of many types of protein aggregation, a striking feature is the highly organised hydrogen-bonded structure between the polypeptide main chain atoms. As a result, once formed, the aggregates are stable and allow for a progressive deposition of protein in tissue with additional quantities of the same protein being converted into more amyloid fibrils (Dobson, 2003). In order to survive, the cell must be able to prevent, repair and/or remove if required, misfolded protein and any aggregates.

## 1.2 Chaperones

Although some newly translated proteins are able to fold spontaneously, a substantial fraction of proteins are inefficient at folding and are particularly vulnerable to misfolding and aggregation, aggravated by the highly crowded cellular environment. Large proteins with complex structures expose their hydrophobic residues to solvent during folding, making them particularly susceptible to non-native interactions that lead to aggregation and prevention of protein functionality. To counteract the non-native interactions, stabilise the unfolded structure and prevent intermolecular contacts forming, cells recruit specialised 'chaperone' proteins (Figure 1.1). A molecular chaperone is defined as "any protein that interacts with and aids in the folding or assembly of another protein without being part of its final structure" (Hartl, 1996). In order to counteract the formation of non-native interactions that can lead to non-functional protein aggregation, a network of molecular chaperones exist that mediate both *de novo* folding and protein unfolding-refolding events required during the protein's life cycle. Aside from their key role in preventing protein aggregation, chaperones are able to function in protein quality control pathways, including protein unfolding and disaggregation, mediation of protein translocation across membranes, ribosomal RNA processing and targeting terminally misfolded proteins for proteolytic degradation via endoplasmic reticulum associated degradation (ERAD) (Bukau and Horwich, 1998, Hartl and Hayer-Hartl, 2002, Kim et al., 2013). In this way, chaperones are able to modulate the assembly and disassembly of protein-protein, protein-DNA and protein-RNA complexes. To maintain a tightly controlled protein-folding pathway, molecular chaperones tend to function cooperatively, ensuring that the various stages in the folding of a protein is successfully completed. Chaperones are able to improve the efficiency of the protein folding process by preventing competing reactions, particularly through hydrogen bonding in the polypeptide main chain as observed in non-reversible protein aggregation. By preventing such non-native interactions, chaperones may increase the overall rate of protein folding to the native state. However, the increase in folding rate is dependent

upon the mechanism of substrate binding and biochemistry of the chaperone; particular classes of chaperone act to maintain the substrate in an unfolded state for translocation across a membrane or to pass the substrate onto the degradation pathway of the cell, discussed in detail within this chapter. Chaperones are able to recognise a wide variety of protein substrates, typically targeting the exposed hydrophobic regions that are internalised within the native structure (Hartl and Hayer-Hartl, 2002).

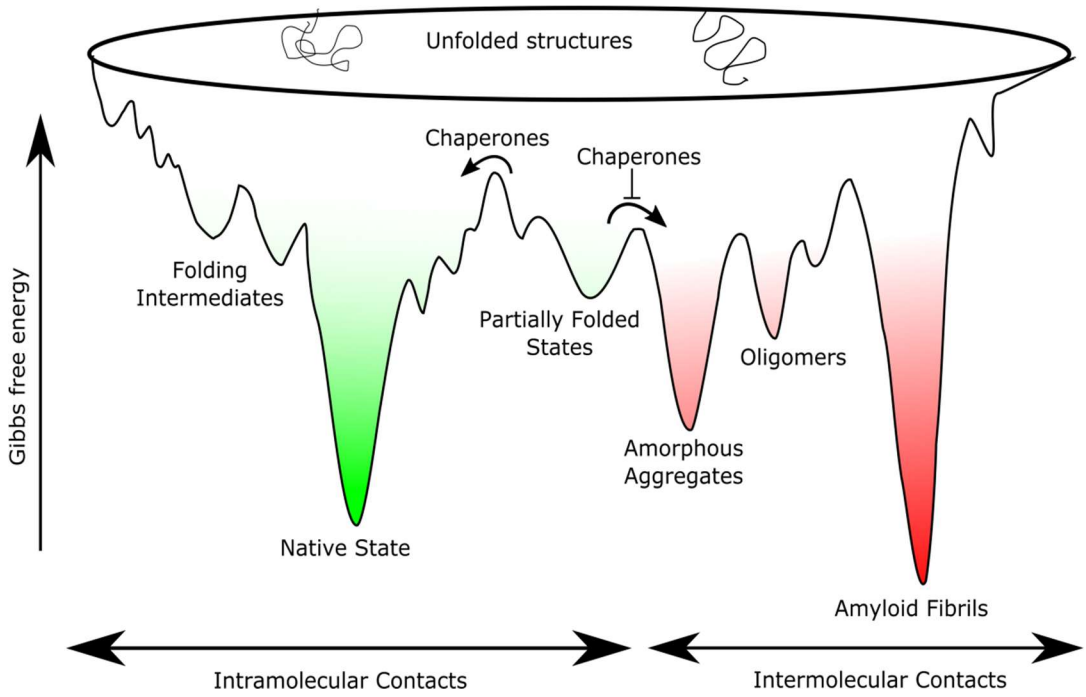



**Figure 1.1) Chaperones mediate folding of partially folded states towards the native state**

Energetically favourable intermolecular interactions (green) are stabilised as the protein progresses down the folding funnel towards the native state. Due to exposure of hydrophobic regions to solvent, proteins can form non-native contacts, resulting in partially folded intermediates/misfolded protein that are trapped in low-energy wells. Chaperones assist these partially folded intermediates in overcoming the free energy barriers and preventing intermolecular contacts (red) forming leading to aggregation. Adapted from publication (Kim et al., 2013).

Many chaperones become induced following heat stress, giving rise to their name 'Heat Shock Proteins' (Hsps) as a result of the heat shock response (HSR). However, not all Hsps are chaperones and not all chaperones are Hsps. In order to further understanding of the HSR, *Saccharomyces cerevisiae* serves as an ideal model system. With its genome easily manipulated and recent advances in transcriptional profiling and proteomics analyses, the study of stress response has progressed at a pace much faster than possible in humans. Despite this, there are certain features of the human HSR that are not within the *S. cerevisiae* model. *S. cerevisiae* has a single heat shock transcription factor (HSF), a homolog of Hsf1 in the human proteome. An

additional three HSFs (Hsf2, Hsf3 and Hsf4) are found in humans that act to mediate the HSR. This disparity has been exploited to utilise yeast as a test system towards understanding protein homeostasis. Some human genes can be made functionally competent in yeast in order to study the difference in the systems HSR in order to elucidate the role of the gene (Takemori et al., 2009). Indeed, understanding of modulation of the HSR will enable better understanding of human diseases linked to protein folding (e.g. Alzheimer's disease and Parkinson's Disease) (Dobson, 2001).

According to Gong and colleagues (Gong et al., 2009), *S. cerevisiae* has 63 known chaperones that can be classified into 8 distinct families dependent upon their molecular weight and sequence homology: Small, Hsp40, Hsp60, Hsp70, Hsp90, Chaperonin-Containing TCP-1 complex (CCT), Prefoldin complex (PFD) and ATPases Associated with diverse cellular Activities (AAA+). Each chaperone family can occupy multiple cellular locations (Figure 1.2).

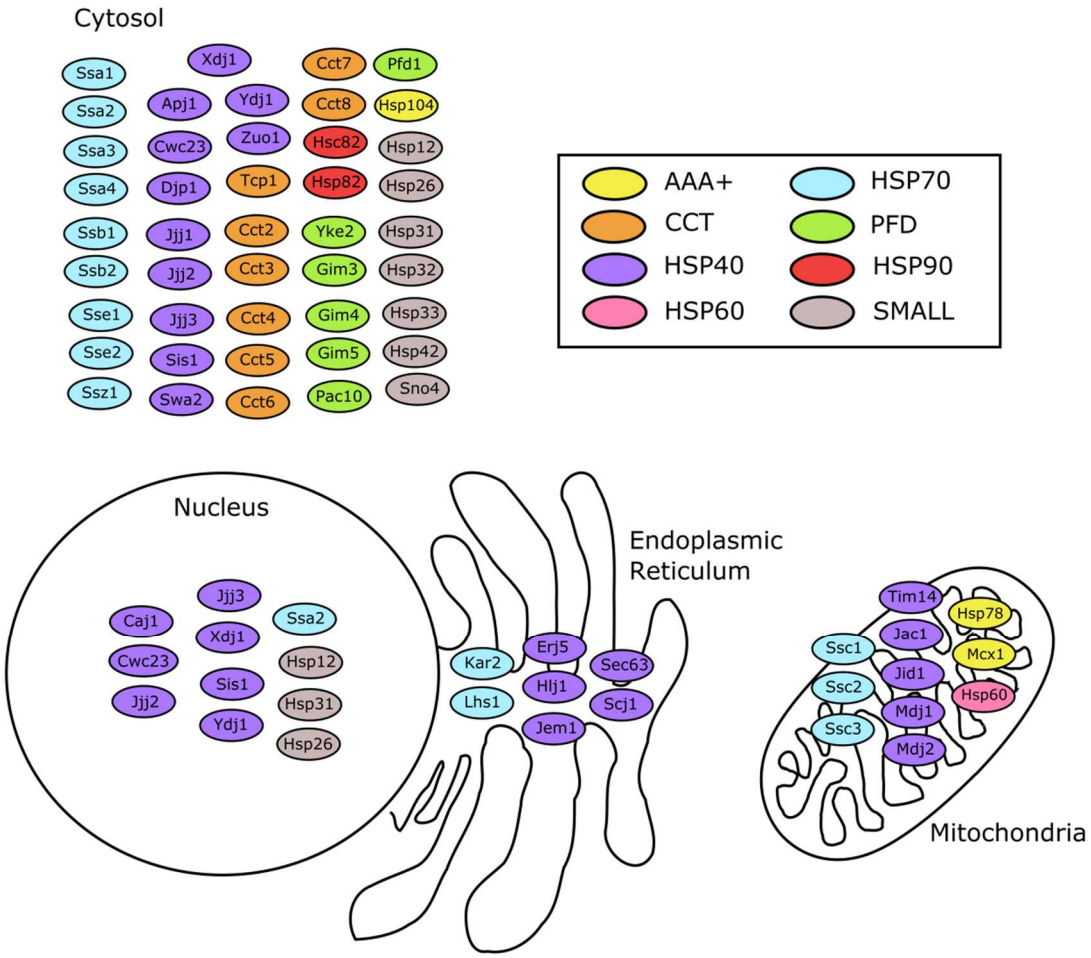



**Figure 1.2) Subcellular locations of 63 known chaperones in *S. cerevisiae***

Sixty-three known chaperones can be classified into eight functional groups according to their molecular weight and sequence homology, with chaperone classes spread across subcellular locations as reported by Gong et al, 2009. The Hsp40 protein Xdj1 has also been proposed to be present in the mitochondria. Cwc23, Jjj2, Jjj3, Xdj1, Sis1, Ydj1, Ssa2, Hsp12 and Hsp31 are observed in both the nucleus and cytoplasm. Hsp26 subcellular location is not known but is predicted to be within the cytoplasm and nucleus. Chaperones may also have the ability to move between subcellular locations. Subcellular locations are as depicted by Gong and colleagues (Gong et al., 2009).

Protein folding is promoted by the majority of chaperones through ATP-regulated cycles of binding and release. Binding to the chaperone blocks aggregation and is able to reduce the concentration of free-folding intermediates, somewhat reducing the crowded nature of the cell. ATP-independent chaperones are able to act as holding proteins, securing the folding protein within a closed environment, buffering aggregation. Efficient folding by the chaperone is achieved when the rate of folding is faster than the rates of aggregation or chaperone binding. If folding is slower, the protein is transferred to a chaperone with different mechanistic properties or, in extreme cases, is transferred to the degradation machinery for removal (Kim et al., 2013). This results in a wide network of co-ordinating chaperones with polypeptide

transfer occurring between classes. By doing so, the cell ensures the protein is fully protected against aggregation from the beginning to the end of its life cycle. This co-operative network has been identified through both system-wide and bioinformatics approaches in both prokaryotic and eukaryotic organisms (Albanese et al., 2006, Calloni et al., 2012, Dekker et al., 2008, del Alamo et al., 2011, Fujiwara et al., 2010, Gong et al., 2009, Kerner et al., 2005, McClellan et al., 2007, Oh et al., 2011, Yam et al., 2008).

A nascent polypeptide chain may begin its route on the chaperone pathway from immediate emergence of its first one hundred amino acids from the ribosome (Kim et al., 2013). Here, the chaperone classes in *S. cerevisiae* are discussed in the order in which a nascent chain may encounter them during its protein-folding pathway. It is important to note that chaperones do not act solely on nascent chains, and indeed play an important role throughout a protein's life cycle.

### 1.2.1 Hsp40 chaperones

Hsp40 chaperones are also known as J-domain proteins (J-proteins) due to their homology with DnaJ from *Escherichia coli* (Greene et al., 1998). J-proteins have the ability to accelerate Hsp70 ATPase activity through its J-domain, thus initiating Hsp70's folding cycle. Within the J-domain lies a four-helix bundle, containing an invariable histidine-proline-aspartic acid (HPD) motif – the Hsp70 interaction site. The co-crystal structure of Hsp70 and a J-protein suggests that the HPD binds Hsp70 near the base of its ATPase domain, thus accelerating the conformational change required for ATP hydrolysis and the subsequent closure of the Hsp70 structure upon binding its substrate protein (Verghese et al., 2012, Jiang et al., 2007).

The J proteins can be localised to particular sites within the cell, with the high J domain concentrations being enough to target Hsp70 to client proteins at these sites, without the need for the J protein to bind to the client itself. In such a case, Zuo1, a Hsp40 co-chaperone, functions in the ribosome-associated complex (RAC) with the Hsp70 Ssz1. Zuo1 is bound to the ribosome exit tunnel with Ssz1, such that as nascent chains emerge, Ssz1 in the RAC is able to interact with the newly synthesised polypeptide chain, preventing unfavourable inter- and intra-molecular interactions forming (Kim et al., 2013, Jiang et al., 2007). Zuo1 is also able to recruit and interact with the Ssb1 and Ssb2 proteins, enabling these Hsp70s to mediate co-translational folding (Huang et al., 2005). Another Hsp40, Jjj1 is able to bind to the ribosome 60S subunit, and can recruit Ssa1 and Ssa2 to the ribosome to mediate a final step in ribosome biogenesis.

Unlike Zuo1 however, Jjj1 does not require the interaction of a second Hsp70 to stimulate the ATPase activity of Ssa-type Hsp70 chaperones. Cells lacking Zuo1 exhibit growth defects that are rescued by overexpression of Jjj1, suggesting functional overlap (Meyer et al., 2007).

Ydj1 is currently the most well-studied class I *S. cerevisiae* J protein. Ydj1 can be farnesylated at the *C*-terminus, allowing for localisation of a sub-population of Ydj1 to the ER membrane (Verghese et al., 2012). Mutant Ydj1, in which the cysteine of the CaaX box for farnesylation is mutated to a serine, demonstrated an increased cystolic localisation and resulted in a temperature-sensitive growth phenotype, thus the farnesylation is required for Ydj1 function at elevated temperatures (Caplan et al., 1992). With the exception of Zuo1, Ydj1 and Sis1 are the most abundant Hsp40 chaperones in *S. cerevisiae* (Ghaemmaghami et al., 2003), with studies demonstrating that knockout of either is tolerated, whilst the loss of both is lethal (Johnson and Craig, 2001). The J-proteins can play specialised roles: the aforementioned Zuo1 and Jjj1 function solely at the ribosomes, whilst the Hsp40 Swa2 is a dedicated adaptor protein with an additional module for recruiting and localising Ssa chaperones to clathrin-coated vesicles, promoting their uncoating (Gall et al., 2000).

### 1.2.2 Hsp70 chaperones

Hsp70 chaperones are ATP-regulated, and require the action of two cofactors in order to complete a folding cycle: a Hsp40 co-chaperone to deliver non-native polypeptides and initiate ATPase activity, and a nucleotide exchange factor (NEF) to exchange the ADP for ATP. HSP70 class chaperones contain an *N*-terminal nucleotide-binding domain (NBD) with ATPase activity and a *C*-terminal substrate-binding domain (SBD), able to recognise 5-7 residue hydrophobic regions. Both the NBD and SBD are connected by a hydrophobic linker region (Kim et al., 2013). The Hsp70 folding cycle begins when a Hsp40 co-chaperone binds and delivers the non-native protein to the Hsp70. Hsp70 is activated and able to carry out ATP hydrolysis via its ATPase domain enabling it to transition into the closed form, in which the client protein is tightly bound and the C-terminal domain folded over it in a lid-like manner (Figure 1.3) (Morano, 2007). Upon nucleotide exchange by an NEF, Hsp70 undergoes a conformational change, releasing the substrate protein and returning to the open conformation.

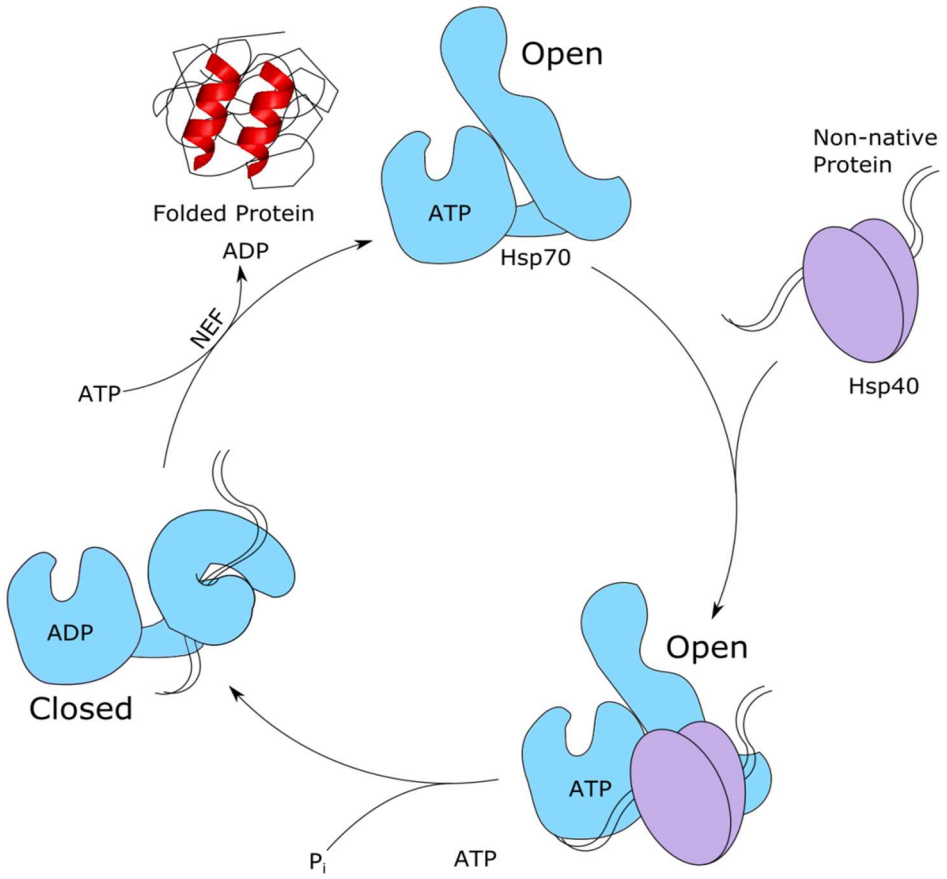



**Figure 1.3) Folding cycle of Hsp70 chaperones**

Through cycles of ATP hydrolysis and ATP rebinding, Hsp70 can mediate folding of non-native polypeptide chains. The co-chaperone Hsp40 stimulates ATP hydrolysis by the Hsp70 NBD, whilst NEFs such as GrpE and Bag can mediate substrate dissociation for release into the cytosol, or passing on to later stages of folding mediated by other chaperones. Adapted from publication (Kim et al., 2013).

Different Hsp70 chaperones have different roles within the cell, each interacting with specific Hsp40 co-chaperones. There are three distinct subfamilies, all of which are apparent in eukaryotes: the classical Hsp70/DnaK; the Hsp110/Sse family and the Grp170/Lhs1 family, distinguished by the length of the loop region between the α helical lid and β sheet SBD in the *C*-terminus (Morano, 2007). The majority of Hsp70 chaperones can interact broadly with substrate proteins in the cytosol, whilst a select few are specialised. For instance, Ssb1 and Ssb2 have ribosome-binding capabilities in order to function in co-translational *de novo* folding of newly synthesised polypeptide chains (Pfund et al., 1998). In comparison, the Hsp70 chaperone Ssq1 resides in the mitochondrial matrix and exclusively mediates the biogenesis of iron-sulphur cluster proteins alongside the co-chaperone Jac1 and the scaffolding protein Isu1 (Voos and Rottgers, 2002).

Hsp70s are involved in both co-translational and post-translational folding. As previously described, Ssz1 is part of the RAC complex, which mediates recruitment of additional Hsp70 chaperones and co-translational folding of nascent chains (Morano, 2007). Other non-ribosome bound Hsp70s may be recruited to the polypeptide in order to stabilise protein domains awaiting inter-domain contacts before natively folding (Morano, 2007). Additional Hsp70 chaperones are able to act as co-chaperones for Ssa and Ssb, the more efficient folding-competent chaperones. The Sse family (Hsp110 homologs) have this function and are unique to eukaryotic cells. Sse chaperones cannot actively refold proteins but act to hold thermally denatured substrates in a protected state such that they may be more effectively renatured by Ssa and Ssb chaperones. The Sse1 Hsp40 chaperone has been demonstrated to work synergistically with the Hsp40 co-chaperone Ydj1 to activate the folding capabilities of Ssa1, with the interaction identified through immunoprecipitation and Native-PAGE analysis (Shaner et al., 2005).

Hsp70 chaperones also have a major role in protein import. In the mitochondria of *S. cerevisiae*, large proteins that are targeted to the mitochondrial matrix must pass through the translocase of the inner membrane (TIM) complex in an unfolded manner. Upon entering the matrix, it must fold into the native state in order to function. A mitochondrial Hsp70, Ssc1, is in complex with the TIM23 complex through interaction with protein Tim44. As the polypeptide emerges from the TIM23 channel, Ssc1 binds and dissociates from Tim44, acting in a trapping and pulling manner (Voisine et al., 1999). Maintaining this interaction with the unfolded polypeptide allows Ssc1 to drag the polypeptide into the matrix whilst protecting from non-

native interactions. Upon the action of an NEF, Mge1, Ssc1 releases the folded polypeptide into the matrix.

### 1.2.3 Hsp60 chaperones

Proteins that are unable to utilise Hsp70 for full native folding are transferred to the chaperonins or the Hsp90 chaperone system. There exist two groups of chaperonins: Group I chaperonins including the GroEL in bacteria, Hsp60 in mitochondria and Cpn60 in chloroplasts; Group II chaperonins comprise the eukaryotic CCT and the archaeal thermosome (Horwich et al., 2007, Kim et al., 2013). In *S. cerevisiae*, the Hsp60 and CCT classes are the chaperonins.

Chaperonins are responsible for protection and post-translational folding of unfolded or partially unfolded proteins and are able to interact with other chaperones in order to fold specific substrates (Verghese et al., 2012). Both classes of chaperonins form double-ring structures, able to fold the protein in a central cavity in a nucleotide-dependent manner. In *S. cerevisiae* a single Group I chaperonin exists, Hsp60, residing within the mitochondrial matrix.

The Hsp60 chaperone is a large complex composed of two seven-membered rings each containing symmetrical subunits of 60 kDa. Hsp60 is able to encapsulate the folding protein via a lid structure, provided by co-operation with a seven membered structure containing identical subunits of 10 kDa, known as Hsp10 in the *S. cerevisiae* mitochondrial matrix (Ellis, 2001b, Frydman et al., 1994, Langer et al., 1992). *S. cerevisiae* Hsp60-Hsp10 is a homolog of the mitochondrial GroEL-GroES complex in *E. coli*; indeed much of the structural knowledge has been obtained via studies on the GroEL-GroES complex (Xu et al., 1997).

Like so many classes of chaperones, the folding cycle of class I chaperonins is ATP-regulated. Each of the subunits in Hsp60 contain an equatorial ATPase domain, an intermediate hinge domain and an apical domain (Horwich et al., 2007). The entrance to the Hsp60 cavity is formed by the apical domains, exposing hydrophobic residues that mediate recognition of proteins up to ~60 kDa with exposed hydrophobic residues. Co-operative binding of ATP to the Hsp60 subunits in the *cis* ring initiates a series of conformational changes that enable association of Hsp10 via the apical domains, followed by substrate release from the hydrophobic binding sites into the Hsp10-capped Hsp60 complex (Kim et al., 2013). The protein is free to fold in the chaperonin cage for ~10 s during which ATP bound to the *cis* ring is hydrolysed to ADP, as demonstrated within the GroES-GroEL complex (Figure 1.4) (Xu et al.,

1997). ATP binding to the *trans* ring induces protein substrate leaving the Hsp60 complex and Hsp10 dissociation, with the protein free to undergo further rounds of Hsp60-mediated, or Hsp70-mediated folding if required (Kim et al., 2013).

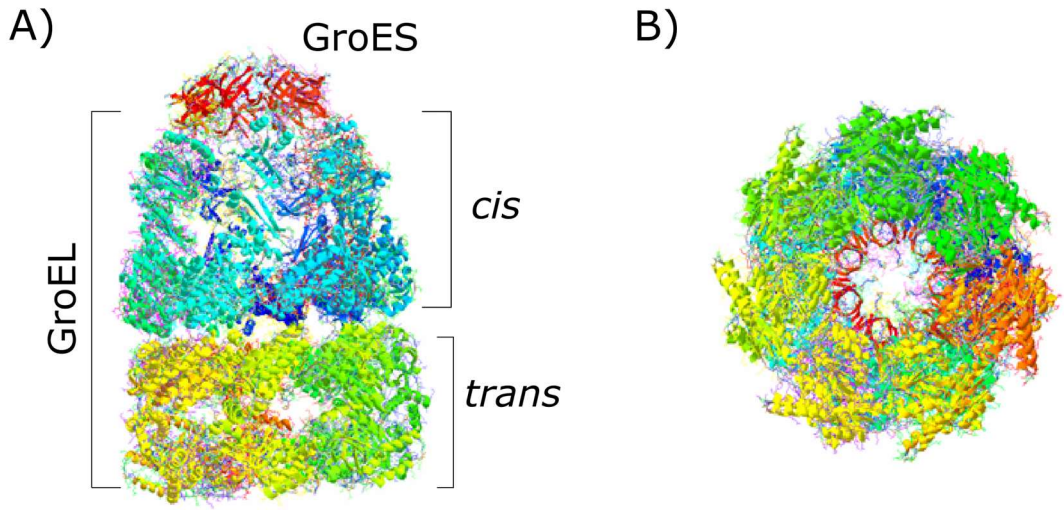



**Figure 1.4) The structure of *E. coli* GroEL-GroES-ADP(7) chaperonin complex; a homolog of the *S. cerevisiae* Hsp60-Hsp10-ADP(7) complex**

A) Side view of the GroEL and co-chaperone GroES complex (GroES in red); B) Bottom-up view of A, showing the 7-membered ring structure. PBD ID: 1AON (Xu et al., 1997). No PDB structure is available for the *S. cerevisiae* Hsp60 complex.

Within the mitochondria, *S. cerevisiae* proteins that are destined to remain within the mitochondrial matrix are transferred from the Hsp70 system to Hsp60 to complete folding (Cheng et al., 1989). Hsp60 may also mediate protein translocation across the mitochondrial membrane if the protein contains a mitochondrial sorting signal, in addition to the *N*-terminal matrix-targeting signal sequence. Whilst the mitochondrial Hsp70, Ssc1, mediates translocation into the mitochondrial matrix, work by (Koll et al., 1992) on Cytochrome $B_2$ observed export mediated by Hsp60 either as import occurs, or following complete import into the matrix. According to Koll and colleagues (Koll et al., 1992), mitochondrial Hsp70 binds the *N*-terminal Cytochrome $B_2$ extension as it protrudes into the mitochondrial matrix during import. The precursor is then delivered to Hsp60 by Hsp70, during which the hydrophobic export sequence associates with the Hsp60 chaperonin and maintains the protein in an unfolded state for export. Hsp60 delivers the precursor to the export machinery through recognition of the hydrophobic export sequence and releases the substrate in an ATP-hydrolysis dependent manner, in this way the protein remains unfolded for export from the mitochondrial matrix.

### 1.2.4 CCT chaperones

The CCT chaperone class (also known as tail-less complex polypeptide-1) consists of eight cytosolic chaperones: Cct1/Tcp1, Cct2, Cct3, Cct4, Cct5, Cct6, Cct7 and Cct8 and are classified class II chaperonins. Together these form the subunits of an eight-membered heteromeric ring structure, with two rings forming the CCT complex, able to participate in the folding of proteins both co- and post-translationally (Verghese et al., 2012).

Each subunit contains an ATP-binding equatorial domain and a substrate-binding apical domain linked by an intermediate domain. Lacking co-chaperones, CCT contains an extra-helical protrusion at the tip of its apical domain, which acts as a lid upon substrate binding, enclosing the substrate within the chaperonin cage. The folding cycle differs from that of class I chaperonins, as ATP-binding induces a conformational change in which the cavity of CCT narrows, as a result of anticlockwise rotation of the intermediate and apical domains of subunits with respect to their equatorial domains, resulting in iris-like lid closure of the helical regions (Booth et al., 2008). Class I chaperonins exist as homomultimers, however, class II chaperonins are composed of eight distinct subunits. A second difference is the lid structure, provided by Hsp10 in class I chaperonins in *S. cerevisiae*, but formed from the helical protrusions of each subunit in class II chaperonins (Figure 1.5) (Llorca et al., 2000, Verghese et al., 2012).

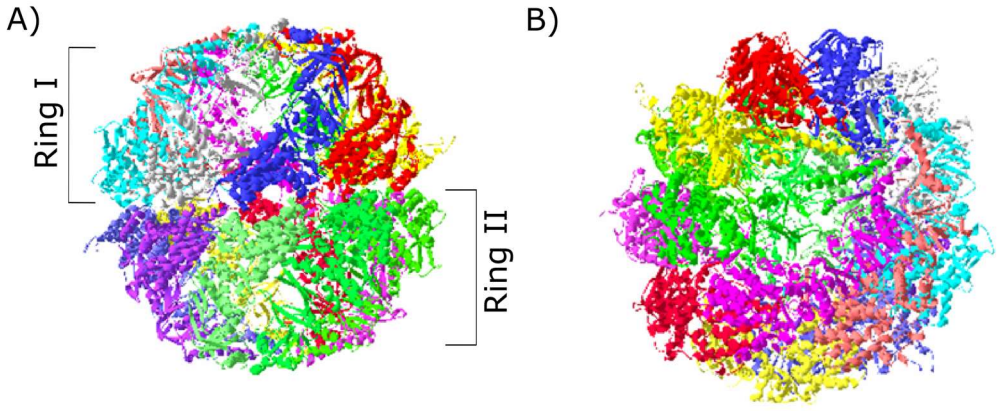



**Figure 1.5) Representation of the eukaryotic CCT complex**

A) Front view of CCT, a double eight-membered ring structure. B) Top-down view of A, demonstrating the eight subunits forming each ring. PDB ID: 4V94 (Leitner et al., 2012).

Approximately 5-10 % of newly synthesised polypeptides interact with CCT, with substrates identified through immunoprecipitation assays including the cytoskeletal proteins actin and tubulin as well as several proteins with β-propellers (Yam et al., 2008). The CCT complex is able to bind to nascent chains and co-operate with Hsp70 to fold multi-domain

structures co-translationally, as has shown to be the case for folding of actin and tubulin (Etchells et al., 2005, Cuellar et al., 2008). In the cytosol, both tubulin and actin share similar structures, with two topological domains connected via a hinge region. Through the use of cryo-electron microscopy, work by Llorca and colleagues (Llorca et al., 2000, Llorca et al., 1999) demonstrated that both tubulin and actin are able to bind to CCT through their topological domains, but to opposite sides of the chaperonin ring. Tubulin was previously thought to bind to five subunits at their helical protrusions in the apical domain, in two different arrangements, therefore being able to utilise all eight of the CCT subunits. However, more recent evidence from Leitner and colleagues (Leitner et al., 2012) using cross-linking with mass spectrometry (MS) has shown that tubulin interacts with four subunits of CCT, specifically with the equatorial domains of subunit 2, 4 and 5, and the apical domain of subunit 7, binding near the negatively charged region of the central cavity. Actin on the contrary, only binds to two subunits, beneath the helical region in the apical domain (Llorca et al., 1999).

Actin folding in cells is a tightly controlled process; in which newly synthesised actin is maintained in a sequestered environment until full native folding is achieved. This environment is regulated from the point of its synthesis at the ribosome, during which select Hsp70s (Ssb1 and Ssb2) are able to bind actin co-translationally due to their ribosome-binding nature. Actin is transferred co-translationally from the ribosome to the CCT chaperonin for sequestered folding within the chaperonin cage, enabled by its iris-like closure of the segments in the apical domains of the subunits. CCT functions in conjunction with the PFD molecular chaperone to facilitate the folding of both actin and tubulin; a PFD oligomer is able to bind to each of CCT's rings, interacting with subunits 1 and 4 in each ring (Martin-Benito et al., 2002). This observation proved a notion put forward by Siegers and colleagues (Siegers et al., 1999), whom demonstrated more efficient, accelerated folding of actin in the presence of PFD. This PFD-CCT interaction is discussed further in Section 1.2.6.

### 1.2.5 Hsp90 chaperones

In eukaryotes, many signalling proteins are transferred to the Hsp90 system from the Hsp70 system. There are two Hsp90 isoforms in *S. cerevisiae* sharing 85 % sequence identity: Hsc82 and Hsp82, with Hsc82 constitutively expressed and Hsp82 induced upon stress (Johnson, 2012, Borkovich et al., 1989). In contrast, higher eukaryotes have four homologs: cytosolic Hsp90, mitochondrial tumour-necrosis factor receptor-associated protein 1, the ER 94 kDa glucose-regulated protein (Grp94) and the chloroplast Hsp90c (Johnson, 2012). Systems

approaches on *S. cerevisiae* identified enrichment for components of the vesicle-mediated and Golgi apparatus transport systems, signal transduction, cell cycle, cytokinesis and bud components in the interactors for Hsp90 chaperones under normal growth (NG) conditions (30 °C). In heat stressed conditions (37 °C) components of the microtubule organising centre were enriched, with the components of cell cycle, signal transduction, cytokinesis and bud components more abundantly represented (McClellan et al., 2007). With such a diverse range of clients, Hsp90 chaperones in *S. cerevisiae* are able to interact directly or indirectly with approximately 10 % of the proteome.

Hsp90 chaperones interact with clients that are partially folded or in a nearly native folded state, with clients lacking sequence and structural similarities and sizes ranging from 14 to 290 kDa (Falsone et al., 2009, Forsythe et al., 2001). Clients have also been shown to interact with either the *N*-terminal, middle or *C*-terminal domain and so Hsp90 appears to be a promiscuous chaperone (Johnson, 2012). Almost-native Hsp90 clients are transferred from the Hsp40/Hsp70 system after failed rounds of folding, as a final attempt to achieve the native structure or to mediate assembly of complex macromolecular structures (Street et al., 2011).

The Hsp90-client interaction is highly regulated by co-chaperones that modulate the ATPase activity of Hsp90, much in the same way Hsp40 co-chaperones regulate the ATPase activity of Hsp70. Alongside Hsp70 and its co-chaperone Hsp40, additional co-chaperones are able to stimulate its ATP hydrolysis and/or client binding activity, with 12 co-chaperones existing in yeast (Table 1.1) (Johnson, 2012, Li et al., 2012). Hsp90 is more abundant than any individual co-chaperone, and many of these co-chaperones compete for the same binding site, thus multiple distinct Hsp90-cochaperone complexes may exist within the cell (Johnson, 2012).

| **Table 1.1) The functions of *S. cerevisiae* Hsp90 co-chaperones**<br>Table taken from publication (Li et al., 2012). | |
|---|---|
| **Co-Chaperone** | **Function** |
| Sti1 | Scaffold for Hsp90/Hsp70 interaction; involved in client protein maturation; inhibition of Hsp90 ATPase |
| Cpr6/Cpr7 | Peptidy-prolyl-isomerase; chaperone; involved in client protein maturation |
| Ppt1 | Phosphatase |
| Sgt1 | Forms complex with Hsp90 and cysteine- and histidine- rich domain proteins |
| She4 | Assembly of myosin fibers |
| Cns1 | Nuclear transport protein |
| Tom70p | Mitochondrial transport protein |
| Tah1 | Forms complex with Pih1 and Hsp90 |
| Aha1 | Stimulates ATPase activity; induces conformational change in Hsp90 |
| Sba1 | Involved in client protein maturation; inhibition of Hsp90 ATPase; chaperone |
| Cdc37 | Kinase-specific co-chaperone; inhibition of the Hsp90 ATPase; chaperone |
| NudC | cysteine- and histidine- rich domain-containing chaperone; dynein-associated nuclear migration protein; plays multiple roles in mitosis and cytokinesis |

The *C*-terminal domain has been demonstrated to play a role in catalysing the ATP-hydrolysis reaction at the *N*-terminal nucleotide-binding domain (Prodromou et al., 1999). Co-chaperones containing a tetratricopeptide repeat (TPR) domain are able to bind to Hsp90; for example Sti1, a co-chaperone that shares 40 % sequence identity with the mammalian Hop, is able to bind the *C*-terminal domain and lock it in such a conformation that prevents its catalytic activity, thereby inhibiting ATP hydrolysis by Hsp90 (Prodromou et al., 1999). Sti1 is able to bind simultaneously to both Hsp90 and Hsp70, acting as a scaffold protein coupling the two chaperones, thus enabling delivery of a client to the Hsp90 chaperone. Hsp90 is then in an 'intermediate' conformation in which client protein is bound but Sti1 blocks ATP hydrolysis. This inhibition is relieved when Sti1 is replaced by other TPR-containing chaperones, for example Cpr6, allowing Hsp90 to adopt a mature conformation in which ATP hydrolysis is enabled and folding of the client protein ensues (Figure 1.6) (Prodromou et al., 1999). Other co-chaperones are able to bind Hsp90 at different sites, with various effects on its activity. In this way, Hsp90 can function in an ATP-dependent folding cycle that is influenced by a complex network of co-chaperone proteins. Despite Hsp90 being widely studied, precisely how it promotes the maturation of client proteins and the mechanism behind its selectivity for the client proteins remains unknown (Verghese et al., 2012).

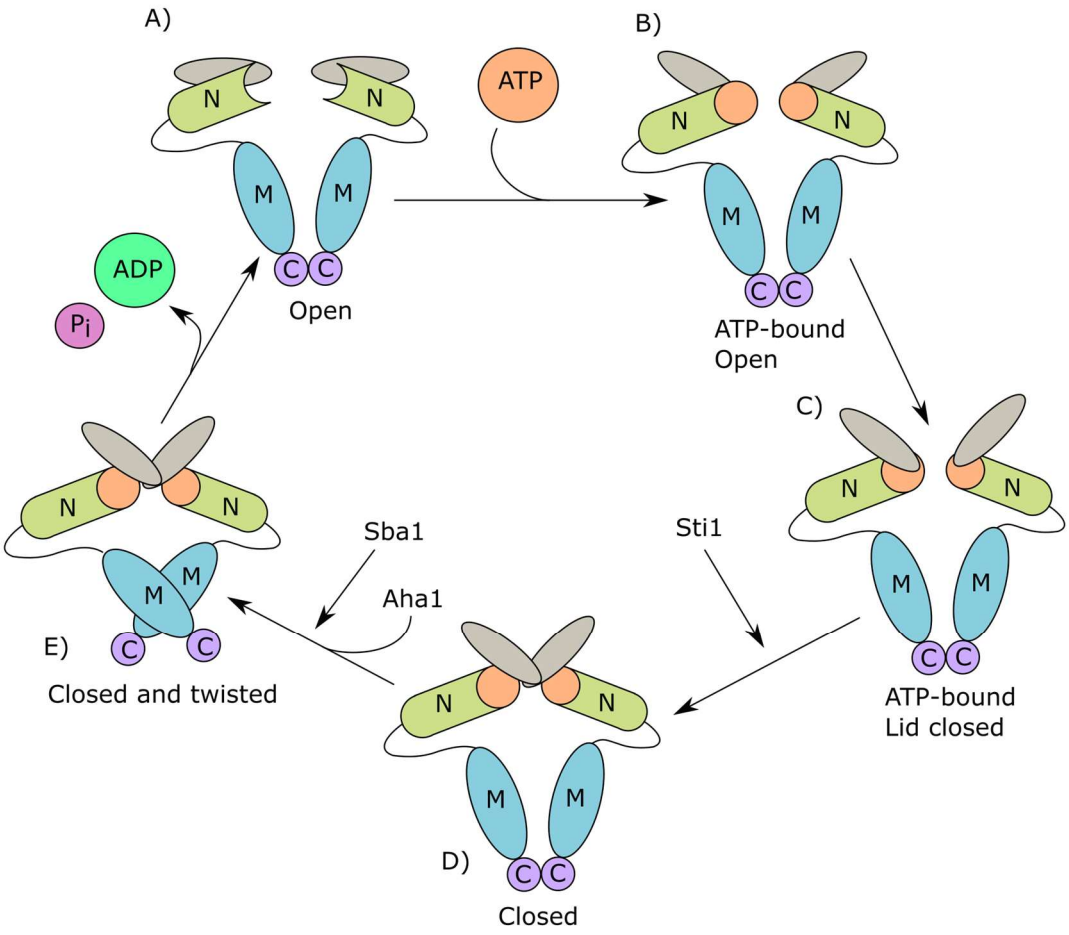



**Figure 1.6) Folding cycle of Hsp90 chaperones**

Starting with an open conformation (A), Hsp90 is dimerised at the *C*-terminal of two monomers. Upon binding ATP, the *N*-terminal domains fold over (B), locking ATP into a binding pocket and forming a closed lid structure (C). The two *N*-terminal are then able to dimerise, twisting and compacting the Hsp90 structure about the middle domain, activating the ATPase activity of the protein (D). Upon completion of folding into the closed and twisted state (E), ATP hydrolysis may be completed, allowing ADP to be released and the protein to revert to the open conformation. Various co-chaperones regulate this cycle at distinct stages. Adapted from publication (Zuehlke and Johnson, 2010).

Co-chaperones of Hsp70 and Hsp90 are also able to escort terminally misfolded proteins to the autophagy or ubiquitin-proteasome degradation pathways (Arndt et al., 2007). Proteins destined for the ubiquitin-proteasome degradation pathway are modified by an attachment of a chain of ubiquitin moieties via the E2 ubiquitin-conjugating enzyme and E3 ubiquitin ligase, with the ubiquitin chains transferred onto lysine residues of the substrate protein (Arndt et al., 2007). A key player in substrate labelling is the E3 ubiquitin ligase - the *C*-terminus of Hsp70-interacting protein (CHIP) which is able to associate with both Hsp70 and Hsp90 through its *N*-terminal TPR (Ballinger et al., 1999, Connell et al., 2001). CHIP possesses a U-box that allows for it to interact with ubiquitin-conjugating enzymes and allows ubiquitylation of substrates for the degradation pathway. In this way, the Hsp70/Hsp90 chaperone acts as the substrate recognition factor, with CHIP able to transiently interact with the substrate - it has been shown to bind substrates with low affinity independently of Hsp70 and Hsp90 chaperones (Demand et al., 2001). Recruitment of the co-chaperone CHIP therefore switches the chaperone pathway from that of protein-folding to protein-degradation (Arndt et al., 2007). The balance between recruitment of pro-degradation CHIP and other pro-folding chaperones allows for regulation of whether a protein is folded or degraded. Muller and colleagues (Muller et al., 2013) demonstrated that differential binding of CHIP or Hop (the mammalian homolog for Sti1) to Hsp90 is dependent upon its *C*-terminal phosphorylation, regulated by various kinases (including Casein Kinase 1, Casein Kinase 2 and Glycogen Synthase 3β). When Hsp90 is phosphorylated, binding to CHIP is prevented, enhancing binding to Hop and a pro-folding pathway. Dephosphorylated Hsp90 at the *C*-terminus is pro-degradation, binding CHIP and thus ubiquitylating the substrate protein. Murata and colleagues (Murata et al., 2001) demonstrated *in vivo* ubiquitylation of thermally denatured proteins captured by Hsp90. CHIP is also able to ubiquitylate Hsp70 in response to heat shock in order to reduce its high levels following stress, as part of the cellular recovery (Qian et al., 2006).

### 1.2.6 PFD chaperones

The PFD (also known as the GimC) complex is a ~90 kDa complex consisting of six subunits. In archaea, PFD is an oligomer of two proteins: two identical α subunits and four identical β subunits (Siegert et al., 2000). However, in eukaryotes PFD is an oligomer of six different proteins: two α-like subunits and four β-like subunits (Vainberg et al., 1998). PFD is localised to the cytosol, with the *S. cerevisiae* subunits termed Yke2, Gim3, Gim4, Gim5, Pac10 and Pfd1 (Leroux et al., 1999, Martin-Benito et al., 2002). Martin-Benito and colleagues (Martin-Benito et al., 2002) demonstrated via electron microscopy that the eukaryotic structure of PFD is similar to that of archaeal PFD as determined by Siegert and colleagues (Siegert et al., 2000) via crystallisation (Figure 1.7): both PFD forms have a jellyfish-like structure with six 'tentacles' consisting of coiled coil regions, with the body consisting of a double β-barrel (Martin-Benito et al., 2002).

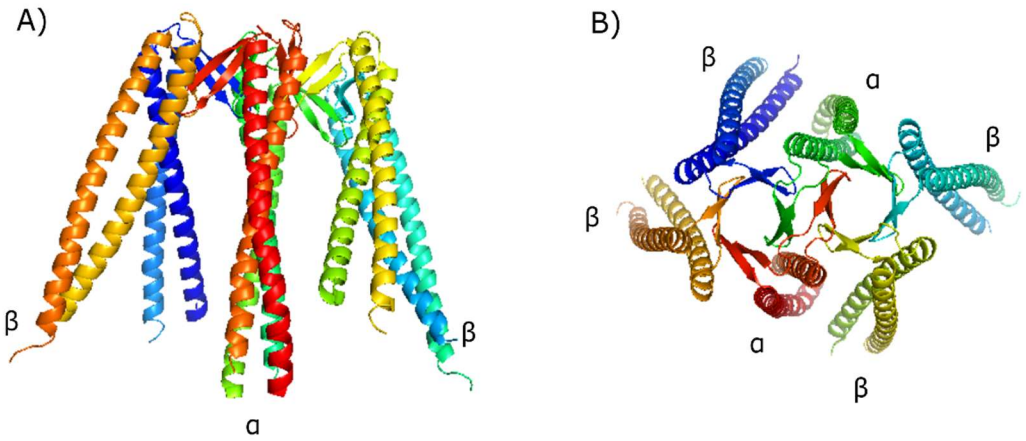



**Figure 1.7) Representation of the Archaeal PFD complex, a homolog of the *S. cerevisiae* PFD**

A) Front view of the PFD complex, in which each α subunit is flanked by two β subunits. The double β hairpin in the α subunit is a contact point for dimerisation giving rise to a jellyfish-like structure with the six subunits forming tentacle-like extrusions from two β-barrels; B) Top view of A showing the six subunits. PDB ID: 1FXK (Siegert et al., 2000).

Vainberg and colleagues (Vainberg et al., 1998) first identified the hetero-oligomeric PFD complex via its association with unfolded β-actin. Disruption of *S. cerevisiae* genes encoding for PFD resulted in actin and tubulin defects, also observed in temperature-sensitive defects in the CCT chaperonin, demonstrating that PFD plays a major role in the cytoskeletal biogenesis pathways. In *S. cerevisiae*, PFD can specifically bind both actin and tubulin, components of the cytoskeleton and microtubules respectively, as a result of specific sequences in the tips of the chaperones' coiled coil regions (Martin-Benito et al., 2002). PFD acts by binding unfolded substrate protein, preventing aggregation before transfer to the CCT complex in a nucleotide-

dependent manner (Vainberg et al., 1998), with actin-folding accelerated at least five-fold in the presence of PFD (Siegers et al., 1999). Martin-Benito and colleagues (Martin-Benito et al., 2002) resolved the structure of this interaction in *S. cerevisiae*, in which PFD can bind to either one or both rings of CCT, interacting with subunits one and four in each ring. The same group also solved the PFD-actin interaction, with the unfolded actin located along the interior of PFD, bound to the tips of the PFD tentacles. Thus, the tentacles act in a way to protect and encapsulate the unfolded protein and deliver actin/tubulin to CCT via a docking-type mechanism. Whilst actin requires only the CCT complex and can reach the native state in an ATP-dependent manner, folding of tubulin to the native state requires further cofactors that act on the folding intermediate released by CCT (Tian et al., 1997).

### 1.2.7 AAA+ chaperones

The AAA+ family of chaperones act to mediate the unfolding and refolding of proteins in response to aggregation, and disassembly of proteins for delivery to the proteasomes, through the degradation pathway (Neuwald et al., 1999). In this way, the AAA+ family can regulate the activities of protein complexes by mediating the degradation or availability of specific components required for function (Confalonieri and Duguet, 1995, Patel and Latterich, 1998). So far, only three members of this family have been identified in *S. cerevisiae*: Mcx1, Hsp78 and Hsp104 (Rottgers et al., 2002). Hsp104 is located within the cytosol, whilst Hsp78 and Mcx1 are located within the mitochondrial matrix (Leonhardt et al., 1993, Sanchez and Lindquist, 1990).

The AAA+ chaperone Hsp104's main function is in mediating the disassembly of protein complexes or aggregates in the cytosol. In *S. cerevisiae*, Hsp104 co-operates with Hsp70 in mediating thermotolerance, in that it can eliminate stress-induced protein aggregates, with an important role in disaggregating the amyloid formed by prions (Glover and Tkach, 2001, Parsell et al., 1994, Chernoff et al., 1995). Glover and Lindquist (Glover and Lindquist, 1998) demonstrated that Hsp104 co-operated *in vivo* with Hsp70 to solubilise aggregated proteins, which are then refolded to their native states by Hsp70.

Hsp104 contains two nucleotide-binding domains (NBD1 and NBD2). NBD1 provides the main hydrolytic activity of Hsp104, although binds ATP with a lower affinity than NBD2. Upon ATP binding to NBD2, the activity of NBD1 is increased in a co-operative manner, and formation of the hexameric complex of Hsp104 is promoted (Wendler et al., 2009, Hattendorf and Lindquist, 2002, Doyle et al., 2007, Schirmer et al., 1998, Schaupp et al., 2007). When ATP is bound to NBD1, the substrate protein is able to bind to Hsp104 (Lum et al., 2004). ATP hydrolysis by NBD1 whilst the non-native protein is bound initiates a conformational change in each subunit, in which the *N*- and *C*-terminal domains are brought closer together, narrowing the central pore and providing a 'crowbar' mechanism to pull apart the protein aggregate (Bochtler et al., 2000, Sousa et al., 2000, Wang et al., 2001a, Wang et al., 2001b). This motion enables *N*- to *C*- terminal threading of the protein aggregate through the central pore, disaggregating the protein (Figure 1.8) (Yokom et al., 2016).

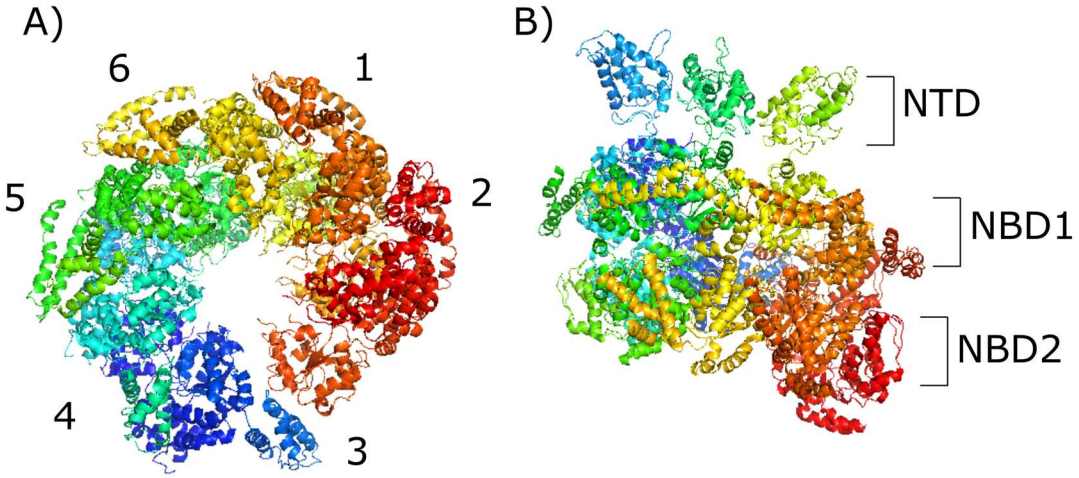



**Figure 1.8) Hexameric *S. cerevisiae* Hsp104 determined by cryo-electron microscopy**

A) Top-Down view of Hsp104 demonstrating the central pore, with each subunit numbered one to six. B) Side view of A, with N-terminal domain (NTD) and two nucleotide binding domains highlighted. PDB ID: 5KNE (Yokom et al., 2016).

### 1.2.8 SMALL chaperones

Small Hsps (sHsps) are ubiquitous and conserved chaperones with a molecular mass between 12 and 43 kDa forming large complexes up to around 900 kDa in order to prevent protein aggregation (Jakob et al., 1993). The number of sHsps can vary from species to species, each with differing sequences and size (Stromer et al., 2003). However, sHsps share characteristic features: a conserved α-crystallin domain of ~90 amino acids; a small molecular mass between 15 and 30 kDa; formation of large complexes; a dynamic quaternary structure

and are induced as a result of stress and chaperone activity in containing aggregation (Haslbeck et al., 2005). Unique to the sHsps, each subunit of the oligomer is able to bind a target protein, thus acting as a 'holdase' to prevent irreversible aggregations in a nucleotide-independent cycle.

The ability for sHsps to form large oligomeric proteins is highly conserved, with the structures of five sHsps being determined by X-ray crystallography or electron microscopy. Hsp26 from *S. cerevisiae*, Hsp16.5 from *Methanocaldococcus jannaschii* and Hsp16.3 from *Myobacterium tuberculosis* all form hollow sphere-like structures containing 24, 24 and 12 subunits respectively (Haslbeck et al., 1999, Kim et al., 1998, Kennaway et al., 2005). Although there is a high degree of sequence diversity, the highest of all the chaperone classes, all typical sHsps have conserved structural domains: the *N*-terminal region of variable length, followed by the α-crystallin domain (a conserved sequence of ~80 amino acids) and a short *C*-terminal region (Narberhaus, 2002).

In *S. cerevisiae*, sHsps such as Hsp26 and Hsp42 are able to bind to unfolded proteins, preventing their aggregation. Under normal conditions, Hsp26 exists as a large inactive oligomer of 12 dimers. Upon heat shock, conformational changes are induced that allow binding of the substrate to each dimer of the Hsp26 complex to suppress heat induced aggregation. It can also reversibly dissociate into the twelve dimers exposing hydrophobic regions, each capable of binding a single substrate protein non-specifically thus preventing their aggregation, which are able to reform the active oligomer with the substrates bound (Haslbeck et al., 1999, Franzmann et al., 2005). The substrate may dissociate spontaneously or will fold into the native state with the help of other chaperones, for example, Hsp70s. Upon restoration of physiological conditions, the Hsp26 will form the inactive complex once again (Haslbeck et al., 1999).

In contrast, the Hsp42 monomer is ~43 kDa and may be expressed constitutively (Wotton et al., 1996) and in response to stress (Gasch et al., 2001), with a similar *C*-terminal domain to other sHsps, but an unusually long *N*-terminal domain with no sequence homology to other sHsps (Haslbeck et al., 2004). Haslbeck and colleagues (Haslbeck et al., 2004) performed structural and functional assays on Hsp42, observing its existence in low concentrations as a 12-16 subunit oligomer, but in high concentrations as a 24-26 subunit oligomer; they also reported Hsp42 to be five to ten times more abundant than Hsp26 in the *S. cerevisiae* cytosol, suggesting it is the general sHsp in *S. cerevisiae* (Haslbeck et al., 2004).

Of the sHsps as classified by Gong and colleagues (Gong et al., 2009), only Hsp42 and Hsp26 are structurally and functionally similar and so are considered 'typical' sHsps. With regards to the remaining sHsps in this class: Hsp12, Hsp31, Hsp32, Hsp33 and Sno4; all are low molecular weight and do not share the conserved α-crystallin domain. As an example, Hsp12 exists as a monomer and exhibits low sequence homology to the sHsp superfamily with a protective function under stress conditions, particularly during heat shock (Welker et al., 2010). Hsp12 is natively unfolded and exists in soluble and membrane-associated forms, with the membrane-associated form levels enhanced during stress and crucial for its protective function. According to Welker and colleagues (Welker et al., 2010), Hsp12 binds to lipids with a negatively charged head group, with the interactions of Hsp12 with membrane lipids influencing the organisation and stability of the membranes. Hsp31, Hsp32, Hsp33 and Sno4 belong to the highly conserved DJ-1 superfamily, including the human DJ-1 involved in Parkinson's disease and cancer. Despite little known about their function, expression is induced in response to diauxic shift, oxidative stress and are required for survival in the stationary phase with impairment of autophagy induction when deleted, indicating potential chaperone roles for these proteins (Miller-Fleming et al., 2014).

Table 1.2 summarises the key properties for each of the 63 known chaperones in *S. cerevisiae* as considered by Gong and colleagues (Gong et al., 2009). The chaperone network is a complex and tightly regulated tool, required for protein homeostasis under both normal and stressed conditions. Following the observation that select chaperones are upregulated in response to stress in order to repair misfolding and prevent protein aggregation, observing the chaperones in response to various stresses has led to a better understanding of the environmental stress response. Additional to the typical chaperone response program, the cell also has a protein quality control feature within the ER, such that upon increased misfolded protein in the ER lumen, the unfolded protein response (UPR) can be activated.

**Table 1.2) Classification and Properties of known chaperones in *S. cerevisiae***

For the 63 known chaperones in *S. cerevisiae* according to (Gong et al., 2009), a summarisation of the in-text information is supplied. The location of each chaperone is as defined by Gong and colleagues (Gong et al., 2009), with the molecular weight as defined in the Saccharomyces Genome Database (available at www.yeastgenome.org). If the chaperone is an oligomer, the subunit molecular weight is given.

| Number | Chaperone | Class | Location | Size (kDa) | Structure | Regulation | Recognises |
|---|---|---|---|---|---|---|---|
| 1 | Ssa1 | Hsp70 | Cytosol | 69.6 | 5' – Nucleotide Binding Domain, Hydrophobic linker region, Substrate Binding Domain - 3' | ATP, co-chaperones | hydrophobic segments |
| 2 | Ssa2 | Hsp70 | Cytosol/Nucleus | 69.4 | | | |
| 3 | Ssa3 | Hsp70 | Cytosol | 70.5 | | | |
| 4 | Ssa4 | Hsp70 | Cytosol | 69.6 | | | |
| 5 | Ssb1 | Hsp70 | Cytosol | 66.6 | | | |
| 6 | Ssb2 | Hsp70 | Cytosol | 66.6 | | | |
| 7 | Sse1 | Hsp70 | Cytosol | 77.3 | | | |
| 8 | Sse2 | Hsp70 | Cytosol | 77.6 | | | |
| 9 | Ssz1 | Hsp70 | Cytosol | 58.2 | | | |
| 10 | Ssc1 | Hsp70 | Mitochondria | 70.6 | | | |
| 11 | Ssc2 | Hsp70 | Mitochondria | 72.4 | | | |
| 12 | Ssc3 | Hsp70 | Mitochondria | 70.1 | | | |
| 13 | Kar2 | Hsp70 | Endoplasmic Reticulum | 74.4 | | | |
| 14 | Lhs1 | Hsp70 | Endoplasmic Reticulum | 99.5 | | | |
| 15 | Apj1 | Hsp40 | Cytosol | 58.8 | Class I, Class II or Class III J protein | ATP binding to Hsp70 required for Hsp40 client transfer | Promiscuous, selective or non-client binding |
| 16 | Cwc23 | Hsp40 | Cytosol/Nucleus | 33.2 | | | |
| 17 | Djp1 | Hsp40 | Cytosol | 48.6 | | | |
| 18 | Jjj1 | Hsp40 | Cytosol | 68.8 | | | |
| 19 | Jjj2 | Hsp40 | Cytosol/Nucleus | 67.4 | | | |
| 20 | Jjj3 | Hsp40 | Cytosol/Nucleus | 20.0 | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 21 | Sis1 | Hsp40 | Cytosol/Nucleus | 37.6 | | | |
| 22 | Swa2 | Hsp40 | Cytosol | 75.0 | | | |
| 23 | Xdj1 | Hsp40 | Cytosol/Nucleus | 51.3 | | | |
| 24 | Ydj1 | Hsp40 | Cytosol/Nucleus | 44.7 | | | |
| 25 | Zuo1 | Hsp40 | Cytosol | 49.0 | | | |
| 26 | Caj1 | Hsp40 | Nucleus | 44.8 | | | |
| 27 | Pam18 | Hsp40 | Mitochondria | 17.9 | | | |
| 28 | Jac1 | Hsp40 | Mitochondria | 21.8 | | | |
| 29 | Jid1 | Hsp40 | Mitochondria | 35.0 | | | |
| 30 | Mdj1 | Hsp40 | Mitochondria | 55.6 | | | |
| 31 | Mdj2 | Hsp40 | Mitochondria | 16.4 | | | |
| 32 | Erj5 | Hsp40 | Endoplasmic Reticulum | 34.2 | | | |
| 33 | Hlj1 | Hsp40 | Endoplasmic Reticulum | 25.0 | | | |
| 34 | Jem1 | Hsp40 | Endoplasmic Reticulum | 75.2 | | | |
| 35 | Sec63 | Hsp40 | Endoplasmic Reticulum | 75.3 | | | |
| 36 | Scj1 | Hsp40 | Endoplasmic Reticulum | 41.5 | | | |
| 37 | Tcp1 | CCT | Cytosol | 60.5 | Eight-membered double ring structure | ATP, Prefoldin (for actin and tubulin) | hydrophobic segments |
| 38 | Cct2 | CCT | Cytosol | 57.2 | | | |
| 39 | Cct3 | CCT | Cytosol | 58.8 | | | |
| 40 | Cct4 | CCT | Cytosol | 57.6 | | | |
| 41 | Cct5 | CCT | Cytosol | 61.9 | | | |
| 42 | Cct6 | CCT | Cytosol | 59.9 | | | |
| 43 | Cct7 | CCT | Cytosol | 59.7 | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 44 | Cct8 | CCT | Cytosol | 61.7 | | | |
| 45 | Hsc82 | Hsp90 | Cytosol | 80.9 | Dimer: N-terminal domain, middle and C-terminal domain | ATP, co-chaperones | hydrophobic segments |
| 46 | Hsp82 | Hsp90 | Cytosol | 81.4 | | | |
| 47 | Yke2 | PFD | Cytosol | 13.3 | 6 subunit oligomer, jelly-fish like structure | ATP-independent | Hydrophobic residues followed by Glutamic acid – Histidine – Glycine - Isoleucine sequence |
| 48 | Gim3 | PFD | Cytosol | 15.2 | | | |
| 49 | Gim4 | PFD | Cytosol | 13.0 | | | |
| 50 | Gim5 | PFD | Cytosol | 18.4 | | | |
| 51 | Pac10 | PFD | Cytosol | 23.1 | | | |
| 52 | Pfd1 | PFD | Cytosol | 12.8 | | | |
| 53 | Hsp104 | AAA+ | Cytosol | 102.0 | Conserved AAA domain | ATP | Aggregated proteins, sequence motif unknown |
| 54 | Hsp78 | AAA+ | Mitochondria | 91.3 | | | |
| 55 | Mcx1 | AAA+ | Mitochondria | 57.9 | | | |
| 56 | Hsp60 | HSP60 | Mitochondria | 60.7 | Two seven-membered rings | ATP, Hsp10 co-chaperone | hydrophobic segments |
| 57 | Hsp12 | SMALL | Cytosol/Nucleus | 11.7 | Monomer | membrane binding | negatively charged residues in lipid head |
| 58 | Hsp26 | SMALL | Cytosol/Nucleus | 23.9 | 24-subunit oligomer | substrate binding, activated in response to stress | hydrophobic segments |
| 59 | Hsp31 | SMALL | Cytosol/Nucleus | 25.7 | unknown | unknown | unknown |
| 60 | Hsp32 | SMALL | Cytosol | 25.9 | unknown | unknown | unknown |
| 61 | Hsp33 | SMALL | Cytosol | 25.9 | unknown | unknown | unknown |
| 62 | Hsp42 | SMALL | Cytosol | 42.8 | 12-16 subunit oligomer, 24-26 subunit oligomer | constitutively active | hydrophobic segments |
| 63 | Sno4 | SMALL | Cytosol | 26.0 | unknown | unknown | unknown |

## 1.3 Response to heat shock

Organisms are constantly challenged by their ever-changing environment, including varying nutrient levels, changes in temperature, and exposure to toxic molecules. Whilst multi-cellular organisms are usually able to alter these conditions by a change in location or physiology, single celled *S. cerevisiae* must be able to adapt or perish. The most common stress experienced by *S. cerevisiae* is temperature. Exhibiting optimal growth between 25 °C and 30 °C, temperatures above 36 - 37 °C cause *S. cerevisiae* cells to activate their protective transcriptional program, the HSR. In eukaryotes, the primary modulator of the HSR is the HSF protein family (so is focussed upon here), with a second transcription factor encoded by the *msn2* and *msn4* genes also contributing significantly to environmental response gene expression (Morano et al., 2012). Both *hsf1* and *msn2/4* gene expression is activated in response to a variety of stress conditions.

Whilst four Hsf isoforms exist in mammals, *S. cerevisiae* exhibits a single HSF equivalent to the mammalian Hsf1 (Akerfelt et al., 2010). *S. cerevisiae* Hsf1 is able to regulate genes under physiological conditions as well as under stress conditions and is essential for cell viability. It is inducibly phosphorylated with trimerisation activating its DNA-binding ability, but may be negatively controlled by the cAMP-dependent kinase via phosphorylation at Hsf1's *C*-terminal regulatory domain (Hashikawa et al., 2006). Hsf1 recognises a pentameric heat shock element (HSE), defined as three repeating units of the sequence 5'-nGAAn-3' (Morano et al., 2012). The number of bases separating each of the repeating units gives rise to three unique HSE types: the perfect (three contiguous inverted nGAAn units); the step (with each nGAAn unit separated by five base pairs); and the gap (two inverted nGAAn units separated from a third by a five base pair gap); each requiring different Hsf1 binding and activation behaviours (Yamamoto et al., 2005, Hashikawa et al., 2006). 'Perfect' HSEs consist of three or more contiguous repeats of the motif and may be induced under normal conditions by Hsf1. Under heat shock conditions, Hsf1 becomes extensively phosphorylated, with a stronger activating ability, allowing for binding to 'step' and 'gap' type HSEs in addition to 'perfect' type HSEs (Hashikawa et al., 2006). 'Gap' HSEs contains a 5-bp gap separating two contiguous motifs from a third, whilst 'step' HSEs contain 5-bp gaps separating each of the three motifs, with the spacer regions allowing for co-operative binding of Hsf1. With the ability to bind to multiple HSE organisations and activate the expression of genes both constitutively and in response to stress, Hsf1 targets are widespread in the *S. cerevisiae* genome, functioning in a broad range of biological processes including protein folding and degradation, detoxification, energy generation, carbohydrate metabolism

and cell wall organisation and are induced in response to a variety of stresses including heat shock, oxidative stress and starvation (Yamamoto et al., 2005, Hashikawa et al., 2006).

In contrast, Msn2/4 is able to bind the stress response element (STRE), identified first through analysis of the cystolic catalase gene (CTT1), with the consensus sequence 5'-TAAGGG-3' (Wieser et al., 1991). Binding of Msn2/4 to STRE is required for the expression of a wide array of genes in response to multiple types of stress (Martinez-Pastor et al., 1996). Together, the Hsf1 and Msn2/4 transcription factors are able to regulate expression of a wide range of target genes in order to elicit the appropriate stress response to the change in conditions.

Heat stress is described as a change in temperature to above 36 – 37 °C, whilst heat shock is considered a change in temperature to above 42 °C. Cells are able to maintain their growth up to temperatures around 41 °C, with cell growth diminished above 42 °C (Yamamoto et al., 2008). Alongside activating a transcriptional response program, various physiological changes occur. At temperatures of 42 °C and above, bulk poly(A)+ RNA becomes stably accumulated in the nucleus, preventing it from being able to be exported for transcription at the ribosome (Saavedra et al., 1996). In contrast, Rip1, a nuclear export factor required for export of stress induced Hsps, is upregulated, enabling their increased export from the nucleus. Another stress response by the cell is the production of processing bodies (P bodies) and stress granules (SGs) which concentrate non-translating mRNA in sequestered pools in the cytoplasm. Under heat shock conditions, these SGs contain translation initiation factors and non-heat shock mRNAs, able to redistribute in the cell following recovery (Parker and Sheth, 2007, Grousl et al., 2009). Consequently, the cell is able to restrict translation of non-heat shock proteins via two mechanisms: blocking mRNA export and redirecting and trapping mRNA in the cytosol away from the ribosome into subcellular complexes (Morano et al., 2012). This allows the cell to focus translation on heat responsive and protective genes only.

In addition to these physiological changes, cells respond to heat shock by dramatically altering their genes expression. As described by Morano and colleagues (Morano et al., 2012), a model for Hsf1 activation includes the HSP chaperones. Experiments involving the forced misfolding of nascent proteins by the proline analog azetidine 2-carboxylic acid resulted in the arrest of the cell cycle at G1, in a manner similar to heat shock (Morano et al., 2012, Rowley et al., 1993). Forced misfolding also resulted in the transcriptional activation of genes closely matching the Hsf1 but not the *msn2/4* regulon (Morano et al., 2012, Trotter et al., 2002). As heat stress at 37 °C does not result in bulk protein aggregation (Nathan et al., 1997), these

experiments suggest the misfolding of newly synthesised polypeptides is sensed due to an increased load on the chaperone proteins. Monomeric Hsf1 is able to interact with Hsp90 under normal conditions, with Hsp90 preventing Hsf1 trimerisation and activation of its DNA binding abilities; therefore, a common activation model for Hsf1 is through liberation of Hsp90 due to increased misfolded protein levels (Anckar and Sistonen, 2011, Zou et al., 1998). However, trimeric Hsf1 is also able to bind Hsp90 via its *C*-terminal modulator domain in heat shocked cells; potentially playing a role in repressing (and attenuating) Hsf1's DNA-binding abilities in response to heat shock (Zou et al., 1998). As such, it is likely that Hsp90 also plays a role in repressing the trimeric Hsf1 as well as regulating the DNA-binding activity (Anckar and Sistonen, 2011). Hsf1 is also able to interact with Hsp70 and its co-chaperone Hsp40 with interactions correlating with increased amounts of Hsp70 (Shi et al., 1998). Binding to Hsp70 does not repress DNA-binding but results in the inhibition of Hsf1 trans-activating capacity, acting as a negative feedback mechanism to ensure that Hsf1 activity is coordinated according to the expression of its target genes (Anckar and Sistonen, 2011). The HSR can also be activated by the cell wall integrity pathway; elevated temperature increases the fluidity of *S. cerevisiae* cell membranes, translating into specific activation of heat-sensing $Ca^{2+}$ channels and a downstream signalling cascade through the Pkc1-MAP kinase pathway resulting in the activation of Hsf1 through hyperphosphorylation (Bromberg et al., 2013, Balogh et al., 2005).

Gene expression abundance profiling has demonstrated that ~10 % of the genome is remodelled in response to stress, with the abundance of ~600 genes decreased, whilst ~300 are induced, in what is described as the environmental stress response (ESR) (Gasch and Werner-Washburne, 2002, Causton et al., 2001, Morano et al., 2012). The HSR is a subset of ESR, and is composed of genes requiring the transcription factors Hsf1, Msn2/4 or both in order to activate their expression during heat shock. Previous studies have demonstrated ~72 genes that are induced upon heat shock at 39 °C (Yamamoto et al., 2008), whilst there are a possible 165 Hsf1 targets containing a HSE (Hahn et al., 2004), indicative of many non-stress gene targets.

The magnitude of the stress response increases with the intensity of the stress. Gasch and Werner-Washburne (Gasch and Werner-Washburne, 2002) observed a longer lasting HSR with greater changes in gene expression in a temperature shift from 25 to 37 °C when compared to a temperature shift from 29 to 33 °C. However, a threshold point exists at which the maximal HSR is achieved; there is no significant difference between the binding of Hsf1 to chromosomal loci between 39 °C and 42 °C (Hahn et al., 2004), suggesting maximal activation of the HSR is ~40 °C, with cellular growth being maintained up to 42 °C. The HSR typically involves the

induction of HSPs and their cofactors to prevent or restabilise misfolded proteins, also including the activation of gene targets involved in oxidant defence, cell wall remodelling and transport (Hahn et al., 2004). Biosynthesis of the glucose disaccharide trehalose is upregulated in response to various stress conditions, particularly under heat shock, with trehalose observed to stabilise the tertiary structure of the *C*-terminal activation domain of Hsf1 allowing for maximal transcriptional activity (Crowe, 2007, Singer and Lindquist, 1998). During stress, proteins denature and adopt partially folded conformations that are prone to aggregation. Trehalose functions against this process by stabilising the proteins in their native states. Proteins that denature are bound by chaperones (HSPs) to prevent aggregation and promote refolding. Trehalose is able to function in this context by reducing aggregation of non-native proteins whilst the chaperones are occupied. Upon degradation of trehalose, the chaperones are able to continue folding the remaining non-native proteins and complete cellular recovery (Singer and Lindquist, 1998).

Elevated temperature represents the primary insult during heat shock. A secondary consequence is the production of reactive oxygen species (ROS), toxic agents that are able to damage a variety of cellular components resulting in lipid peroxidation, protein oxidation and DNA modification through the resulting oxidative stress (Morano et al., 2012). Mitochondrial respiration is considered the primary source of ROS via the process of oxidative phosphorylation (Murphy, 2009). During ATP generation, electrons are passed along the electron transport chain to molecular oxygen, forming water. Leakage of these electrons from the respiratory chain can result in the reduction of oxygen, generating the ROS. As oxygen is also the terminal acceptor in the ER during oxidative protein folding, the ER is a secondary source of ROS (Tu and Weissman, 2004).

The inherent quantities of expressed proteins within a cell can define various functional stages of the cell. When stress is encountered, in order for the cell to survive, the abundance of protective proteins, in particular chaperones as discussed above, must increase in order to aid both cell recovery and cell survival. In order to enable understanding of the role of chaperones in response to stress, it is paramount to be able to detect the quantities of individual chaperones with high sensitivity and accuracy. Experimental methods used for quantifying proteins are discussed in the next section.

## 1.4 Methods for quantification of proteins in cell samples

Previous proteomic and transcriptomic studies have characterised chaperone upregulation in response to various stress conditions. Proteomic studies have typically used stable-isotope labelling of amino acids in cell culture (SILAC) approaches, pulse labelling with $^{35}$S-methionine and semi-quantitative western blots to measure the proteome directly, whilst northern blots and DNA microarrays have inferred transcriptome changes (Jun et al., 2012, Causton et al., 2001, O'Connell et al., 2014, Finka et al., 2015, Piper et al., 1994, Jarnuczak et al., 2015, Ghaemmaghami et al., 2003, Shui et al., 2015). Indeed, global transcriptomic studies by both Causton and colleagues and Gasch and colleagues (Causton et al., 2001, Gasch and Werner-Washburne, 2002) observed significantly changing gene expression levels, and has been utilised as a major indicator of changing protein abundances in response to environmental stress. However, it is widely reported that correlation between the absolute copy per cell value of a protein and its respective mRNA expression level is modest (Vogel and Marcotte, 2012), with post-transcriptional, post-translational and protein degradation rates altering protein expression levels. In contrast, proteomics analyses have generally used relative quantification rather than defining changes in absolute protein levels. Current quantitative methods in proteomics include both gel-based and MS-based applications, including relative and absolute quantification strategies involving the use of enzymatic, chemical and metabolic labelling and label free approaches (Figure 1.9) (Nikolov et al., 2012). Relative quantification workflows involve comparing the amount of proteins or whole proteomes between samples and yielding a quantitative ratio or relative change whilst absolute quantification workflows provide detail about the absolute amount or the concentration of a protein within the sample.

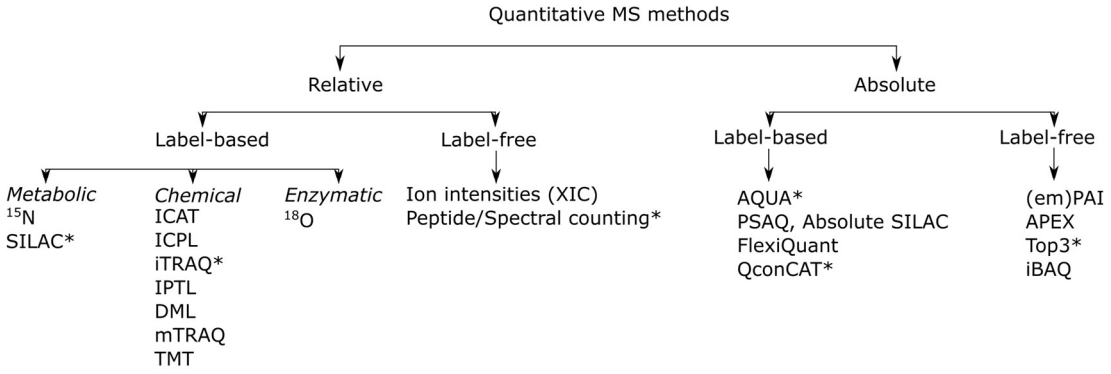



**Figure 1.9) Quantitative MS methods**

MS strategies for quantitative proteomics may be label based or label free and relative or absolute in nature. Those with an asterix are further discussed in this chapter. Abbreviations are as follows: SILAC – stable isotope labelling of amino acids in cell culture; ICAT – isotope-coded affinity tag; ICPL – isotope-coded protein label; iTRAQ – isobaric tag for relative and absolute quantification; IPTL - isobaric peptide termini labelling; DML – dimethyl labelling; mTRAQ – mass differential tags for relative and absolute quantification; TMT – tandem mass tagging; AQUA – peptides for absolute quantification; PSAQ – protein standard absolute quantification; (em)PAI – exponentially modified protein abundance index; APEX – absolute protein expression; iBAQ – intensity based absolute quantification. Adapted from publication (Nikolov et al., 2012).

### 1.4.1 Gel-based protein quantification techniques

Epitope tagging strategies may be used in order to relatively or absolutely quantify proteins. Such epitope tags include green fluorescent protein and the tandem affinity protein tag, allowing for both purification of the proteins of interest and detection. In these strategies, collections of modified strains are generated that incorporate a quantifiable tag that can be detected via immunoblotting or fluorescence. In work by Ghaemmaghami and colleagues (Ghaemmaghami et al., 2003) proteins were tagged with the tandem affinity purification (TAP) tag and subject to western blotting using an antibody against the calmodulin binding peptide present within the TAP tag. Image analysis is used for the relative intensity of each protein to be measured, giving an indicator of the relative protein abundance in each sample. Using standards of known concentration also bound to a TAP tag, the absolute concentration of each protein could be determined. However, such epitope strategies do not quantify the endogenous protein as the host is genetically manipulated and relies heavily on the specificity of antibodies used to detect the protein of interest, making the use of MS strategies a much more attractive concept.

Some gel-based techniques may be co-ordinated with MS to identify the protein band of interest, with one approach used being two-dimensional polyacrylamide gel electrophoresis

(2D-PAGE) (Figure 1.10). In the first dimension, proteins are separated by their isoelectric point, whilst in the second dimension they are separated by their electrophoretic mobility with sodium dodecyl sulphate (SDS). Combining these two separation techniques allows for separation of up to 10,000 protein species, therefore allowing a global quantitative proteome analysis (Rabilloud, 2012). Protein detection and quantification in a 2D-gel is achieved through the use of visible or fluorescent dyes, commonly the Coomassie or silver stain techniques, in which the signal intensity of the dye corresponds to the relative abundance of the protein. Due to the use of commercially available immobile pH gradients, 2D-PAGE is able to offer a robust technique towards sample proteome comparison, with protein identification achieved through MS. In order to identify bands of interest to excise, comparative image analysis is performed during which data is acquired, spots are detected and relatively quantified through the use of image processing software, and then gel matching is performed to identify those significantly changing in response to various biological stresses or events. However, determination of significantly changing proteins is victim to the problem of false positives, as variation is introduced via biological, technical and sample variability. Despite statistical tools utilised (including the use of false discovery estimates), without the use of internal standards, it is difficult to ascertain whether a signal change is due to variability as a result of sample variability, staining linearity and homogeneity or is a true reflection of a changing protein abundance. The variation can be reduced using two-dimensional difference in-gel electrophoresis (2D-DIGE) in which up to three different protein samples are labelled covalently with different spectrally resolvable fluorescent dyes before being simultaneously separated in a single gel (Figure 1.10b) (Unlu et al., 1997). In this way, technical variations as a result of staining homogeneity are minimised, and the use of a pooled internal standard (typically a mixture of the same amount of all samples included in the study) as a reference allows for comparison to different gels, improving the reproducibility and statistical analysis (Alban et al., 2003).. Finally, upon identification of spots of interest within the gel, protein spots are excised and subject to in-gel proteolytic digestion prior to a mass spectrometric experiment in which spectra are mapped to peptides and thus the parent protein. 2D-PAGE also offers the ability to identify and relatively quantify protein isoforms and proteins with post translational modifications as the use of non-denaturing gels allows for separation of the complete proteins. However, a major limitation of the gel technique is that spots of interest may contain multiple proteins of similar molecular weight or isoelectric point and so are comigrated on the gel. Therefore, absolute quantification performed on this spot would be prone to error.

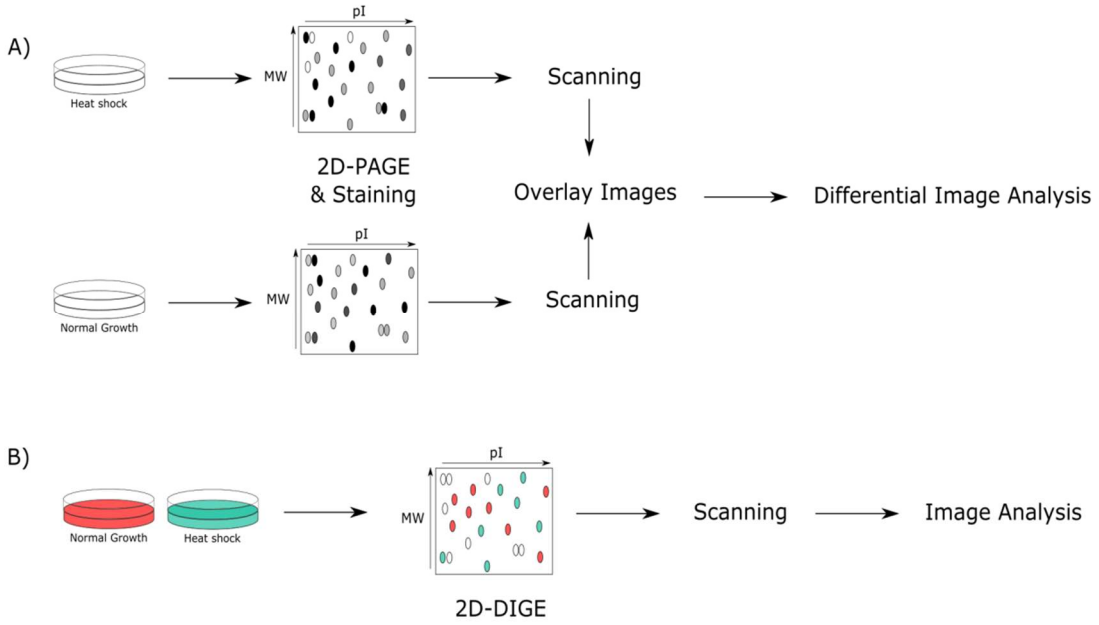



**Figure 1.10) Relative quantification via 2D-PAGE methods**

A) Unique samples are run on 2D-PAGE gels to separate proteins based on isoelectric point and molecular weight. After staining, gels are scanned, overlaid and subject to differential image analysis to identify differentially expressing proteins. The intensity of the stain should correlate linearly with the abundance of the protein, enabling for relative protein abundance determination. B) An extension of 2D-PAGE is 2D-DIGE, involving the use of multiple samples with fluorscent dyes unique to the sample attached to each protein. Samples are combined and ran on the same gel often with an internal standard to allow for relative quantification, and determination of differential protein expression. Typically, proteins with no abundance change appear as 'white' spots. Spots of interest from both methods may be excised and subject to digestion and MS for identification.

### 1.4.2 MS-based relative quantification strategies

Relative MS-based quantification workflows may utilise either label-free or labelled samples. Label-free methods can use peptide ion intensities or spectral counting methods, whilst label-based methods also include the use of metabolic, chemical or enzymatic labelling of proteins towards deducing relative abundances. MS measures the mass to charge ratio of ions, with the signal detected related to the number of ions hitting the detector (Nikolov et al., 2012, Washburn et al., 2001). Due to the differing physiochemical properties of peptides, their respective signals in a mass spectrometer are not comparable between differing molecular species. Instead, quantification is performed by comparing signals of the same species between different samples, or comparison of those that differ only in isotopic composition therefore having identical physiochemical properties (Wilm, 2009).

Label-free quantification relies heavily on reproducible sample handling, separation by liquid chromatography and intensity measurements, in order to be able to perform quantification of large numbers of samples without the requirement of any labelling modification (Nikolov et al., 2012). Label-mediated quantification offers a means of quality control between samples. Artificial incorporation of heavy isotopes (for example, $^{13}C$, $^{15}N$ and $^{18}O$) induces a mass shift of the peaks corresponding to the peptide and its isotopic pattern in a mass spectrum without altering the behaviour of peptides and proteins in the LC-MS system (Wilm, 2009). The isotopic intensity ratios of peaks corresponding to labelled and non-labelled samples reflects directly the relative abundance ratios between the same isotopic compositions (Nikolov et al., 2012). Label-free techniques include peptide/spectral counting and ion intensities, whilst label-based techniques include SILAC and the use of isobaric tag for relative and absolute quantification (iTRAQ) (Ong et al., 2002, Ross et al., 2004).

#### 1.4.2.1 Peptide/Spectral Counting & Ion intensities

Unlike 2D-PAGE that allows for fractionation of proteins in-gel prior to identification via MS, “shotgun” proteomics incorporates digesting complex protein samples to produce a collection of peptides for identification and quantification by MS. Typically, in a liquid chromatography tandem MS (LC-MS/MS) experiment, samples are digested to produce peptides, which are then subject to liquid chromatography to allow separation before being subject to ionisation to produce precursor ions and fragmentation, to produce product ions. Ions are searched against spectral libraries to enable peptide (and thus protein) identifications. However, particularly for complex samples, the number of co-eluting ions can exceed the number of ions for which tandem mass spectra can be acquired. Data acquisition is therefore biased towards the higher abundance ion signals and so higher abundance peptides are more routinely identified (Liu et al., 2004). To improve the chromatographic resolution and limit the level of co-elution that occurs, multi-dimensional liquid chromatography or high pressure liquid chromatography may be used (Washburn et al., 2001). In this data-dependent acquisition mode, more abundant proteins are identified by multiple peptides, whilst low abundant proteins are identified by only one or two (Washburn et al., 2001). Work by Liu and colleagues (Liu et al., 2004) demonstrated that the probability of a protein being identified is directly related to a proteins abundance, with the number of spectra acquired from a complex peptide mixture being the most invariant property of MS analysis. This probability can be improved if one normalises for the number of potentially observable peptides. This takes into account the fact that for the same number of molecules, larger proteins and proteins with many peptides

in the preferred mass range will generate more observed peptides, whilst smaller proteins may be identified by fewer peptides (Ishihama et al., 2005). The number of peptides observed can also be used for relative quantification. Exponentially modified Protein Abundance Index (emPAI) scores are calculated from the experimentally observed and the theoretically possible numbers of peptides in a protein mixture after digestion and MS analysis, with scores roughly proportional to the protein abundance (Ishihama et al., 2005). Therefore, relative protein abundance may be determined either by the number of peptides identified or the number of acquired MS/MS spectra for a protein. A relative comparison of different samples may be achieved by comparing these numbers without the requirement for labelling, however LC-MS experiments must be reproducible and remains less robust than label-mediated strategies. Quantification using spectral counting is severely limited by the normalisation and statistical evaluation (as described above) required in order to achieve accurate quantification. This accuracy is significantly decreased for proteins with only a few observable peptides, as well as when the quantitative changes between experiments are small. Instead, chromatographic intensities may be utilised towards relative protein abundance. In this way, comparing the signal intensity between different samples for an identical peptide demonstrates the relative protein changes between the two samples. However, the variation between measurements of the peak intensities of peptides from the same sample across technical replicates should be recorded and any normalisation applied; whilst any variation in retention time should also be considered in order to achieve optimal alignment between samples.

Absolute quantification at the protein level may be determined by averaging the peptide intensities or scores observed for each protein. emPAI scores may be used to determine protein contents in molar and weight fraction percentages, with scores determined by the number of theoretically possible and observed peptides for a protein (Ishihama et al., 2005). Alternatively, intensity-based absolute quantification (iBAQ) allows for absolute quantification of proteins using their respective peptide intensities. In this technique, the protein intensity is calculated as the sum of all identified peptide intensities for the protein, with the summed protein intensity divided by the number of theoretically observable peptides in order to account for protein length effects (Schwanhausser et al., 2011). The IBAQ intensities for proteins can be converted into absolute copy numbers by comparison with the iBAQ intensities of known protein standards.

Absolute amounts of individual proteins or protein classes can be determined using the proportion of their MS signals to the total MS signal (Wisniewski et al., 2012). These values can be transformed into protein copy numbers if the total number of cells in the sample is known, usually through cell counting. However, during cell counting a small sample is taken for visualisation and then scaled up according to the total volume of culture. If cells are not uniform in size, the scale up may result in incorrect counting. Indeed, according to Wisneiwski and colleagues (Wisniewski et al., 2014), "a 25% variation in the diameter of a sphere-shaped cell corresponds to a 2-fold change in protein volume". The amount of DNA in the sample is proportional to the number of cells. DNA is packaged into chromatin by histones, with the combined mass of histones equal to the mass of DNA. Subsequently, the mass of histones is proportional to the mass of DNA and therefore the number of cells. Therefore, the mass spectrometric signal of histones is used as a proxy for the number of cells without the need for cell counting and may be used towards determining absolute protein copy numbers. However, this approach is weakened in accuracy as the total MS signal depends on the depth of the proteomic analysis, with histones contributing some of the most intense peptides and so the fractional abundance of histones may be overestimated. Using fractionation approaches, the depth of proteomic analysis may be increased and therefore a more robust histone abundance determined. Secondly, a major pre-requisite for this approach is that the sample is whole-cell and eukaryotic, where the chromatin fraction is not over- or underrepresented due to sample handling prior to mass spectrometry analysis (Wisniewski et al., 2014).

#### 1.4.2.2 SILAC

In SILAC approaches, isotopically labelled analogs of essential amino acids instead of the natural abundance amino acid are added into the culture medium such that the labelled amino acid is incorporated into each newly synthesised protein chain. After a number of cell doublings, every instance of the particular amino acid will have been replaced by the isotopically labelled analog (Ong et al., 2002). As there is no physiochemical difference between the labelled and unlabelled amino acid, cells behave identically to the control population grown with the natural abundance amino acid. Observed peak ratios for isotopic analogs are highly accurate when determining the relative abundance of a peptide or protein as there are no chemical differences between the species and they are analysed in the same experiment.

Isotopically labelled cells may then be treated in a particular way, for instance subject to heat shock, and protein populations from both samples are harvested. As one sample is

isotopically labelled, proteins and peptides from that treatment can be categorically identified. Extracts can thus be mixed directly prior to MS while still being able to deduce provenance of the peptides. In this way, one ensures that any steps following extraction and mixing (proteolysis, liquid chromatography etc.) are identical between samples and therefore removes any between-sample variability, improving accuracy of the quantification (Figure 1.11). Although the labelling choice is dependent upon the aims of the experiment, commonly isotopic heavy-labelled analogs of lysine and arginine are used, so that in a tryptic digest, every peptide produced has (at least) a single label incorporated. However, when applied to an entire proteome, the financial cost of labelling is likely to be high. In addition, due to the metabolic labelling requirement, the number of peptides per experiment is doubled, resulting in an increase in spectral complexity (Merrill et al., 2014). Combining the SILAC and isobaric tagging methods, neutron encoding (NeuCode) SILAC relies on the mass defects of atoms and their isotopes. Mass defects may be induced with many elements and their isotopes, for example, $^{12}C$ for a $^{13}C$ atom produces a mass difference of + 3.3 mDa (Hebert et al., 2013). Neutron signatures are embedded via the use of selected amino acids during protein synthesis; instead of using heavy and light labelled peptides (traditional SILAC), NeuCode SILAC incorporates isotopologues of the heavy labelled peptides. As the NeuCode SILAC pairs are closely spaced in terms of their mass to charge ratio, both isotopologues are co-isolated, fragmented and mass analysed together to produce MS/MS spectra that are identical to non-multiplexed samples under non-resolution settings. Thus, the encoded signatures are concealed and the spectral matching is unaffected. Under high resolution however, the MS/MS fragments are separated according to the isotopologues, allowing for identification and quantification of the peptide (and sample). As a single precursor peak is observed containing all quantitative information using NeuCode SILAC, it avoids the known issues with traditional SILAC regarding sample complexity and peptide identifications (Hebert et al., 2013).

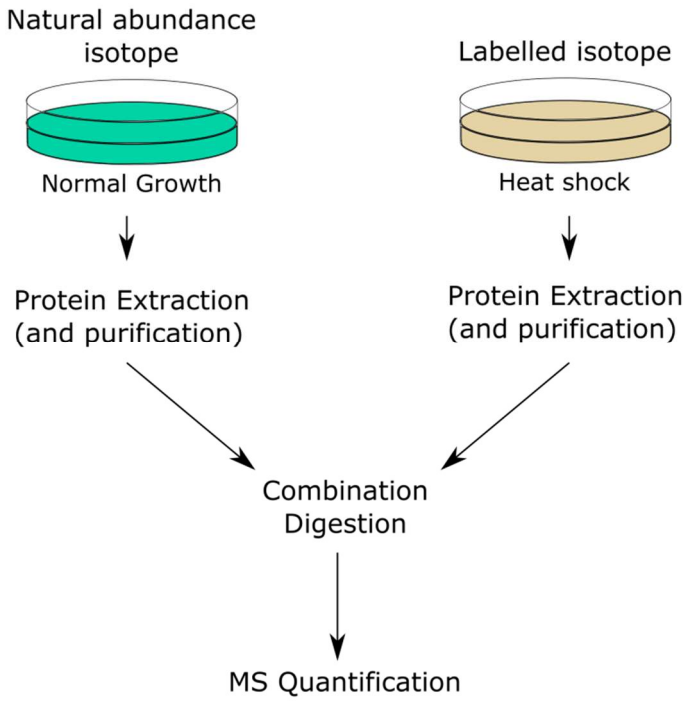



**Figure 1.11) Traditional SILAC Quantification strategy**

Isoptically labelled analogs of natural abundance amino acids are incorporated such that after a few rounds of cell doublings, complete labelling of proteins is achieved, with the sample identical in behaviour to the control. Stress (or other constraint) may be applied to the labelled sample, and following protein extraction, both samples may be combined, subject to proteolysis and quantified via mass spectrometry.

#### 1.4.2.3 **iTRAQ and TMT**

SILAC allows incorporation of a mass difference detectable within a mass spectrometer. However, SILAC requires metabolic labelling of samples, and so is restricted to small organisms and cell culture. Multiple datasets can be combined after separate analyses; however, it is likely that differential peptide and protein identifications are identified between each set of experiments, adding a level of difficultly to accurately comparing these datasets. In contrast, for iTRAQ, a multiplexed set of reagents are used, placing isobaric mass labels at the *N*-termini and lysine side chains of peptides in a digest mixture, thereby virtually all kinds of protein sample (including clinical samples) can be chemically labelled (Ross et al., 2004). Each of the reagents used are isotopically labelled differentially, such that derived peptides are indistinguishable chromatographically and are isobaric, but yield signature ions following fragmentation that can be used to identify and quantify individual peptides from the multiplexed set. The iTRAQ label consists of a reporter group, a balancing group and an amine-reactive group, with an isobaric mass of 145 Da for all 4-plex labels (Beck et al., 2012). Recent advancements have produced 8-plex iTRAQ labels, allowing for the labelling of eight samples

(Choe et al., 2007, Karp et al., 2010); however here, focus is on the 4-plex labelling. The reporter group has a mass of 114 to 117 Da, whilst the balancing group is given a mass to total 145 Da for the two groups, with the balancing group lost through neutral loss after collision-induced fragmentation. The subsequent peptide-derived fragment ions are identical in mass (isobaric) whilst the reporter ions are differential in mass due to the presence of the reporter group (Figure 1.12). Therefore, the precursor ions appear as a single precursor signal in MS survey scans, but are selected in MS/MS allowing for quantification of multiplexed peptides based on the relative peak areas of the low molecular weight reporter ions. A similar approach is the use of 2-plex, 6-plex and 10-plex tandem mass tags (TMTs), involving incorporation of one ($^{13}C$ – 2-plex) or five ($^{13}C$ or $^{15}N$ – 6-plex and 10-plex) stable isotopes to perform relative protein quantification between two and up to six samples. Like iTRAQ, during CID, TMTs release reporter ions allowing the provenance of the sample to be identified. As quantification is performed only at the MS/MS level, chemical noise is reduced upon fragmentation offering a higher precision compared to quantification performed at the MS level (Beck et al., 2012). However, due to the commercial availability and the requirement for labelling, iTRAQ and TMT is typically of higher financial cost than label-free methods and efficient fragmentation must be performed in order to allow accurate quantification. Secondly, underestimation of iTRAQ and TMT ratios can arise due to co-eluting peptides with similar *m/z* values, which are co-selected during ion selection and co-fragmented during collision induced dissociation. These co-eluting peptides will have a 1:1 ratio across the reporter ion tags (as is required for normalisation in experiments) and so will contribute a background equally to each of the reporter ion signals and diminish the computed ratios (Karp et al., 2010). Triple-stage mass spectrometry and gas-phase purification have been suggested to alleviate this problem by further separating the co-eluting/co-fragmenting peptides (McAlister et al., 2014, Wenger et al., 2011, Ting et al., 2011)

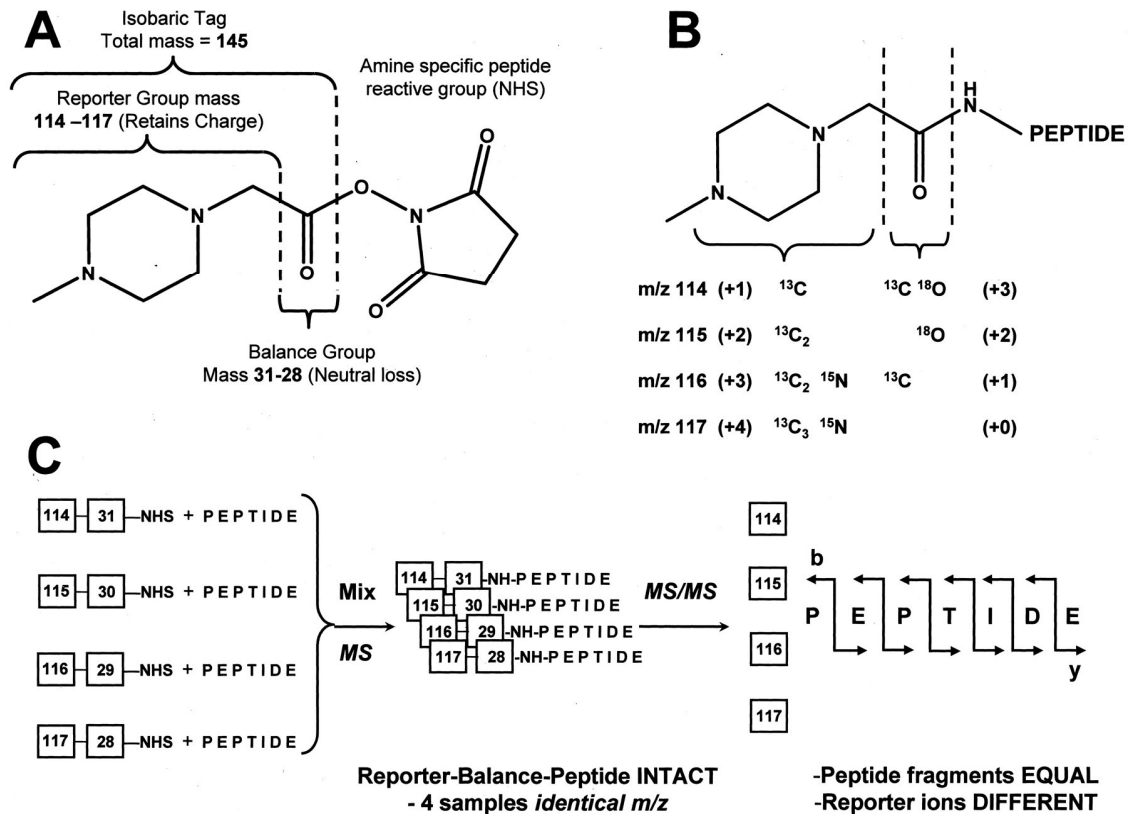



**Figure 1.12) 4-plex iTRAQ**

Up to four samples are mixed, each containing unique iTRAQ reagents composed of a reporter group, balancing group and amine-reactive group (A and B). When intact, each identical peptide has an equal *m/z*, regardless of the reporter group/balancing group attached. After fragmentation, the balancing group is lost due to neutral loss, and reporter ions corresponding to the reporter group give differential *m/z* values, whilst the peptide *b* and *y* ions *m/z* are identical across samples (C). Taken from publication (Ross et al., 2004).

### 1.4.3 MS-based absolute protein quantification strategies

Absolute quantification determines the absolute concentration or amount of protein in terms of mass, or copy number in a sample mixture. By knowing the absolute amount, one is able to determine protein level fold changes across a multitude of samples and conditions, and enable comparison to non-identical peptides and proteins, determine protein stoichiometries in protein complexes and inform modelling approaches, particularly in interaction modelling networks. Again, both label-mediated and label-free approaches can be used.

#### 1.4.3.1 Top3

In Top3 analysis, a known quantity of intact unlabelled standard protein is spiked into the sample mixture prior to digestion. Each subsequent tryptic peptide is analysed by LC-$MS^E$ and identified via database searching. The 'top 3' tryptic peptides with the most intense signals are identified and averaged for each protein, including those belonging to the standard protein (Silva et al., 2006b). The universal signal response factor is calculated using the internal standard proteins, such that the observed 'count' is divided by the known concentration of internal standard. The universal signal response factor is then utilised in determination of the absolute quantification of all remaining proteins: each averaged signal is divided by the universal signal response factor to give the corresponding absolute protein concentration (Nikolov et al., 2012).

#### 1.4.3.2 Selection of peptides for absolute quantification

In label-mediated strategies for relative quantification, the ratio of labelled ('heavy') to endogenous unlabelled ('light') protein between two samples is determined in order to provide a relative fold change of the particular protein in response to a changing state. The labelling of a protein does not affect its ability to be proteolysed by endoproteases such as trypsin, with the respective heavy and light peptides produced demonstrating identical behaviour within an LC-MS system. Labelling is usually introduced by chemical, enzymatic or metabolic approaches into the test sample to be compared directly to a control. The peptides for absolute quantification (AQUA) strategy uses synthesised isotopically labelled standard peptides as surrogates for protein quantification rather than labelling a sample in entirety.

To begin selection of representative surrogate peptides for use as AQUA peptides for quantification, proteins of interest are defined. Several peptides are then selected for use as standards for these proteins of interest based on an array of features denoted in previous experiments or literature: ionisation efficiency; potential modification of amino acid residues; chromatographic elution and ability to be detected in the mass spectrometer. Each standard peptide is chemically synthesised and labelled with $^{15}C$ or $^{15}N$ amino acids (Kirkpatrick et al., 2005). In this way, both the heavy and endogenous peptide share identical physiochemical properties but can be distinguished in the mass spectrometer by a defined mass shift. Known amounts of each AQUA peptide is added to the sample mixture during digestion, before being subject to liquid chromatography and selected reaction monitoring, discussed later in this chapter). Firstly the ratio between amounts of internal AQUA peptide standard and the identical

analyte peptide can be determined by exploiting the difference in mass shift, and as the amount of AQUA peptide is known, the amount of the analyte peptide can be calculated (Kirkpatrick et al., 2005) (Figure 1.13).

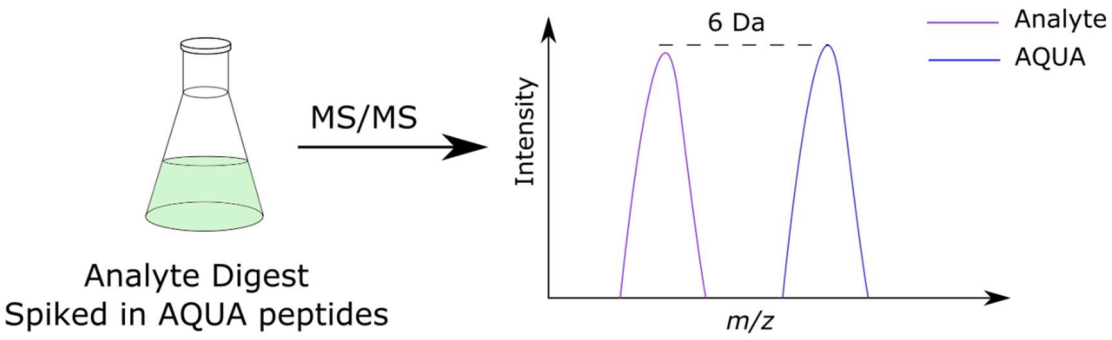



**Figure 1.13) Absolute quantification via AQUA peptide**

Known amounts of labelled AQUA peptides are added to a tryptic digest of sample. As analyte and labelled peptides share identical physiochemical properties co-elution occurs, however the peptide precursor ion and product ions (the labelled amino acid for the latter) will have a mass shift of ~6 Da. The ratio between AQUA and analyte peptide is calculated, and using the known amount of AQUA peptide, the absolute abundance of the peptide can be determined.

As with all strategies, the use of AQUA peptides has advantages and disadvantages. In this strategy, the absolute peptide abundance is used as a surrogate for the absolute protein abundance. Using additional AQUA peptides per protein will improve the accuracy of absolute protein quantification, however will incur higher costs. In addition, absolute quantification via the use of AQUA peptides requires the user to know which proteins to target prior to the experiment, and so discovery experiments are required to elect suitable target proteins whilst label-free methods do not require this. However, as there is a standard incorporated for every peptide to be quantified there exist multiple standard data points for every peptide including a chromatographic retention time, peak shape, and transition profile. Therefore, any sub-standard peptides may be excluded from quantification and will not contribute to incorrect absolute values. The use of multiple AQUA peptides per protein will protect against any losses in this way and demonstrate the level of agreement between peptides belonging to the same parent protein. However, the AQUA strategy does not control for protease efficiency. As only the analyte protein is required to be digested prior to an MS experiment, any issues with cleavage will be present in the analyte protein only. Therefore, as quantification can only be performed using identical standard-analyte peptide combinations, missed cleavage in the analyte peptide counterpart only will result in erroneous quantification.

#### 1.4.3.3 QconCAT

As exemplified in the description of AQUA above, the use of proteotypic (i.e. peptides that are only found in a single protein, therefore serves to identify the protein) peptides as surrogates for the protein of interest has emerged as a key strategy for absolute protein quantification. When multiple proteins are targeted, the number of standard peptides to be synthesised and purified undoubtedly increases, increasing the experimental resources and time required to perform absolute quantification.

In the QconCAT strategy, multiple proteins are targeted by quantification peptides (Q-peptides) that are selected on the same basis as AQUA peptides. Q-peptides are concatenated in a 1:1 ratio within a bacterial expression construct (QconCAT) to produce a single protein (Rivers et al., 2007). Each construct contains on average 25-30 Q-peptides, which if two Q-peptides target a single protein, allows for absolute quantification of up to 15 proteins. Usually these Q-peptides are arginine- or lysine- terminated at the *C*-terminus in order to represent the proteotypic peptides following tryptic digestion of the protein(s) of interest. Additional to the Q-peptide concatemer, flanking sequences are added. Namely, an initiator codon, a purification tag and protective sacrificial regions protecting the assembly of true Q-peptides from exoproteolytic attack during expression. Internal standard peptides are added within the construct (typically glu-fibrinopeptide) to be used for quantification of the QconCAT construct (Figure 1.14). A gene encoding for the QconCAT protein is constructed and ligated using an expression vector into *E. coli*. The *E. coli* QconCAT strain is grown in isotopically-labelled media (typically heavy arginine and lysine such that every tryptic peptide is isotopically-labelled) and purified through the use of the purification tag. In this way, a standard targeting up to 15 proteins is created and purified using a single purification protocol. If many target proteins are selected, multiple constructs may be created. Use of the QconCAT constructs avoids having to purify and quantify many individual isotopically-labelled peptides, as is required in AQUA.

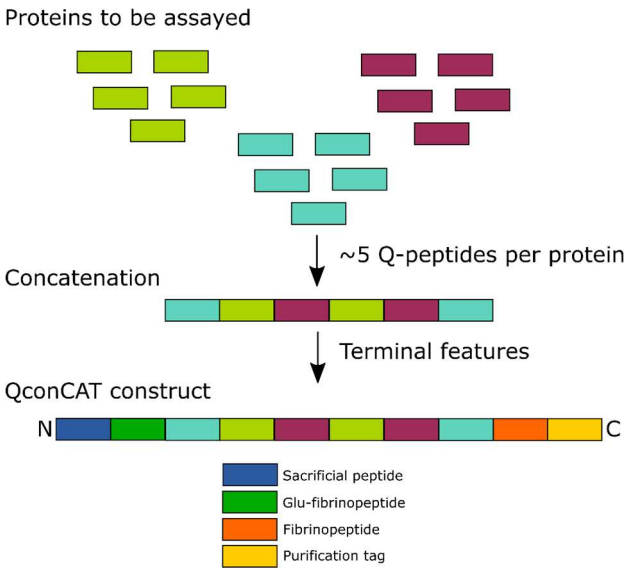



**Figure 1.14) Design of a QconCAT construct**

Selected Q-peptides representing target chaperones are concatenated in an order that maximises likelihood of tryptic cleavage as determined by MC:pred, with terminal features added to allow purification and quantification of the QconCAT standard.

The QconCAT is quantified in absolute terms via its own internal standard, and known amounts are added to the cell lysate containing the proteins of interest. Following tryptic digestion, both analyte (unlabelled) and standard (labelled) proteotypic peptides are produced. As each Q-peptide within the QconCAT is in a 1:1 ratio, the quantity (moles) of each Q-peptide is equal to that of the spiked QconCAT. As the proteins of interest are preselected, selected reaction monitoring (SRM) experiments targeting both the standard and analyte peptides enable for quantification of the ratio between the labelled and unlabelled form of the peptide targeting the protein of interest. As the Q-peptide quantity is known, this is used to calculate the absolute quantification of the endogenous peptide. Typically, multiple Q-peptides are used to target a single protein, and so an average of multiple peptide-level values are used to determine the target protein-level abundance.

More recently, improvements to the QconCAT strategy have been performed. Namely, flanking sequences to each Q-peptide are added to imitate the cleavage site of the native unlabelled peptide. In this way, the digestion efficiency and release of both standard and analyte peptides are expected to be identical and absolute quantification is more reliable. However, work by Brownridge and Beynon (Brownridge and Beynon, 2011) has demonstrated that with stringent selection criteria and an optimised digestion protocol, both standard and

analyte peptides are digested to completion and so such flanking regions are not always required.

As with AQUA peptides, a limitation with the QconCAT strategy is that the proteins of interest must be known prior to the quantification experiment. Secondly, in order to elect suitable peptides for quantification, investigatory experiments or modelling approaches must be performed such that each standard peptide is known to be observed in the mass spectrometry experiment workflow to be used. As both the QconCAT strategy and the AQUA strategy select few peptides to use as standards for each protein, the protein sequence coverage is low, limiting the statistical reliability of the quantification in a single biological replicate.

#### 1.4.3.4 Protein Standards for Absolute Quantification

Protein Standards for Absolute Quantification (PSAQ) shares the notion of isotope labelling a standard to match the analyte of interest (Brun et al., 2007). However, unlike the AQUA strategy and QconCAT strategy, a whole protein identical in sequence to the protein of interest is isotopically labelled and added to the sample prior to MS analysis. As the standard protein is identical in sequence to the analyte protein and is added to the sample prior to digestion, protease digestion is controlled between standard and analyte removing quantitiative error as a result of incomplete digestion. All observable analyte peptides in an MS workflow also have a standard counterpart, albiet with a mass difference according to the label used. Like the AQUA and QconCAT strategy, the known amounts of standard are used to determine absolute quantification values for each protein of interest. However, as there is a single PSAQ standard for every protein of interest, there is a proportionally large molecular biology effort to create each standard whilst for QconCAT standards, multiple proteins may be targeted with one concatemer.

## 1.5 MS proteomics workflow

Although the basis for MS-based proteomics experiments are somewhat similar, in that every workflow includes sample preparation, mass analysis and data processing, workflows can be tailored to the experimental requirements and the needs of the user. Indeed, it is imperative to understand the various sample preparation, analysis techniques and MS instruments before processing data collected. A wide variety of processing tools exist that give dissimilar results for

the same set of data dependent upon the parameters input, minor processing changes can result in major differences in the conclusions made. The techniques and instruments used within these studies are further discussed here.

Two main goals of proteomics are to identify proteins derived from complex mixtures extracted from cells and to quantify expression levels of those identified proteins. Sample preparation is perhaps the most crucial stage of any proteomics workflow, as the quality of protein samples is critical to generating accurate and informative data. Any proteins that are lost in the sample preparation stage are consequently lost from the entire proteomics experiment and so the workflow for sample preparation must be optimised towards the aims of the experiment. There is no single 'standard' workflow, however there are common steps towards obtaining the 'perfect' sample.

### 1.5.1 Sample preparation

Sample preparation for MS begins with cell lysis and protein extraction. Proteins are heterogeneous and vary widely depending on charge, size, hydrophobicity and structure. In addition, many undergo post-translational modifications and so may have a variety of moieties attached. Therefore, isolating the proteins can be extremely challenging, with the reagents (salts, buffers and detergents) added to lyse the cells often potentially interfering with proteome analysis (Andrecht and von Hagen, 2008). Sample preparation workflows used in MS-based proteomics therefore often include multiple steps such as sample desalting, concentration, subfractionation, and further separation by gel electrophoresis and/or chromatography in order to reduce sample complexity (if required) and remove contaminants likely to interfere with analysis (Barbour et al., 2008).

Proteins are denatured, solubilised and cysteine residues are subject to reduction and alkylation, typically with dithiolthreitol (DTT) and iodoacetamide respectively to irreversibly reduce disulphide bonds and thus destabilise tertiary structure. Proteins are then digested by an endoprotease either in-gel after separation, or in solution, with trypsin most commonly used as it yields peptides that are highly amenable to MS analysis. In-gel digestion protocols offer an attractive method of removing detergents during electrophoresis, however the effectiveness is limited due to poor accessibility of the endoprotease to proteins within the gel matrix. Alternatively, proteins are digested within solution with the selected protease. Digested samples (in-gel or in-solution) may then be subject to separation by liquid chromatography to

separate peptides prior to electrospray ionisation MS (ESI-MS). During digestion, trypsin cleaves peptide bonds that are *C*-terminal to the basic residues arginine and lysine unless followed by a proline residue, typically generating peptides of between 800 to 4000 Da.

### 1.5.2 Electrospray Ionisation

Although multiple ionisation methods exist, ESI (Fenn et al., 1989) is focussed on in this thesis. In ESI, a soft-ionisation technique, the peptide analyte mixture is diluted in a suitable solvent (commonly $CH_3CN$) and sprayed from a fine capillary needle in the presence of a strong electric field (3-5 kV $cm^{-1}$). A mist of ionised droplets approximately 10 µm in diameter is produced, with the surface charge depending on whether the needle carries a positive or negative voltage potential, with the interface plate of the MS instrument acting as a counter electrode to the needle (Barbour et al., 2008). In positive ion mode, positive ions are enriched at the surface of the droplet, repelling each other whilst being pulled by the electric field. When this exceeds the surface tension of the liquid, a "Taylor Cone" is formed, in which the tip of the cone extends into a micron-size filament until it reaches the Raleigh limit, determined by the Coulomb forces of the accumulated positive charges and the surface tension of the solution (Barbour et al., 2008). As the volatile solvent evaporates, Coulombic repulsion increases between the positive charges, thus exceeding the Raleigh limit and causing the droplet to burst into nanometer-sized daughter droplets. Further solvent evaporation from the droplets continues until all ions are completely desolvated. Nitrogen gas entering the region where ionisation occurs inhibits neutral molecules from entering the high-vacuum region of the mass spectrometer whilst charged ions proceed into the high-vacuum chamber (Barbour et al., 2008). During this entire process, very little energy is retained in the analyte, causing it to remain stable in the gas phase. Depending upon the availability of basic amino acid sites, peptides may become multiply charged. Multiply charged ions extend the mass range of analysis in proportion to the multiplicity of charging, allowing for use of mass analysers with only a modest *m/z* range.

### 1.5.3 Tandem mass spectrometry

During tandem mass spectrometry, peptide ions are further fragmented into precursor ions via CID or electron transfer dissociation (ETD). The signals of fragment ions are recorded via the detector within the mass spectrometry and are able to provide information regarding the primary structure of a polypeptide or protein as well the relative intensity. The electrospray ionisation process of polypeptides typically produce a series of multiply charged peptide ions, with protonation occuring at basic residues within the peptide sequence (Arg, Lys, His, $NH_2$-terminus). These peptide ions may be fragmented, generating product ion series dependent upon the fragmentation method used (Figure 1.15).

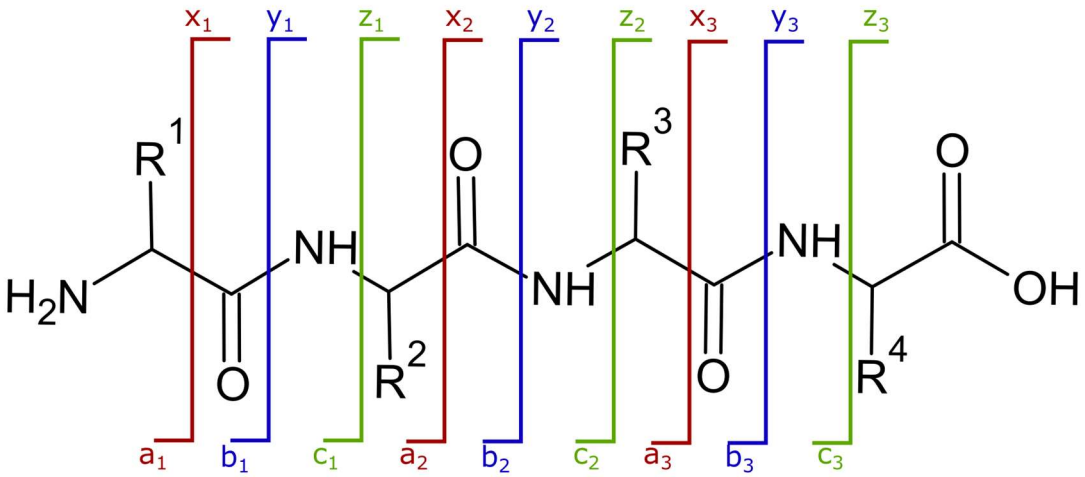



**Figure 1.15) Fragmentation of a peptide into product ions**

$R^n$ refers to the R group for each amino acid within the peptide sequence. Depending on the fragmentation method used, different ion series will be produced. CID typically produces b and y ion series whilst ETD typically produces c and z ion series. Negative electron transfer dissociation typically produces a and x ion series (not discussed within this thesis).

In the process of collision induced dissociation, ions are collided with a neutral atom or molecule in the gas phase, subsequently dissociating the ion (Equation 1.1), where $AB^+$ is the ion that collides with the neutral molecule M and dissociates. Due to a rapid redistribution of the internal energy, the process is equivalent to heating the precursor ions. Upon reaching the fragmentation threshold, the weakest bonds are cleaved preferentially, regardless of their location in the peptide backbone or the side chain, generating y- and b- type fragment ions. However, fragmentation is not equal along the backbone; cleavage near glumatic acid, aspartic acid and proline residues are preferred, with the loss of post translational modifications occuring frequently (Sobott et al., 2009).

$$AB^+ + M \rightarrow A + B^+ + M$$

*(Equation 1.1)*

In contrast, during ETD, an electron is added through an ion-ion reaction, resulting in the fragmentation of multiply charged peptide cations by electron transfer to them (Figure 1.16) (Syka et al., 2004). ETD induces fragmentation along the peptide backbone in a sequence independent manner. ETD takes advantage of the energy released in the exothermic capture of an electron, allowing for fragmentation without the redistribution of energy (Sobott et al., 2009). This process results in a rapid bond cleavage mainly at the sites where the originally formed radical was trapped, allowing modifications linked to the peptide chain to remain in place and for a more complete series of c- and z- type ions to be generated in comparison to CID. Therefore, a major advantage of ETD in MS-based proteomics is the ability to localise the exact sites of peptide post-translational modifications that can be missed by CID (Kim and Pandey, 2012).

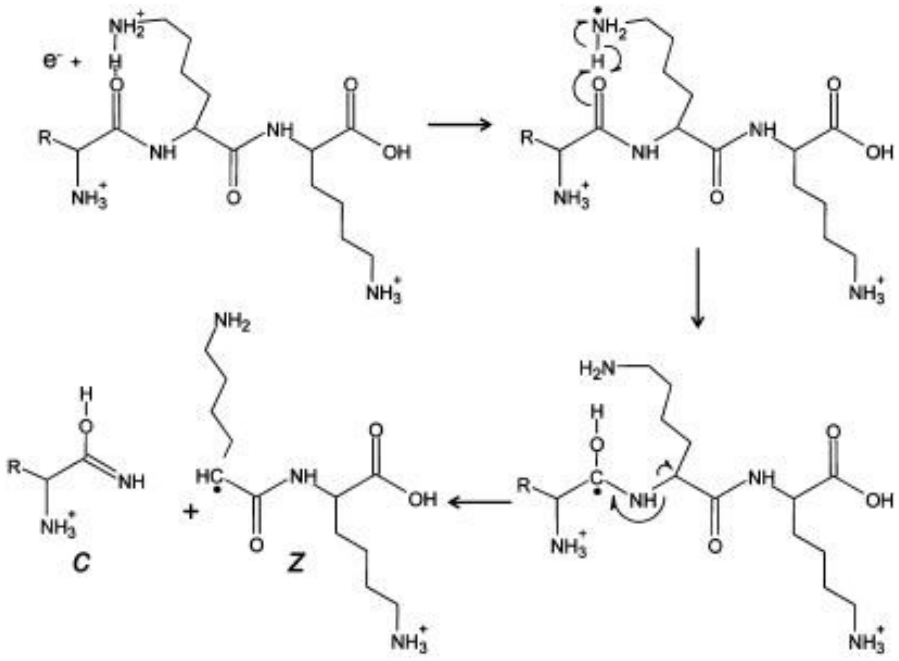



**Figure 1.16) Electron transfer dissociation of a multiply-charged peptide ion**

Transfer of an electron to the carbonyl groups of the peptide backbone leads to a homologous series of complementary fragments of c- and z- ion types due to fragmentation of the N-$C_\alpha$ bond. Figure taken from publication (Syka et al., 2004).

The dissociation method used therefore is dictated by the aims of the experiment. ETD is used extensively with proteins and peptides for sequence analysis, particularly for the identification of post translational modification sites, and is advantageous for top-down

proteomics. In shotgun (bottom-up) proteomics studies, CID is the most commonly used fragmentation method, during which fragment ions are used to identify (and quantify) proteins of interest. During the studies described in this thesis, CID was used to generate b- and y-fragment ions towards quantifying the yeast proteome.

Together the precursor and product ions may be used to elicit information regarding the parental protein. During data-dependent acquisition methods as is typical for 'shotgun' proteomics experiments, a number of precursor ions generated by ESI are selected to undergo ion activation and thus fragmentation, typically by CID. The fragment ions generated are then mass analysed to identify and quantify the peptides and proteins from which they originate. However, as the selection of peptides for ion activation is based on their ion intensity there is a bias towards the more abundant peptides/proteins, limiting the depth to which the proteome can be analysed (Holman et al., 2012). An alternative acquisition method is data-independent acquistion (including $MS^{Excess}$ and SRM), during which the precursor ion is selected according to a protein of interest, rather than according to its intensity. The selected precursor ion is then fragmented, giving rise to a number of fragment ions. The $MS^{Excess}$ and SRM methods are used within this thesis.

#### 1.5.3.1 $MS^E$ experiment

In an $MS^{Excess}$ ($MS^E$) experiment (Figure 1.17), the identities of the peaks of interest are not known; therefore data for all detectable precursor and product ions are collected. To avoid need for second or third analyses, both precursor and fragment ions are generated by the mass spectrometer simultaneously – all the ions are transmitted from the ion source, fragmented and recorded as a fragment ion spectrum. No pre-selection of the ions occurs prior to the fragmentation so no data is lost during the experiment (Silva et al., 2006a). In this way, an $MS^E$ experiment allows for all data regarding detectable peptides/proteins to be gathered in a single untargeted run, which may be analysed post-acquisition to extract both the chromatographic and mass spectrometric information on the tryptic peptides, and provide quantitative information regarding the parent proteins.

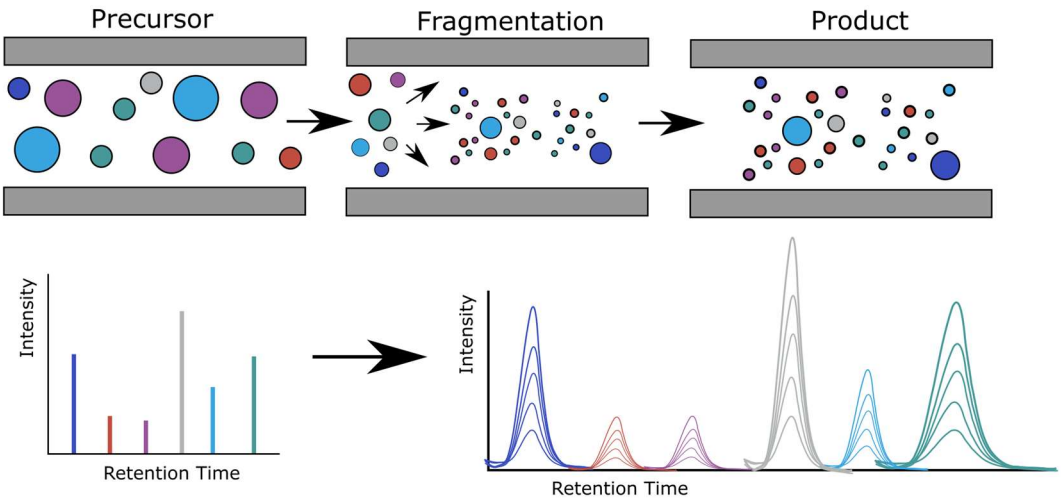



**Figure 1.17) MS$^E$ experimental workflow**

All detectable precursor ions are fragmented without prior selection. All of the resulting fragment ions are recorded as fragment ion spectra, such that no data is lost.

### 1.5.3.2 Selected reaction monitoring

During SRM, precursor and product ions are selected and isolated (Figure 1.18), with *b* or *y* product ions from low-energy CID pre-selected according to their *m/z* value and relative intensity. SRM is performed on a tandem-quadrupole (TQ) instrument, with the first quadrupole (Q1) acting as a mass filter to isolate the precursor ion of choice. The second quadrupole (Q2) acts as a second mass to charge (*m/z*) filter, increasing the selectivity without the requirement for very high resolution and mass accuracy (Brownridge et al., 2012). Each quadrupole is separated by a collision cell in which fragmentation of the precursor ions take place, however in older instruments fragmentation takes place in an RF-only quadrupole, giving rise to the name ‘triple quadrupole'. As the duty cycle of the instrument is solely dedicated to analysis of the selected ions, sensitivity in an SRM experiment is the highest of all the mass spectrometric methods (Brownridge et al., 2012).

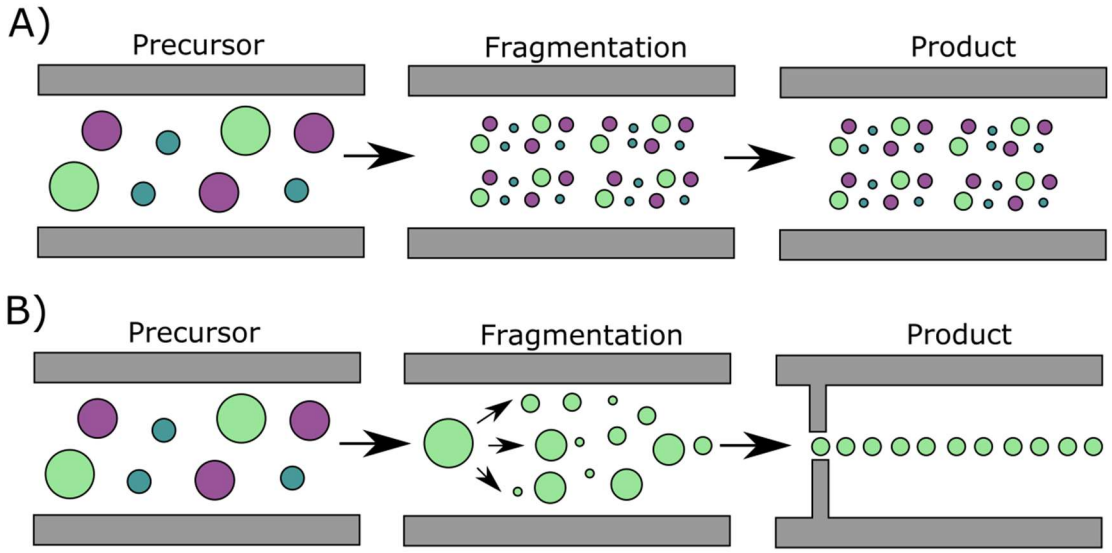



**Figure 1.18) Selected reaction monitoring following transition design and selection**

A) In order to select the optimal transitions for an SRM experiment, an $MS^2$ experiment is carried out in which all the precursor and product ions are selected and measured. Precursor and product pairings resulting in a single 'transition' are selected. For QconCAT transition design of each Q-peptide, three product ions are selected from one precursor ion fragmentation resulting in three transitions. B) In a TQ instrument, a single precursor ion is selected and fragmented by collision induced dissociation into its constituent product ions. Product ions of interest are selected for in a *m/z* range to produce chromatograms unique to the peptide of interest.

As an alternative to SRM, parallel reaction monitoring (PRM) may be performed (although is not done so within this thesis). During parallel reaction monitoring (Figure 1.19), a target precursor ion is isolated and fragmented, with all the resulting product ions detected. In this way, more product ions can be matched to the target peptide and confirm its identity, increasing the specificity of the experiment over SRM methodology (Ronsein et al., 2015). As no selection occurs in the second quadrupole, much less method development is required compared to SRM – only the precursor ion mass needs to be known for isolation for fragmentation.

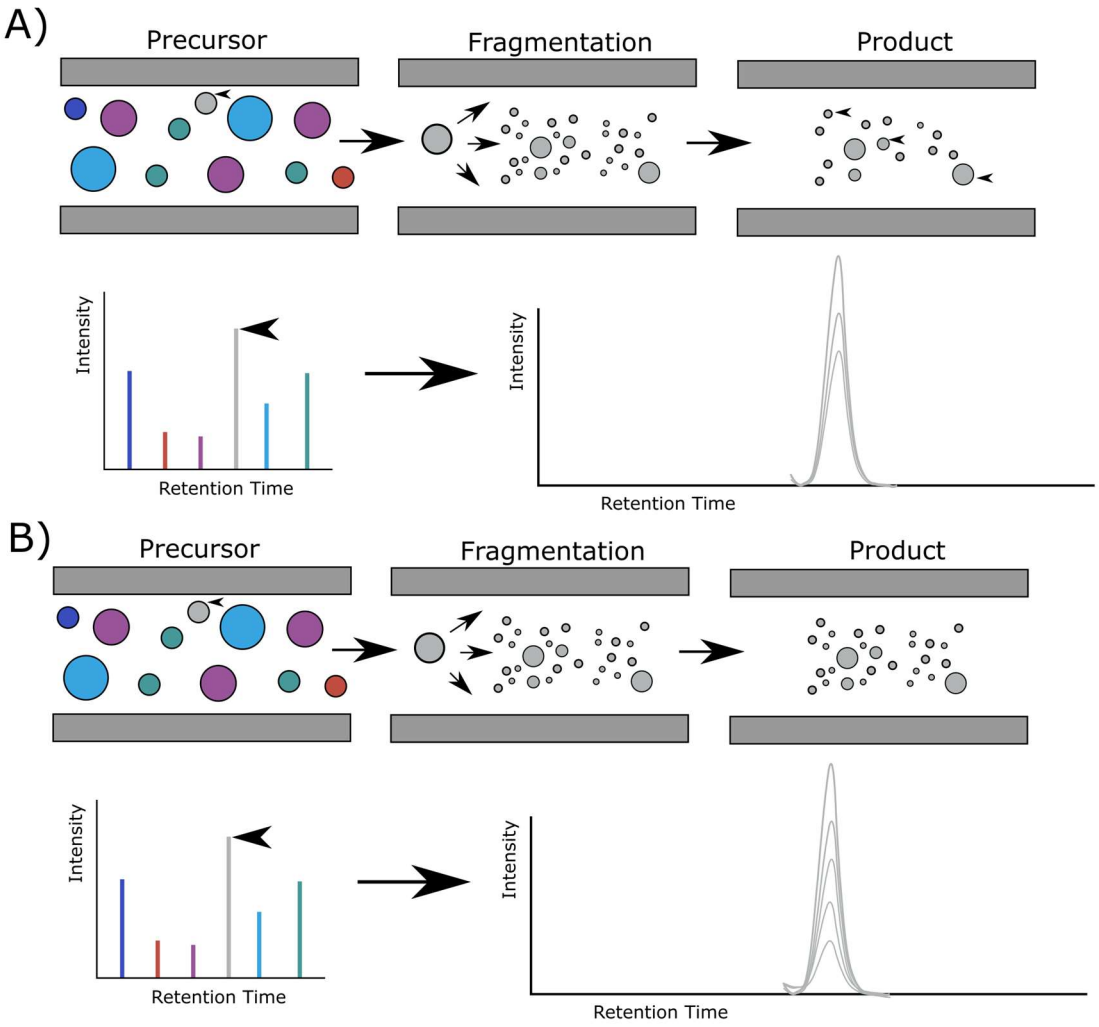



**Figure 1.19) SRM and PRM methodology**

A) During SRM, a single precursor ion of interest is isolated for fragmentation. Selected product ions are recorded to produce a product ion spectra. B) During PRM, a precursor ion of interest is isolated and subject to fragmentation. All resulting product ions are recorded to produce the product ion spectra. Selected ions are denoted by a small black arrow.

### 1.5.4 Instrumentation

As the process of ion generation by ESI is continuous, the source is typically coupled to quadrupole or ion trapping instruments. A mass spectrometer typically consists of three components: the ion source, a mass analyser and a detector. The ion source as described above, converts the analyte into charged ions which are passed into a high-vacuum mass analyser. The differences in mass of the ions allows the mass analyser to differentiate the ions by their mass to charge ratio (*m/z*). Mass analysers include quadrupoles, time of flight and ion trap mass analysers. The detector measures the signal of each ion present, thus providing data to a

connected computer system about the analyte. Different MS instruments have unique combinations of ion sources, mass analysers and detectors suitable for the demands of different proteomics experiments. I discuss further details of different mass analysers and instruments used in this thesis here.

#### 1.5.4.1 Quadrupole Mass Analysers

In quadrupole mass analysers, electric fields are used to separate ions according to their *m/z* ratio as they pass along the central axis of four parallel equidistant electropoles that can have fixed direct current (DC) and alternating radio frequency (RF) voltages applied to them. Depending upon the magnitude of these voltages, only ions of certain *m/z* ratios are allowed to pass the entire length of the quadrupole into the detector. When an ion is caused to oscillate with a trajectory whose amplitude exceeds the distance between the quadrupole tunnel radii, it will collide with a rode and discharge or pass out of the mass analyser and not be detected (Miller and Denton, 1986) (Figure 1.20). This enables selection of an ion with a particular *m/z* or a range of *m/z* values by continuously varying the applied voltages. Typically, DC and RF voltages are altered according to a linear relationship, often referred to as the 'scan' function. The slope of this line is referred to as the quadrupole 'gain' – the rate of change of DC against RF voltages, whilst the magnitude of the initial DC voltage applied is termed the quadrupole offset. The resolution of a quadrupole can be improved by increasing the gain and offset, however results in a lower sensitivity of the mass analyser. Removal of the DC voltage allows the scan line to become horizontal, in this way ions of all *m/z* values are allowed to travel through the quadrupole which acts solely to keep the ions focussed and permit transmission from one part of instrument into another.

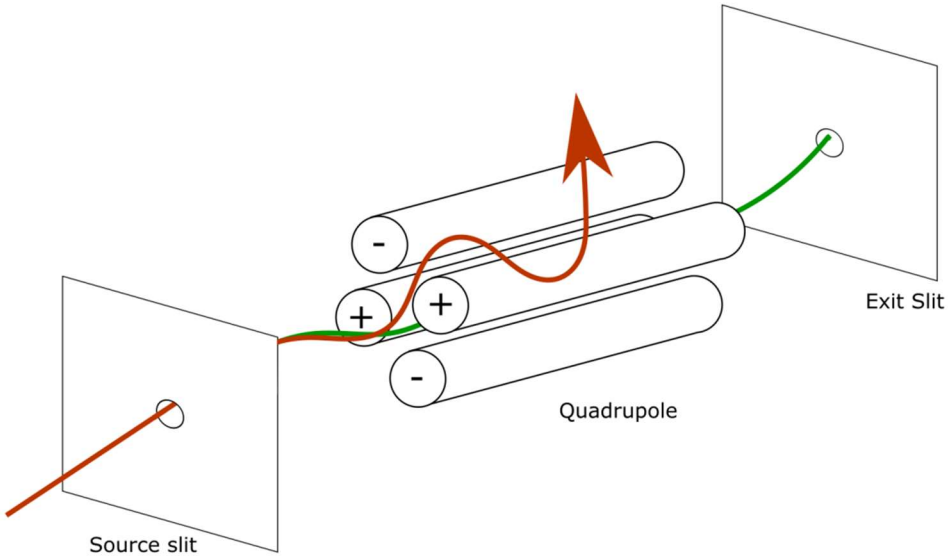



**Figure 1.20) A quadrupole mass analyser**

By alternating the voltages applied to the rods, only ions with a selected *m/z* are able to pass through to the detector in a stable trajectory (green). Ions with unstable trajectories will collide with the rods, lose their charge and/or pass out of the mass analyser so are not detected (red).

#### 1.5.4.2 Time of Flight Mass Analyser

In a time of flight (ToF) mass analyser, packets of ions are accelerated at a defined time (in bursts). The ions then pass down a tube of set length, and as there are no ion focussing devices, the ions pass in a straight line at constant speed towards the detector (Figure 1.21) (Guilhaus, 1995). The speed at which the ions pass through the tube is dependent upon both their mass and charge – for an ion of a given charge, the greater the mass the slower it will travel towards the detector, as the velocity is inversely proportional to the mass for an ion with the same kinetic energy. At the detector, the *m/z* of the ion is calculated using the time taken to reach the detector, the length of the flight tube, the charge of an electron and the potential used to accelerate the ion. The resolution in a ToF mass analyser is limited by two factors. The flight time of an ion is dependent upon its mass and charge. Ions with very small *m/z* differences will not be separated well in short flight tubes, making it harder to differentiate at the detector. Secondly, not all ions of the same *m/z* arrive at the detector simultaneously due to a distribution of the kinetic energies they acquire in the accelerating voltage, and not all begin accelerating from the same point. As they do not arrive at the detector simultaneously, spectral resolution is compromised. Resolution in instruments with a ToF mass analyser can be improved through the use of a reflectron device or 'ion mirror' – a series of electrostatic lenses that create a homogenous electrostatic field at the end of the usual flight path of the ions, with the same polarity as the ions. Positive ions are quickly slowed within the electrostatic field and come to

a brief halt before being accelerated in the opposite direction. Ions with greater kinetic energy will enter the reflectron to a greater depth and spend longer within the device (having a longer flight time) than ions with a lower kinetic energy. Therefore, isobaric ions are able to group together and reduce their distribution of flight times, thus enhancing the resolving power of the instrument. An example of this is seen in the Waters G2 Synapt instrument below.

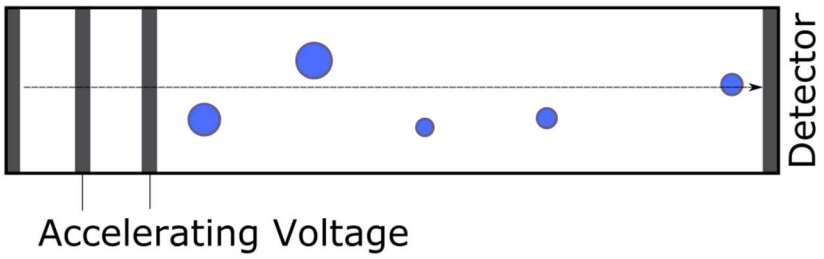



**Figure 1.21) The Time of Flight Mass Analyser**

Ions pass in a straight line at constant speed towards the detector. The flight time to the detector is determined by the ions *m/z* .

#### 1.5.4.3 Ion Trap Mass Analyser

Ion trap mass analysers use oscillating electric fields to trap ions in a controlled manner, many types of ion traps exist however I focus upon the quadrupole ion trap. In a typical quadrupole ion trap, a ring electrode with a hyperbolic inner surface is neighboured by two electrically common hyperbolic end-cap electrodes (Figure 1.22). In order to trap ions, a sinusoidal potential is applied to the ring electrode, whilst the end-cap electrodes are either set to a constant continuous current or maintained at an oscillating alternating current. Combinations of the ring and end-cap electrode potentials may be used to trap all ions within or above a specific *m/z* range, trap ions of a particular *m/z* and eject ions of specified *m/z* values (March and Todd, 2005). The trap has the capacity for only a limited number of ions at any one time before repulsive charges cause excess ions to be ejected, resulting in low resolution. This is controlled by determining the optimum number of ions to inject within the trap. After the ions are injected into the trap from the ion source, a suitable RF voltage on the ring electrode sets them to stable trajectories. In the case of tandem MS, voltage is applied causing the trapped ions to oscillate with a higher energy, causing them to collide with the background gas and undergo fragmentation. The *m/*z of the ions are then determined by knowing the voltage required for their ejection from the ion trap to the detector through an opening in the end-cap electrode. Damping gases such as helium may be used to increase the resolution and sensitivity of the mass analyser. Such gases cool the trapped ions via gentle collisions, causing the

respective ions oscillations to dampen and the ion to become focussed near the centre where the trapping fields are closest to the ideal field.

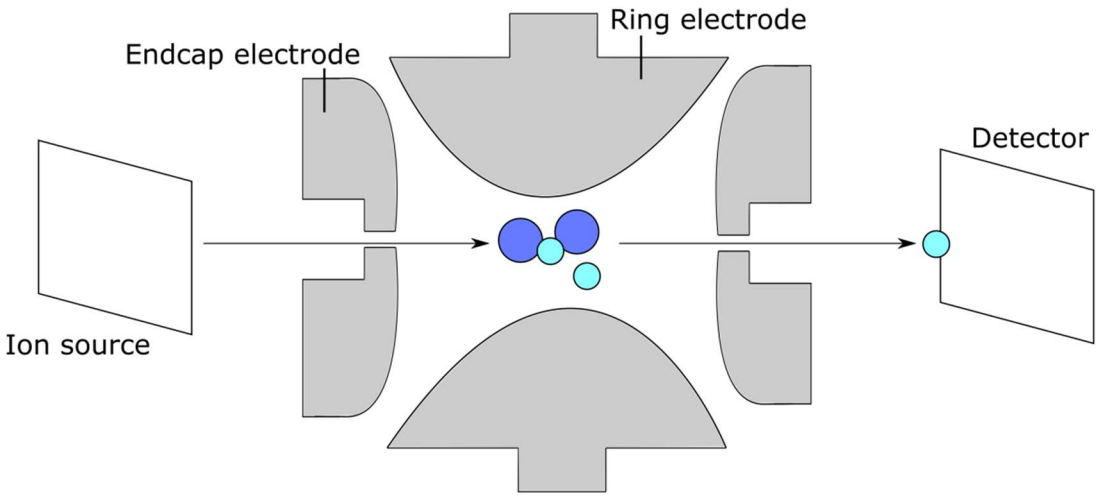



**Figure 1.22) Quadrupole ion trap mass analyser**

Ions entering the quadrupole are confined to stable trajectories via a suitable RF voltage on the ring electrode. To fragment ions in tandem MS, selected ions are excited and undergo collisions that ultimately fragment the ions. Ions are analysed by ejection from the ion trap upon exiting their oscillatory trajections.

#### 1.5.4.4 Orbitrap mass analyser

Traditional ion trap mass analysers have disadvantages including limited linear range, reduced charge capacity or high complexity. In contrast, the orbitrap mass analyser has high performance in terms of resolution, mass accuracy, space charge capacity, although are significantly more expensive. In an orbitrap, a central electrode is stretched along the axis of an outer electrode cylinder enclosing the trapping volume (Figure 1.23). A voltage is applied to the central and outer electrode, such that injected ions cycle around the central electrode whilst oscillating along the horizontal axis. If a high quality signal and resolution is required, the ions of each mass are injected in short bursts. Ion introduction can be performed after modification of the electric field by a field compressor at the injection port. In addition, although operating in a pulsed fashion, orbitrap mass analysers can be coupled to continuous ion sources. To make this possible, ions are stored in an ion storage device (C-trap) prior to entry into the orbitrap. The curved linear ion trap is one of the most efficient ways of injecting ions into the ion trap. The curved modification of the ion trap forces the ion beam to traverse a curved path prior the mass analyser, reducing the incidence of photons and other interfering particles from reaching the mass analyser (and detector), substantially improving the signal to noise ratio. The curved trajectory means that ions can be focussed into narrow areas to be ejected as compact bursts

of ions, required for high resolution to be achieved by the orbitrap mass analyser. Orbitrap mass analysers are able to operate in two modes: Fourier transform and mass selective instability mode (Perry et al., 2008). In Fourier transform mode, the frequencies of the oscillations of ions about the axis are detected to measure the image current, generated as the ions oscillate axially back and forth. The Fourier transformation of this signal can be used to produce the required mass spectrum, yielding the highest mass resolution that the orbitrap can achieve. In mass selective instability mode, ions are ejected and collected on a detector as with conventional mass spectrometers. This mode is typically used when ions with certain *m/z* are required to be stored for tandem MS analysis, or high intensity signals from unwanted compounds can be ejected to improve the dynamic range.

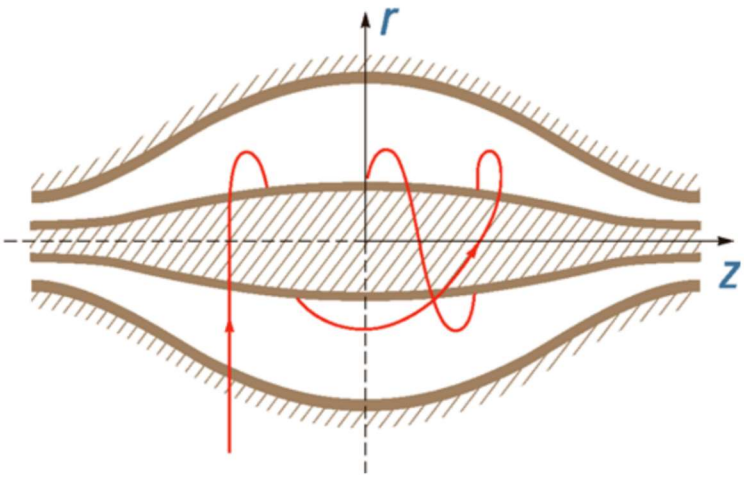



**Figure 1.23) The orbitrap mass analyser**

Pulsed injections of ions of pre-selected mass oscillate about the central electrode. In fourier transform mode, the image current is detected whilst in mass selective instability mode, ions with selected *m/z* are ejected and analysed by a detector. Figure supplied by vendor, available at www.thermofisher.com.

Different MS instruments incorporate unique combinations of mass analysers and vendor-specific technologies to create instruments adapted to the requirements of various experiments. Here I discuss the technologies behind three instruments used for the work presented in this thesis.

#### 1.5.4.5 Waters G2 Si mass spectrometer

The Waters G2 Si mass spectrometer consists of four principal components: a source with a StepWave™ ion guide, a quadrupole, a Triwave™ device and a time of flight mass analyser (Figure 1.24). Sample directly from the liquid chromatography column, or other solvent delivery system, is introduced at atmospheric pressure into the ESI ionisation source in which positively

charged ions corresponding to peptide precursor ions are produced. The resulting ion beam passes through the source sampling orifice, undergoing expansion. G2-Si instruments contain StepWave transfer optics, in which the conjoined ion guide first focusses the ion and directs it to the second stage, a narrow bore ion guide. As the design is off-axis, any neutral materials entering the source sampling orifice are extracted from the system, therefore improving the efficiency of ion transfer. The remaining ions are focussed into the quadrupole during which they are separated according to their *m/z* ratio. Mass separated ions pass into the Triwave region (Figure 1.18b), in which they undergo collision induced dissociation (CID). The Triwave device incorporates three Travelling-Wave™ (T-Wave) ion guides: the first ion guide traps, accumulates and releases ions, the second T-wave guide acts as a high-efficiency transfer device in Time-of-Flight mode or as a travelling wave ion mobility device, whilst the third transfers ions to the time of flight mass analyser via a transfer lens, although can also be used for CID. In the QuanTof™ region, ions are analysed by a ToF mass analyser. This instrument can be used for data-dependent analysis (DDA) or for data-independent (DIA) in an approach termed $MS^E$, in which all precursor ions present in a given scan are simultaneously fragmented by CID and all product ions generated are analysed. Proprietary software then realigns product ions with the precursor ion from which it was derived based in part on the intensities of both precursor and product ions prior to peptide identification.

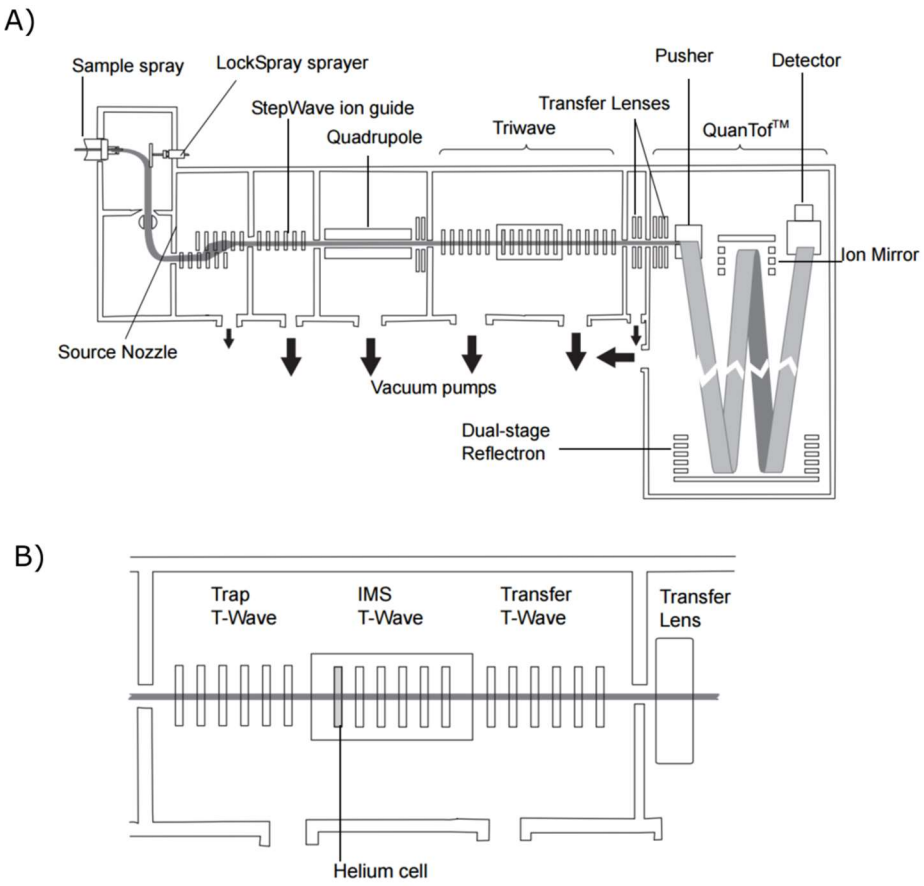



**Figure 1.24) The Waters G2 Si mass spectrometer**

A) The Waters G2 Synapt mass spectrometer consists of four principal components: a source with a StepWave ion guide, a quadrupole, a Triwave device (B) and a ToF mass analyser with a reflectron. Figure supplied by the vendor, available at http://www.waters.com.

#### 1.5.4.6 Xevo TQ-S mass spectrometer

The Waters Xevo TQ-S mass spectrometer contains a StepWave component and two quadrupole mass analysers separated by a ScanWave collision cell (Figure 1.25). The quadrupole is able to filter precursor ions of select *m/z* into the ScanWave collision cell. In the collision cell, precursor ions are subjected to fragmentation by CID. Waters ScanWave technology improves sensitivity when the instrument is run in spectral mode (when in SRM mode, sensitivity is already high), the ScanWave allows ions within the collision cell to be accumulated and then separated according to their *m/z*. When the release of these ions is synchronised with the scanning of the second quadrupole mass analyser, the signal intensity of the full scan product ion spectra is enhanced. The Xevo TQ-S instrument is typically used for SRM experiments, in which the first quadrupole mass analyser is set to stably transit a single *m/z* value corresponding to the ionised peptide of interest. After fragmentation in the collision cell of the precursor ion, only specific product ions of select *m/z* pass through the second

quadrupole with a stable trajectory. The combination of a single precursor and product *m/z* value is defined as a transition, with multiple transitions collected in an experimental run. In this way, SRM is the 'golden' technique for absolute quantitative proteomics: it offers high selectivity, low background signal and is able to detect proteins of low abundance.

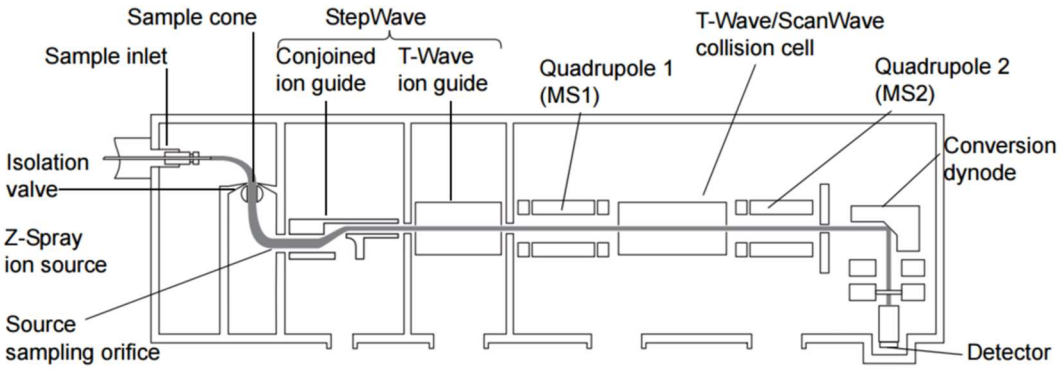



**Figure 1.25) Waters Xevo TQ-S mass spectrometer**

The Xevo TQ-S is comprised of two quadrupole mass analysers. In SRM mode, only precursor ions of particular *m/z* are sent through to the collision cell. In the second quadrupole, only selected fragment ions are sent through to the detector. Figure supplied by the vendor, available at http://www.waters.com.

### 1.5.4.7 ThermoScientific Q-Exactive HF Mass Spectrometer

Hybrid instruments based on the orbitrap mass analyser are very popular in proteomics due to their high resolution and mass accuracy capabilities. The Q-Exactive HF mass spectrometer consists of an atmospheric pressure ion source, a stacked ring ion guide, an injection flatopole in the source region, an advanced active beam guide ejecting solvent droplets and other neutral species to prevent them from entering further into the instrument, a segmented quadrupole mass filter, a C-trap, a higher energy collisional dissociation cell and an ultra-high field (HF) orbitrap mass analyser (Figure 1.26) (Scheltema et al., 2014). The injection flatopole is equipped with ion selection capabilities, providing a low resolution selection mechanism for the removal of unwanted ions prior to higher resolution removal in the segmented quadrupole mass filter. The ultra-high field orbitrap mass analyser can be used to increase the resolution of the instrument through permitting higher frequency of ion motion. In a typical experiment on this instrument, sample is injected and subject to ESI. Neutral ions and other unwanted ions are removed during focussing, with ions filtered as required in the quadrupole mass filter. In DDA mode, *m/z* isolations windows are utilised to filter precursor ions at this stage. Ions are then passed through to a curved ion trap, which stores and compresses the ion population prior to injection into the orbitrap mass analyser. The ability of

the C-trap to store ions combined with the mass selection capability of the quadrupole allows for multiplexing of different ion populations (for example, fragment ions of two or more distinct precursor ions) (Scheltema et al., 2014). As ions from the C-trap enter the orbitrap analyser off-axis, the axial oscillations are initiated without the need for additional excitement, facilitating the use of Fourier transform on the image current to produce the corresponding mass spectrum. As the orbitrap is accessible to ions before fragmentation, it is able to analyse both precursor and product ions.

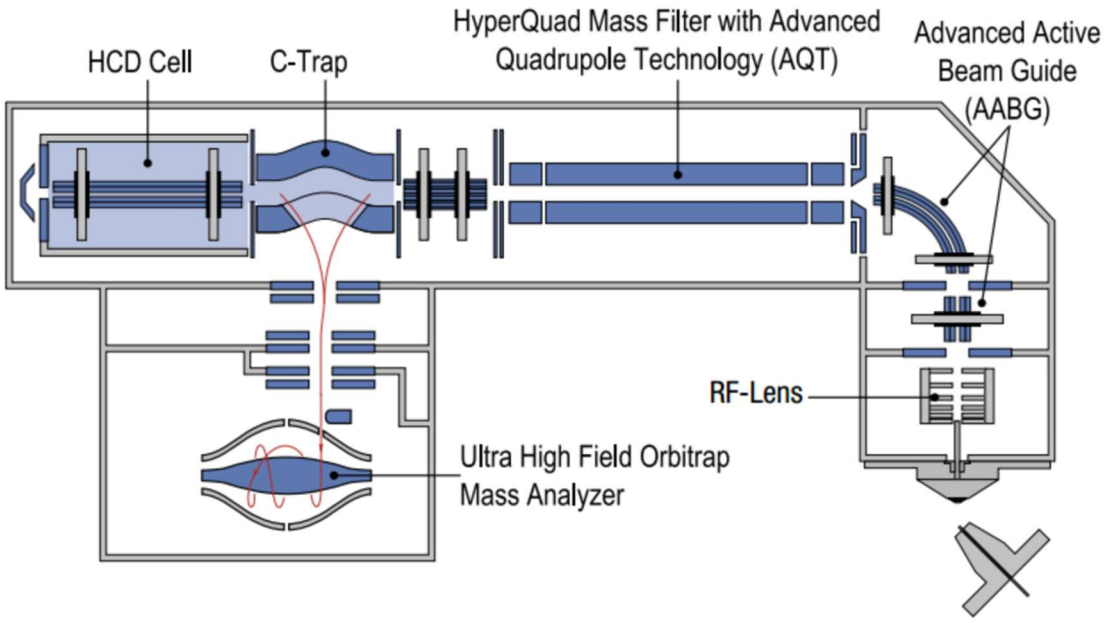



**Figure 1.26) ThermoScientific Q-Exactive HF mass spectrometer**

This hybrid instrument contains a quadrupole mass filter, a curved ion trap, a high energy collisional dissociation cell and an ultra high field orbitrap mass analyser. Figure supplied by the vendor, available at http://www.thermofisher.com.

### 1.5.5 Data processing

In a typical MS proteomics experiment, the tandem mass spectra containing peptide fragment ion information, together with peptide ion mass and intensity data are searched against protein sequence databases. Peaks are extracted from the raw data, and the corresponding mass values are counted or scored in a way that allows the best-matching peptide or protein to be identified. Results are typically exported in a table format containing a list of identified proteins with the corresponding protein identification score and/or an indication of probability of correct 'identification'. Relative protein abundance information may also be recorded. If stable isotopes were used, peptide (and ultimately thus protein) ratios are determined using the ratio of heavy to light isotopes in the same run. Data collected from various MS experiments must be analysed in such a way that the number of true peptide identifications are optimised and the number of false identifications made are reduced as well as ensuring the measurements made in each replicate are comparable to others. A number of bioinformatics data processing tools are available for use depending on the experiment performed: SearchGUI using the PeptideShaker platform for untargeted experiments (Vaudel et al., 2011); MaxQuant for the analysis and quantification of large scale mass-spectrometric data (Cox and Mann, 2008); mProphet for the analysis of SRM data for peptide identifications (Reiter et al., 2011); as well as vendor-specific data analysis software including Protein Lynx Global Server (PLGS, Waters).

A common theme to all bioinformatics data processing tools is the searching of peptide spectra against a sequence database. The masses calculated from the spectra are matched to known masses of proteolytic peptides in the database. However, additional parameters are included in the database search to improve the likelihood of true matches identified, with parameters unique to the processing tool utilised. Firstly, the number of missed cleavages within the peptide determines which spectra it is matched to. A missed cleavage occurs when the protease is unable to cleave at a known cleavage site due to steric hindrance, or the presence of flanking residues which may compromise catalytic efficiency due to electrostatic interactions, as is discussed by Lawless and Hubbard (Lawless and Hubbard, 2012). Secondly, post-translational modifications and modifications due to chemical derivatisation contribute to the complexity of the search, as there is often uncertainty about whether a modification is present or not. There are three classes of modification: modification at a specific amino acid residue at a specific terminus; modifications that affect one or other of the termini independently of the residues located there; and modifications that affect specific amino acid

residues independently of its location within the peptide (Perkins et al., 1999). Such modifications may be stoichiometric or non-stoichiometric. As an example of non-stoichiometric modification, a peptide with three methionine residues may be subject to oxidation at each methionine residue, with each residue equally susceptible. However, all instances of the peptide may not be modified on the same methionine residue giving rise to different tandem mass spectra from precursor ions of the same *m/*z. Processing software attempts to match the experimental data to the sum of the contributions from all possible modifications of this peptide that fall within the mass window allocated in order to provide an accurate match. Thirdly, the mass accuracy window - the difference between the monoisotopic calculated exact mass and the measured accurate mass, supplied as a parameter to the database search undoubtedly affects the proportion of true identifications made. If the mass accuracy window is set too wide, the identification statistics are observed as worse than they need to be, whilst if the mass accuracy window is set too narrow, the data is subject to false negatives, as true identifications are missed (Cox and Mann, 2009). A final parameter for MS/MS data is the ion series present in the spectra. Fragment ion data must be matched to calculated values for an ion series that is well represented in the experimental data. The ion series produced is dependent upon the experimental conditions (Perkins et al., 1999). Additionally, precursor ions with high charge states (over 2+) have a potential for multiple charge states for each ion series produced, complicating the accuracy of the matching algorithm.

The likelihood of a 'correct' peptide match is assessed through the use of randomised or reverse database searches, to give indication as to the accuracy of the identifications made by each processing tool (Elias and Gygi, 2010). Specifically, MaxQuant and mProphet use the false discovery rate (FDR) approach (Benjamini and Hochberg, 1995) to estimate the number of incorrect target peptide spectrum matches (PSMs). In the target-decoy competition method by Elias and Gygi (Elias and Gygi, 2010) a concatenated database comprising of target and decoy sequences is created and used to search spectra. For this database, it is generally assumed that there is a 1:1 correlation between target and decoy sequences (for every target sequence, there is a reversed decoy sequence). Every spectrum is searched against a database containing both target and decoy sequences to identify potential matches. Individual search engines assign an internal score to each candidate peptide spectrum match and typically, the top-scoring PSM is retained. Globally, all PSMs are then ranked and this list is used to determine a threshold at which a predetermined FDR is reached.

In most cases, the ratio of target and decoy sequences in the database is 1:1; therefore there should be a 1:1 ratio of target and decoy hits for the random (most likely false) identifications. Each PSM is scored according to a confidence in the hit. Incorrect hits can also be scored highly. All peptide hits are sorted by descending score, and the number of true positive hits are estimated using Equation 1.2, where *TP* is the number of true positives, *FP* is the number of false positives, *T* is the total hits and *d* is the number of decoy hits.

$$TP + FP = T - d$$

*(Equation 1.2)*

As the ratio of target and decoy sequences in the databases is 1:1, one can assume that the number of decoy hits is equal to the number of false positives (FP = decoy hits). Therefore, the global FDR can be estimated using Equation 1.3, where *FP* is the number of false positives, and *TP* is the number of true positives (Jones et al., 2009).

$$FDR = \frac{FP}{TP + FP}$$

*(Equation 1.3)*

Usually, experimentalists are concerned only in the spectra that matches a target peptide. Therefore, the target-only variant of the target-decoy competition method removes all decoy identifications from the reported lists and adjusts the FDR estimate accordingly (Equation 1.4) (Keich et al., 2015). A score threshold is then selected based upon the desired statistic threshold. Typically, the score threshold is set so that it must pass a 1 % FDR (that is, only 1 in 100 hits are false positives). To assign statistical scores to each individual peptide spectrum match, a Q-value is used, defined as the minimal FDR value at which the dataset includes the PSM as a match.

$$FDR = \frac{FP}{TP}$$

*(Equation 1.4)*

Finally, for tools that are able to perform protein quantification, the relative abundance of peptides of a given length in a digest depends on the length of the peptide and thus emPAI score. In a tandem MS experiment, large peptides with many fragments of similar intensities are identified with high confidence, whilst shorter peptides yield few predominant fragment

ions resulting in more intense MS signals (Lange et al., 2008). Many tools incorporate complex algorithms that attempt to normalise these factors, as well as removing any systematic errors encountered during the experiment. Here, the processing tools MaxQuant and mProphet are further described.

#### 1.5.5.1 MaxQuant

Cox and Mann (Cox and Mann, 2008) created MaxQuant, a set of computational algorithms for the analysis of bottom-up proteomics data. MaxQuant utilises Andromeda (Cox et al., 2011), an algorithm that performs probability-based scoring when determining peptide identifications. Using probability based scoring methods allow for the use of a single rule to judge whether a result is significant or not, particularly when guarding against false positives. MaxQuant detects the peaks in each MS scan by fitting a Gaussian peak shape to the three central raw data points. The two-dimensional peaks are arranged into 3D peaks by joining adjacent MS scans along the *m/z*-retention time plane only if the centroid masses of the 2D peaks differ by less than 7 ppm. The centroid mass of the 3D peak is calculated using the intensity-weighted estimates of the 2D peaks. An edge is inserted into the 3D peak if two peaks in the 2D data differ in mass by the isotopic mass of an average amino acid within bootstrap errors. MaxQuant then determines the longest consistent sub-graphs, using the isotope patterns as a potent noise filter (‘de-isotoping’), to determine centroid masses of each peak (Figure 1.27).

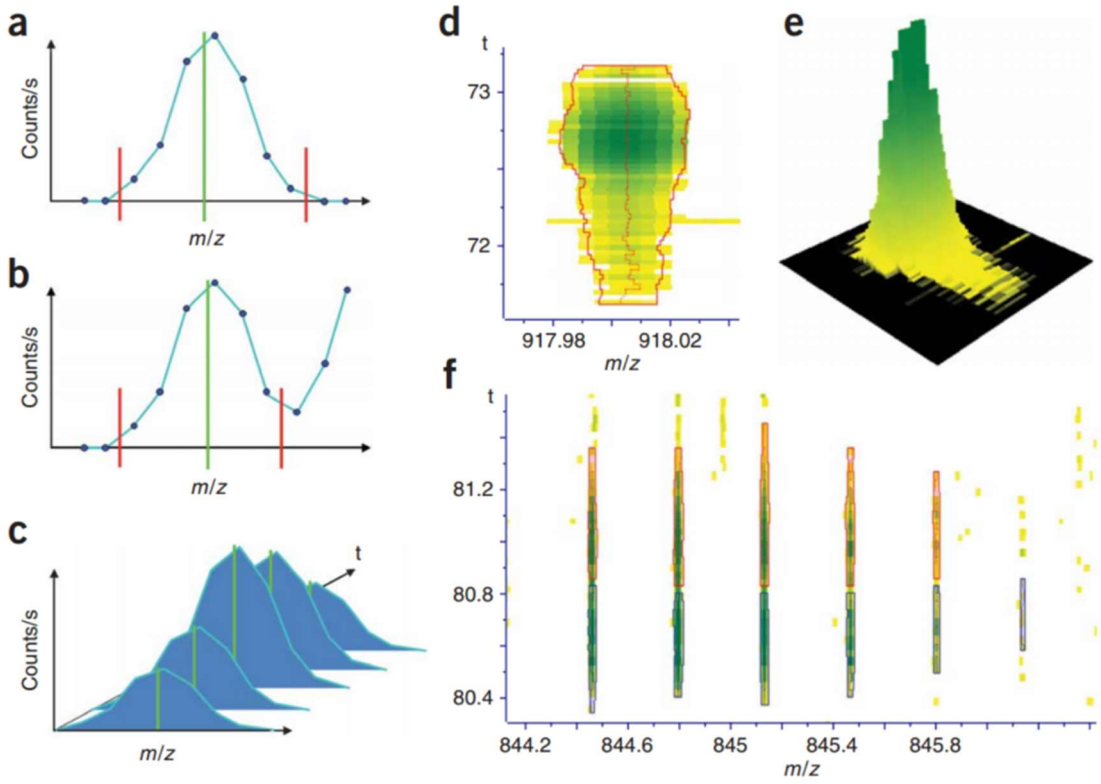



**Figure 1.27) MaxQuant peak detection and intensity-weighted mass calculation**

A) 2D feature extraction; B) 2D peaks broken up at local intensity minima; C) Alignment of adjacent MS scans in the *m/z*-retention time window plane; D and E) 3D peak construction, with green as the highest intensity peaks from the 2D MS spectra; F) Isotopic pattern achieved through de-isotoping of the 3D peak. Figure taken from the MaxQuant publication (Cox and Mann, 2008).

The centroided mass values with the associated intensity values are submitted to Andromeda for peptide identification by searching against a sequence database. Andromeda calculates the probability that the observed match between the experimental data set and each sequence database occurs by chance, with the match with the lowest probability reported as the best match. A significance threshold is used (typically $p < 0.05$), such that all matches with a probability score less than the threshold are accepted as true. The validities of the probabilities calculated are tested by repeating a search against a decoy database. In the decoy database, the overall amino acid composition, number of entries and distribution of entry lengths are identical to the original database but with random or reversed sequences. To increase the number of peptides that can be used for quantification, peptide identifications can be transferred to unsequenced or unidentified peptides by matching their mass and retention time, known as "match between runs". Here, different LC-MS runs are made comparable by retention time alignment in a hierarchical clustering method. The clustering typically connects LC-MS runs of the same or neighbouring fractions (if a fractionated experiment) or replicate runs, as these are most similar. Unidentified LC-MS features are then assigned to peptide identifications in other runs that match based on their accurate masses and aligned retention times.

For MaxQuant's label free quantification (MaxLFQ), fractionated data is adjusted using delayed normalisation (as I do not perform fractionation I do not review it here, a thorough description is available in the MaxLFQ publication (Cox et al., 2014)). For unfractionated samples, and normalised fractionated samples, pair-wise ratios between peptides common to a sample are used to determine MaxLFQ intensities (Figure 1.28). Briefly, the peptide ratio between any two samples is determined using only the peptide species that are present in both (Figure 1.28a-c). The protein pair wise ratio is determined by the median peptide ratios between the two samples. MaxQuant requires two peptide ratios in order for a given protein ratio to be considered valid. This results in the construction of a triangular matrix, containing all pair-wise protein ratios between any two samples, giving the maximal possible quantification information (Figure 1.28d). A least-squares analysis is performed to reconstruct the abundance profile satisfying the individual protein ratios based on the sum of squared differences (Figure 1.28e) before the whole profile is rescaled to the cumulative intensity across samples, preserving the total summed intensity for a protein over all samples (Figure 1.28f) (Cox et al., 2014).

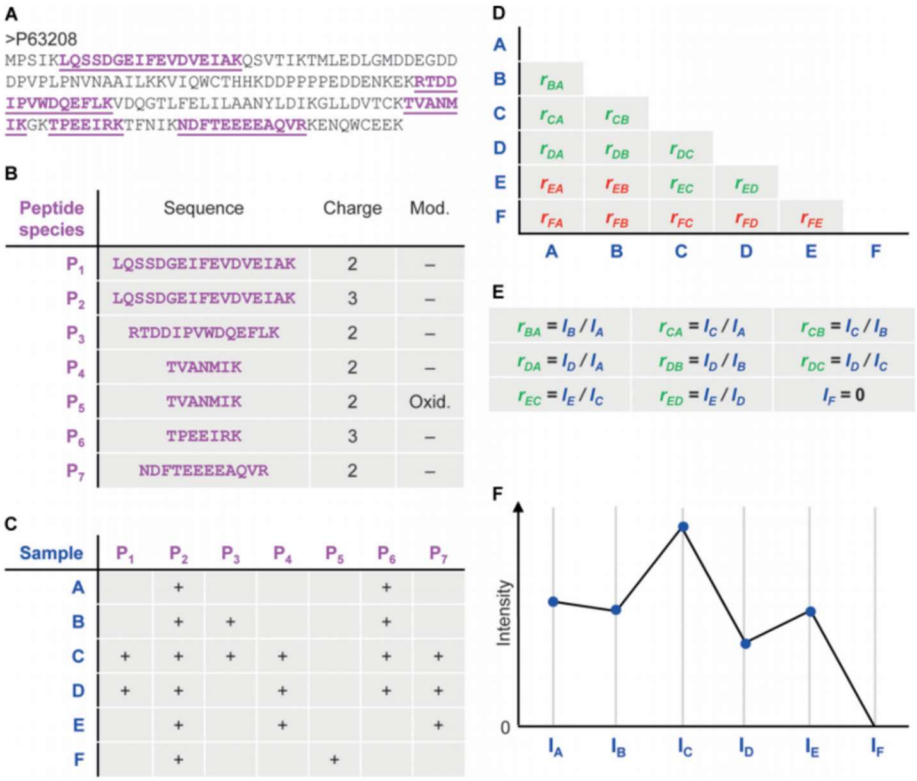



**Figure 1.28) The MaxLFQ algorithm**

A) Example protein sequence; B) Peptide species identified in the resulting mass spectra across all samples; C) For each sample (A-F), the presence of each peptide species is determined; D) Pairwise peptide ratios are determined for samples containing the peptide species in both samples, with the pairwise protein ratio between two samples calculated as a median of all pairwise peptide ratios between the two samples; E) System of equations solved for optimisaton of the abundance profile to satisfy all protein ratios; F) Resulting protein abundance profile for one protein. Figure taken from the MaxLFQ publication (Cox et al., 2014)

### 1.5.5.2 mProphet

In SRM experiments, only sets of predefined peptides are detected and quantified. However, to ensure reliable quantification, collected spectra must be confidently assigned to the target peptide. Although graphical user interface software such as Skyline exists to visualise and determine the quality of spectra through a dot product score, Skyline does not provide confidence scores in the form of FDRs when assigning spectra. Using a decoy-transition approach, mProphet assigns a confidence measure to each peak group for quality control. Peak groups are collections of transitions targeting the same peptide, with the transition peaks sharing retention time profiles and peak shapes (Figure 1.29). These features are assessed, resulting in a series of sub-scores for each peak group.

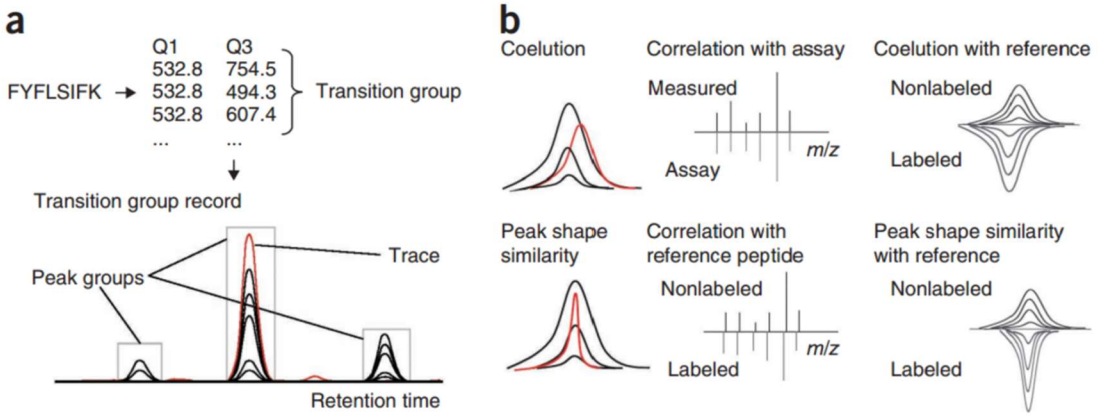



**Figure 1.29) mProphet peak group scoring parameters**

A) Transitions targeting a single peptide are defined as a transition group, for which the corresponding peaks should share the same features (retention time profile and shape) as a peak group. B) If any standard peptides are used, the peak group for the standard peptide will share the same retention time profile and peak shapes as the target peptide, however will differ in *m/z* according to the label. Figure taken from the mProphet publication (Reiter et al., 2011).

mProphet uses a decoy approach to determine the statistical error rates. However, unlike database searching where spectra are searched against decoy sequences, mProphet assigns scores to the transition decoy list gathered during the SRM experiment. This decoy transition list is generated by adding random integers to the precursor and product *m/z* values of the target transition list. Typically, a random integer between 3 and 10 is subtracted from the precursor *m/z* whilst for the product *m/z* a random integer between -5 and 5 is added, such that the difference in *m/z* between precursor and product is *m/z* >5. All other attributes, for example the retention or dwell time, are copied from the corresponding target transition groups. In this way, the decoy transitions are different from classical background noise, instead they are real signals from peptides that have similar precursor masses and have fragment ions in the selected *m/z* window. Each decoy peak group is given a score using the same method as for target peak group scoring. The target and decoy scores are then combined and split into a training dataset and a test dataset. Sets of true (high scored) and false (decoy) peak groups from the training set are used to refine the model iteratively, such that the weights of the scores are converged to a point that there is discrimination between the true and false peak groups. The peak groups are then reranked between each transition group record, and only the top-ranked peak group is considered as a true peak. The normalised discriminant scores (mProphet score) of the top ranked peak groups is used to derive the FDRs across the dataset. Again, only peaks with a score that passes a 1 % FDR are given as true peaks.

### 1.5.6 The CoPY Project

The Census of the Proteome of Yeast (CoPY) project has been undertaken by those in the laboraties of Prof. Claire Eyers (University of Liverpool), Prof. Rob Beynon (University of Liverpool), Prof. Simon Hubbard (University of Manchester) and Prof. Chris Grant (University of Manchester), often cited here as "Brownridge et al.". In this project, the group sought to develop QconCATs for absolute quantification of the entire proteome of yeast under normal conditions with the most recent development resulting in copy per cell values for roughly 1800 proteins, obtained through the use of up to two Q-peptides per target protein in an SRM strategy (Lawless et al., 2016). As part of this work, modelling approaches could be used to elicit the role of different classes of protein on the cell, with focus upon the chaperones and their clientele published in 2013 (Brownridge et al., 2013). This work highlighted the role of individual classes of chaperone, with conclusions including small type chaperones being the most abundant under chemostat conditions and the CCT complex interacting with approximately 20% of the yeast proteome. Absolute quantification was achieved for 51 of the 63 chaperones, with chaperone abundances ranging from 250 to 440,000 copies per cell. These chaperone-client interaction studies were also published by Lawless and Hubbard (Lawless and Hubbard, 2014) describing in further detail the methodology. In terms of QconcAT design, the same groups performed studies on the features of a peptide to be detectable in an LC-ESI-MS experiment (Eyers et al., 2011) and for a peptide to be fully cleaved by trypsin; both of which are predictor tools that are utilised within this work towards selection of suitable Q-peptides, however were not used towards Q-peptide selection in the 2013 chaperome publication (Brownridge et al., 2013).

### 1.5.7 Aims and Objectives

To extend the work on chaperones by Brownridge and colleagues, I elected to design a series of QconCATs targeting each of the 63 chaperones in *S. cerevisiae*, using up to five Q-peptides per protein where possible, and considering peptides according to the predictor values for detectability in an LC-ESI-MS experiment and propensity to be miscleaved. By increasing the number of Q-peptides targeting each protein, more than 51 chaperone proteins may be quantifiable, if previous quantification was not possible due to issues with detectability of both the standard peptides targeting the chaperone. Secondly, using additional Q-peptides per protein will allow for a more robust measurement of the copy per cell value, as multiple values are likely to center around a single value.

Samples grown under conditions of normal growth and heat shock (42 °C, 30 minutes) will be analysed using the QconCAT strategy to determine absolute copy numbers for the targeted chaperones, and by label free experiments to determine quantitative values for the proteome. This data will be used to inform modelling approaches to elicit the chaperone, proteome and 'chaperome' response to heat shock. In order to obtain copy per cell values for all proteins, modelled copy per cell values will be obtained via a regression between the absolute copy per cell values and the label free intensities observed for chaperones. Copy per cell values for chaperones and their client proteins under both conditions will be applied to the modelling approaches used by Lawless and Hubbard (Lawless and Hubbard, 2014) to establish any significant changes in the network as a result of heat shock.

# Chapter 2 - Materials & Methods

*Sections of these methods were originally published in the journal 'Proteomics': Mackenzie, R. J., Lawless, C., Holman, S. W., Lanthaler, K., Beynon, R. J., Grant, C. M., Hubbard, S. J. & Eyers, C. E. 2016. Absolute protein quantification of the yeast chaperome under conditions of heat shock. Proteomics DOI: 10.1002/pmic.201500503. Full details of buffer solutions are available in Appendix 1.*

## 2.1 Design of chaperone QconCATs (ChapCATs)

The sequences of 63 known chaperones in *S. cerevisiae* were subjected to *in silico* tryptic digestion, and the limit peptides analysed for suitability as Q-peptides for chaperone QconCATs (ChapCATs), according to criteria outlined previously (Brownridge et al., 2012, Carroll et al., 2011, Holman et al., 2012). Peptides must be sequence-unique to the protein and proteome under investigation, and not known to be post-translationally modified according to dbPTM (http://dbptm.mbc.nctu.edu.tw/) (Lee et al., 2006). Propensity for the peptide to undergo missed cleavage in the native protein sequence was evaluated using MC:pred (http://king.smith.man.ac.uk/mcpred/), recording scores for both the *N*-terminal and *C*-terminal bond (Lawless and Hubbard, 2012). Likelihood of detection in a liquid-chromatography-electrospray ionisation MS (LC-ESI-MS) experiment was assessed via CONSeQuence (Eyers et al., 2011), available at http://king.smith.man.ac.uk/CONSeQuence/. Candidate peptides were omitted if their sequence contained any of the following features: dibasic sequences; Asn-Gly motifs or contiguous Gln (2-5) residues; < 5 amino acids; or were reported to have a post-translational modification. Brownridge and colleagues (Brownridge et al., 2013) designed QconCATs towards 63 chaperones, with two Q-peptides elected per target chaperone. From this set, 31 Q-peptides were retained for use in the final ChapCAT set. A combination of CONSeQuence score and MC:pred scores enabled potential Q-peptides to be ranked, with the top five (where possible) Q-peptides selected per chaperone. For the chaperone proteins Hsp31, Sno4 and Hsp33, no unique quantotypic peptides (*i.e.* fully tryptic peptides suitable for use as quantification standards) were identified. Therefore, non-unique Q-peptides were selected representing the summed protein group. Non-unique but potential Q-peptides were also observed for the protein pairs Ssa1:Ssa2 and Ssb1:Ssb2 and selected as quantification standards owing to few unique alternatives; a combination of both unique and non-unique Q-peptides were used to improve quantification reliability.

Q-peptides were assigned to a ChapCAT such that an individual ChapCAT targeted chaperones of the same general chaperone class, as used by Gong and colleagues (Gong et al., 2009). A total of 10 ChapCATs were designed, each targeting 6 to 8 chaperones and containing 25-37 Q-peptides. The constituent Q-peptides were concatenated *in silico* within a ChapCAT for maximal likelihood of completion of tryptic cleavage, determined using MC:pred (Lawless and Hubbard, 2012). The concatenated Q-peptides were surrounded by sequences required for the quantification and purification of the ChapCAT construct. At the *N*-terminus of each ChapCAT construct, a sacrificial peptide is added (MAGR) followed by a glu-fibrinopeptide sequence (EGVNDNEEGFFSAR). At the *C*-terminus of the construct, a fibrinopeptide sequence (GVNDNEEGFFSAR) and a sacrificial peptide and histidine purification tag is added (LAAALEHHHHHH). The final ChapCAT design was used to direct the design of a gene, codon-optimised for expression in *E. coli*; the gene was then ligated into the *Nde1* and *BamHI* sites of a pET21a expression vector to yield the ChapCAT plasmid with ampicillin resistance (PolyQuant GmbH, Germany).

## 2.2 Expression and purification of ChapCATs in *E. coli*

### 2.2.1 Preparation of *E. coli* glycerol stocks

*E. coli* BL21 (λDE3) $F^-$-*dcm ompT hsdS*$_B$ ($r_B^-$ $m_B^-$) *gal* were transformed with 5 ng of ChapCAT plasmid DNA as previously described (Brownridge et al., 2012). Cells were grown on Luria Agar (LA) (LA as in Appendix 1) plates containing 50 µg $mL^{-1}$ ampicillin and incubated overnight at 37 °C. Single colonies were spread on fresh LA plates containing 50 µg $mL^{-1}$ ampicillin. 2 mL of pre-warmed luria broth (LB) containing 100 µg $mL^{-1}$ ampicillin was inoculated with a single transformed colony from the inoculation plate culture and grown for 8 hours (37 °C, 160 rpm). To prepare the glycerol stocks, 1 mL of culture was added to 0.5 mL sterile 60% (v/v) glycerol, vortexed and stored at -80 °C.

### 2.2.2 Expression testing of ChapCAT constructs in Luria Broth and Minimal Media

To test protein expression from the ChapCAT constructs in *E. coli*, 2 mL of prewarmed LB (LB as in Appendix 1) containing 100 µg $mL^{-1}$ ampicillin and 1 % glucose (w/v) was inoculated with each of the ChapCAT glycerol stocks from -80 °C storage and grown overnight (37 °C, 160 revolutions per minute (rpm)). One mL from the overnight cultures was diluted in 50 mL LB

containing 100 µg $mL^{-1}$ ampicillin and grown at 37 °C for 2 hours at 160 rpm until an $OD_{600}$ of 0.6 was reached as determined by an Eppendorf BioPhotometer Plus (Thermo Fisher Scientific, Altringham, UK). Expression of ChapCAT protein was induced via addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and cells grown for a further 5 hours at 37 °C, 160 rpm). Samples were taken at time intervals prior to IPTG addition and activation of ChapCAT expression, with a final sample taken 5 hours after expression induction. ChapCAT002, ChapCAT006 and ChapCAT010 were subjected to a Q-peptide reshuffle in order to maximise likelihood of generating detectable protein and the constructs resynthesised due to a lack of expression when testing in LB media (PolyQuant, Germany).

Upon successful expression in LB media, ChapCAT glycerol stocks were used to inoculate 5 mL starting minimal media (minimal media as in Appendix 1) containing 100 µg $mL^{-1}$ ampicillin and grown overnight at 37 °C, 160 rpm. Four mL of culture was used to inoculate 200 mL M9 minimal media lacking stable isotope labelled amino acids (L-Lysine and L-Arginine) and grown to an $OD_{600}$ of 0.6 before being induced with 1 mM IPTG and grown for a further 5 hours at 37 °C. Expression was assessed via SDS-PAGE and western blot analysis. ChapCAT002 and ChapCAT009 expression levels were optimised by varying the $OD_{600}$ at which IPTG was added (0.6 or 0.8); addition of 10 mM benzyl alcohol 20 minutes prior to IPTG addition; varying the IPTG concentration (0.4 mM or 1 mM) used for induction and the incubation time following induction (3 or 5 hours).

### 2.2.3 SDS-PAGE and Western Blot analysis

Samples were mixed 1:5 with 5X SDS sample buffer containing 5% β-mercaptoethanol (5X SDS sample buffer as in Appendix 1) and heated (95 °C, 5 min). Samples were run in 10 or 15 well NuPage 4-12% 1.0 mM Bis-Tris Precast Gels (ThermoFisher Scientific, UK) or on 12 % polyacrylamide gels. Molecular weight markers used were in the range 10 to 260 kDa (CAT: 1610375, CAT: 1610375, Bio-Rad, UK).

Proteins were transferred to Amersham Protran 0.45 µm Whatman nitrocellulose membrane (GE Healthcare, UK) using a XCell II Blot Module (Life technologies) in transfer buffer for 1 hour at 35 V (transfer buffer as in Appendix 1). The membrane was incubated in 1 X PBS-Tween and 5 % (w/v) milk powder for 1 hour, washed 3 times in 1 X PBS-Tween before being incubated with 1:5000 primary anti-His (Mouse) antibody:1X PBS-Tween (CAT 05-949-KC, Merck, Millipore, Hertfordshire, UK) (10 X PBS-Tween as in Appendix 1). The membrane was

then washed 3 times for 15 minutes in 1X PBS-Tween buffer before being incubated in 1:2000 secondary anti-Mouse (Goat) antibody:1X PBS-Tween (CAT: A4416, Sigma-Aldrich, Merck, Millipore, Hertfordshire, UK). ChapCAT expression was confirmed via chemiluminescent signal using SuperSignal™ West Pico Chemiluminescent Substrate (CAT: 34078, ThermoFisher Scientific, UK) on film.

### 2.2.4 Heavy labelled ChapCAT expression and purification

ChapCAT glycerol stocks were used to inoculate 5 mL starting minimal media containing 100 µg $mL^{-1}$ ampicillin and grown overnight at 37 °C, 160 rpm. 4 mL of culture was used to inoculate 200 mL M9 minimal media (minimal media as in Appendix 1) containing 10 mg $mL^{-1}$ of the stable isotope labelled amino acids (L-Lysine and L-Arginine) and grown according to their appropriate conditions as determined via expression testing (Section 2.2.3). Following IPTG induction and growth of the heavy labelled cultures, ChapCAT proteins were purified from the *E. coli* inclusion bodies (IB) as previously described (Brownridge et al., 2012). Briefly, cultures were centrifuged at 3500 rpm for 15 minutes at 4 °C. Pellets were resuspended carried in 2.5 mL Binding buffer (Binding buffer as in Appendix 1) and subject to sonication on ice at 7 X 10 seconds with 20 seconds rest. Cell lysate was centrifuged (10,000 g, 30 minutes, 4°C) with the soluble fraction (SF) retained for analysis and the pellet containing IB resuspended in 1 mL Binding buffer. Centrifugation (10,000 xg, 30 minutes, 4 °C) and resuspended in 500 µL guanidine buffer (guanidine buffer as in Appendix 1) isolating the IB proteins (including ChapCAT) and debris. Centrifugation (10,000 xg, 30 minutes, 4 °C) yielded ChapCAT proteins isolated within the supernatant (IB fraction).

The ChapCAT proteins were purified from other IB proteins using Ni-MAC™ His-Tag purification columns (CAT: 71658, Novagen, Merck, Millipore, UK), as previously described (Brownridge et al., 2012), utilising the histidine purification tag at the *C*-terminus of the construct. Briefly, Ni-NTA spin columns were equilibrated with 600 µL of guanidine buffer and centrifuged (1600 rpm, 2 minutes, 4 °C). The inclusion body supernatant fraction was loaded and centrifuged (1600 rpm, 5 minutes, 4 °C) and the flow through ('FT') collected. Columns were washed three times with 600 µL guanidine buffer containing 10 mM Imidazole, with centrifugation (1600 rpm, 2 minutes, 4 °C) after each wash. ChapCAT was eluted from the column in two elution steps with 300 µL elution buffer containing 250 mM imidazole (elution buffer as in Appendix 1) with centrifugation after each elution step (1600 rpm, 2 minutes, 4 °C). For each ChapCAT, elution fractions were pooled and dialysed in Amicon Ultra 2 mL MWCO 30

kDa membrane (MilliPore, CAT: UFC203024PL) against 50 mM ammonium bicarbonate overnight at 4 °C before being concentrated through a 0.5 mL 3K Amicon Ultra Centrifugal Filter (CAT UFC500308, Merck, Millipore) to a final volume of ~200 µL. 10 µL samples were taken for SDS-PAGE and Coomassie stain analysis as described above to determine approximate ChapCAT concentrations against Bovine Serum Albumin standards (Sigma). To each dialysed sample, equivalent volumes of RapiGest$^{TM}$ (Waters, UK) were added to a final concentration of 0.2 % (w/v), with the final ChapCAT concentration being half that approximated.

## 2.3 Preparation of *S. cerevisiae* samples

*S. cerevisiae* (EUROSCARF accession number Y11335 BY4742; *Mat ALPHA*; *leu2Δ0*; lys2*Δ0*; *ura3Δ0*; *arg3::KanMX4*) was grown in C-limited F1 medium (Hayes et al., 2002), such that 10 $gL^{-1}$ of glucose was the only carbon source. To meet auxotrophic requirement of the strain, 0.5 mM arginine and 1 mM lysine were introduced into the F1 medium. A 5 mL pre-culture inoculated with a single *S. cerevisiae* colony was incubated at 30 °C for 24 hours prior to inoculation of eight biological replicates of 50 mL F1 medium. Samples were grown overnight (30 °C) to an $OD_{600}$ of 2. To prepare the heat shock (HS) samples, four of the eight biological replicates were removed and placed in a water bath with shaking at 42 °C for 30 minutes (Saavedra et al., 1996). Subsequently, individual samples were aliquoted (15 mL) and cell counts recorded using an Auto M10 Cellometer® (Nexcelom, Manchester) prior to centrifugation (4000 rpm, 4 °C, 10 minutes). The pellet was resuspended in 250 µL breakage buffer (50 mM ammonium carbonate with 1 EDTA-free protease inhibitor tablet per 10 mL of buffer). An equivalent volume of acid washed glass beads were added and sample subject to bead beating at 4 °C for 15 X 30 seconds with 1 minute rest. Lysed sample was then subject to centrifugation (13,000 rcf, 4 °C, 10 minutes), supernatant stored on ice, and then pellet resuspended in 250 µL fresh breakage buffer. Using a hot needle to pierce the extraction vial, the vial was centrifuged over a fresh eppendorf tube (2000 rpm, 4 °C, 15 minutes) and the flow through collected and combined with the supernatant from the previous step. The total volume was determined and aliquoted such that each aliquot contained approximately 25,000,000 cells. For label-free quantification, the protein concentration was determined by Bradford Assay (Multiskan FC Microplate Photometer, Thermo Scientific, UK). Aliquots were frozen at -80 °C.

## 2.4 ChapCAT-only digestion

In one eppendorf tube, 25 pmol each of ChapCAT 1, 2, 3 and 4 were made to a final volume of 40 µL with 25 mM Ammonium Bicarbonate. To a second eppendorf tube, 25 pmol each of ChapCAT 5, 6, 7, and 8 were made to a final volume of 40 µL with 25 mM ammonium bicarbonate. The ChapCAT mixes within each tube were subject to tryptic digestion. The proteins were denatured using 2.5 µL of 1 % (w/v) Rapigest™ (Waters, Manchester, UK) in 25 mM ammonium bicarbonate, followed by incubation at 80°C for 10 minutes. Samples were then reduced via addition of 2.5 µL of 9.2 mg $mL^{-1}$ DTT and incubation at 60 °C for 10 minutes; and alkylated with 2.5 µL of 33 mg $mL^{-1}$ iodoacetamide and incubated at room temperature in the dark for 30 minutes. Trypsin (Sigma, Poole, UK, proteomics grade) was reconstituted in 50 mM acetic acid to a concentration of 200 µg $mL^{-1}$ and 2.5 µL added to the sample followed by an incubation at 37 °C for 4.5 hours. An additional 2.5 µL of 200 ug $mL^{-1}$ reconstituted trypsin was added to the sample after 4.5 hours and incubated at 37 °C for a further 16 hours (overnight). The digestion was terminated and RapiGest™ was removed by acidification with 3 µL of trifluoroacetic acid and incubation at 37 °C for 45 minutes followed by centrifugation (15,000 xg for 15 minutes).

## 2.5 ChapCAT-analyte mixture digestion

Each 25,000,000 cell aliquot as produced in Section 2.3, was mixed with 22.5 pmols of ChapCAT and made up to 150 µL by addition of 25 mM ammonium bicarbonate. The proteins were denatured using 10 µL of 1 % (w/v) Rapigest™ (Waters, Manchester, UK) in 25 mM ammonium bicarbonate, followed by incubation at 80°C for 10 minutes. Samples were then reduced via addition of 10 µL of 9.2 mg $mL^{-1}$ DTT and incubation at 60 °C for 10 minutes; and alkylated with 10 µL of 33 mg $mL^{-1}$ iodoacetamide and incubated at room temperature in the dark for 30 minutes. To allow quantification of the ChapCAT, 10 µL of 2.15 pmol $µL^{-1}$ glu-fibrinopeptide (Waters, UK) was added to each digest. Trypsin (Sigma, Poole, UK, proteomics grade) was reconstituted in 50 mM acetic acid to a concentration of 200 µg $mL^{-1}$ and 10 µL added to the sample followed by an incubation at 37 °C for 4.5 hours. An additional 10 µL of 200 ug $mL^{-1}$ reconstituted trypsin was added to the sample after 4.5 hours and incubated at 37 °C for a further 16 hours (overnight). The digestion was terminated and RapiGest™ was removed by acidification with 3 µL of trifluoroacetic acid and incubation at 37 °C for 45 minutes followed by centrifugation (15,000 xg for 15 minutes).

To determine digestion efficiency in both the ChapCAT standard and analyte, yeast aliquots containing the equivalent of 25,000,000 cells and 22.5 pmoles of ChapCAT (as determined approximately by comparison with BSA standards following SDS-PAGE analysis) were subjected to tryptic digestion as described in the previous paragraph. At 0 min, 1 min, 2 min, 5 min, 10 min, 20 min, 50 min, 120 min, 240 min, 270 min (enzyme top up) and 1230 min a 10 µL portion of sample was removed and incubated with 10 µL of 5 % (v/v) TFA to terminate proteolysis.

## 2.6 Mass Spectrometry

### 2.6.1 $MS^E$ experiments

To determine efficiency of isotope labelling to select the top 7 transitions, each ChapCAT digest was analysed by LC-MS using a nanoAcquity UPLC™ system (Waters, Manchester) coupled to a Synapt™ G2-Si mass spectrometer (Waters, Manchester) in $MS^E$ mode. One µL of sample (corresponding to approximately 50 fmol) was loaded onto the trapping column (Symmetry C18, 5 µm packing material, 180 µm x 20 mm, Waters, Manchester) using partial loop injection, for 3 minutes at a flow rate of 5 µL $min^{-1}$ with 99.9 % A (0.1 % formic acid) : 0.1 % B (99.9% ACN, 0.1 % formic acid). The sample was resolved on an analytical column (nanoAcquity UPLC© HSS T3 C18 75 µm x 150 mm, 1.7 µm column, Waters, Manchester), using a gradient of 97 % A : 3 % B to 60 % A : 40 % B over 60 minutes; then washed 60 % A : 40 % B (DIA manner) applying a collision energy ramp of 15 to 40 eV for increased energy scans. The data was processed and database searched using ProteinLynx Global Server v2.5 (Waters), using a low energy threshold of 100 and an elevated energy threshold of 20, with the processed spectra searched against a database created from the sequences of ChapCAT 1 to 8, using fixed modifications for carbamidomethyl modification of cysteine and $^{13}C_6$ modification of arginine and lysine. The database search results were converted into a spectral database using Skyline (MacLean et al., 2010) with the top seven most intense product ions identified according to both the $MS^E$ spectrum results and identification in various downloaded spectrum libraries: the yeast SRM Atlas (available at http://www.srmatlas.org/) library, the yeast NIST library (available at http://chemdata.nist.gov/), and the yeast GPM library (http://www.thegpm.org/). Specifically, tryptic peptides generate doubly-charged precursor ions, unless they contain a histidine or internal [K/R]P sequence facilitating the production of triply charged ions. Under CID conditions, only y and b ion series types are prevalent, with y ions being predominant (Holman et al., 2012). Series product ions (of $m/z$ >400) were thus preferentially selected, and

only b ions that were observed in the $MS^E$ experiment (regardless of visualisation in the spectral libraries) may be selected.

To check efficiency of digestion of yeast and to quantify the individual ChapCATs, each ChapCAT-analyte digest was analysed by LC-MS using a nanoAcquity UPLC™ system (Waters, Manchester) coupled to a Synapt™ G2-Si mass spectrometer (Waters, Manchester) in $MS^E$ mode. One µL of sample (corresponding to approximately the protein equivalent of 100,000 cells) was loaded and run using identical instrument parameters to those described above. The data was searched against a sequence database created from the sequences of ChapCAT001 to ChapCAT008, with fixed modifications for carbamidomethylation of cysteine and $^{13}C_6$ labelling of lysine and arginine using ProteinLynx Global Server v2.5 (Waters, Manchester). The ChapCAT was quantified via integration of the extracted ion chromatogram (XIC) of the ChapCAT heavy glu-fibrinopeptide standard (*m/z* 788.8) compared to the exogenously added light internal standard glu-fibrinopeptide (*m/z* 785.8) (Equation 2.1).

$$ChapCAT\ Concentration\ (fmol\ \mu L^{-1}) = \frac{Heavy\ signal\ intensity\ *\ light\ concentrati}{light\ signal\ intensity}$$

*(Equation 2.1)*

### 2.6.2 SRM experiments for absolute quantification

Each digested ChapCAT in a NG background and HS background was analysed using a NanoAcquity UPLC™ system coupled to a Xevo TQ(-S) triple quadrupole mass spectrometer (Waters, Manchester) in which the top seven most intense fragment ions as identified using database searches above were targeted. The mass spectrometer was operated in unscheduled SRM mode with Q1 and Q3 operating at unit resolution. Based on these analyses, the three transitions with the greatest S/N ratio (as calculated in Skyline (MacLean et al., 2010)) were selected for the final scheduled SRM analysis and quantification and retention times scheduled. The instrument was set to acquire 15 data points over a 30 s chromatographic peak within a 3 minute window. The final transition list was divided to achieve a minimum dwell time of 50 ms and each NG and HS sample analysed with all subsequent transition lists. For final protein quantification, sample volumes containing the equivalent of 200,000 cells with 0, 0.1, 1 or 10 fmol (via serial dilutions) of ChapCAT were analysed. The sample closest to a 1:10 ratio between Q-peptide and analyte XIC was selected for final quantification. A set of decoy transitions,

created as described by Reiter and colleagues (Reiter et al., 2011) were run against the yeast digest under identical instrument parameters.

For the digestion time course, a scheduled SRM experiment was performed on sample volumes equivalent to 200,000 cells using a nanoAcquity UPLC™ system coupled to a Xevo™ TQS triple quadrupole mass spectrometer. Transitions and scheduling windows were identical to those used for absolute quantification.

### 2.6.3 Label-free experiment

Fresh digests (1 µg) of the same NG and HS samples (not containing ChapCATs) were subject to label-free quantification by DDA using a Dionex UltiMate™ 3000 HPLC system coupled to a Q-Exactive HF mass spectrometer with an EASY-Spray™ column and source (ThermoScientific, Hemel Hempstead). Sample was loaded onto the trapping column (ThermoScientific PepMap® 100 C18, 300 µm x 5 mm) using partial loop injection, for 7 minutes at a flow rate of 9 µL $min^{-1}$ with 97.9 : 2 : 0.1 % (v/v) $H_2O$ : ACN : TFA. The sample was resolved on an analytical column (EASY-Spray™ C18 75 µm x 500 mm, 2 µm column) using a gradient of 96.2 % A (0.1 % formic acid) : 3.8 % B (79.9 % ACN, 20 % $H_2O$; 0.1 % formic acid) to 50 % A : 50 % B over 90 minutes; then washed at 1 % A : 99 % B before re-equilibrating to starting conditions at a flow rate of 300 nL $min^{-1}$. Following a full MS scan between *m/z* 350 to 2000 (mass resolution of 60,000 FWHM at *m/z* 200), a DDA top-16 method $MS^2$ analysis was performed with a target value of 1 x $10^5$ ions determined with automatic gain control. Precursor ions were isolated with an isolation window of *m/z* 1.2, with scans acquired at a mass resolution of 30,000 FWHM at *m/z* 200 and dynamic exclusion of 20 s. Three biological replicates were analysed for NG samples (NG1, NG3 and NG4 – NG2 was lost) and four biological replicates (HS1, HS2, HS3, HS4) for the HS samples.

## 2.7 Data Processing and Analysis

To determine digestion efficiency and the generation of limit Q-peptides from the standard and the related endogenous analyte peptide, data acquired for each time point was processed with Skyline (MacLean et al., 2010) and a report detailing the total peak area for each peptide at each time point exported. Each peak area was transformed to a percentage by dividing by the maximum peak area for the label type for the particular peptide. The pseudo-first order rate kinetics were modelled using the 'nls' function in the statistical software package

R, as was previously carried out (Brownridge and Beynon, 2011). For each Q-peptide, the rate constant (*k*) was determined for both the standard and analyte, allowing for determination of the respective digestion half-lives (*ln2/k*). Digestion was deemed complete after five half-lives.

Reiter and colleagues' (Reiter et al., 2011) mProphet was used to determine peak areas for both the unlabelled target peptides and isotope-labelled ChapCAT Q-peptide internal standards; copies per cell (cpc) values were calculated using the measured area ratios and the known quantities of Q-peptides according to Equation 2. Production of decoy transitions and the subsequent quantification workflow is described in previous literature (Brownridge et al., 2013). To avoid observed issues with peak group detection, an in-house Perl script was developed that set a retention time window +/- 30 seconds *in silico* either side of the maximum peak intensity for the peak group, through curation of the merged target and decoy. mzXML files (Perl script available in Appendix 2).

The raw spectral data (complete with associated *m/z* values and peptide sequences) were input into Skyline (MacLean et al., 2010), an open source tool for the analysis of targeted data. Using Skyline's automatic peak selection, each peak selection for every Q-peptide was manually verified in every sample and biological replicate analysed. A report file from Skyline, containing the corresponding retention time and peak width for every optimally selected peak was exported. An in-house Perl script that narrowed the retention time window to 1 minute centred around the retention time of the peak selected by Skyline was created. mProphet first organised each raw file into separate transition groups, thus each group contained three heavy transitions for the standard Q-peptide, three light transitions for the analyte peptide and six decoy transitions (three decoy light and three decoy heavy). The in-house script matched the peptide transition group to the Skyline report file, and determined the new one minute retention time window using the mean retention time of the heavy and light transitions on a biological replicate basis, adjusting the raw files accordingly (Figure 2.1). Adjusted raw files were subject to reprocessing again with mProphet, such that the any peak groups considered by mProphet for scoring would be within one minute of the expected 'true' peak retention time. In this way, any decoy transitions that were still within one minute of the expected true peak could be considered to determine any false positive peak matches. All peak groups used for absolute quantification still passed a 1 % mProphet FDR using the adjusted retention time windows.

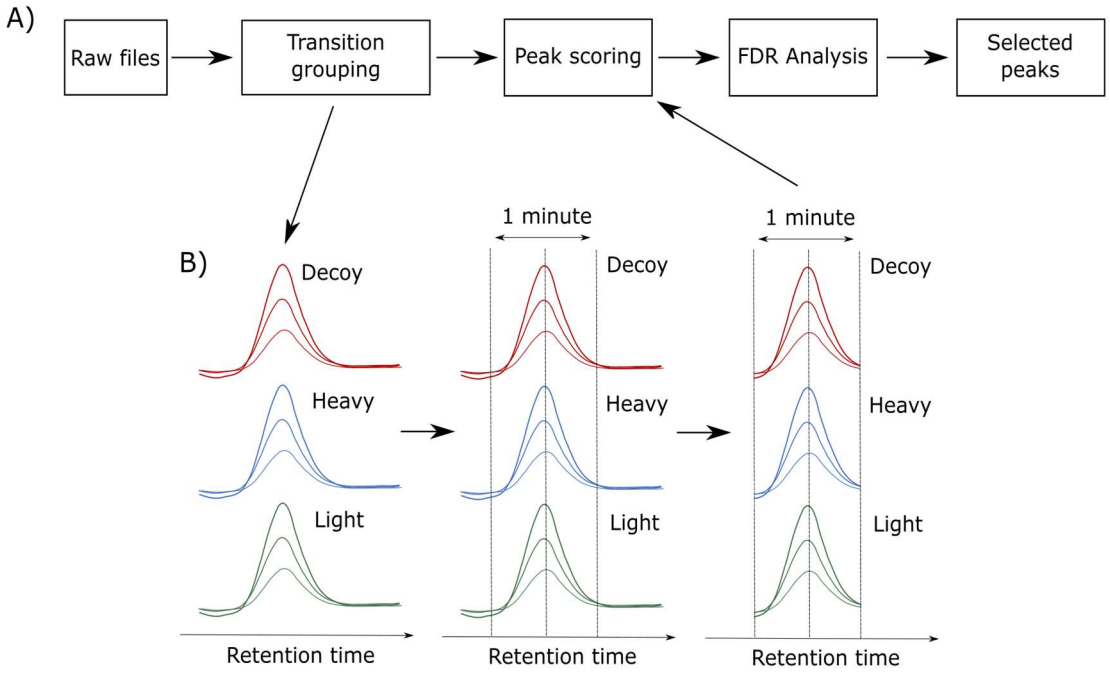



**Figure 2.1) *In silico* retention time window alteration**

A) The typical process for mProphet peak scoring and FDR analysis; B) Retention time windows of transition groups were altered to 1 minute using a skyline report file. Altered files were then processed as usual. Only three decoy transitions are illustrated.

The subsequent *.mzXML* files were then processed as previously described (Brownridge et al., 2013), with mQuest search parameters contained within the file 'copy.params' (Appendix 3). The resulting standard and analyte peak areas for a peptide were used to perform absolute quantification, resulting in a peptide cpc value (Equation 2.2).

$$CPC = \frac{\frac{Heavy\ peak\ area}{Light\ peak\ area} \times ChapCAT\ Concentration\ (mol) \times (6 \times 10^{23}\ copies)}{Cell\ count}$$

*(Equation 2.2)*

Selected PaxDB (Wang et al., 2012) datasets were downloaded from the *S. cerevisiae* subset (available at http://pax-db.org/) in order to compare to the calculated NG cpc values for chaperones. As the preferred unit is parts per million (ppm), the copies per cell value for each protein was adjusted assuming 60,000,000 proteins per cell (Equation 2.3) (Prof. Rob Beynon 2016, personal communication, 23rd August).

$$PPM = \frac{CPC}{60{,}000{,}000} \times 1{,}000{,}000$$

*(Equation 2.3)*

For label-free quantification, acquired data were processed with MaxQuant (v1.5.2.8) (Cox and Mann, 2008) with peptides identified using the Andromeda search engine (Cox et al., 2011), searching the entire *S. cerevisiae* protein sequence database (canonical and isoform. fasta downloaded from UniProt – http://www.uniprot.org.downloads/, accessed April 2015 containing 6721 entries), additional to a reverse decoy database and a database of known contaminants as available within the MaxQuant software. MaxQuant default search parameters were used, specifying two missed cleavages and LFQ minimum ratio count set to one. Additionally, the 'requantify' and 'match between runs' options were selected. The 'proteinGroups.txt' file was then manually filtered such that proteins had to be observed with a non-zero MaxLFQ intensity (Cox et al., 2014) in at least three biological replicates and quantified using at least two unique peptides. Final MaxLFQ intensities for a protein were calculated as the median MaxLFQ intensity across biological replicates for samples obtained under the same growth condition. Protein identifications that passed a 1 % FDR against the decoy database were deemed true positive matches and the corresponding Q-values for each protein identified recorded.

Condition-dependent linear regression was performed using the R package 'aLFQ' (Rosenberger et al., 2014) to predict cpc values from the median MaxLFQ intensities modelled on the cpc values for chaperones identified in both the SRM and unfractionated label-free datasets. Modelled copy per cell (mod-cpc) values were determined for 1644 proteins identified in both NG and HS samples.

## 2.8 Chaperone interactions, protein volume and workload

To create a high quality interaction dataset, *S.* cerevisiae interaction datasets from IntAct (downloaded 24th September 2015, available at http://www.ebi.ac.uk/intact/) (Hermjakob et al., 2004), String (version 10, available at http://string-db.org/) (Szklarczyk et al., 2015) and BioGrid (version 3.4.128, available at http://thebiogrid.org/) (Stark et al., 2006) were downloaded. String interactions were subset such that only interactions above a String score of 0.7 were listed. Using these datasets, a new high quality dataset was created containing only chaperone:non-chaperone interactions. Furthermore, each interaction had its reciprocal

interaction listed in one of the downloaded databases and was reported in at least two of the databases. The final high quality dataset contained 2761 interactions for 60 chaperones and 1412 non-chaperone proteins.

To determine the protein volume encountered by each chaperone ($V_c$) the mod-cpc values of all non-chaperone proteins interacting with the chaperone of interest in a condition dependent manner were summed (Equation 2.4), where n depicts the number of interacting non-chaperone proteins. For any non-chaperone protein interacting with multiple chaperone proteins, it was assumed that non-chaperone client proteins were shared equally between the interacting chaperones. The protein volume ($V_c$) for a chaperone *c* was estimated using the total of all mod-cpc values for all *n* substrates.

$$V_c = \sum_{1}^{n} modcpc_n$$

*(Equation 2.4)*

To calculate the synthesis rate ($k_{syn}$) of each protein in the interaction dataset, degradation rates ($k_{deg}$) determined by Christiano and colleagues were used, assuming the cell was at steady state ($\frac{dmodc_n}{dt} = 0$) (Equation 2.5) (Christiano et al., 2014, Lawless and Hubbard, 2014). Twenty-seven non-chaperone proteins were reported to have a half-life of over 100 hours, thus were removed from the dataset due to their stability. Thirty-one non-chaperone proteins were not included within the degradation database, so were given the geometric mean $k_{deg}$ of 0.00149 $min^{-1}$. Of the 635 non-chaperone proteins, only 608 were used in chaperone workload ($F_c$) calculations (Equation 2.6).

$$k_{syn} = modcpc_n * k_{deg}$$

*(Equation 2.5)*

$$F_c = \sum_{1}^{n} k_{syn}$$

*(Equation 2.6)*

To determine any significant changes in protein volume for a chaperone, an unpaired Wilcoxon test between all mod-cpc values for its interactors between NG and HS conditions was performed. To determine any significant changes in workload for a chaperone, an unpaired

Wilcoxon test between all $k_{syn}$ values for its interactors between NG and HS conditions was performed.

# Chapter 3 - Absolute quantification of *S. cerevisiae* chaperones under conditions of normal growth via the QconCAT strategy

The *S. cerevisiae* chaperome can be defined as the whole set of chaperone-chaperone and chaperone-client interactions, which has been largely characterised in qualitative terms by affinity pulldown MS (Gong et al., 2009). Given the basic chaperome topology is known; one approach to understanding cellular stress response is via modelling these interactions under various stress conditions (Mihalik and Csermely, 2011). This study used mRNA levels as a proxy for protein abundance whilst others have used MS to generate protein weights in network approaches to rationalise hydrogen peroxide stress in *S. pombe* (Lehtinen et al., 2013). However, neither of these generated absolute protein copy per cell values which can be more informative in modelling approaches and when estimating the stoichiometry of interacting proteins. Here, I have deployed the "QconCAT strategy" developed in Liverpool and Manchester (Brownridge et al., 2013, Lawless et al., 2016) to absolutely quantify yeast chaperones to more accurately determine the effect of stress on the *S. cerevisiae* chaperome.

Absolute protein quantification via QconCATs is carried out at the peptide level using quantotypic peptides (i.e. fully tryptic peptides suitable for use as quantification standards) with standardisation of signal intensity to known quantities of an isotope-labelled reference Q-peptide, identical in sequence to the surrogate peptide for the protein of interest. The ratio between the analyte and the heavier internal reference can then be used to quantify the unknown analyte. I used the QconCAT approach where artificial genes can be designed *de novo* to mediate synthesis of novel proteins containing equimolar assemblies of the Q-peptides. These Q-peptides are arginine or lysine terminated at the *C*-terminus to represent the tryptic peptides derived from digestion of the analyte proteins. Other features may be added to the QconCAT construct, including an initiator codon, a purification tag, and protective sacrificial regions before the gene is transformed into a bacterial expression system (Rivers et al., 2007). The QconCAT construct is not identical to the analytical protein as the QconCAT construct is an assembly of tryptic peptides from multiple proteins, therefore the QconCAT and analyte do not share the same physicochemical properties until both are fully proteolysed to the limit peptides. It is therefore critical for accurate quantitation that digestion is complete in both the QconCAT and the analyte protein (this is discussed in detail in Chapter 5). Following proteolysis, both the standard and analyte counterpart of a Q-peptide will behave identically in an LC-MS system, but will have a mass difference according to the label used, allowing for absolute quantification if the concentration of standard is known

## 3.1 *In silico* design of the ChapCATs

For each of the 63 chaperones in *S. cerevisiae*, listed by Gong and colleagues (Gong et al., 2009), theoretical tryptic digestions were performed to elucidate the potential peptides available for use as Q-peptides. As outlined by Brownridge and colleagues (Brownridge et al., 2012), the suitability for a peptide to be used as a Q-peptide is determined by several features. Briefly, peptides must be unique to the chaperone of interest and ideally isoform-specific; the sequences surrounding the peptide should not be flanked by dibasic sequences (RR, KK, RK, KR) as these tend to be (mis)cleaved at one or the other of the basic residues splitting the peptide signal between two variants. The sequence 'NG' (asparagine-glycine) within a peptide is prone to rapid but variable non-enzymatic deamidation, increasing the peptide mass by 0.98 Da, and is hence avoided. Similarly, the sequence 'DP' (aspartic acid-proline) which is reportedly prone to nonenzymatic cleavage. In addition, peptides with an *N*-terminal glutamine ('Q') may have partial conversion of the *N*-terminus to pyroglutamic acid, resulting in a loss of 17 Da from the peptide mass and is avoided, as are peptides with contiguous glutamine (three to five) residues. Finally, Q-peptides with sequences less than five amino acids were discarded. Using the online tool MC:pred (available at http://king.smith.man.ac.uk/mcpred/) (Lawless and Hubbard, 2012), the propensity for the peptide to undergo missed cleavage in the analyte protein sequence was evaluated, giving values for both the *N*- and *C*-terminus. Selected peptides should also ionise and be readily detected; the likelihood of detection in a liquid chromatography – electrospray ionisation – MS (LC-ESI-MS) experiment is governed by various parameters, including degree of secondary structure, charge, isoelectric point and degree of hydrophobicity. I utilised the online tool CONSeQuence (available at http://king.smith.man.ac.uk/CONSeQuence/) (Eyers et al., 2011) to give a consensus prediction for each Q-peptide to be detected in an LC-ESI-MS experiment. Finally, I searched each protein of interest against the protein post-translational modification database dbPTM (available at http://dbptm.mbc.nctu.edu.tw/) (Lee et al., 2006) to identify tryptic peptides reported with a post-translational modification and avoided these. Previous best practice, in part dictated by cost considerations, selected two Q-peptides to target each protein of interest. To this end, I selected up to five Q-peptides where possible to target each chaperone as undoubtedly increasing the number of Q-peptides would increase the robustness of the absolute quantification performed. As demonstrated in Table 3.1, this was not possible for a number of chaperones due to the sequence feature limits I adhered to, with a combination of MC:pred and CONSeQuence scores allowing for ranking of potential Q-peptides. The selected Q-peptides were concatenated into ten chaperone QconCATs (ChapCATs), depending on their target chaperones' classification group (Table 3.2). Q-peptides

were ordered within the ChapCATs to maximise likelihood of tryptic cleavage as determined by MC:pred scoring, codon optimised for expression in *E. coli* and ligated into an expression vector by PolyQuant GmBH, Germany (see Methods Section 2.1). Full details of the Q-peptides selected are provided in Appendix 4.

**Table 3.1) Example of selection of Q-peptides for the chaperones Hsp31 and Erj5**

Potential Q-peptides were initially ranked via their CONSeQuence score, followed by a manual interpretation of both terminal MC:Pred scores, which in combination with CONSeQuence allowed for selection of the final Q-peptides. Broadly, CONSeQuence scores were as high as possible, whilst McPred scores were below 0.5. Columns 3 to 7 are binary, with the value '1' meaning 'true'. Those highlighted in green were selected for use as Q-peptides. The Hsp31 Q-peptide FGWDEHSLAK (coloured orange) is a good candidate for a Q-peptide, however was classified a 'C'-type peptide by Brownridge and colleagues and so not selected – discussed later in this section. Key: a.a. – amino acids; NG/DP – asparagine-glycine or aspartic acid-proline sequences; QQQ+ - contiguous glutamine runs; PTM – post-translational modification.

| **Chaperone** | **Peptide** | **Di-basic** | **< 5 a.a.** | **NG / DP** | **QQQ +** | **PTM** | **Chaps targeted** | **McPred Native *N*-terminal** | **McPred Native *C*-terminal** | **CONS-eQuen-ce** | **Rank** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Hsp31 | DSDFNK | 0 | 0 | 0 | 0 | 0 | 1 | 0.83 | 0.54 | 0.203 | |
| Hsp31 | SIDALK | 0 | 0 | 0 | 0 | 0 | 1 | 0.41 | 0.83 | 0.212 | |
| Hsp31 | TGRPLIEGK | 0 | 0 | 0 | 0 | 0 | 1 | 0.71 | 0.38 | 0.232 | |
| Hsp31 | VLLALTSYNDVFYSDGAK | 0 | 0 | 0 | 0 | 0 | 1 | 0.78 | 0.43 | 0.457 | |
| Hsp31 | NLATVEDVAK | 1 | 0 | 0 | 0 | 0 | 1 | 0.6 | 0.67 | 0.458 | |
| Hsp31 | DLQDIASEIYANGGVVAAVCHGPAIFDGLTDK | 1 | 0 | 1 | 0 | 0 | 1 | 0.68 | 0.94 | 0.519 | |
| Hsp31 | FGWDEHSLAK | 0 | 0 | 0 | 0 | 0 | 1 | 0.42 | 0.44 | 0.523 | |
| Hsp31 | EGFEVDFVSETGK | 0 | 0 | 0 | 0 | 0 | 1 | 0.77 | 0.42 | 0.528 | |
| Hsp31 | SITGFTDVGETILGVDSILK | 0 | 0 | 0 | 0 | 0 | 1 | 0.38 | 0.4 | 0.562 | 4 |
| Hsp31 | DFLNGQDETDFK | 0 | 0 | 1 | 0 | 0 | 1 | 0.44 | 0.47 | 0.608 | |
| Hsp31 | EVNADDYQIFFASAGHGTLFDYPK | 0 | 0 | 0 | 0 | 0 | 1 | 0.53 | 0.41 | 0.619 | 3 |
| Hsp31 | TGVFVVEALHPFNTFR | 1 | 0 | 0 | 0 | 0 | 1 | 0.43 | 0.72 | 0.675 | |
| Hsp31 | YLAPVGPWDDYSITDGR | 0 | 0 | 0 | 0 | 0 | 1 | 0.42 | 0.4 | 0.722 | 2 |
| Hsp31 | LVTGVNPASAHSTAVR | 0 | 0 | 0 | 0 | 0 | 1 | 0.4 | 0.41 | 0.782 | 1 |

Continued over the page

| Chaperone | Peptide | Di-basic | < 5 a.a. | NG / DP | QQQ + | PTM | Chaps targeted | McPred Native *N*-terminal | McPred Native *C*-terminal | CONS-eQuen-ce | Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Erj5 | SLVVR | 0 | 0 | 0 | 0 | 0 | 1 | 0.48 | 0.42 | 0.155 | |
| Erj5 | VIYSR | 1 | 0 | 0 | 0 | 0 | 1 | 0.4 | 1 | 0.18 | |
| Erj5 | QLTFK | 0 | 0 | 0 | 0 | 0 | 1 | 0.39 | 0.54 | 0.181 | |
| Erj5 | SVGSAGK | 0 | 0 | 0 | 0 | 0 | 1 | 0.39 | 0.77 | 0.186 | |
| Erj5 | YDGNQTK | 1 | 0 | 0 | 0 | 0 | 1 | 0.8 | 0.69 | 0.194 | |
| Erj5 | YHPDK | 0 | 0 | 0 | 0 | 0 | 4 | 0.78 | 0.95 | 0.203 | |
| Erj5 | MNGYWKPALVVLGLVSLSYAFTTIETEIFQLQNEISTK | 0 | 0 | 1 | 0 | 0 | 1 | - | 0.34 | 0.211 | |
| Erj5 | LQNSSTK | 0 | 0 | 0 | 0 | 0 | 1 | 0.49 | 0.53 | 0.211 | |
| Erj5 | EEIITDSK | 1 | 0 | 0 | 0 | 0 | 1 | 0.77 | 0.59 | 0.218 | |
| Erj5 | MELPNGK | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0.4 | 0.224 | |
| Erj5 | NCLFWR | 0 | 0 | 0 | 0 | 0 | 1 | 0.45 | 0.46 | 0.239 | 5 |
| Erj5 | TWFLLAFIWIVVNIGQYIISIIQYR | 0 | 0 | 0 | 0 | 0 | 1 | 0.51 | 0.48 | 0.286 | 4 |
| Erj5 | IENFISQCK | 0 | 0 | 0 | 0 | 0 | 1 | 0.78 | 0.43 | 0.361 | |
| Erj5 | GGFYFSR | 0 | 0 | 0 | 0 | 0 | 1 | 0.42 | 0.63 | 0.386 | |
| Erj5 | YGPDMNFYK | 0 | 0 | 0 | 0 | 0 | 1 | 0.34 | 0.47 | 0.388 | 3 |
| Erj5 | QQDDTNGLGVK | 0 | 0 | 1 | 0 | 0 | 1 | 0.43 | 0.39 | 0.453 | |
| Erj5 | IPASVWNMTFGK | 0 | 0 | 0 | 0 | 0 | 1 | 0.46 | 0.39 | 0.509 | 2 |
| Erj5 | IYDYYLQNGFPNYDFHK | 0 | 0 | 1 | 0 | 0 | 1 | 0.68 | 0.42 | 0.521 | |
| Erj5 | FSDVYVVEPDGSETLISPDTLDKPSVK | 0 | 0 | 0 | 0 | 0 | 1 | 0.42 | 0.45 | 0.603 | 1 |
| Erj5 | LNLATQILSNSSNR | 1 | 0 | 0 | 0 | 0 | 1 | 0.64 | 0.76 | 0.653 | |

**Table 3.2) Q-peptides targeting each chaperone were assigned to a ChapCAT construct according to the chaperones' classification**

Ten ChapCATs were created, with sizes ranging from 36.66 kDa to 58.50 kDa. Non-unique Q-peptides were used to target the protein groups Hsp32_Sno4_Hsp33, Ssa1_Ssa2 and Ssb2_Ssb1. For Ssa1, Ssa2, Ssb1 and Ssb2 few unique Q-peptides were available, but were used in conjunction with non-unique Q-peptides to improve reliability of quantification.

| ChapCAT | Chaperones (Gene) | Classes | Length (bp) | MW (kDa) |
|---|---|---|---|---|
| 001 | Hsp82, Hsc82, Hsp60, Mcx1, Hsp78, Hsp104 | Hsp90, Hsp60, AAA+ | 1392 | 49.13 |
| 002 | Yke2, Gim3, Gim4, Gim5, Pac10, Pfd1 | PFD | 975 | 36.66 |
| 003 | Hsp32_Sno4_Hsp33, Hsp31, Hsp42, Hsp26, Hsp12 | Small | 1158 | 41.91 |
| 004 | Ssa1, Ssa2, Ssa1_Ssa2, Ssa4, Ssz1, Sse2, Ssq1 | Hsp70 | 1395 | 49.34 |
| 005 | Ssb1, Ssb2, Ssb2_Ssb1, Sse1, Ssc1, Ssc3, Lhs1 | Hsp70 | 1344 | 47.46 |
| 006 | Cct2, Cct3, Cct4, Cct5, Cct6, Cct7 | CCT | 1401 | 49.20 |
| 007 | Ssa3, Kar2, Cct8, Tcp1, Ydj1, Xdj1 | Hsp70, CCT, Hsp40 | 1650 | 58.50 |
| 008 | Apj1, Sis1, Djp1, Zuo1, Swa2, Jjj1 | Hsp40 | 1251 | 45.46 |
| 009 | Jjj2, Jjj3, Caj1, Cwc23, Mdj1, Mdj2 | Hsp40 | 1350 | 50.01 |
| 010 | Jac1, Jid1, Scj1, Hlj1, Jem1, Sec63, Erj5, Pam18 | Hsp40 | 1500 | 55.40 |

In summary, 288 Q-peptides were selected to target the 63 chaperones in *S. cerevisiae*. Of the 63 chaperones, 43 were targeted by five unique Q-peptides, eight were targeted by four Q-peptides, three were targeted by three unique Q-peptides, two were targeted by two Q-peptides and a single chaperone was targeted by a single Q-peptide (Figure 3.1). For Hsp32, Sno4 and Hsp33, no unique quantotypic peptides were identified; therefore, five non-unique Q-peptides were selected to represent the summed group. Non-unique but potential Q-peptides were also observed for the protein pairs Ssa1 and Ssa2, and Ssb2 and Ssb1, and selected as quantification standards owing to few unique alternatives; both unique and non-unique Q-peptides could therefore be used to improve quantification reliability. I compared the selected peptides to those selected for use in a previous study (Brownridge et al., 2012), discussed in the next section. Thirty-one of these peptides were used in both studies and are discussed in terms of quantification later in this chapter.

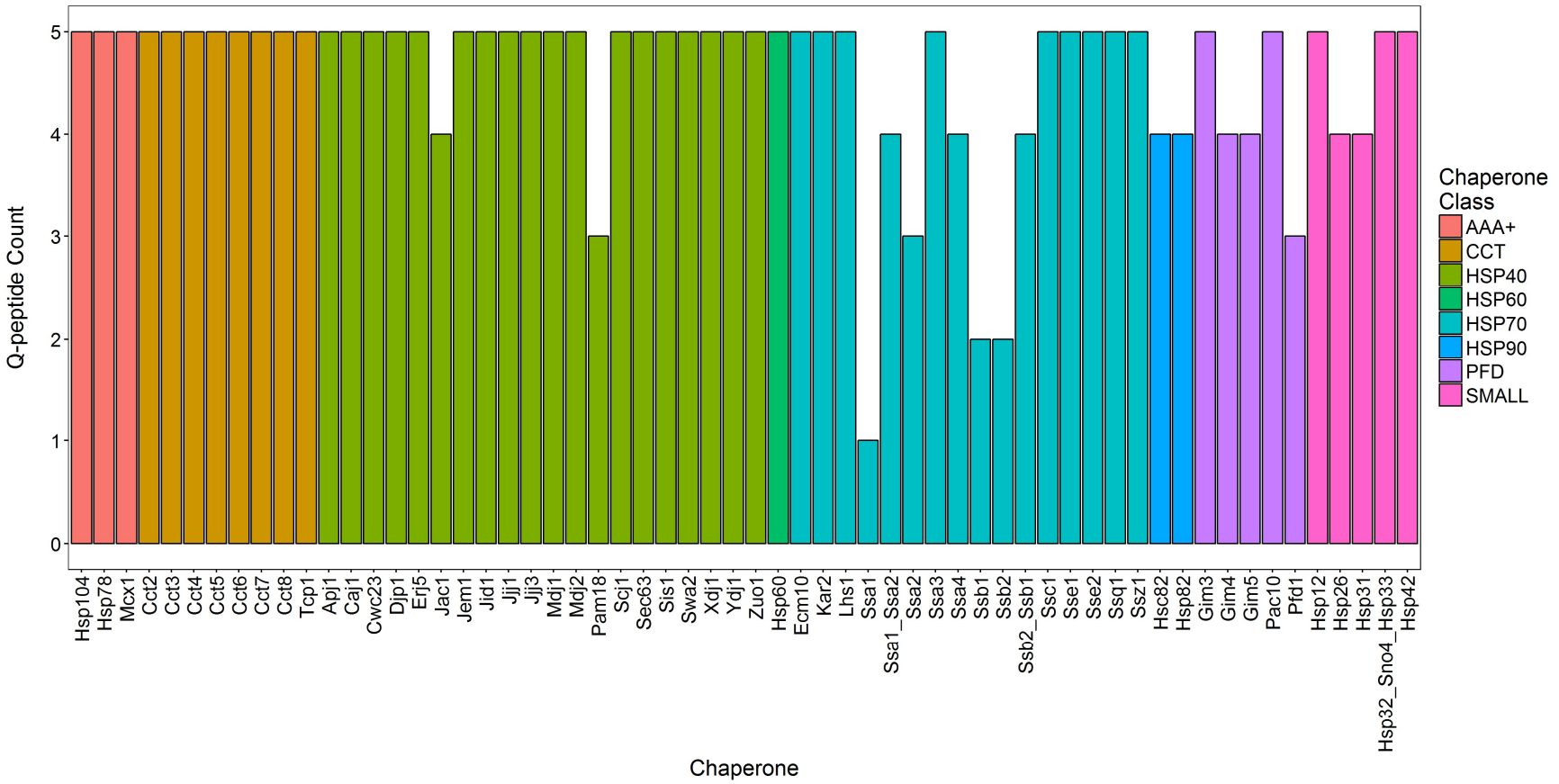



**Figure 3.1) Up to five Q-peptides are selected to target a chaperone in *S. cerevisiae***

Fourty-three chaperones were targeted by five unqiue Q-peptides, whilst others were also targeted by non-unique Q-peptides due to the low number of unique quantotypic peptides available.

## 3.2 Theoretical comparison with two Q-peptide QconCAT constructs

As previous attempts to quantify the chaperones in *S. cerevisiae* yielded absolute copy per cell values for 51 of the 63 chaperones (Brownridge et al., 2013) using two Q-peptides per chaperone, I questioned whether theoretically the use of up to five Q-peptides could improve the ability to quantify all 63 chaperones. Notably, during the design phase for the two Q-peptide set, termed 'CopyCATs', MC:pred scores were not yet formally taken into account. To determine any differences in predicted missed cleavages between the CopyCATs and my ChapCATs, the terminal MC:pred scores for each Q-peptide within the analyte were calculated using MC:pred (Lawless and Hubbard, 2012). The maximum of the two terminal scores for each peptide was selected, and the distributions of these maximum peptide scores were compared. I defined a 'good' MC:pred score as one below 0.5, meaning that the peptide is not predicted to be subject to missed cleavage by trypsin at the corresponding terminus. For CopyCAT Q-peptides, 78.57 % had a maximum MC:pred score below 0.5, whilst 77.78 % of ChapCAT Q-peptides had a maximum MC:pred score below 0.5 (Figure 3.2). The distributions of maximum MC:pred scores between the two sets were not significantly different (*U* test, $p = 0.323$, $p < 0.05$), confirming that the ChapCAT set were no better (or worse) in terms of missed cleavage potential.

Brownridge and colleagues (Brownridge et al., 2013, Lawless et al., 2016) classified individual Q-peptides in terms of suitability for absolute quantification. Class 'A' peptides were observed in both heavy (standard) and light (analyte) counterparts; class 'B' peptides were only observed as heavy standard, with the light analyte peptide being below the limit of detection of the mass spectrometer and/or failing to pass a 1 % FDR, whilst class 'C' peptide standards were not observed by SRM above a minimal signal/noise ratio. As Brownridge and colleagues (Brownridge et al., 2013, Lawless et al., 2016) used only class 'A' peptides to determine absolute copy per cell values, I sought to observe if the 'A' peptides had on average a lower MC:pred value than the 'B' and 'C' types. Performing a Mann Whitney Wilcoxon test (*U* test, $p < 0.05$) between each of the classification types, I observed no significant difference (Figure 3.3c). Therefore, it is apparent that even without the explicit use of MC:pred for selecting Q-peptides, the likelihood of complete tryptic digestion remains equivalent between the two datasets and no apparent bias was present in the old data set. Although the original empirical selection criteria for two Q-peptides proved to be valid, as I am targeting more Q-peptides per chaperone requiring a deeper selection of candidate peptides, MC:pred was used as extra criterion for selection of Q-peptides. Comparison of these datasets indicated that in terms of the analyte

peptide propensity to be fully tryptically cleaved, the new peptide selection would yield similar results (discussed in Chapter 4).

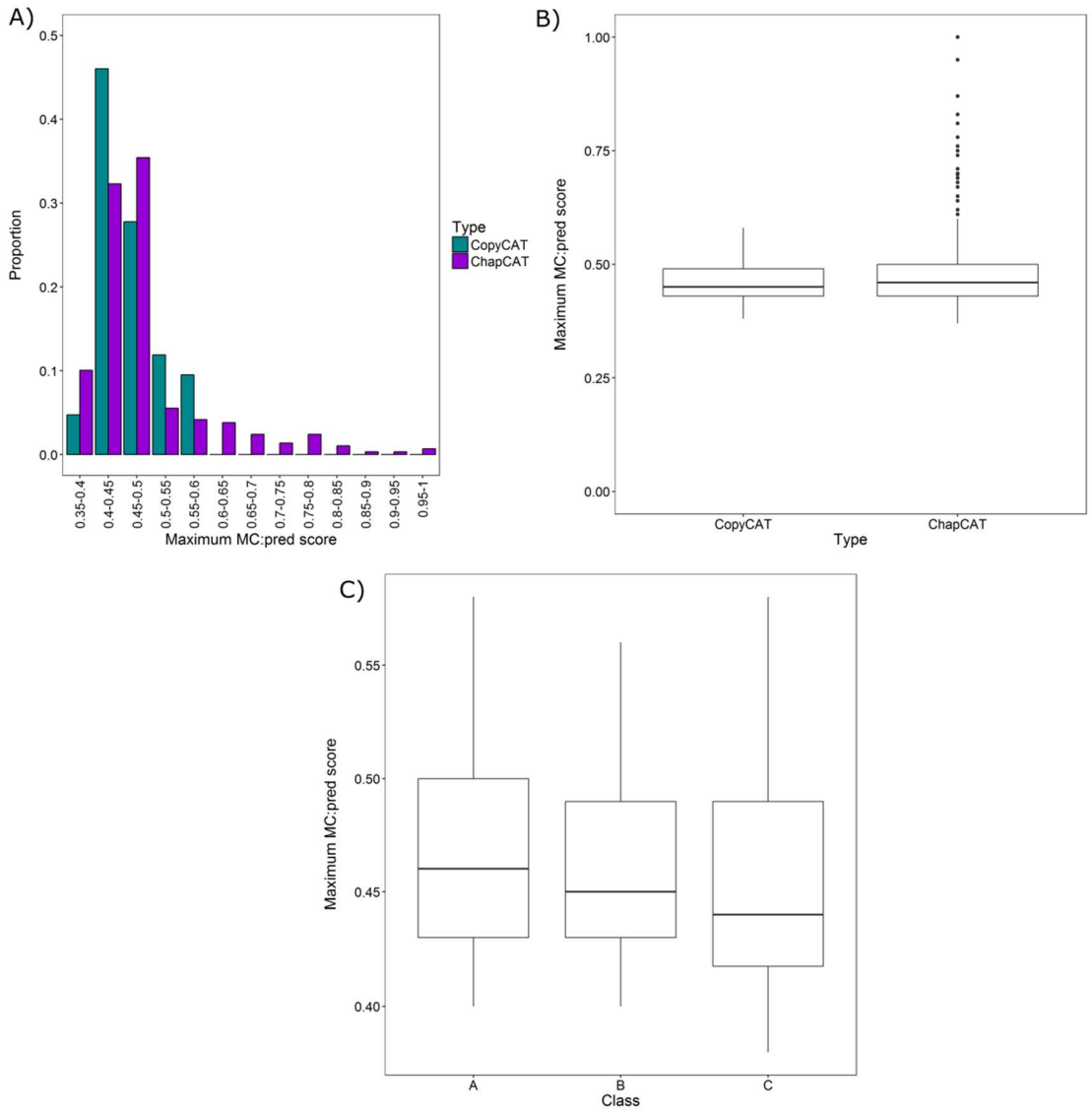



**Figure 3.2) Comparison of analyte maximum MC:pred scores for CopyCAT and ChapCAT Q-peptides**

A) A lower proportion of ChapCAT Q-peptides had a high MC:pred score due to the requirement to select additional Q-peptides fo each chaperone, up to a maximum of five; However, despite this, as shown in B) the distributions of maximum MC:pred scores for CopyCAT and ChapCATs did not significantly differ (*U* test, $p = 0.323$); C) Although CopyCAT class 'C' Q-peptides were of fewest number (16) in the dataset, there is a wider distribution of MC:pred scores. Regardless, there is no significant difference between each of the Q-peptide classification types in terms of their maximum MC:pred scores. A- vs. B- type: $p = 0.8736$, A- vs. C- type: $p = 0.2189$, B- vs. C- type: $p = 0.2813$ (*U* test, $p < 0.05$).

The online tool CONSeQuence (Eyers et al., 2011) provides a ranked score to each Q-peptide representing the propensity of that Q-peptide to be detected in an LC-ESI-MS

experiment. Like MC:pred, CONSeQuence scores were not originally used as advisory scores when designing the first chaperone CopyCATs (Brownridge et al., 2013, Lawless et al., 2016). A Q-peptide with a high CONSeQuence score indicates an increased likelihood of detectability compared to lower ranked Q-peptides. When comparing the CONSeQuence scores of ChapCAT Q-peptides to CopyCAT Q-peptides I observed a wider range of scores for ChapCAT Q-peptides with a lower median value (0.499 compared to 0.573 for CopyCATs), resulting in a significantly different distribution (*U* test, $p < 0.05$, $p = 0.0005$) (Figure 3.3). According to Brownridge and colleagues' A, B and C classifications, one would anticipate A- and B- type Q-peptides to have a higher CONSeQuence score than C-type Q-peptides. Interestingly, there is no significant difference between each of the classes of Q-peptide for the chaperone CopyCAT dataset, despite the median CONSeQuence scores being 0.593, 0.515 and 0.615 for A-, B- and C-type Q-peptides respectively. This data suggests that despite stringent filtering protocols to identify the best potential standards, even highly scored Q-peptides may still be unobservable in an LC-ESI-MS experiment. To further investigate this, I searched for each of the CopyCAT and ChapCAT Q-peptides in PeptideAtlas (Desiere et al., 2006), an online repository of peptides previously observed in a tandem MS experiment. For ChapCATs, 164 Q-peptides (56.9 %) were previously observed, with 74 Q-peptides (59.7 %) for CopyCATs. Again, this is unsurprising and suggests the more *ad hoc* design principles used for the original QconCATs were still effective, and the additional requirement to select more for the ChapCAT project requires slightly less optimal peptides to be picked. In terms of the classification by Brownridge and colleagues for the CopyCAT Q-peptides, 75.0 % of A-type, 31.3 % of B-type and 43.8 % C-type peptides were previously observed in a tandem MS experiment according to PeptideAtlas (Desiere et al., 2006). The fact that C-type peptides have been previously observed highlights inherent challenges in reproducibility for LC-MS/MS experiments, an issue that I contend with in Chapter 6.

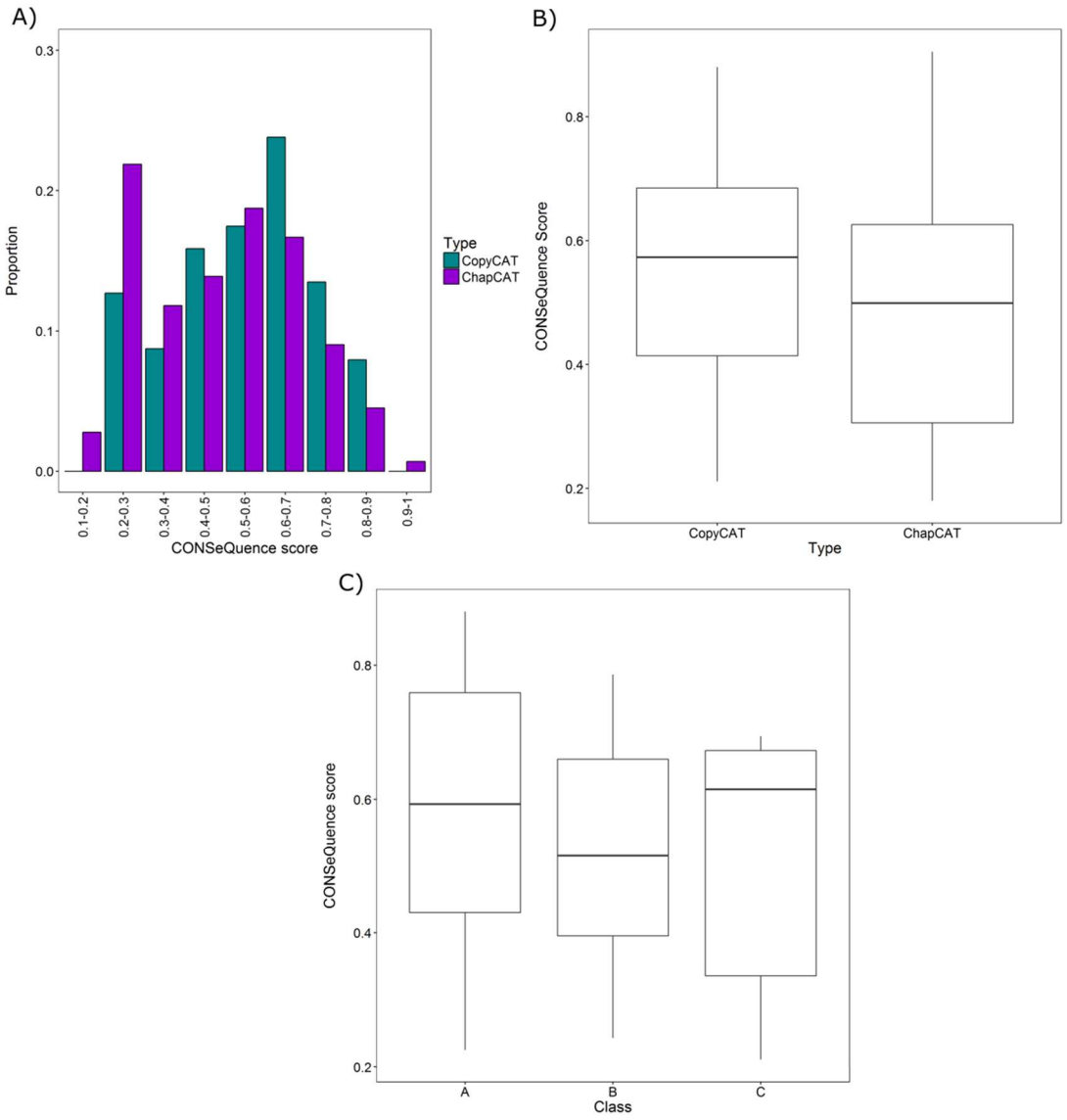



**Figure 3.3) Comparison of CONSeQuence scores for CopyCAT and ChapCAT Q-peptides**

A) ChapCAT Q-peptide CONSeQuence scores appear to have a bimodal distribution with a large proportion having a low CONSeQuence score (0.2 to 0.3). Despite this, ChapCAT Q-peptides also occupy the larger CONSeQuence scores (above 0.9). B) The distribution of CONSeQuence scores for ChapCAT Q-peptides is spread over a wider range compared to CopyCAT Q-peptides, with an overall lower median score (*U* test, p = 0.0005). C) There is no significant difference between the ranks of CONSeQuence scores for distinctly classified CopyCAT Q-peptides. A- vs. B-type: p = 0.1285; B- vs. C-type: p = 0.9090 and A- vs. C-type: 0.3280 (*U* test, p < 0.05).

For a ChapCAT to be used as a standard, it must be abundantly expressed and isotopically labelled in *E. coli;* purified from the inclusion body fraction utilising its *C*-terminal histidine tag and a known amount mixed with the analyte and subject to complete proteolysis via trypsin. This mixture is then subject to various MS experiment workflows to determine the labelling efficiency and quantify the ChapCAT standard, perform transition selection and determine the respective retention times of each Q-peptide. Finally, a scheduled SRM experiment is

performed in which absolute abundances of each native (light-labelled) Q-peptide, and subsequently the chaperone copy per cell values, are determined. These steps are summarised and shown in schematic form in Figure 3.4 and are discussed in further detail in this Chapter.

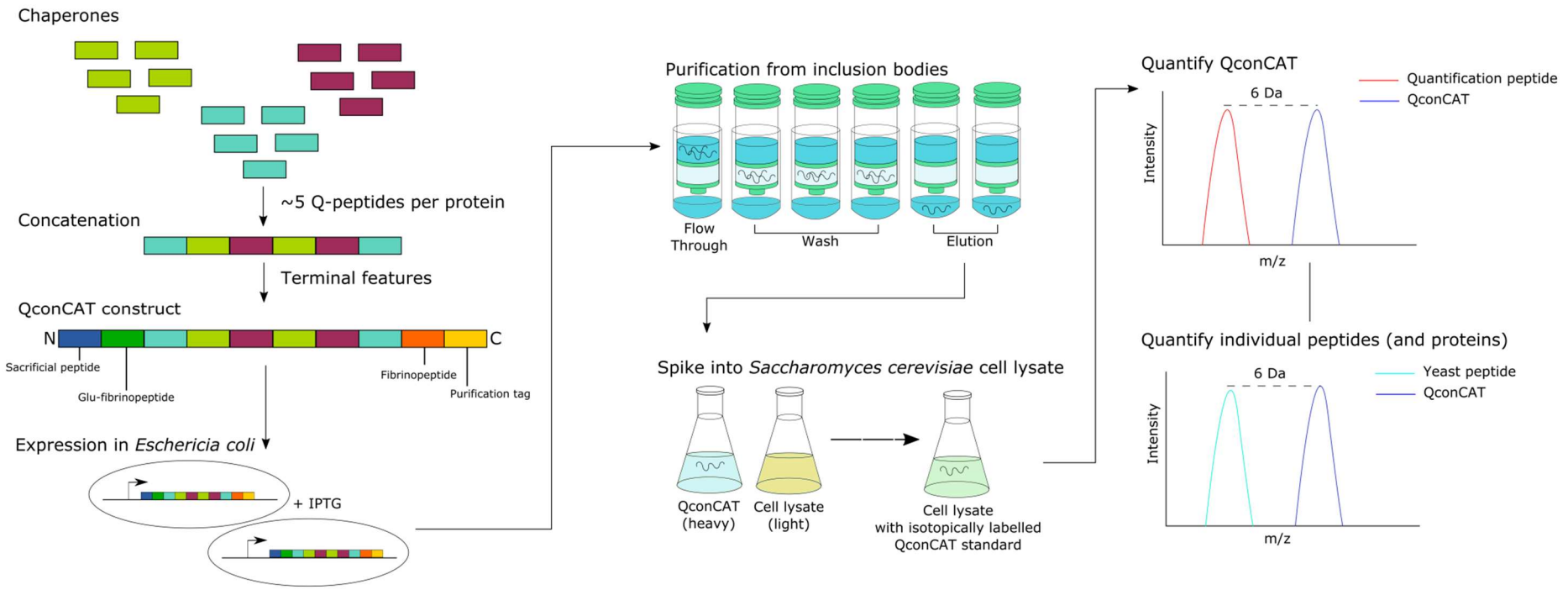



**Figure 3.4) Steps involved in the production of ChapCATs for use towards absolute quantification**

Chaperone Q-peptides are selected and concatentated into a chaperone QconCAT (ChapCAT) synthesised for optimal expression in *E. coli*. After induction of isotopically-labelled ChapCAT with IPTG , ChapCAT protein is isolated within the inclusion bodies and may be purified through the use of nickel affinity chromatography. Purified isotopically-labelled is mixed with yeast analyte sample and subject to mass spectrometry experiments in which both the ChapCAT standard and analyte peptides are quantified, allowing for determination of absolute copy per cell values for the target chaperones.

## 3.3 Expression and purification of ten ChapCATs

Expression of each of the ten ChapCATs in *E. coli* following addition of IPTG was tested in LB medium prior to heavy labelling and purification (Figure 3.5). Western blotting demonstrated a repeated lack of expression for ChapCATs 2, 6 and 10, resulting in a redesign by reshuffling the constituent Q-peptides for these ChapCATs (see Methods).

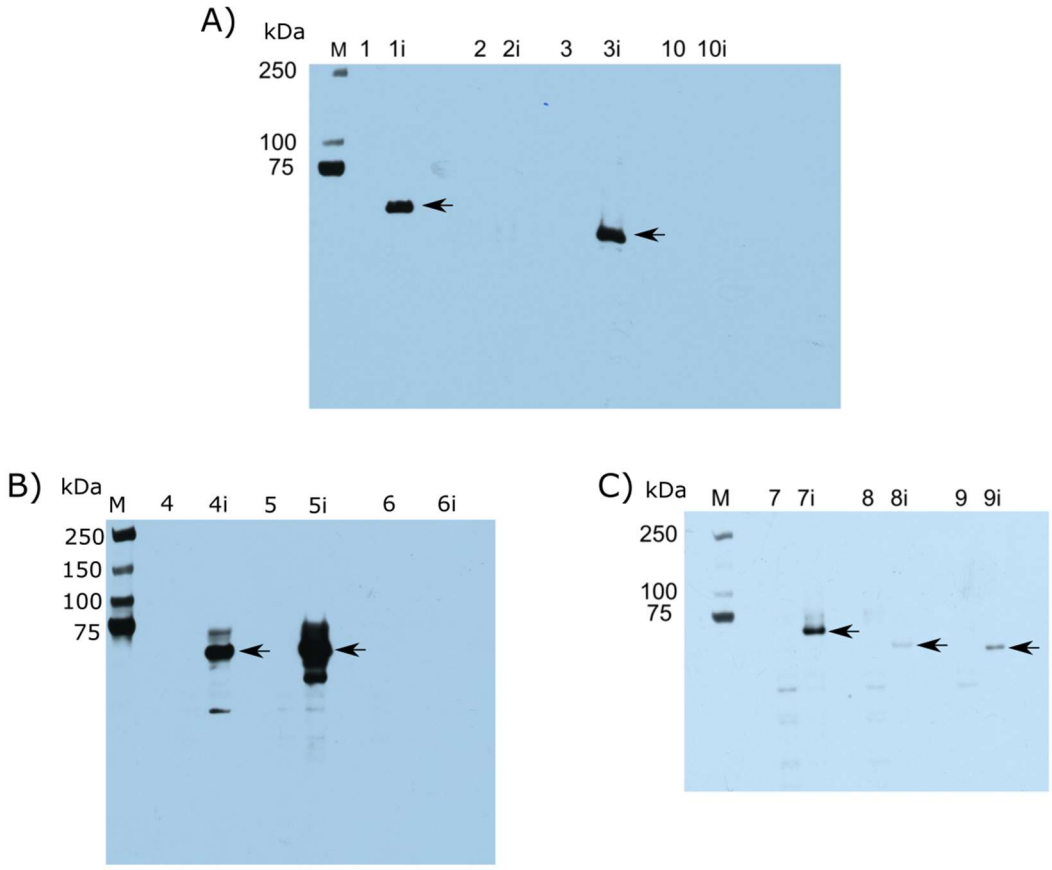



**Figure 3.5) Western blot expression testing in LB medium**

A) ChapCATs 1 and 2 expressed following IPTG induction and five hours growth ('i')whilst ChapCAT 10 failed and required a redesign. B) ChapCATs 4 and 5 expressed following IPTG induction whilst ChapCAT 6 failed so required a redesign. C) ChapCATs 7, 8 and 9 were successfully expressed in LB medium. Primary antibody: α-His, Secondary antibody: α-Mouse.

The reshuffled ChapCATs 2, 6 and 10 (named 2.1, 6.1 and 10.1 respectively) were subject to a second round of expression testing in response to IPTG addition in LB medium (Figure 3.6). ChapCAT 10.1 failed to express, and the decision was made not to reshuffle and resynthesise, and concentrate on the other successful ChapCATs. ChapCAT 2.1 and 6.1 were successfully expressed in LB medium and so could be tested in minimal medium.

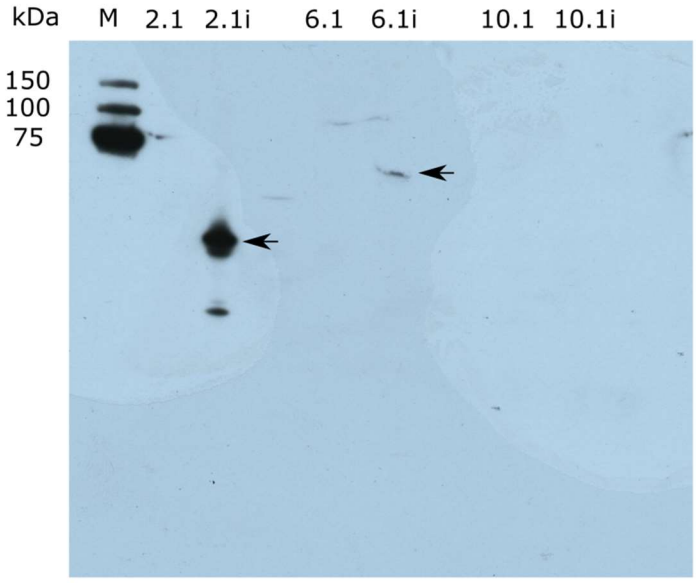



**Figure 3.6) Western blot for LB medium expression testing of reshuffled ChapCAT constructs**

Following IPTG induction, after five hours growth ('i') ChapCAT 2 and 6 demonstrated expression in LB medium whilst ChapCAT 10 expression was still absent, with low levels of expression observed for ChapCAT 6.1. 'M' depicts the marker lane. Primary antibody: α-His, Secondary antibody: α-Mouse.

After successful expression of all ChapCATs (excluding ChapCAT 10) I proceeded expression and purification of ChapCAT with heavy labelled arginine and lysine. Each ChapCAT was subject to expression testing in minimal medium, with expression induced with 1 mM IPTG and samples taken five hours after induction (Figure 3.7). An SDS-PAGE gel containing both expression samples in minimal media for the original ChapCAT 6 and reshuffled ChapCAT 6.1 was run, with again only ChapCAT 6.1 showing expression (Figure 3.7f).

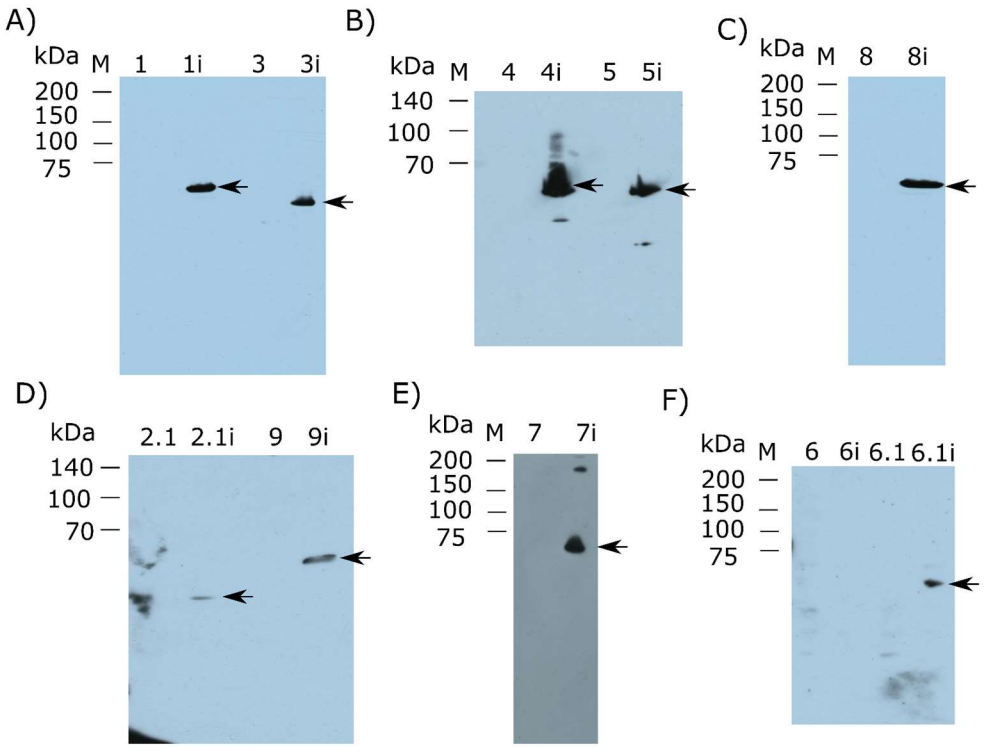



**Figure 3.7) Western blot for ChapCAT expression in heavy-labelled minimal media**

All ChapCATs expressed following addition of 1 mM IPTG and growth for five hours ('I'). For demonstration of a known failed ChapCAT, panel E contains expression data for the original ChapCAT 6 and the reshuffled ChapCAT 6.1, with only ChapCAT 6.1 being successfully expressed in heavy-labelled minimal media. 'M' depicts the marker lane, with ChapCAT1 as an example (A), 1 represents *E. coli* prior to IPTG induction, whilst 1i represents *E. coli* after IPTG induction and a further 5 hours growth. Primary antibody: α-His, Secondary antibody: α-Mouse.

For ChapCATs 2.1 and 9, I observed fairly low expression compared to others, and attempted to optimise expression levels of these ChapCATs by introducing benzyl alcohol, altering the concentration of IPTG for induction and altering the growth time before (allowing for growth to an $OD_{600}$ of 0.6 or 0.8) and after induction with IPTG (up to 3 or 5 hours) (Figure 3.8). Adding benzyl alcohol approximately 30 minutes prior to induction by IPTG initiates a stress response in the *E. coli* cells, resulting in up-regulation of chaperones able to prevent degradation of the unfolded ChapCAT protein (de Marco et al., 2005). ChapCAT 2.1 expression was observed one hour after induction for each condition tested. However, 3 hours after induction with 1 mM IPTG after growth to an $OD_{600}$ of 0.8, ChapCAT levels appeared reduced indicative of degradation, whilst induction with 0.4 mM IPTG shows the most intense band at this time point (Figure 3.8b). For ChapCAT 9, proteolytic products were observed after growth to an $OD_{600}$ of 0.6, potentially as a result of degradation, with whole ChapCAT expression levels consistent between 1 and 5 hours after induction with both 0.4 mM and 1 mM IPTG (Figure 3.8c). In contrast, when grown to an $OD_{600}$ of 0.8, ChapCAT 9 band intensity appeared to

increase with post-induction growth time (Figure 3.8d). Subsequently, ChapCAT 2.1 and ChapCAT 9 induction protocols differed from the remaining ChapCATs (Table 3.3).

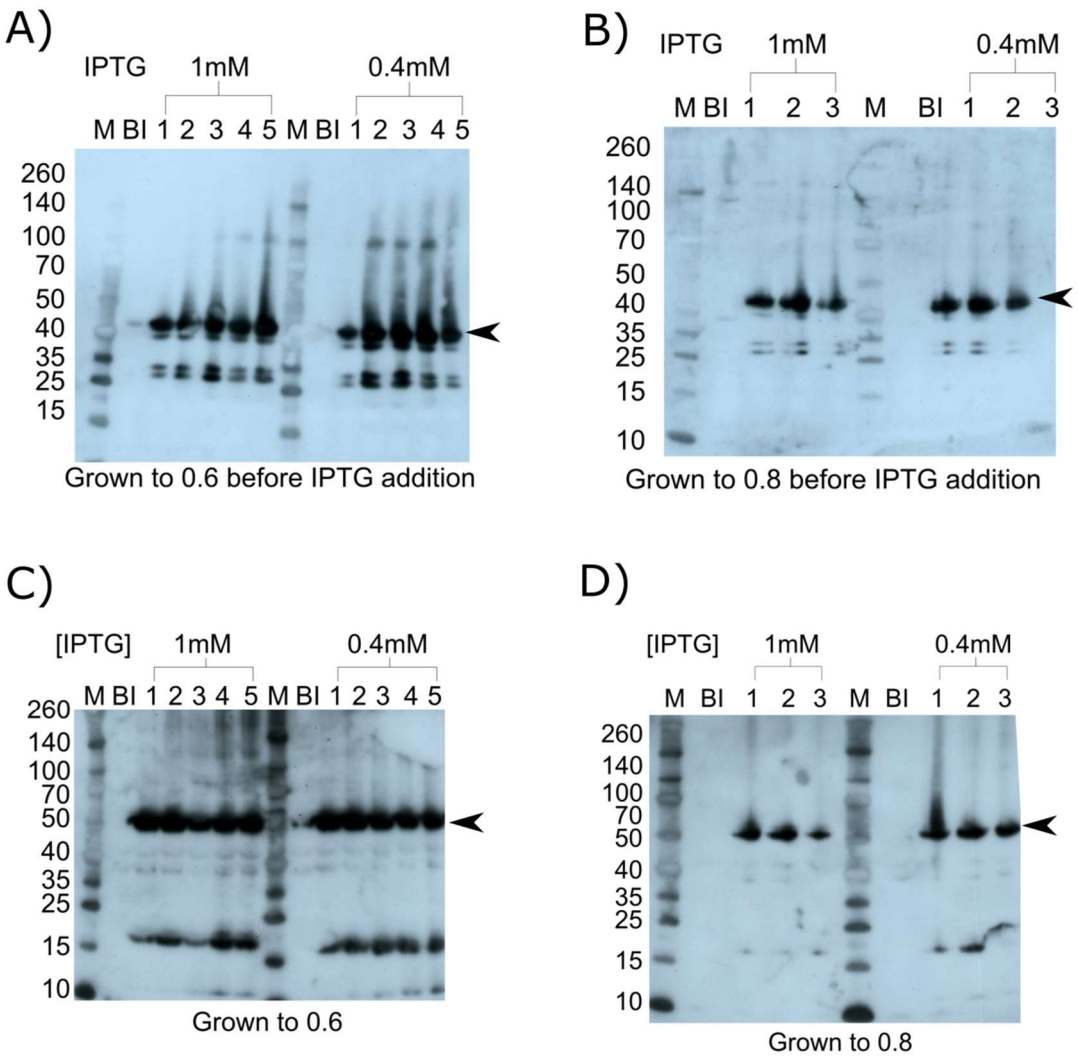



**Figure 3.8) Western blot analyses of minimal media expression testing for ChapCAT 2.1 and 9**

A and B) *E. coli* strains containg the ChapCAT 2.1 gene are grown to an $OD_{600}$ of 0.6 and 0.8 respectively, induced with either 1 mM or 0.4 mM IPTG and samples taken every hour up to 3 or 5 hours post-induction respectively. C and D) *E. coli* strains containing the ChapCAT 9 gene are grown to an $OD_{600}$ of 0.6 and 0.8 respectively, induced with either 1 mM or 0.4 mM IPTG and samples taken every hour up to 3 or 5 hours post-induction respectively. Benzyl alcohol was added approximately 30 minutes prior to induction with IPTG for both ChapCATs. Key: M – Marker (in kDa); BI - Before Induction; Arabic numerals - hours after induction. Primary antibody: α-His; Secondary antibody: α-Mouse.

**Table 3.3) Expression protocol for ChapCATs for minimal medium (testing in LB medium)**

| ChapCAT | $OD_{600}$[a] | Benzyl Alcohol[b] | IPTG (mM)[c] | Incubation Time (hrs)[d] |
|---|---|---|---|---|
| ChapCAT001 | 0.6 | None | 1 | 5.0 |
| ChapCAT002*[e] | 0.8 | Yes | 0.4 | 3.0 |
| ChapCAT003 | 0.6 | None | 1 | 5.0 |
| ChapCAT004 | 0.6 | None | 1 | 5.0 |
| ChapCAT005 | 0.6 | None | 1 | 5.0 |
| ChapCAT006* | 0.6 | None | 1 | 5.0 |
| ChapCAT007 | 0.6 | None | 1 | 5.0 |
| ChapCAT008 | 0.6 | None | 1 | 5.0 |
| ChapCAT009[e] | 0.8 | Yes | 1 | 3.0 |
| ChapCAT010*[f] | NA | NA | NA | NA |

[a] Growth of cell culture to desired $OD_{600}$ in 50 ml prior to induction with Imidazole.

[b] If used, 10mM Benzyl Alcohol is added 20 minutes prior to induction with Imidazole.

[c] Imidazole Concentration used for expression of ChapCAT.

[d] Incubation time for cell cultures subsequent to induction of ChapCAT expression.

[e] ChapCAT constructs were subject to expression testing by varying conditions to optimise expression.

[f] ChapCAT010 did not express following a redesign of the construct.

*Q-peptides were subject to a reshuffle and construct resynthesised consequent to null expression.

QconCATs are able to form insoluble inclusion bodies (Beynon et al., 2005) and so were isolated via low-speed centrifugation, with the collection of samples for the starting material (SM), the soluble fraction (SF) and the inclusion body fraction (IB). Nickel-affinity chromatography was then used on the IB to capture the ChapCAT on column via the *C*-terminal Histidine tag and subsequently elute the ChapCAT from the column upon addition of 250 mM Imidazole (example in Figure 3.9).

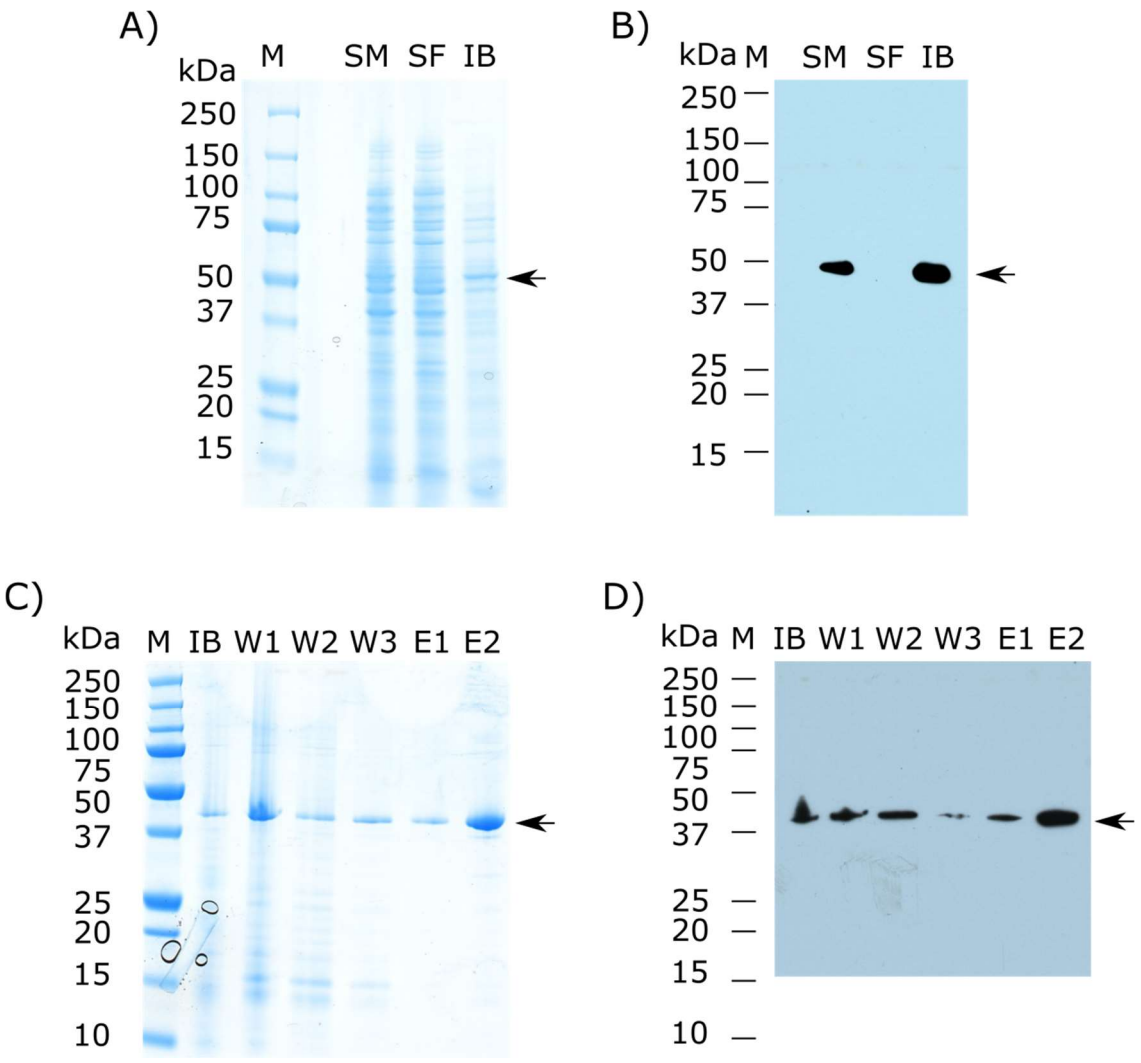



**Figure 3.9) SDS-PAGE and Western blotting analysis of ChapCAT 1 purification from the inclusion bodies**

A and B) SDS-PAGE and western blot analysis (respectively) of the cell lysate starting material (SM), soluble fraction (SF) and the inclusion body fraction (IB). C and D) SDS-PAGE and western blot analysis (respectively) of the purification of ChapCAT 1 from the IB fraction. Primary antibody: α-His, Secondary antibody: α-Mouse. Key: M - Marker; W – Wash; E – Elution;

For the majority of ChapCATs, protein was present in both the flow through and wash stages of the purification. However, despite these losses, adequate ChapCAT amounts were eluted from the column in the two elution steps with 250 mM imidazole (Figure 3.10). This mitigated against futher timing consuming purification optimisation experiments. However, for ChapCAT 7 (Figure 3.11g) SDS-PAGE analysis demonstrated low yields of ChapCAT, potentially due to less efficient expression in *E. coli* compared to other ChapCATs.

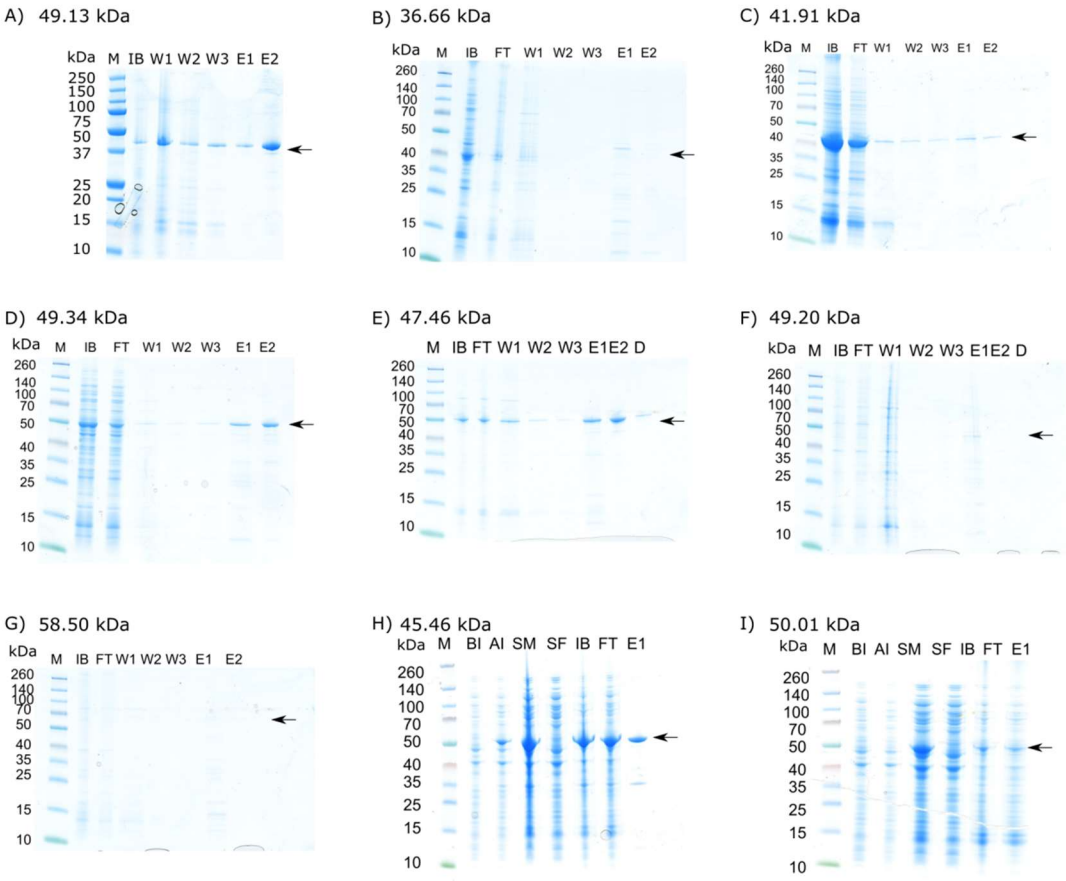



**Figure 3.10) SDS-PAGE gels for ChapCAT purifications**

A) ChapCAT 1 purification; B) ChapCAT 2 purification; C) ChapCAT 3 purification; D) ChapCAT 4 purification; E) ChapCAT 5 purification; F) ChapCAT 6 purification; G) ChapCAT 7 purification; H) ChapCAT 8 purification; I) ChapCAT 9 purification; ChapCAT sizes are provided. Key: M – Marker; SM – Starting Material; SF – Soluble Fraction; IB – Inclusion Body; FT – Flow Through; W – Wash; E – Elution; D – Dialysis;

The pooled elution fractions for each ChapCAT were subject to dialysis such that ChapCAT proteins were retained in a buffer (50 mM ammonium bicarbonate) suitable for tryptic digestion. Typically, pooled fraction volumes were equivalent to ~600 µL which were then increased to 2 mL during the dialysis process. As such, ChapCAT concentrations were too low and were subjected to concentration to a volume of ~200 µL resulting in concentrations of purified ChapCAT ranging from 200 to 1600 ng $\mu L^{-1}$. To the concentrated ChapCAT volume, an equivalent volume of RapiGest$^{TM}$ was added. The final ChapCAT concentrations were determined using BSA standards with approximated ChapCAT concentrations ranging from 100 to 800 ng $\mu L^{-1}$. At this point, ChapCAT 9 was deemed too low to quantify in an MS experiment and was not continued.

Western blotting analysis of the final purified ChapCAT samples identified the presence of proteolytic products for ChapCAT 2, 4 and 5, containing at least the *C*-terminal histidine-tag (Figure 3.11). The effect of this proteolysis is discussed further in this chapter in terms of quantification. Contaminants observed at the end of purification and concentration are irrelevant in this instance as Q-peptides are subject to transition selection, with each transition targeting only the peptide sequence of interest. By specifically targeting each unique Q-peptide, contaminants do not affect the final absolute copy per cell value determined for each Q-peptide (and thus chaperone). Finally, as each ChapCAT quantification is determined via in an internal glu-fibrinopeptide standard sequence, the presence of contaminants does not affect the quantification of the ChapCAT standard.

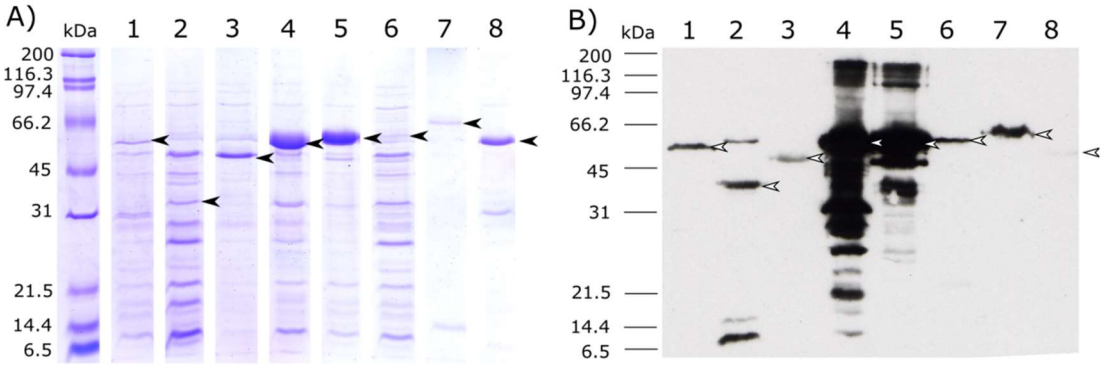



**Figure 3.11) SDS-PAGE and western blotting analysis of purified and concentrated ChapCATs**

A) Ten µL of each ChapCAT was loaded into a freshly-cast gel. B) Western blotting demonstrated the presence of proteolytic products with a His-tag in ChapCATs 2, 4 and 5. Primary antibody: α-His, Secondary antibody: α-Mouse.

## 3.4 Labelling efficiency

Naturally, proteins synthesised by *E. coli* are not isotopically labelled, allowing for determination of the labelling efficiency of each ChapCAT. Complete labelling of all Q-peptides results in only the heavy-labelled (*m/z* +3 if doubly charged) monoisotopic peak present in a chromatogram of the $MS^E$ data collected. I selected three Q-peptides for each ChapCAT at random, and inspected their chromatograms (Figure 3.12), noting 100 % labelling in all cases. On this basis, I was reassured that any absolute quantifications performed using these Q-peptides would not need to be adjusted.

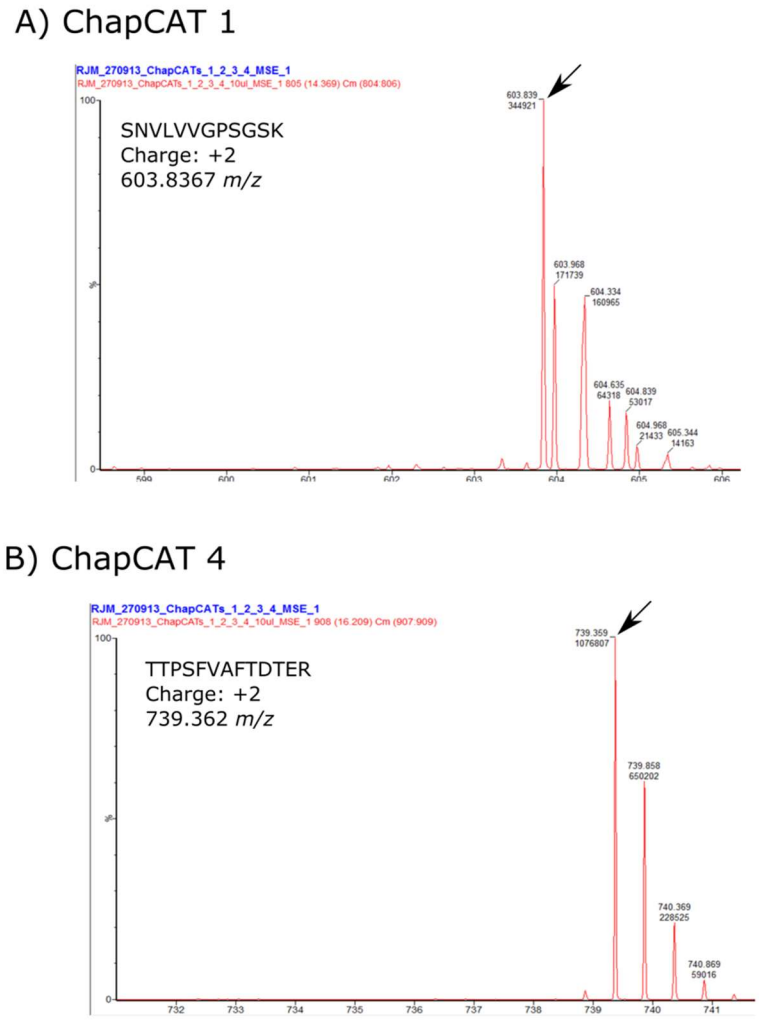



**Figure 3.12) Example of 100 % labelling observed in ChapCAT 1 and ChapCAT 4.**

One can determine the labelling efficiency by observing the precursor ion peak at the expected heavy labelled *m/z* ratio for doubly charged ions. If an isotopic peak is present approximately *m/z* 3 lower, the non-labelled version of the peptide is present within the ChapCAT standard sample. To demonstratre 100 % labelling efficiency, the chromatogram from Q-peptide from ChapCAT 1 (A) and a Q-peptide from ChapCAT 4 (B) are shown. Black arrows indicate the monoisotopic peak of the heavy-labelled ion.

## 3.5 Digestion

Each biological replicate was subject to cell counting and separated into ~25,000,000 cell aliquots prior to tryptic digestion with a ChapCAT standard. For each biological replicate, three aliquots were digested: a ChapCAT-containing digest (CCD) and two yeast only digests (YOD) for use in ChapCAT dilution at later stages. To confirm completeness of digestion, SDS gels were run for samples taken pre- and post-addition of trifluoroacetic acid (TFA) for the digestion of all aliquots of the four biological replicates of NG yeast combined with ChapCAT 8 (Figure 3.13). As expected, no protein bands are present in the SDS-PAGE analysis of both pre- and post-TFA addition, with trypsin precipitating out of solution after TFA addition, demonstrating complete proteolysis. To examine completeness of digestion for individual Q-peptides, a digestion time course was also carried out on fresh yeast samples, further discussed in Chapter 5.

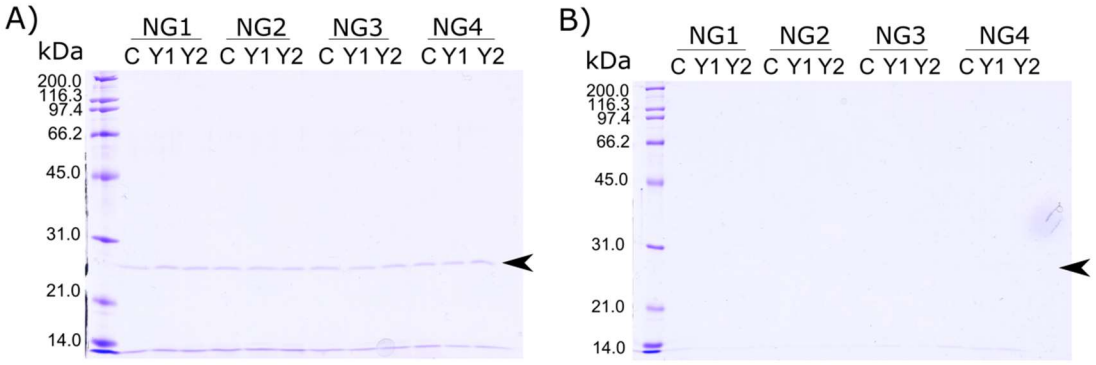



**Figure 3.13) SDS-PAGE analysis of pre- and post-TFA addition for digestion protocol**

Key: C – CCD sample, Y – YOD sample; A) Prior to TFA addition, a single band corresponding to trypsin is observed. No other protein bands are present. B) Following TFA addition, the ~23 kDa band corresponding to trypsin is no longer visible.

## 3.6 Quantification of ChapCAT

To determine the final ChapCAT concentration and labelling efficiency in the CCD aliquot, an $MS^E$ experiment on the Synapt G2-Si was conducted on all samples. To determine the final ChapCAT concentration, the signal intensity corresponding to 100 fmol $\mu L^{-1}$of light labelled glu-fibrinopeptide (*m/z* 785.5) was compared to the signal intensity of the heavy labelled glu-fibrinopeptide (*m/z* 788.8) contained within the ChapCAT construct (Figure 3.14). In Figure 3.15, this would equate to 28.32 fmol $\mu L^{-1}$ of ChapCAT 8 in the CCD aliquot for the NG1 sample. Final ChapCAT concentrations for all CCD aliquots ranged from 13 fmol $\mu L^{-1}$ to 83.38 fmol $\mu L^{-1}$.

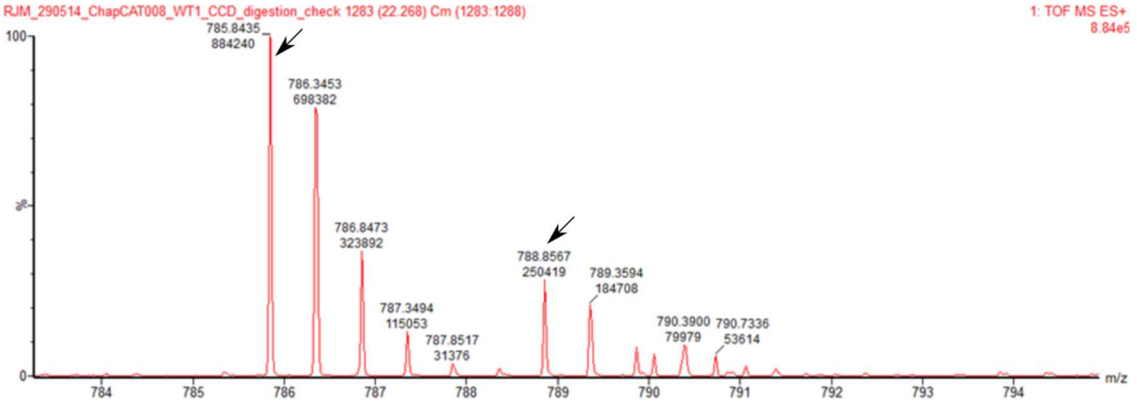



**Figure 3.14) Quantification of ChapCAT 8 using the internal glu-fibrinopeptide standard**

100 fmol $\mu L^{-1}$ light labelled glu-fibrinopeptide was spiked into a ChapCAT 8 CCD digest containing heavy labelled (*m/z* +3, +6 Da) glu-fibrinopeptide. The respective intensities are used to determine final ChapCAT concentration within the CCD digest.

## 3.7 Absolute quantification of *S. cerevisiae* chaperones in batch-grown cultures under NG conditions

Following transition selection and retention time scheduling (Figure 3.15), absolute chaperone copy per cell values were determined. Four ChapCAT dilutions were examined, for each yeast analyte biological replicate, containing 10 fmol, 1 fmol and 250 amoles of ChapCAT and a yeast only sample. Each SRM experiment was conducted on a yeast extract generated from approximately 200,000 cells. The run with the analyte:ChapCAT or ChapCAT:analyte ratio closest to 1:10 was selected for use towards absolute quantification (Figures 3.16e & 3.16f). Exported reports were subject to mProphet (Reiter et al., 2011) processing and Q-peptides were required to pass a 1 % FDR. Although three minute retention time windows were collected, only one minute windows were processed *in silico* in order to improve mProphet peak selection (discussed further in Chapter 4).

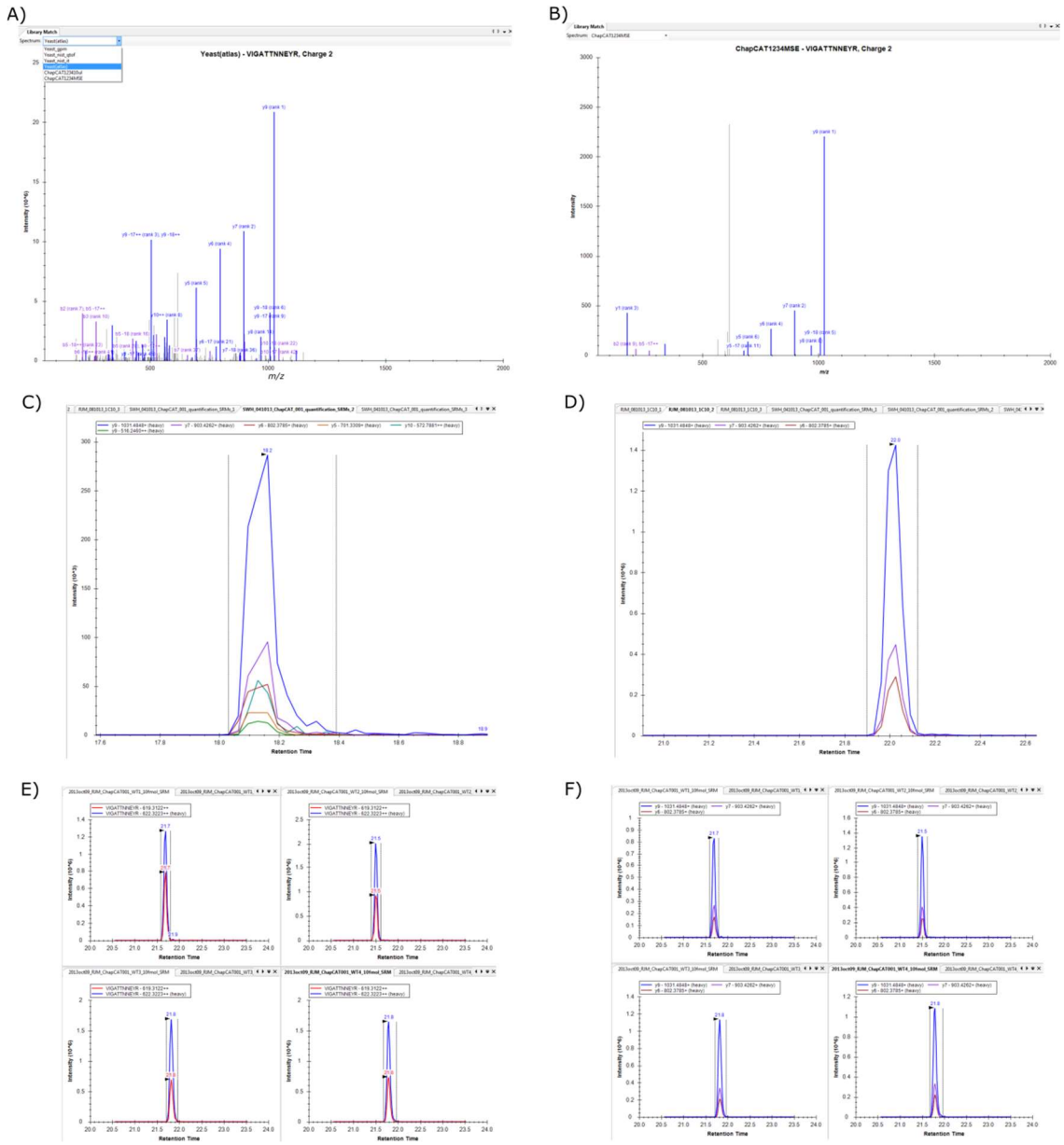



**Figure 3.15) Selection of top 3 transitions for SRM design, retention time scheduling and scheduled SRM experiments for use towards absolute quantification.**

Peptide: VIGATTNNEYR, ChapCAT: ChapCAT 1, Chaperone: Hsp104. A) Peptide product ions are searched against known spectral databases, including SRM atlas (shown). B) Following an $MS^E$ experiment, the top seven most intense ions are selected. C) An unscheduled SRM experiment is performed on a digest contaning only ChapCAT, targeting these seven ions, with the top three most intense ions selected as the final transitions for use towards absolute quantification. D) A ChapCAT and yeast digest is subject to an unscheduled SRM experiment targeting the top three transitions, allowing for determination of the the retention time in the respective condition. Yeast digests containing 10 fmol, 1 fmol and 250 amoles of ChapCAT are subject to scheduled SRM using the retention times determine in D, with both heavy and light isotope variants targeted (E). The concentration matching closest to 1:10 (analyte:ChapCAT) is used towards absolute quantification of the light analyte peptide (F – heavy labelled variant is shown). Full page figures are given in Appendix 5.

Using a classification system similar to that of Brownridge and colleagues (Brownridge et al., 2013, Lawless et al., 2016), peptides not observed in an SRM experiment were classified 'C' peptides. Peptides for which the yeast analyte counterpart was below the limit of detection were defined as 'B1' class. Peptides that failed the 1 % mProphet FDR cut-off were classified 'B2'. In both cases, an upper limit of detection was defined for the analyte peptide (and parent protein). All peptides that otherwise passed these criteria were classified 'A'. However, these were subject to additional quality control processing steps: comparison to the analyte peptide cpc values for the same parent chaperone (discussed in Chapter 4), reliable mProphet peak selection confirmation (discussed in Chapter 4) and digestion efficiency testing (discussed in Chapter 5). Peptides that failed these quality control steps were classified into the subset 'A2' with those passing all steps classified into the highest quality subset 'A1'. Subsequently, 88 peptides were classified 'A1' whilst 28 were classified 'A2' (Figure 3.16). 'A1' peptides were used for final absolute quantification of the analyte peptides and parent chaperone proteins.

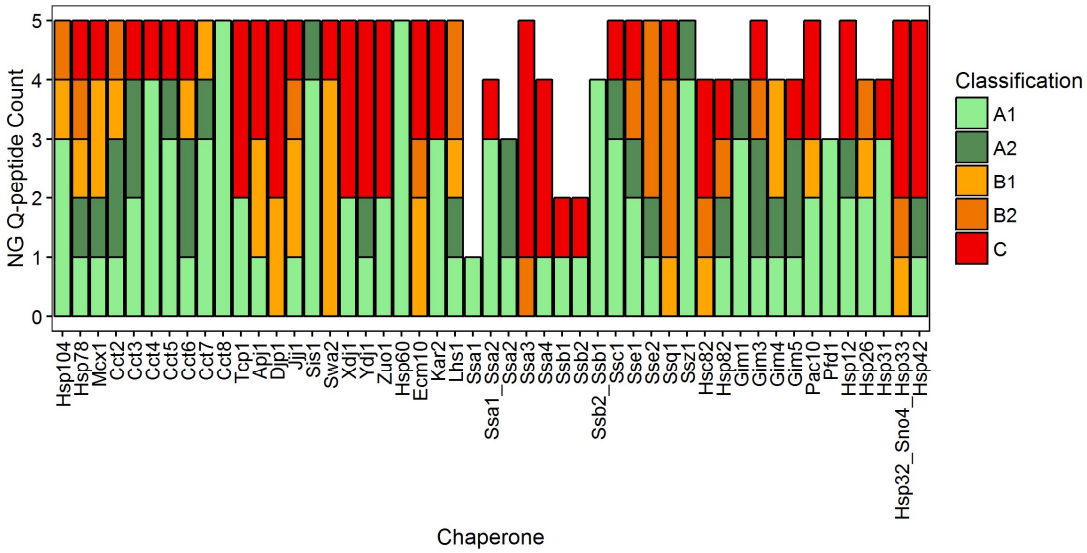



**Figure 3.16) Classification of Q-peptides under NG conditions**

88 peptides were classifed 'A1'; 28 peptides were classified 'A2'; 27 peptides were classified 'B1'; 20 peptides were classified 'B2' and 58 peptides were classified 'C'. From the 88 'A1' class peptides, absolute abundances for 40 chaperone proteins were determined in NG conditions.

Firstly, to observe SRM ChapCAT reproducibility, I compared the absolute copy per cell values including both 'A1' and 'A2' classes at the peptide level across biological replicates (Figure 3.17). I did not include peptide biological replicate cpc values that did not pass a 1 % mProphet FDR (discussed in Chapter 4. An unpaired Mann-Whitney Wilcoxon test (*U* test, $p < 0.05$) was performed between the cpc values for a ChapCAT biological against the cpc values

for each of the remaining biological replicates for that ChapCAT under the same conditions. Subsequent p-values are available in Appendix 6. According to these tests, for ChapCAT 5 under NG conditions, biological replicate 3 is significantly different to the remaining three biological replicates. Despite this, robust coefficient of variances (rCVs) for combined 'A1' and 'A2' class peptides within ChapCAT 5 under NG conditions ranged from 9 to 45 %, with a median rCV (rCV) of 10.24 %.

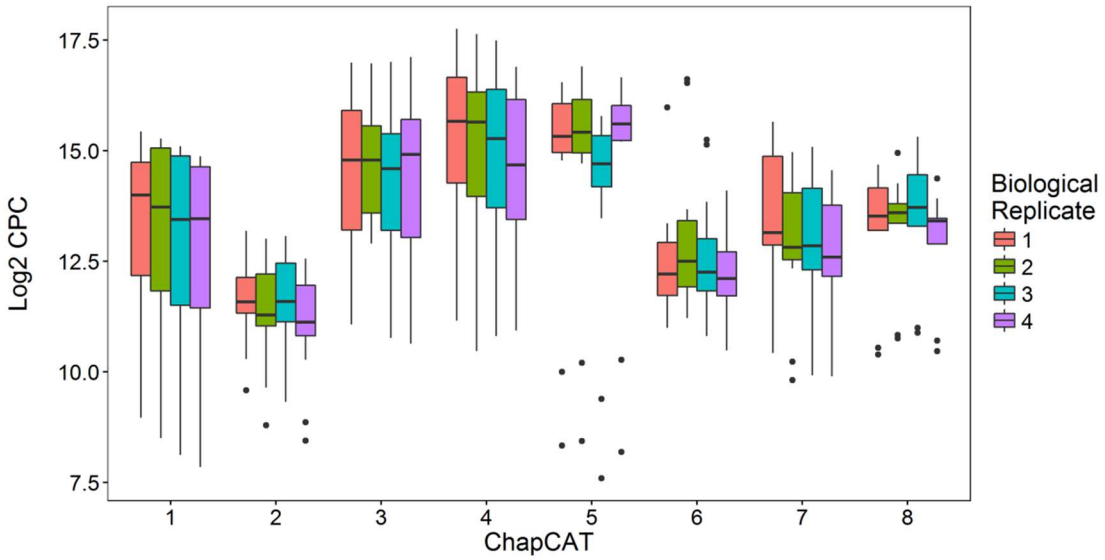



**Figure 3.17) Spread of cpc values across biological replicates for each ChapCAT under NG conditions**

Each SRM experiment was run as an independent combination of ChapCAT, biological replicate and condition. Q-peptide biological replicates that did not pass a 1 % FDR are not included here.

Secondly, to examine whether proteolytic products observed in Figure 3.12 were adversely affecting absolute quantification, I compared both the classification and copy per cell values determined for each peptide within the proteolysed ChapCAT construct (Figure 3.18). If proteolysis was unequal throughout the purified ChapCAT proteins, one would expect different abundances of the *N*- and *C*-terminus of the construct that would result in erroneous ChapCAT and analyte peptide quantification. By observing the peptide classification as a function of the peptide location within the ChapCAT construct ('order' - a numerical *N*- to *C*-terminal sequence) between one that did not proteolyse (Figure 3.19a) and one observed to proteolyse (Figure 3.19d), I observed no preference in the classification of the peptide with regards to the location within the construct. Proteolysis therefore appeared to have no effect on the suitability of the peptide towards absolute quantification. In terms of absolute abundance, if proteolytic products were present unequally, that is, there are more proteolysed *N*-terminal Q-peptides than *C*-terminal, the absolute abundance of a peptide would differ between the termini. Again, comparing a non-proteolysed ChapCAT (Figure 3.19b) and a proteolysed ChapCAT (Figure

3.19e), I observed whether Q-peptides targeting a common parent protein held different absolute abundances respective of their order within the QconCAT construct. Strangely, for the non-proteolysed ChapCAT 1, lower abundances were observed at the *C*-terminus (higher order Q-peptides) but demonstrated good agreement between two of the Q-peptides targeting Mcx1. The apparent disparity between the *N*-terminal peptide and *C*-terminal peptides targeting Hsp82 is likely a direct result of chromatography and the ability of each Q-peptide to be observed in an SRM experiment. As discussed previously, no two peptides share the same properties and will exhibit unique features in an LC-MS system. For the proteolysed ChapCAT 2.1, there appeared to be no relationship between the location of the Q-peptide within the ChapCAT construct and the absolute abundance determined. Finally, to ensure that the raw intensity of the Q-peptide was not affected by its location in the ChapCAT, I observed the XICs for both the standard (ChapCAT) Q-peptide and the analyte (yeast) counterpart. If the proteolytic products were in unequal abundance, the standard XIC in terms of the location of the Q-peptide would show differing patterns of intensity to the analyte XIC. I did not observe this to be the case in either the non-proteolysed or proteolysed ChapCAT constructs (Figures 3.19c and 3.19f respectively), and so deemed any proteolytic products to be present in equal abundance, therefore would not affect the calculation of absolute abundance of the chaperone. This was also the case for ChapCAT 4 and ChapCAT 5 (Appendix 7).

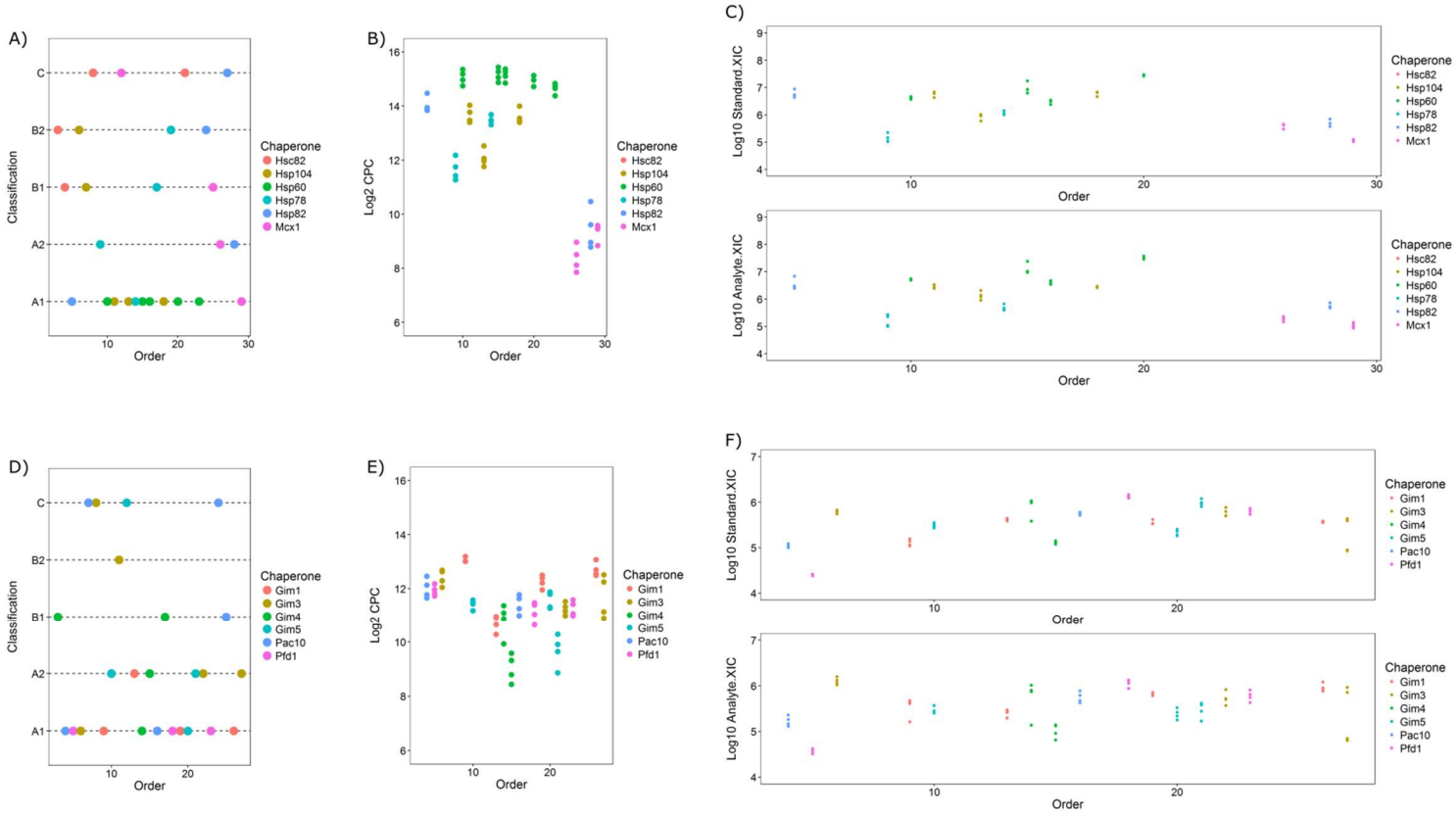



**Figure 3.18) Comparing non-proteolysed and proteolysed ChapCAT Q-peptide classifications, copy per cell values and XICs**

A – C) For the non-proteolysed ChapCAT 1, figures demonstrate the Q-peptide classification, absolute abundance and XICs respectively in response to the Q-peptides location (order) within the ChapCAT construct. D – F) For the proteolysed ChapCAT 2.1, figures demonstrate the Q-peptide classification, absolute abundance and XIC respectively in response to the Q-peptides location within the ChapCAT construct. For panels B, C, E and F, only 'A' class Q-peptides are shown.

Absolute quantification at the peptide level was highly reproducible, with a median rCV of 15.23 % for the combined 'A1' and 'A2' class Q-peptides. Using 'A1' class Q-peptides alone, a median protein-level rCV of 18.47 % was obtained (Figure 3.19). The chaperone protein pairs Ssa1, Ssa2 and Ssb1, Ssb2 were targeted by Q-peptides that were both unique to the protein, and unique to the protein pair. As such, it was anticipated that the abundance of the protein pair would be the sum of the unique protein abundances, and therefore could be used alongside unique Q-peptides to determine chaperone abundance. For the protein pair Ssa1_Ssa2, the NG absolute abundance was 125,000 cpc, whilst unique protein abundances were 90,000 cpc (determined via a single unique Q-peptide) and 78,000 cpc (determined via 3 unique Q-peptides) for Ssa1 and Ssa2 respectively. The sum of the unique protein abundances was higher than the combined non-unique protein pair abundance and as such, protein cpc values for these proteins were determined using unique Q-peptides only. For the protein pair Ssb1_Ssb2, NG absolute abundance was 84,000 cpc, whilst unique protein abundances were 45,000 cpc and 58,000 cpc for Ssb2 and Ssb1 respectively. Again, the sum of the unique protein abundances was higher than non-unique protein data; again, only unique peptide derived values were used. There were no unique Q-peptides available targeting the proteins Hsp32, Sno4 and Hsp33, with the non-unique Q-peptide absolute abundance below the limit of detection of the mass spectrometer and failing to pass a 1 % mProphet FDR ('B1' and 'B2' class peptides respectively).

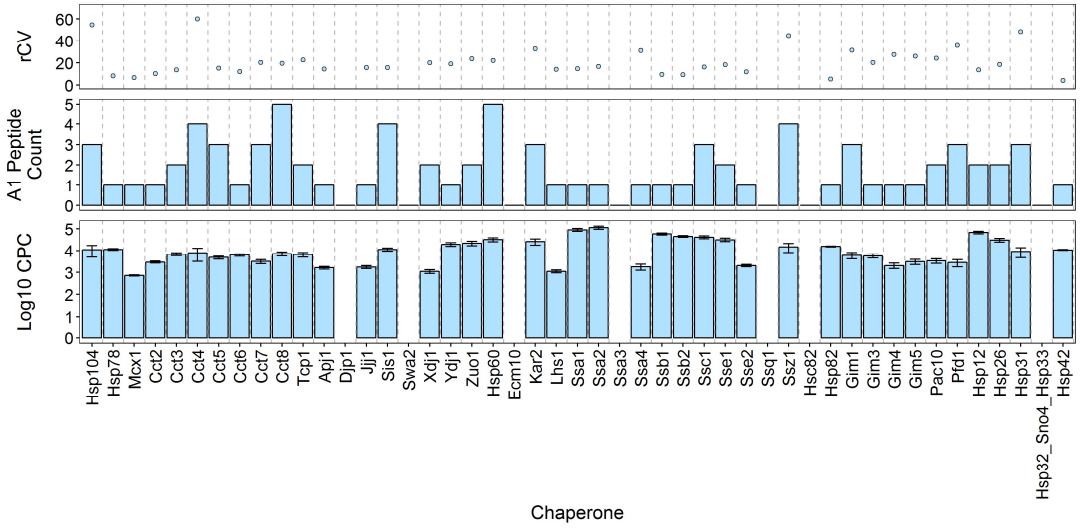



**Figure 3.19) rCV, A1 Peptide Count and cpc values under NG conditions**

Cpc values were obtained for 40 chaperones under conditions of NG. rCV values are below 40, with minor exceptions.

Whilst absolute cpc values could be determined for 40 chaperones of the 49 targeted, an upper limit could be defined for a further 9 chaperones/chaperone groups using the median of the biological replicate upper limits for the 'B1' Q-peptide with the highest value targeting the chaperone. In this latter set, non-unique 'B1' class Q-peptides targeted the protein group Hsp32_Sno4_Hsp33, defining an upper limit of ~700 cpc for the summed total of these proteins. For every chaperone targeted by the eight ChapCAT constructs, an absolute cpc value could be determined either as an upper limit or as an absolute measure of its abundance in NG conditions, with chaperone cpc values ranging from 500 cpc to 114,000 cpc. The Hsp70 class chaperones made up the majority of cellular chaperone abundance, whilst Hsp90 chaperones occupied the lowest proportion (Table 3.4).

**Table 3.4) Total cpc for each chaperone class under NG conditions**
Total cpc was determined as a sum of all 'A' class chaperones cpc values.

| Class | Number of A-class chaperones | Total cpc (3 s.f.) | Percentage of total chaperone cpc (%) (3 s.f.) |
|---|---|---|---|
| HSP90 | 1 | $1.54 \times 10^4$ | 2.11 |
| PFD | 6 | $2.29 \times 10^4$ | 3.15 |
| AAA | 3 | $2.30 \times 10^4$ | 3.15 |
| HSP60 | 1 | $3.20 \times 10^4$ | 4.39 |
| CCT | 8 | $4.51 \times 10^4$ | 6.19 |
| HSP40 | 6 | $5.91 \times 10^4$ | 8.11 |
| SMALL | 6 | $1.19 \times 10^5$ | 16.4 |
| HSP70 | 13 | $4.12 \times 10^5$ | 56.5 |

## 3.8 Comparison to other proteomics techniques towards determination of protein abundance

Quantification at the protein level may be determined using relative or absolute techniques. I compared these ChapCAT-determined cpc values to another absolute SRM attempt also using the QconCAT strategy but using up to 2 Q-peptides per chaperone (Brownridge et al., 2013, Lawless et al., 2016) – herein termed 'CopyCAT' cpc values. Irrespective of the number of Q-peptides selected, the Spearman's rank correlation coefficient was 0.799 and 0.901 at the peptide level respectively (Figure 3.20).

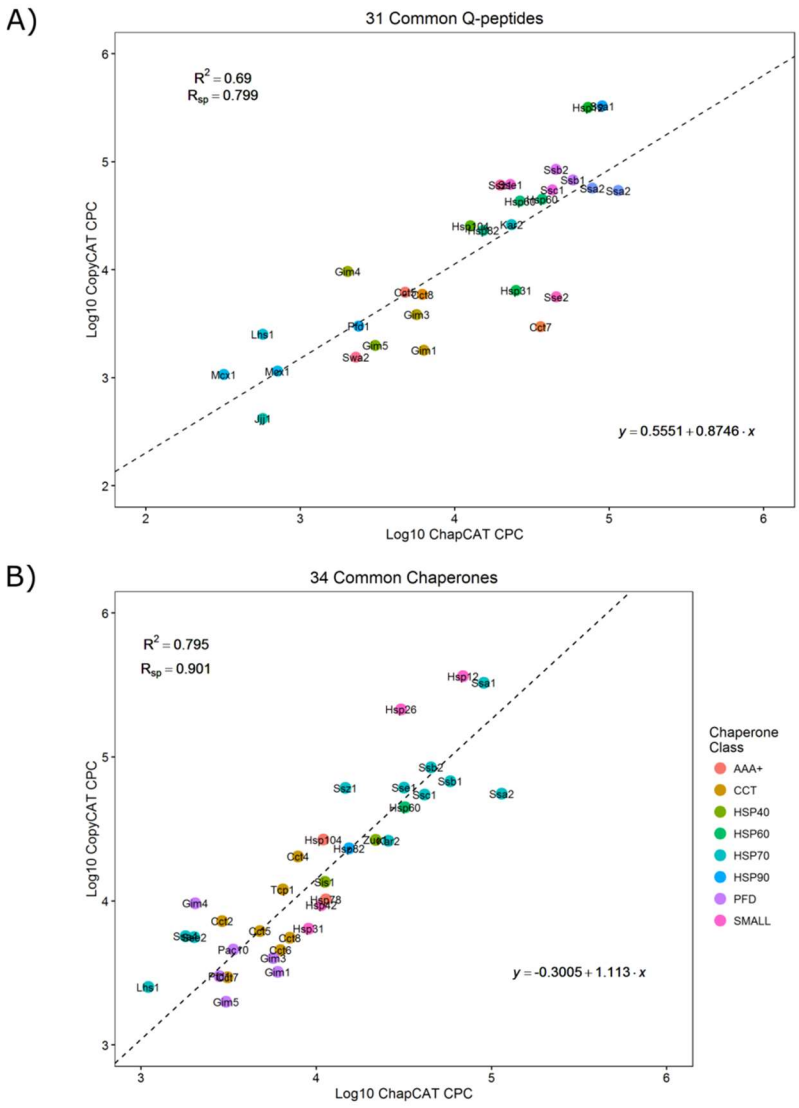



**Figure 3.20) Comparsion between CopyCAT- and ChapCAT- determined chaperone abundances (SRM QconCAT strategies)**

A) 31 Q-peptides were considered 'A' in both datasets and so were used towards absolute quantification. The peptide level cpc values were correlated, resulting in an $R^2$ value of 0.69 and a Spearman's correlation coefficient of 0.799. B) 34 chaperones were quantified in both datasets, with ChapCAT-determined cpc values calculated using additional (up to five) Q-peptides. $R^2$ = 0.795, Spearman's correlation coefficient = 0.901.

As was performed by Lawless and colleagues (Lawless et al., 2016) I compared my chaperone cpc values to yeast data sets available in the PaxDB database (Wang et al., 2012). In order to rescale my cpc values to ppm, I assumed 60,000,000 total protein molecules per cell.

Epitope tagging strategies, such as that developed by Ghaemmaghami and colleagues (Ghaemmaghami et al., 2003) typically involve the attachment of a high-affinity epitope tag to the protein of interest for identification and purification. In this instance, Ghaemmaghami and colleagues introduced an epitope tag, similar to that of the tandem affinity purification tag, into

the open reading frame of all annotated proteins in *S. cerevisiae*. As such, each protein could be identified and purified before being analysed via western blotting methods using a polyclonal antibody to the epitope tag. Absolute quantification was achieved through the use of quantitative western blotting, in which internal and external standards were included in the SDS-PAGE gel to determine copy values for each protein. Thirty-six chaperones were common to both the ChapCAT-derived cpc dataset and the copy values determined using this epitope tagging strategy (Figure 3.21a). Whilst I observed a low $R^2$ value of 0.339 demonstrating a lack of linear relationship between the two, there is a modest relationship between the ranks of the chaperone ppm values (Spearman's rank correlation coefficient was 0.580). As demonstrated by Lawless and colleagues (Lawless et al., 2016) there is reduced similarity between techniques used to determine absolute copy per cell values between quantitative western blotting methods and MS and has been reported in other dataset comparisons (Brownridge et al., 2013).

Using SILAC, de Godoy and colleagues (de Godoy et al., 2008) were able to perform metabolic incorporation of heavy-labelled ($^{13}C_6/^{15}N_2$) lysine into diploid and haploid cells allowed for the pairing of peptides with their light-labelled counterparts after sample digestion, fractionation, MS and processing with the MaxQuant (Cox and Mann, 2008) software package. Peptides are paired according to a mass shift of 6 Da, the difference in mass between the heavy-labelled and light-labelled lysine, and the SILAC protein ratio determined (via the median of all peptide SILAC ratios) is used for quantification of the protein in both haploid and diploid yeast. To express the protein SILAC ratio in ppm, it was rescaled by the sum total. Of the 45 chaperones common to the haploid SILAC data and my ChapCAT ppm, agreement was improved over epitope tagging strategies ($R^2$ = 0.659 and $R^2$ = 0.580 respectively) (Figure 3.21b).

PPM values determined via spectral counting methods (Figure 3.21c) demonstrated the best agreement with the ChapCAT ppm dataset, with 46 common chaperones ($R^2$ = 0.796). PeptideAtlas builds are the result of aggregation over multiple proteomics experiments uploaded to PeptideAtlas (Deutsch et al., 2008, Deutsch, 2010), with each dataset reprocessed via standardised database searches and peptide scoring methods. The authors of PaxDB (Wang et al., 2012), exported these builds and were able to determine ppm values for proteins using a combination of peptide spectral counts, actual and estimated peptide coverage of a protein and the estimated likelihood of detection of a peptide based on its length. As such, the ppm values determined using PeptideAtlas builds are likely more reliable, having been gathered from multiple proteomics experiments submitted to the PeptideAtlas repository.

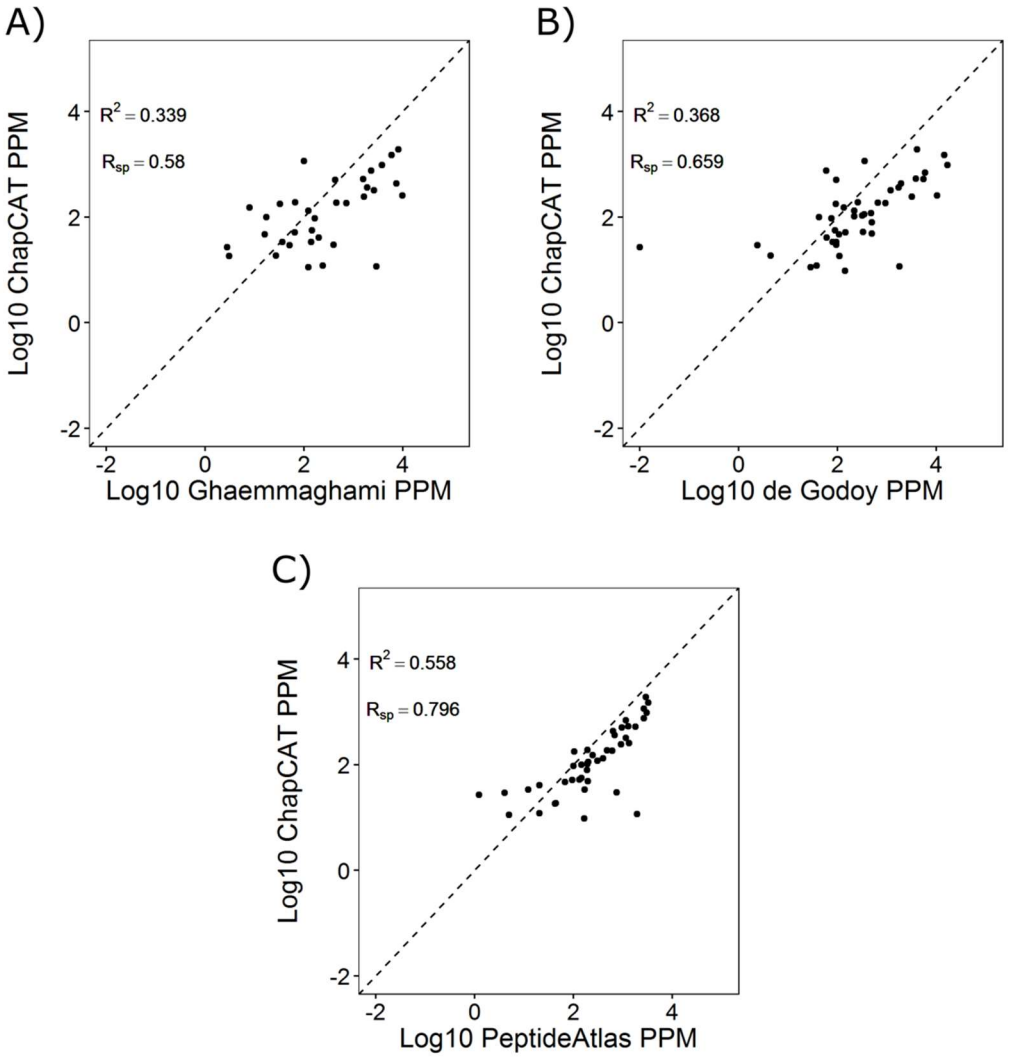



**Figure 3.21) Comparison of the PPM determined via ChapCAT-SRM methodology to epitope tagging, SILAC and spectral counting methods**

ChapCAT-determined PPM values were compared to ppm determined by epitope tagging, Ghaemmaghami (A), SILAC, de Godoy (B) and spectral counting methods, PeptideAtlas (C).

Together, these results suggest best agreement for protein quantification between similar methods. Highest agreement was observed between variations of the QconCAT strategy towards absolute abundance, whilst the lowest agreement was observed with epitope tagging strategies and gel-based approaches.

# Chapter 4 - Quality control management of selected reaction monitoring data

SRM is a valuable tool for detecting and quantifying proteins from highly complex samples in a targeted mode. In this mode, only predefined peptides are detected and quantified by virtue of isolation of selected precursor and product *m/z*. The combination of a precursor and product *m/z* constitutes a transition, with multiple transitions targeting a single peptide in an SRM assay. In my QconCAT experiment, I select three transitions per peptide, such that the top three most intense product ion signals (as determined by a prior $MS^E$ and unscheduled SRM experiment, see Methods) are detected and their resulting spectra used towards peptide quantification. By targeting the same transitions for a peptide in each SRM experiment, reproducibility is high and quantification is robust, enabling for consistent analysis of the peptide/protein of interest across multiple samples.

As discussed in Chapter 1, the proteomics workflow for generating absolute quantitative data is long, with many requirements needing to be met to ensure the reproducibility and accuracy of the data acquired. Any protein lost during sample preparation will not be quantified and will result in inaccuracies. However, even after data acquisition, steps are required to ensure protein quantification is accurate and reliable. In an untargeted experiment, spectral data are searched against protein sequence databases to identify proteins in the sample. In an SRM experiment, as the protein is already known, alternative processing must be performed to ensure the spectra observed represents truly 'what I think I am seeing'. mProphet, developed by Reiter and colleagues (Reiter et al., 2011), is a tool used to validate peptide peaks. Like database search validation, the use of decoys is involved. In complex samples, peak groups can occur in which multiple peaks can co-elute and match the relative intensities of the targeted peptide by chance. 'Decoy' transitions are created, in which the retention time and dwell times are kept the same as 'target' transitions, but random integers to the precursor and product ion *m/z* are added such that there is a difference of *m/z* >5 in a single transition. In a separate SRM experiment, the yeast only sample is analysed but with the decoy transitions targeted. As such, false 'decoy' peak groups and true 'target' peak groups are present in the final spectral data collected. Each peak group is scored according to co-elution, peak shape similarity, intensity and correlation of relative intensities between peak groups, and if standard peptides are used, a correlation of the light peak group versus the heavy peak group. As the false peak groups are known by virtue of their randomised *m/z*, the proportion of false discoveries in the spectral data are known and may be filtered out according to a 1 % FDR.

In these studies, I used a combination of both manual verification and mProphet processing to create a high quality, reproducible, quantitative dataset. Selection of Q-peptides was performed using stringent criteria and the ChapCATs were subject to multiple quality control steps to check for proteolysis and labelling efficiency prior to an SRM experiment (Chapter 3). Furthermore, to ensure the digestion efficiency of each Q-peptide as both analyte and standard is complete and comparable, I performed digestion time course experiments, assessing the suitability of the peptide to be used towards protein-level absolute quantification (Chapter 5). In this chapter, I describe additional quality control steps carried out after data acquisition to ensure all resulting protein-level absolute quantifications are reliable and reproducible.

## 4.1 The classification scheme

I built on the classification scheme of Brownridge and colleagues (Brownridge et al., 2013) to define the suitability of various peptides towards quantification (Table 4.1). Simply, peptides that were not observable in an SRM experiment were classified as 'C'; peptides that were below the limit of detection in the light analyte counterpart or failed to pass a 1 % mProphet FDR were classified as 'B', whilst peptides that were observable as both the light (analyte) and heavy (standard) counterparts were suitable for use towards absolute quantification of the parent protein and classified as 'A'. I then extended this classification scheme; 'B' type peptides were divided into 'B1' and 'B2' types depending upon whether a peptide was too low in the light counterpart to quantify (B1) or had failed a 1 % mProphet FDR (B2). I also extended the classification of 'A' type peptides to 'A1' for peptides that were used for absolute quantification, or 'A2' for peptides that I did not deem suitable for quantification due to various additional erroneous factors not considered by previous quality control steps. In this chapter I further discuss the classification process for 'A1' and 'A2' type peptides.

| Table 4.1) Classification scheme used for peptides | |
|---|---|
| **Classification** | **Description** |
| A1 | Passes all quality control steps, used towards absolute protein-level quantification |
| A2 | Passes a 1 % mProphet FDR but manual interjection required to prevent erroneous result |
| B1 | Below the limit of quantification |
| B2 | Failed a 1 % mProphet FDR |
| C | Not observed in the SRM experiment |

## 4.2 'B1' peptides

Peptides were classified as 'B1' if their light counterpart lay below the limit of detection of the mass spectrometer, and as such their transitions were indistinguishable from the background noise. In my example here, the peptide NALLDQYEYIFK targeting Mcx1 is classified as a 'B1' due to its lack of unlabelled analyte counterpart signal in both NG and HS conditions (Figure 4.1). I was therefore able to use the ChapCAT concentration to estimate the upper limit for the peptide-level abundance.

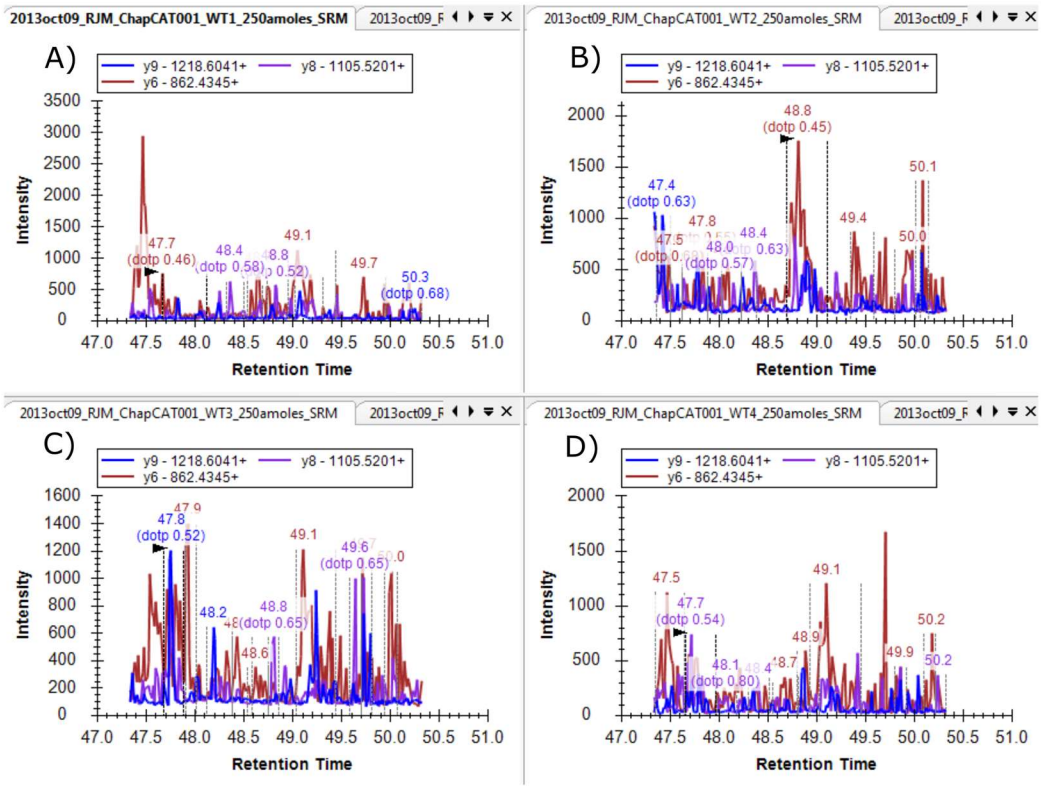



**Figure 4.1) Light transitions for the peptide NALLDQYEYIFK in NG conditions**

A) Biological replicate NG1; B) Biological replicate NG2; C) Biological replicate NG3; D) Biological Replicate NG4; The peptide should be observed as a single peak in the retention time window between 47.5 and 50.5 minutes. The black arrow on each figure represents the peak best matching the peptide according to Skyline.

## 4.3 *In silico* retention time window alteration & B2 peptides

An issue I encountered with mProphet during processing was the problem of selecting the 'top-scoring' peak group for peptides. In the example in Figure 4.2, all four biological replicates for each Q-peptide passed the 1 % mProphet FDR. However, upon closer inspection, single biological replicates are clear outliers amongst the remaining data for the same peptide. mProphet scoring depends largely on the degree of similarity between the peak groups for the standard heavy-labelled peptide counterpart and the peak groups for the analyte unlabelled counterpart. In this way, as identical product ions are used for both standard and analyte, the peak shape should have a high degree of similarity. However, if a peptide ion does not give a very high signal, likely due to physiochemical properties hindering its ability to be detected in an LC-MSMS experiment, noisy spectra can often be selected as the true peak. Using the example, for the peptide SITGFTDVGETILGVDSILK targeting Hsp31, I analysed the spectra with mProphet using a selected spike in of 1 fmol of ChapCAT, which gave a signal intensity ratio of 1:2 standard:analyte. The signal of both the light and heavy peptide counterpart was low and peaks were not smooth. Therefore, in this case, three NG biological replicates appeared to have background noise selected as the best scoring peak groups and therefore would likely underestimate the true abundance of the Hsp31 peptide (Figure 4.2). The higher intensity peak group was only selected by mProphet in the second biological replicate. All top-scoring peak groups in this case passed a 1 % mProphet FDR and so were deemed true matches. In this case, the ratio of heavy to light signal in each selected peak group was similar across biological replicates and so the rCV of the peptide was 14 %. This example demonstrates the ease with which incorrect data can be overlooked and so manual verification must still be used to confirm reasonable peaks are selected. I observed the peak selection for biological replicates containing 10 fmol of spiked in ChapCAT, as an increased standard signal would likely provide a better mProphet score for the peak group.

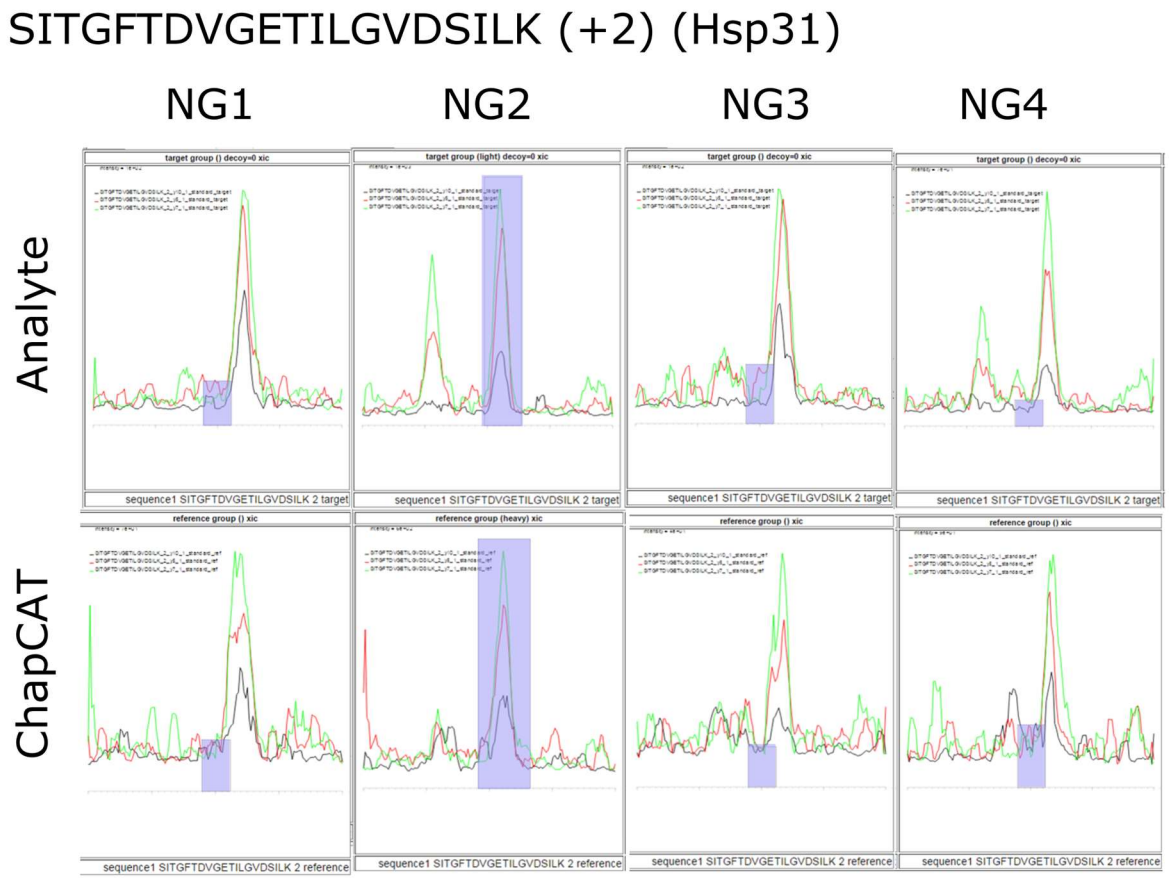



**Figure 4.2) mProphet peak group selection**

The best scoring peak group according to the mProphet scoring system are highlighted in purple boxes. With 1 fmol of ChapCAT spiked in to all biological replicates, the biological replicate NG2 for a Q-peptide targeting Hsp31 appears to have a different peak group selected as the true peak compared to other biological replicates (background noise appears to have been selected for the remaining three biological replicates). All selected peaks passed a 1 % FDR via mProphet processing.

In addition, to reduce the likelihood of mProphet selecting peaks based on similarity in noise across samples, I introduced an *in silico* retention time window reduction. To optimise peak group selection, I used a second processing software, Skyline (MacLean et al., 2010), to give indication as to the best matching peak. Skyline automatically selects the best peak for a peptide according to a heuristic measure of intensity, co-elution and the presence of matching isotopic standards, however it does not perform any FDR analysis. For product ions, the most intense peak groups typically correspond to the true signal of the peptide as the peptide is considered the main source of all three targeted product ions within a single retention time, this can be confirmed via co-elution with the standard peptide. Peak widths for peptide peaks across all ChapCATs as both standard and analyte counterpart were lower than 1 minute (Figure 4.3).

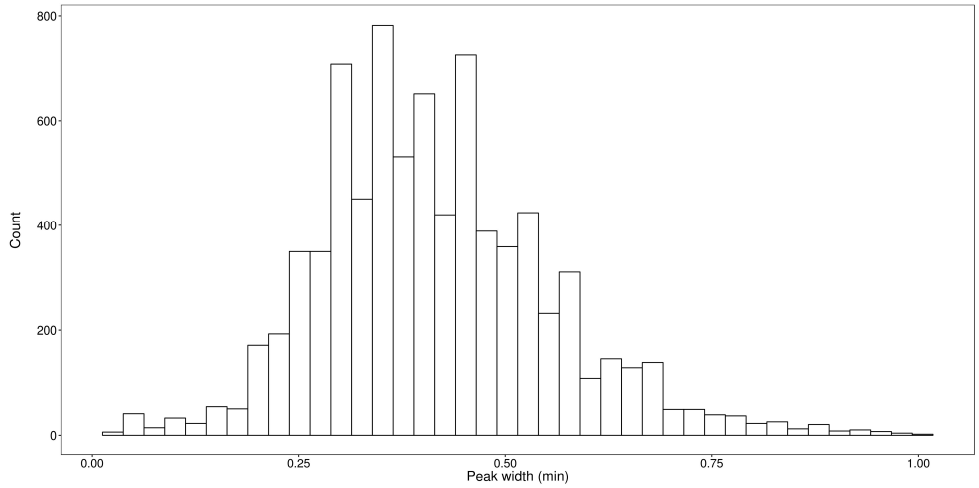



**Figure 4.3) Peak widths for all standard and analyte peptides**

Determined using Skyline, all peptides had a peak width below one minute regardless of biological replicate and ChapCAT sample concentration. Data presented is that for all 'A' and 'B2' type peptides, across four biological replicate for all ChapCAT spike ins.

This approach improved the peak selection by mProphet, as observed for the previous example used (Figure 4.4), with the rCV for the peptide SITGFTDVGETILGVDSILK decreasing to 5.59 %. However, it must be noted that for this peak selection, spectra corresponding to 10 fmol of spiked in ChapCAT was used, and so as the standard Q-peptide signal is of higher intensity, with scoring and peak shape was undoubtedly better. Regardless, I used this *in silico* retention time technique for every peptide standard and analyte counterpart to confirm optimal peak selections were performed. All selected peak groups were still required to pass a 1 % mProphet FDR according to mProphet scoring.

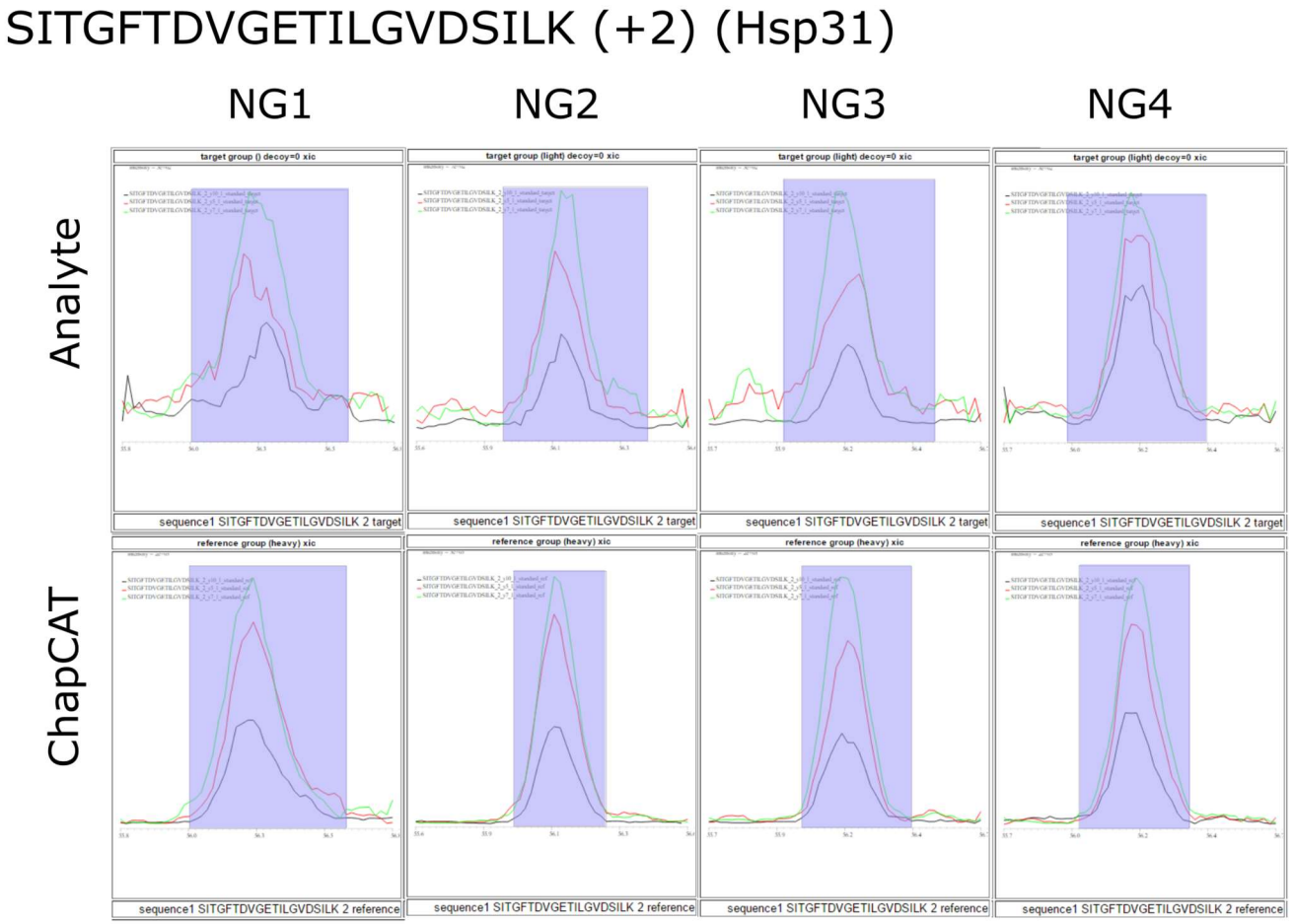



**Figure 4.4) Defining an *in silico* retention time window improves mProphet peak selection.**

With an altered retention time window and a 10 fmol ChapCAT spike in for this particular peptide, mProphet's peak group selection abilities improved. mProphet scores are 1.85, 2.95, 2.97, 2.98 for NG1, NG2, NG3 and NG4 respectively. All peak groups selected passed a 1 % mProphet FDR.

Often in cases I observed single biological replicate peak groups fail to pass a 1 % mProphet FDR. For instance, a single biological replicate (NG1) of the peptide LSGGVAVIR (targeting Hsp60) failed the 1 % mProphet FDR cut-off and so was not used for absolute quantification of Hsp60. Upon closer inspection of the peak groups selected, it appeared that mProphet selected a low intensity peak on the shoulder of a high intensity peak group (Figure 4.5). In this case, mProphet has given a higher score to the peak selected in NG1 than the higher intensity peak group, consequently identifying this peak as the true peptide signal. As this biological replicate for this peptide did not pass a 1 % mProphet FDR, the absolute abundance for Hsp60 was determined under NG conditions using only three biological replicates from this peptide. Under conditions of NG, 8 (6.9 %) 'A' peptides had only three biological replicates pass a 1 % mProphet FDR. In contrast, under conditions of HS, 15 (15.3 %) 'A' peptides had only three biological replicates pass a 1 % mProphet FDR. All protein quantifications were performed using peptides where at least three biological replicates passed the 1 % mProphet FDR threshold.

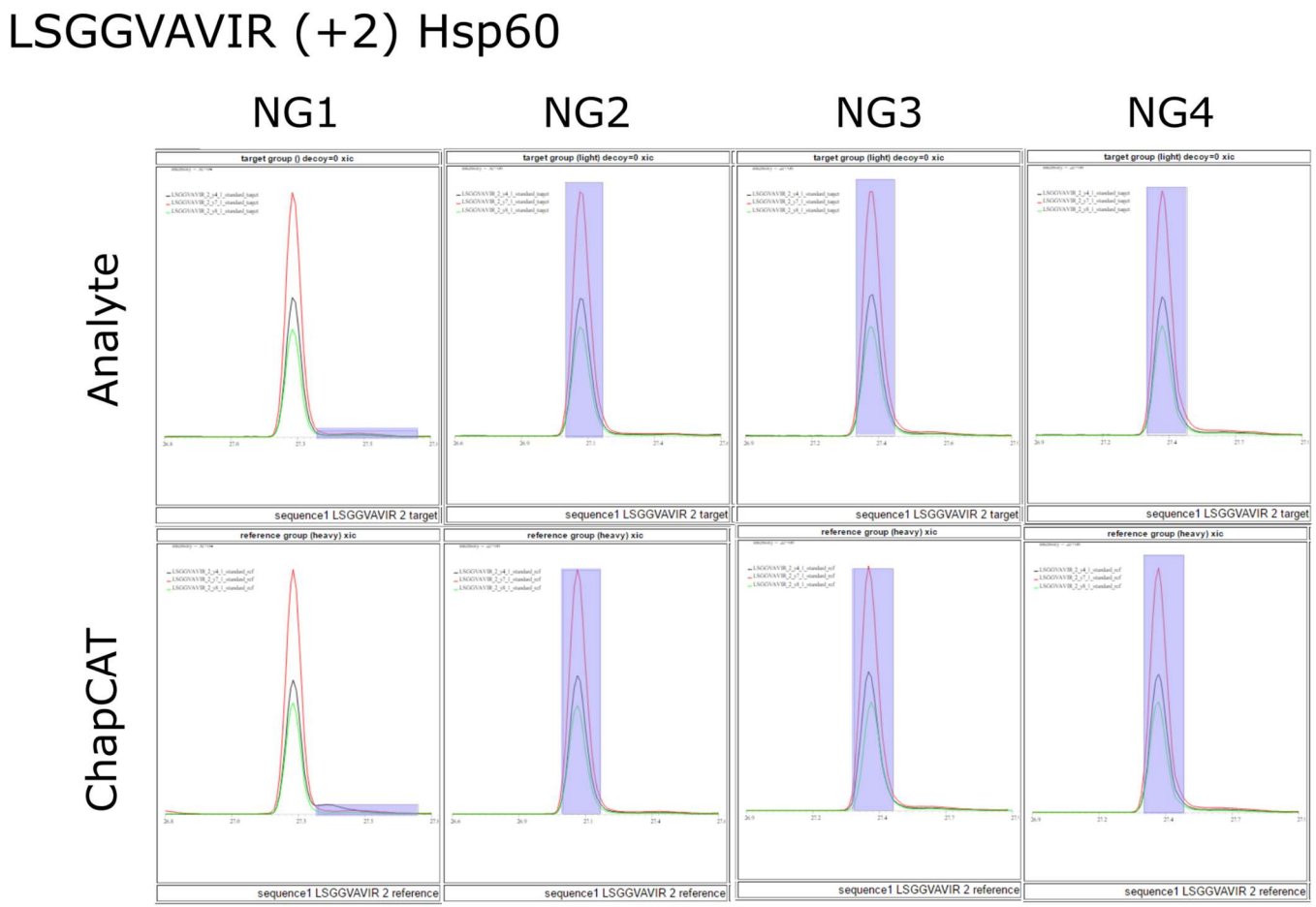



**Figure 4.5) Single biological replicates failing a 1 % mProphet FDR with mProphet scoring**

mProphet selects a low peak as the true peak signal for the peptide LSGGVAVIR targeting Hsp60 in NG1. This peak did not pass a 1 % mProphet FDR, and so was not used towards absolute quantification of the protein.

## 4.4 ‘A1’ and ‘A2’ type peptides

I classified the ‘A’ peptides according to the decision tree in Figure 4.6, typically beginning with peptides with an rCV over 30 but identifying other ‘A2’ candidates during the process. For all peptides considered, I first calculated the peptide-level abundance. The resulting rCVs and differences in abundance between peptides targeting the same parent protein, and their respective response to HS, offered a good indicator of any quality issues that may have been overlooked by both Skyline and mProphet. Such issues are further discussed below. I also considered digestion efficiencies in both the yeast and ChapCAT protein and is presented in Chapter 5.

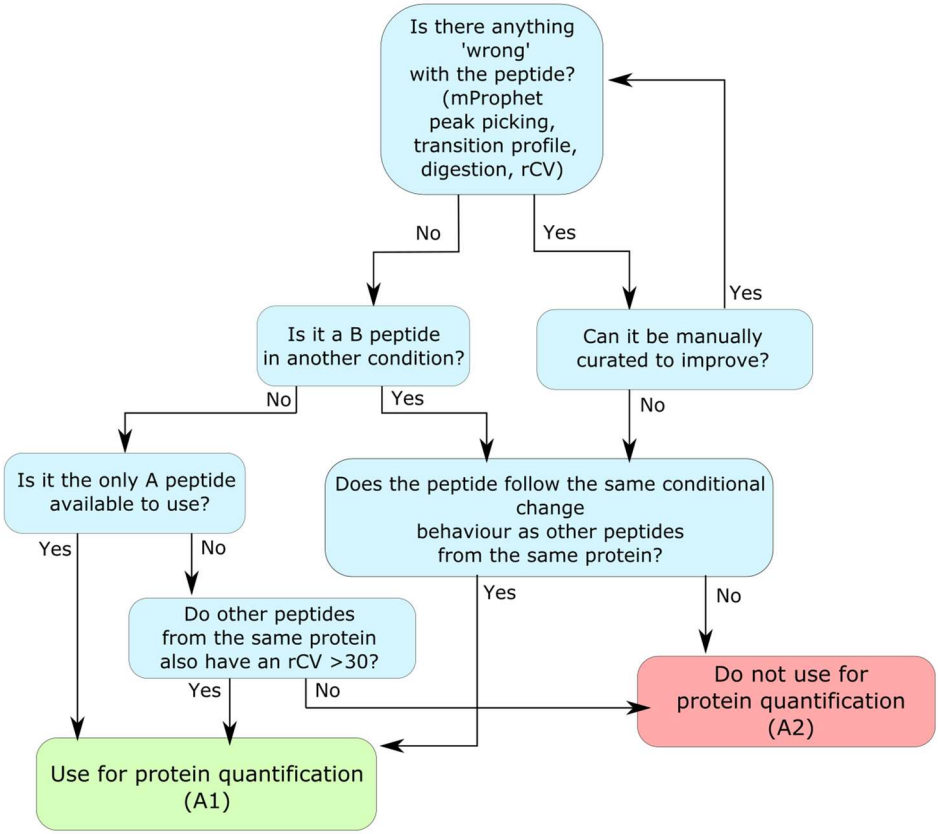



**Figure 4.6) Decision tree for classification of 'A' class peptides for suitability for quantification**

For each Q-peptide that was observed as both a light (yeast) and heavy (ChapCAT) variant, quantification may be performed. However, various situations decree that erroneous quantification will follow if the Q-peptide exhibits particular traits. Such peptides were sorted into 'A1' and 'A2' classifications, and protein quantification performed as a median of all biological replicates of all 'A1' Q-peptides targeting a particular protein.

To demonstrate the process of 'A2' peptide categorisation, I use the peptides targeting Hsp26 (Table 4.2). Firstly, the peptide DETLDDWFDNDLSLFPSGFGFPR has an rCV of 104 % under conditions of HS. If the high rCV is indicative of a problem with the biological replicate, I would observe a high rCV for all peptides targeted by this ChapCAT. The median rCV for the 11 'A'-type peptides targeted by ChapCAT 3 was 10.95 %, ruling out the biological sample as the source of variation.

**Table 4.2) Example of A1 and A2 Q-peptide selection for Hsp26**
Key: Chap – chaperone; Class – classification; rCV – robust coefficient of variance (%). Peptides discussed are highlighted in orange.

| Peptide | Chap | Class NG | Class HS | Median NG | Median HS | rCV NG | rCV HS |
|---|---|---|---|---|---|---|---|
| DETLDDWFDNDLSLFPSGFGFPR | Hsp26 | B1 | A2 | 695.8 | 287796.7 | 4.5 | 104.5 |
| LLGEGGLR | Hsp26 | A1 | A1 | 30470.2 | 219777.9 | 12.9 | 13.3 |
| SVAVPVDILDHDNNYELK | Hsp26 | A1 | A1 | 31646.8 | 260537.3 | 16.8 | 17.1 |
| VVVPGVK | Hsp26 | B2 | A2 | 695.8 | 233225.7 | 4.5 | 7.8 |

To clarify the root of the issue, I observed the spectral data for the peptide DETLDDWFDNDLSLFPSGFGFPR in Skyline (Figure 4.7). The biological replicate HS2 (Figure 4.7b) did not pass a 1 % mProphet FDR and had not been included in the abundance calculation. As this is a scheduled experiment, it may be that the retention time window of 3 minutes was not enough to capture the signal of the peptide in this biological replicate. Indeed, I observed a signal for each counterpart towards the end of the 3 minute window for other biological replicates. This retention time shift was not observed for any other peptides belonging to this ChapCAT in this biological replicate. Concerning biological replicates HS1, HS3 and HS4, it is evident that the light peptide signal differs between biological replicates. As the signal changes between each sample, I questioned the ability of this peptide to be digested to completion in the yeast protein, with time course experiments demonstrating incomplete digestion. This was also the case for the peptide VVVPGVK, and so both were classified as 'A2' in HS (results presented in Chapter 5). As neither of the two peptides from the NG conditions passed the 1 % mProphet FDR threshold, or were below the limit of detection, they were not used for absolute quantification in this condition.

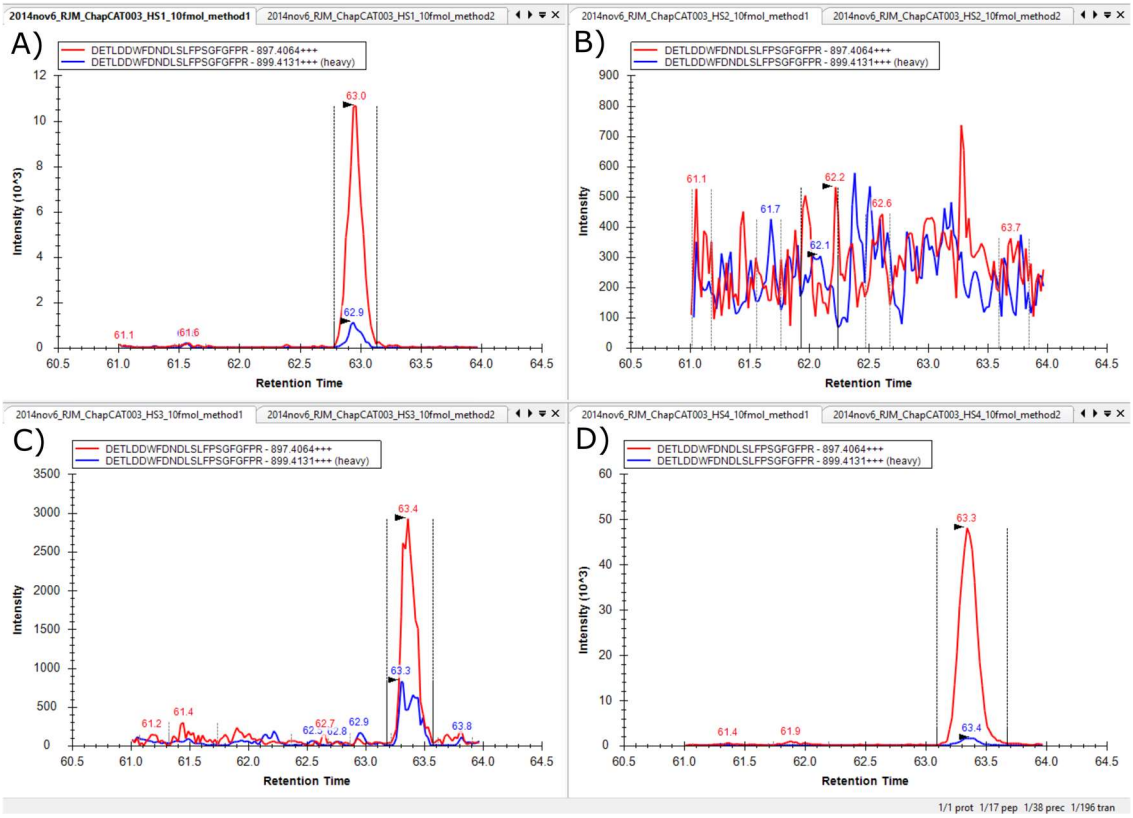



**Figure 4.7) Analyte and standard signals for the peptide DETLDDWFDNDLSLFPSGFGFPR**

A) HS1; B) HS2; C) HS3; D) HS4. The HS2 biological replicate failed to pass a 1 % mProphet FDR, with both light and heavy signals deviating between the remaining three biological replicates. Red: analyte spectra; Blue: standard spectra.

In some cases, there is a difference in the transition profiles of a peptide between its heavy and light counterpart. For example, the peptide FSDDECILIK (targeting Tcp1) has three product ions: y9, y8 and y7. Comparing the analyte and standard intensities of these ions, I observed y7 to be the most intense product ion in the analyte counterpart, whilst y9 is the most intense product ion in the standard counterpart (Figure 4.8). This was the case for the peptide under both NG (Figure 4.8a and b) and HS (Figure 4.8c and d) conditions, and could have resulted in an erroneous quantification. In instances where a single product ion shows such a shift in the relative intensity, I have removed it from both standard and analyte data prior to processing with mProphet. In this particular example, y7 was removed. Across the ChapCAT dataset, 21 transitions were removed from NG and HS spectra due to similar instances (listed in Appendix 8).

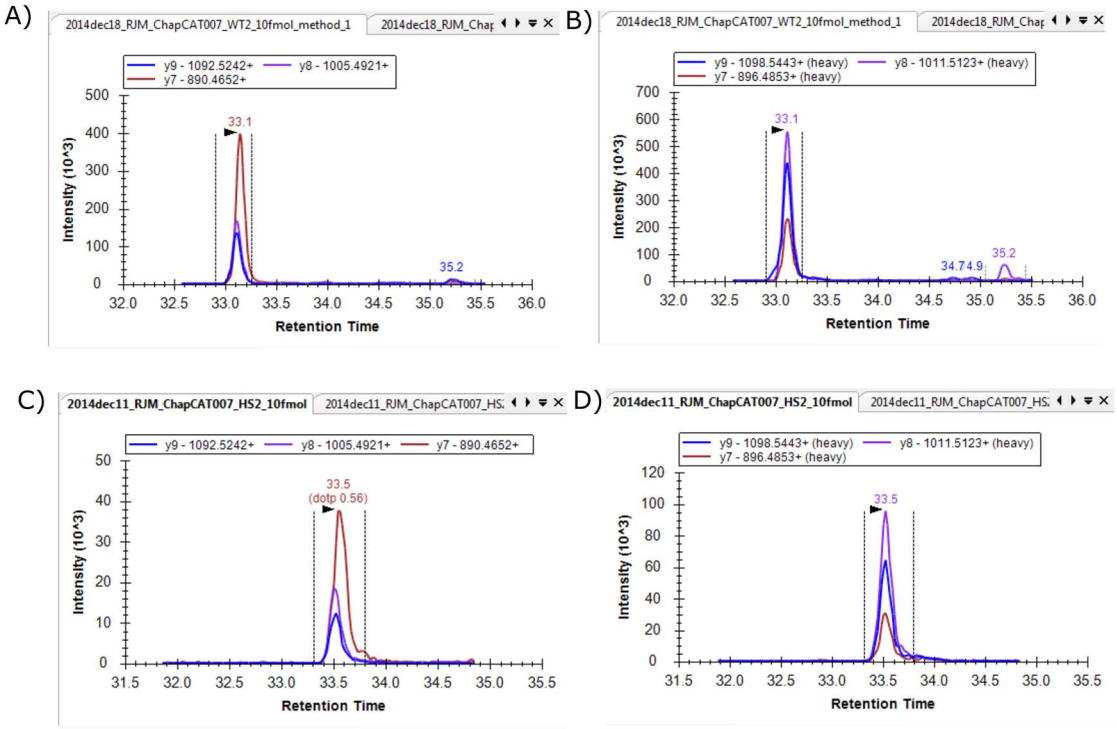



**Figure 4.8) Transition profiles for FSDDECILIK targeting Tcp1.**

I observed different relative intensities for product ion y7 between heavy and light counterparts regardless of sample condition. A) Analyte counterpart, NG conditions; B) Standard counterpart, NG conditions; C) Analyte counterpart, HS conditions; D) Standard counterpart, HS conditions.

I also experienced cases where only two peptides were potentially suitable for quantification of a particular chaperone, but the absolute abundances of each peptide differed by more than two fold in a single condition (Table 4.3). Under NG conditions, the median absolute abundances for the peptides ALAEFLFDDESNVIR and ITDTALVSAAVLSNR were 3083 and 11295 cpc respectively, a difference of 3.6 fold, despite both having modest peptide-level rCV values. Using a similar rule to that used by Brownridge and colleagues (Brownridge et al., 2013), in such situations I always elected the higher abundance peptide to be used towards absolute protein quantification, classifying the lower abundance peptide as 'A2'. Known issues with a peptide that may cause erroneous quantification, for example, the presence of post-translational modification on some (but not all) of the species, missed cleavage and loss of sample, it is rare for this to cause an increase the signal of the peptide, and so lower abundance peptides were classified as 'A2' in these such instances.

**Table 4.3) Example of A1 and A2 Q-peptide selection for Hsp78**
Key: Chap – chaperone; Class – classification; rCV – robust coefficient of variance (%). Peptides discussed are highlighted in orange

| Peptide | Chap | Class NG | Class HS | Median NG | Median HS | rCV NG | rCV HS |
|---|---|---|---|---|---|---|---|
| AHPDVSK | Hsp78 | C | C | NA | NA | NA | NA |
| ALAEFLFDDESNVIR | Hsp78 | A2 | A1 | 3083.1 | 8551.6 | 23.0 | 12.2 |
| ITDTALVSAAVLSNR | Hsp78 | A1 | A1 | 11294.9 | 13598.9 | 8.3 | 21.1 |
| NTIIVMTSNIGQDILLNDTK | Hsp78 | B1 | B1 | 695.8 | 2708.9 | 4.5 | 0.5 |
| TALIDGLAQR | Hsp78 | B2 | B2 | 695.8 | 677.2 | 4.5 | 0.5 |

Finally, in cases where there are more than 2 'A' peptides were available with varying abundances I used a heuristic approach to determine which to use for absolute quantification of the target protein. For instance, if a peptide is below the limit of quantification in NG conditions, it is either due to the low abundance of the parent protein, or the ability of the peptide to be detected in an LC-MS system. If a protein were too low in abundance, one would expect other peptides targeting the same protein to also have a low abundance. It then follows that if the protein was quantifiable under HS conditions, the abundance change for all peptides between NG and HS should be similar. This is demonstrated using peptides targeting Gim1 (Table 4.4). Under conditions of HS, the peptide LETQLQENK was below the limit of quantification ('B1' class), with the limit of quantification calculated as below 2784 cpc using 1 fmol of ChapCAT spiked in. This suggests that Gim1 was lower in abundance in response to HS. However, the remaining target peptides for Gim1 had an average fold change of 1.4 in response to HS, and a median peptide-level abundance of ~5000 cpc. This comparison suggested erroneous quantification would result from using LETQLQENK, and so it was classified 'A2' under NG conditions.

**Table 4.4) Selection of A1 peptides targeting Gim1**
Key: Class – Peptide classification; CPC – median copies per cell across peptide biological replicates; FC – fold change in response to heat shock. The example peptide is highlighted in orange.

| Gene | Peptide | Class. NG | Class. HS | CPC NG | rCV NG | CPC HS | rCV HS | FC (HS:NG) |
|---|---|---|---|---|---|---|---|---|
| Gim1 | IVNEEFDQLEEDTPVYK | A1 | A1 | 5041.00 | 15.18 | 7169.45 | 10.03 | 1.42 |
| Gim1 | LETQLQENK | A2 | B1 | 1763.76 | 16.40 | <2784 | NA | NA |
| Gim1 | LTGNVLLPVEQSEAR | A1 | A1 | 6315.56 | 10.28 | 7218.79 | 22.29 | 1.14 |
| Gim1 | YQQLQNELEEFIVAR | A1 | A1 | 8267.41 | 0.57 | 12774.33 | 22.95 | 1.55 |

Using the decision tree in Figure 4.6, the additional quality control steps described here and digestion efficiencies presented next in Chapter 5, of the 116 'A' peptides under NG

conditions, 28 were classified ‘A2’ (24.1 %) and not used towards protein-level absolute quantification. Under HS conditions, 14 ‘A’ peptides (14.3 %) of the 98 were classified ‘A2’ and not used towards absolute quantification of the target proteins. Protein-level absolute quantifications on a condition basis were calculated using the median peptide-level cpc value across all ‘A1’ peptides across all biological replicates that passed a 1 % mProphet FDR.

# Chapter 5 - Investigating the digestion efficiencies of standard and analyte Q-peptides in a QconCAT construct

As an additional quality control step in ensuring only suitable peptides are used for quantification, the digestion of both standard and analyte peptides from the ChapCAT and parent protein respectively must be complete. In this way, spectra observed are matched to the tryptic peptide only and represent the entire peptide sample, rather than experiencing a potential loss of peptide signal due to a missed cleavage. In this chapter, I present the result for a digestion time course experiment, enabling for the modelling of the rate of peptide release from the protein as either a standard or analyte counterpart.

Prior to MS analysis, protein samples are subject to digestion with (typically) the endopeptidase trypsin, although other proteases may be used depending on the aims of the experiment. Routinely, trypsin cleaves peptide bonds in polypeptides after lysine and arginine residues (Arg-X, Lys-X), unless followed by a proline (Arg-Pro, Lys-Pro), in a hydrolytic reaction resulting in the release of fragments termed peptides. If digestion of all susceptible peptide bonds goes to completion, and all tryptic sites are cleaved within a protein, the attendant peptides are termed 'limit' peptides (Figure 5.1), since no further endopeptidic cleavage is possible. Any potential tryptic sites that are not cleaved are termed 'missed cleavages'. Exopeptidases are able to catalyse the cleavage of the terminal peptide bond, however, as an endopeptidase, one of the most common missed cleavages by trypsin is in the form XXXXXKK or XXXXXRR, where 'X' denotes any amino acid as it is only able to perform cleavage of nonterminal peptide bonds. Most common proteomic workflows require this proteolytic step prior to analysis of the limit peptides via MS, with the majority of products from a tryptic digest having a minimum of two protonatable sites (N-α amino group and the C-terminal basic residue), so can generate $[M+2H]^{2+}$ ions. These ions enhance the generation of gas phase fragmentation products. Limit peptides are typically 10-15 amino acids in length, with those detectable typically in the mass range of 1000 to 3000 Da. This is aligned to the typical *m/z* range of mass analysers in mass spectrometers used in proteomic studies (Brownridge and Beynon, 2011).

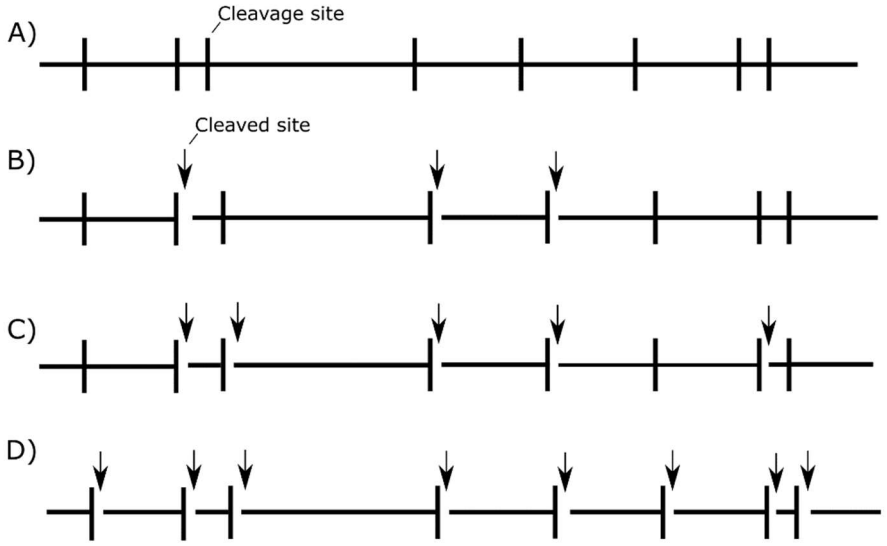



**Figure 5.1) The production of limit peptides during the digestion process**

A) Prior to addition with a protease, a protein will be fully intact with cleavage sites unbroken. B) After addition of protease, high propensity cleavage sites are broken, resulting in the product of limit peptides and peptides containing missed cleavages. C) Further digestion takes place, and more sites are cleaved producing more limit peptides. D) Finally the protein is fully digested, and all limit peptides are produced - all potential cleavage sites have been cleaved by the protease.

The proteome offers a complex space for proteolysis, with residues up to three positions away from the tryptic cleavage site affecting the affinity of the protease for the protein molecule (Lawless and Hubbard, 2012, Siepen et al., 2007). For trypsin, whilst many of the cleavage sites are efficiently and completely cleaved, those with acidic residues close to the cleavage bond could form salt bridges with the basic arginine or lysine residues, preventing recognition and cleavage by trypsin (Siepen et al., 2007). For qualitative proteomic applications, this does not present too much of a problem, although it can contribute to false identifications and can even be exploited as a means to improve identification quality (Stead et al., 2006). Most protein identification strategies used in proteomics are therefore tolerant to a small number (usually one or two) of missed cleavages. This gain is more crucial in peptide mass fingerprinting during which only the mass is obtained from the peptide (Brownridge and Beynon, 2011). Moreover, since missed cleaved peptides are common (30 – 40% of peptides detected in an experiment is typical) and they are more likely to be unique to a given proteoform (therefore helping a confident identification), they are not ignored.

For quantification strategies, however, the absence of complete digestion can present more of a problem. For absolute quantification, peptides must ideally be stoichiometric with the parent protein to enable accurate quantification of the protein. This is less of an issue in most relative quantification strategies since equivalent peptide signal is compared between

runs. In this case, if the digestions protocols and outcomes are the same, relative changes can be calculated. However, for absolute strategies, where the surrogate peptides generated from the protein lead to signal split into multiple overlapping peptides due to missed cleavage, the signal will be attenuated and quantification underestimated (Lawless and Hubbard, 2012). For absolute quantification, known amounts of isotopically labelled peptide (either an AQUA peptide or derived from a QconCAT construct) is compared to an identical analyte peptide. Therefore, it must be assumed that the labelled and unlabelled versions of the peptide are proteolysed to the same extent to allow accurate quantification. In the QconCAT strategy, the primary sequence context will differ between the standard and analyte peptide, as the standard peptide is neighbour to other tryptic standard peptides targeting other proteins in the same QconCAT. Therefore, proteolysis between the ChapCAT and analyte protein could differ, with excision of the peptide occurring at different rates.

I attempted to minimise missed cleavage issues by use of prediction software for the propensity of each peptide to mis-cleave in the standard and analyte protein (MC:pred) (Lawless and Hubbard, 2012). Various other feature limits (discussed in Chapter 3) were also utilised to avoid the chance of missed cleavage in either the analyte or ChapCAT proteins. For instance, the presence of a glutamic acid or aspartic acid residue in P1' and P2' (using the nomenclature P5-P4-P3-P2-P1-P1'-P2'-P3'-P4 (Schechter and Berger, 1967), with P1 as the lysine/arginine cleavage residue) is pronounced in peptides with known missed cleavages, demonstrating the difficulty in cleaving dibasic sites, and in areas where there is potential to form a salt bridge between acidic side chains in P1' or P2' and the basic site at P1 (always arginine or lysine). Hubbard and Lawless (Lawless and Hubbard, 2012) also observed these dibasic sites present in peptides found to be miscleaved, and highlighted the inability of trypsin to cleave peptides if an arginine or lysine exists at the *N*-terminus of the peptide.

Whilst different rates in the digestion process will be observed for analyte and standard peptides, this conflicts with the principal that digestion is complete for both peptides. Consequently, since this could lead to inaccuracies in my ChapCAT quantifications, I examined this in more detail to assess an impact. To determine complete digestion of both the analyte and ChapCAT, I performed a digestion time course experiment for all ChapCATs, taking 10 µL samples at selected time intervals during the digestion process. Each sample was subject to SRM analysis using the same transitions used for absolute quantification, and enabled for determination of the completion of proteolysis for both the ChapCAT and the analyte protein.

## 5.1 Modelling digestion rate kinetics

Typically, a second order reaction has two reactant species, here the protein (ChapCAT/Analyte) and the endoprotease trypsin. The protein to be digested is present to begin with at high concentration, whilst trypsin concentration does not reduce significantly throughout the digestion time course. Considering the ChapCAT/Analyte as 'S' (Substrate) and trypsin as 'B', the rate of change of ChapCAT/Analyte concentration ([S]) over time due to proteolysis can be observed using Equation 5.1, in which *k* is a reaction constant.

$$\frac{-d[S]}{dt} = k[S][B]$$

*(Equation 5.1)*

As trypsin concentration is not reduced over time, $[B]_0 \approx [B]$, I could assume pseudo-first order rate kinetics, so when integrated Equation 5.1 becomes Equation 5.2:

$$[S] = [S]_0 e^{-kt}$$

*(Equation 5.2)*

Using the relationship between substrate ('S') protein and peptide ('P') product (Figure 5.2), I could determine the amount of peptide at time 't' using Equation 5.3.

$$[P]_t = [S]_0 - [S]_t$$

*(Equation 5.3)*

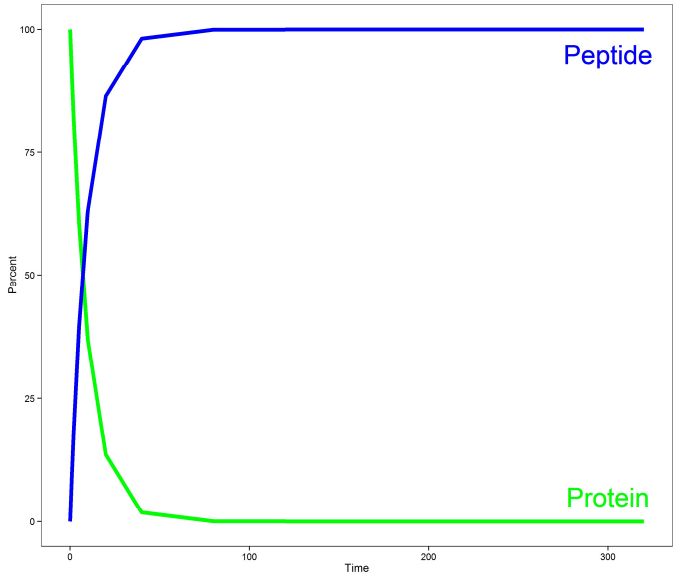



**Figure 5.2) Relationship between protein and peptide during the digestion process**

The relationship between peptide and protein is inversely linear, as protein is proteolysed by trypsin, equivalent amounts of tryptic peptide is produced. Therefore, the equation $Peptide_t = Protein_0 - Protein_t$ can be used to deduce the amount of peptide present.

To convert these amounts into a percentage, I determined the ratio between the intensity of peptide at time 't' and the maximum peptide intensity achieved across all time points (Equation 5.4).

$$\%\ P_t = \frac{Peptide\ peak\ area\ at\ t}{Maximum\ peptide\ peak\ area} \times 100$$

*(Equation 5.4)*

Rearranging equations 5.2 and 5.3 to demonstrate $P_t$ as a percentage, I was able to use Equation 5.5 to model the digestion process for each Q-peptide analysed.

$$P_t = 100 - ((100 - P_0) \times e^{-kt})$$

*(Equation 5.5)*

Using non-linear modelling, values of *k* for each peptide for both analyte and standard counterpart could be determined. I assumed digestion is complete within five half-lives, equalling ~97 % of peptide release from the protein, with the 3 % accounting for any error in the model. The half-life was calculated using Equation 5.6.

$$t_{1/2} = \frac{\ln(2)}{k}$$

*(Equation 5.6)*

I determined the degree to which a peptide had completed digestion for both the standard and analyte counterpart in the time allocated (1230 minutes) (Figure 5.3). Typically, standard-analyte peptide digestion exhibits various scenarios: fast release in both standard and analyte peptide, with digestion going to completion within the time period allocated (Figure 5.3a). Fast standard release, but slow analyte release with both going to completion (Figure 5.3b). Fast analyte release, and slow standard release but both going to completion (Figure 5.3c). Slow analyte and slow standard release with both going to completion (Figure 5.3d). Slow analyte and standard release, without going to completion (Figure 5.3e) and slow standard release, even slower analyte release without going to completion (Figure 5.3f). Typically, peptides with *k* values of less than $2.18 \times 10^{-13}$ $min^{-1}$ were not proteolysed to completion.

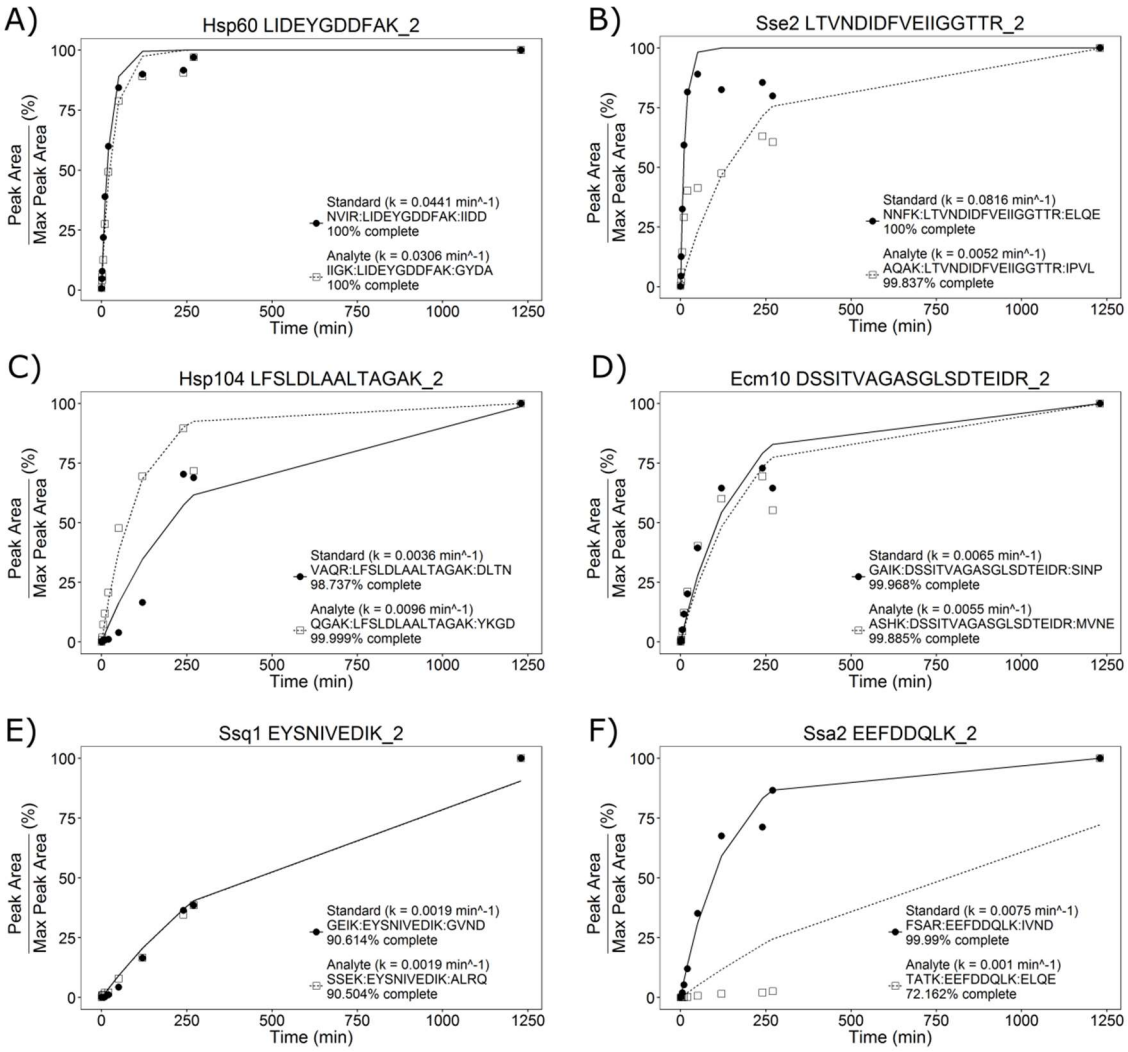



**Figure 5.3) Digestion Time Course Analysis**

Using pseudo-first order rate kinetics, a limit peptide's digestion progress can be estimated. With digestion completion occurring at five half-lives I could determine which peptides were likely to lead to incorrect quantification. I observed various digestion behaviours: A) Fast release in both standard and analyte peptide, with digestion going to completion within the time period allocated. B) Fast standard release, but slow analyte release. Both go to completion. C) Fast analyte release, slow standard release. Both go to completion. D) Slow analyte and slow standard release. Both go to completion. E) Slow analyte and standard release, without going to completion. However, progress is similar and so likely to produce accurate quantifications. F) Slow standard release, even slower analyte release without going to completion. In this instance, release of the analyte peptide was so slow, it was unable to be modelled using pseudo-first order rate kinetics and so the dotted line, indicating the predicted release of analyte peptide according to pseudo-first order rate kinetics, did not match the data acquired. For visual aid, I have included the context of the cleavage sites for each peptide.

## 5.2 Determining complete proteolysis of standard/analyte peptides

For every peptide that was not digested to completion according to this modelling strategy in either standard of analyte counterpart, I determined that use of the peptide towards absolute quantification would result in unreliable peptide (and therefore protein) cpc values. When comparing the digestion times to completion for each peptide counterpart, I found the majority (91 %) of peptides were considered to be fully digested in both standard and analyte within 1230 minutes (the total time allocated for digestion). An additional five peptides (3 %) were not fully expected to be complete in the time allocated but shared similar digestion rates and could be used to provide reliable quantification values. Despite this only one peptide in this latter set had an 'A1' classification. Ten peptides (6 % in cream boxes) were not deemed digested to completion in either standard or analyte and so were removed from use towards quantification (Figure 5.4). Peptide *k* values and digestion times are available in Appendix 9.

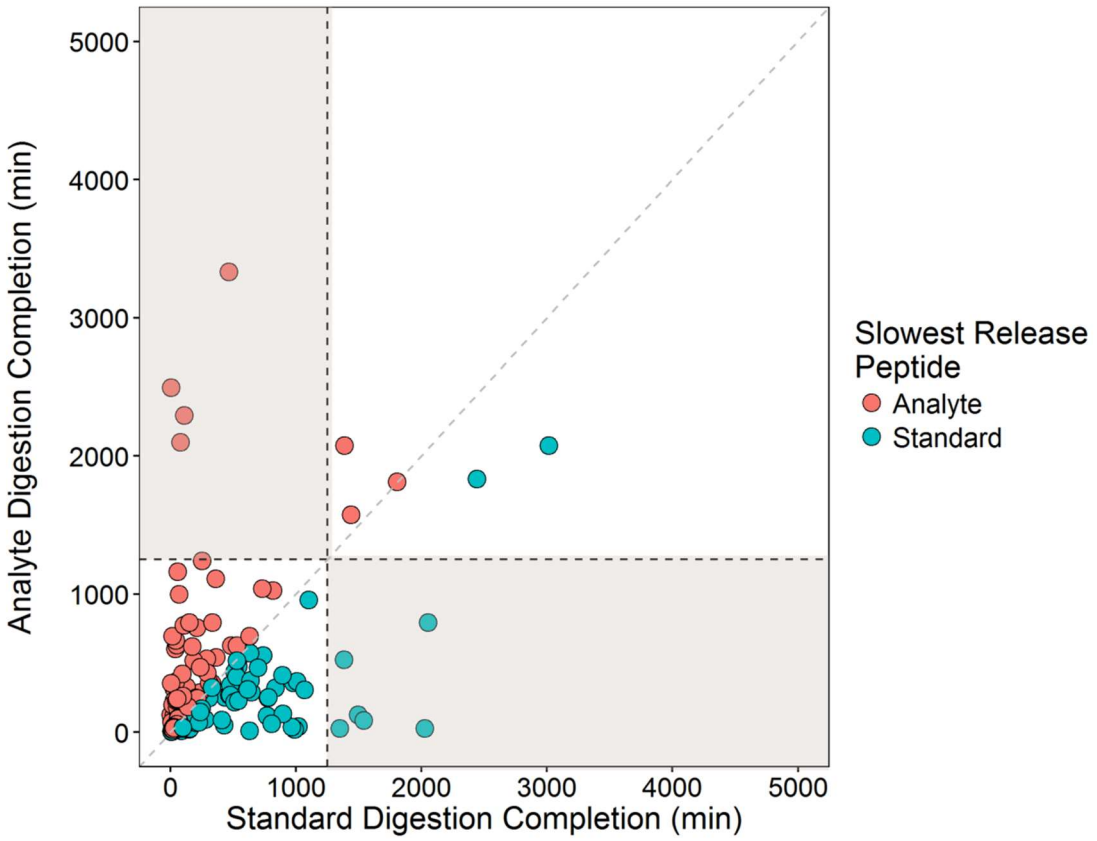



**Figure 5.4) Comparsion of standard and analyte proteolysis**

Ten peptides did not proteolyse at similar rates when comparing the standard and analyte counterparts (upper-left and bottom-right dashed regions). To prevent errors in absolute quantification, any 'A'-type Q-peptides in this set were classified 'A2' and not used for absolute quantification under either NG or HS conditions.

To investigate the ability of the MC:pred software to predict incomplete proteolysis, I correlated the maximum MC:pred score (between the *N*-terminal and *C*-terminal scores) for standard and analyte counterparts with their respective *k* values. The hypothesis is that an expected negative linear relationship would be observed; as *k* is higher implying faster cleavage, MC:pred scoring should be lower, implying less propensity for missed cleavage. Typically, we expect MC:pred scores to be higher for the analyte peptides as a low MC:pred score is a feature I select Q-peptides upon during the design phase when ordering them in the construct. For standard Q-peptides, correlating the maximum MC:pred score and associated *k* constant resulted in a Spearman's Rank Correlation coefficient of -0.28; whilst for analyte peptides, the Spearman's Rank Correlation coefficient was -0.15. Therefore, there appears to be no clear relationship between the predicted (maximum) MC:pred value and the modelled *k* constant for a peptide. As MC:pred scoring is a prediction based solely on the sequence of Q-peptides focussing on residues flanking the cleavage site, additional factors are in play that improve/hinder the digestion efficiencies of peptides. This includes the use of various buffers and reagents, including RapiGest$^{TM}$, that aim to improve the complete digestion times of a protein as well as an optimised digestion protocol. Together, this data demonstrates that although a way to ensure equivalent digestion completion in the QconCAT and analyte peptides, linker sequences are not necessarily required. A combination of prediction technologies (although should not be uniquely relied upon), stringent quality control in classification of 'A'-type Q-peptides, an optimised digestion protocol and a generous digestion time limit is adequate to ensure only peptides that achieve complete proteolysis are used for absolute quantification.

I do not consider missed cleavage or digestion completion to be an issue in my experiments. I have carefully examined the data gathered from digestion time course experiments of every targeted peptide in both analyte and standard, and believe the vast majority of my reactions went to completion, with 91 % of peptides going to completion. Furthermore, additional problematic peptides were removed from quantification experiments as described in Chapter 4. Together, these quality control steps ensured reliable and accurate quantification of chaperone proteins was performed under conditions of NG and HS. Performing absolute quantification allowed for me to begin to characterise the chaperone response to HS, and is discussed further in the next chapter.

# Chapter 6 - SRM-normalised label free quantification of the *S. cerevisiae* proteome under conditions of heat shock

*This work was originally published in the journal 'Proteomics': Mackenzie, R. J., Lawless, C., Holman, S. W., Lanthaler, K., Beynon, R. J., Grant, C. M., Hubbard, S. J. & Eyers, C. E. 2016. Absolute protein quantification of the yeast chaperome under conditions of heat shock. Proteomics DOI: 10.1002/pmic.201500503. An extended version is presented in this chapter. Aside from additional discussions, this version also includes an assessment of the differences in Q-peptide classification between NG and HS conditions and a discussion of the response of Hsps/CLIP-type cystolic chaperones to HS.*

Under conditions of HS, the elevated temperature can shift the conformational equilibrium of proteins to more aggregation-prone states in which the exposed hydrophobic regions of the unfolded protein are able to interact, leading to protein aggregation and preventing functionality (Vabulas et al., 2010). The *S. cerevisiae* cell senses the increase in temperature through both an increase in the fluidity of the cell membranes (Anckar and Sistonen, 2011, Bromberg et al., 2013) and liberation of Hsf1 by Hsps due to accumulated misfolded protein in the cytosol. Hsf1, the primary modulator of the HSR, is able to bind to HSEs located in the promoter region of target genes, either repressing or activating the expression of the gene. Together with the activation of the Msn2/4 transcription factor, *S. cerevisiae* cells are able to elicit the HSR. This includes induction of Hsps in order to manage protein aggregation (via Hsf1); accumulation of the storage carbohydrates trehalose and glycogen (via Msn2/4); transient cell cycle arrest at G1 through inhibition of the cyclins Cln1 and Cln2 (via Hsf1) and thermotolerance against future stress through activation of the Pkc1-MAP kinase pathway (via Msn2/4) (Mensonides et al., 2005, Trotter et al., 2001). During HSR, the primary role of chaperones (HSPs) is considered the protection of hydrophobic surfaces of misfolded and aggregated proteins. Misfolded proteins may be unfolded and later folded when favourable conditions return, whilst terminally misfolded proteins may be directed to the ubiquitin-proteasome pathway for degradation through the action of the E3 ubiquitin ligase CHIP. Without the upregulation of chaperones, cellular protection during and recovery after HS is not possible, making chaperones fundamental cellular effectors of the HSR.

Chaperone upregulation in response to various stress conditions has been characterised in previous proteomic and transcriptomic studies. Transcriptomic studies typically observe the level of mRNA encoding for chaperones, however it is widely reported that mRNA levels do not

correlate well with protein levels and so should be used with caution when inferring protein-level changes (Vogel and Marcotte, 2012). Proteomic studies provide a more accurate depiction of protein-level changes in response to various conditions; however, they have generally used relative quantification rather than defining changes in absolute protein levels. Although I had previously quantified absolute protein abundance (copies per cell) for over 50 chaperones, the studies were only performed under normal conditions in batch (Chapter 3) and chemostat cultures (Brownridge et al., 2013, Lawless et al., 2016). Regardless, using known substrate interactions (Gong et al., 2009) and a simple model, it is estimated that ~62 % of total protein folding flux in the chemostat-grown cell is chaperone-mediated (Brownridge et al., 2013).

Given that prior proteomic studies of the *S. cerevisiae* HSR have been either incomplete or 'relative' in nature, I extended my previous QconCAT SRM-based absolute quantification study (Chapter 3) to chaperones under HS (42 °C, 30 minutes) conditions.

## 6.1 Comparison of QconCAT standards between NG and HS conditions

In order for a peptide to be quantified and cpc values defined, it must be observable in both the heavy-labelled standard (ChapCAT) and light (unlabelled) native yeast sample. I reused the previous classification system (Chapter 3 and 4) to determine quantification suitability of each Q-peptide. Of the 221 Q-peptides in the 8 ChapCATs that were purified, 84 Q-peptides were 'A1' class (38 %); 14 were 'A2' class (6.3 %); 31 were 'B1' class (14 %); 17 were 'B2' class (7.7 %) and 75 were 'C' class for the HS treated yeast (Figure 6.1) (full peptide classification table provided in Appendix 10). As discussed in Chapter 3, absolute protein abundances were determined by the median cpc across all biological replicate values for all 'A1' class Q-peptides targeting a particular chaperone. For particular HSP70 chaperone groups (Ssa1_Ssa2 and Ssb1_Ssb2) degenerate peptides were not used in the final quantification as unique peptides to each constituent chaperone were available. This was not the case for Hsp32_Sno4_Hsp33.

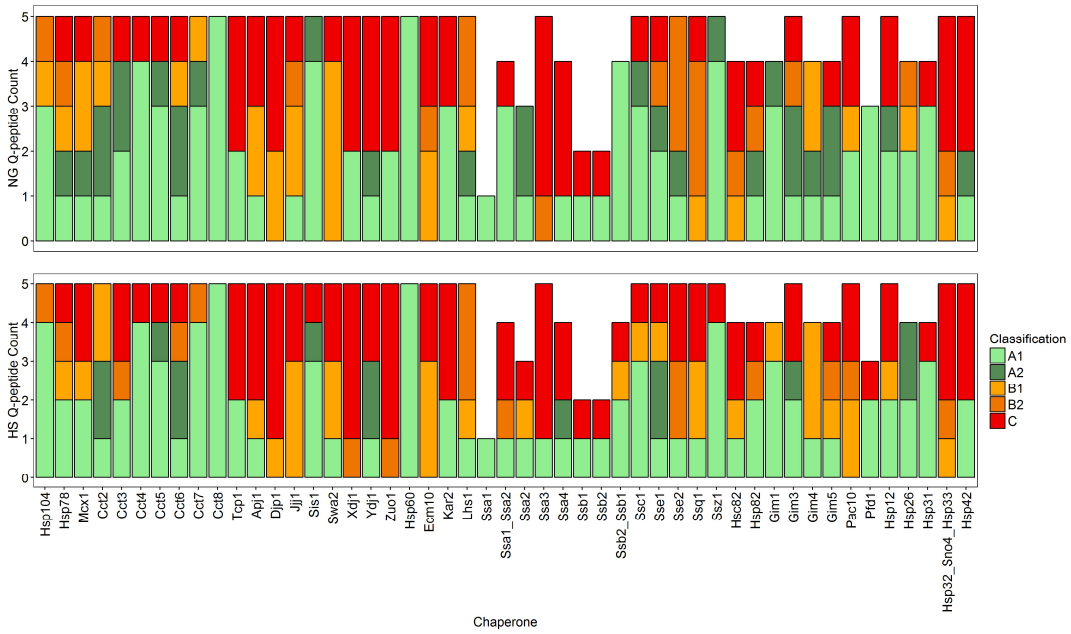



**Figure 6.1) Classification of Q-peptides on a per protein basis under NG and HS conditions**

For particular HSP70 chaperone groups (Ssa1_Ssa2 and Ssb2_Ssb1), degenerate peptides were not used in the final quantification as unique peptides to each constituent chaperone were available. This was not the case for Hsp32_Sno4_Hsp33.

I compared this HS dataset to the NG dataset (as presented in Chapter 3) as both datasets were constructed using identical ChapCAT standards and quality control techniques (Chapters 4 and 5) to determine absolute abundances for 49 chaperones. At the peptide level, 88 Q-peptides were classified as 'A1' under NG conditions, whilst 84 were classified as 'A1' under HS conditions. Sixty-nine peptides were classified as 'A1' under both conditions, suggesting these peptides were excellent Q-peptide standards: both the standard and light peptide were observed in the LC-MS experiment with high confidence in both conditions. However, alternative stress conditions could change the level of analyte and could bring the classification to a 'B' in future experiments. A comparison of the peptide classifications in each condition is made in Figure 6.2, with Appendix 11 depicting peptides that do not share the same classification between conditions.

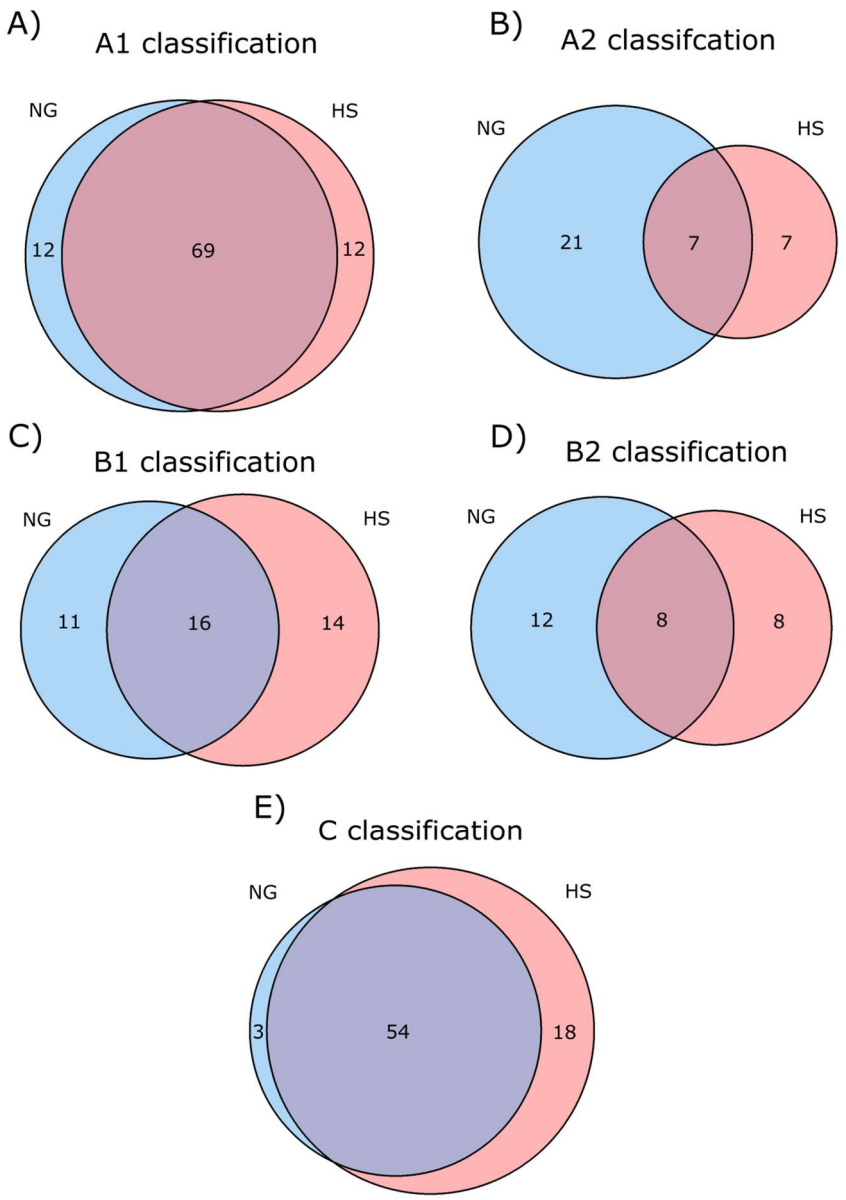



**Figure 6.2) Comparison of Q-peptide classifications between NG and HS conditions**

A) 69 Q-peptides were observed as both standard and analyte with high confidence, classified 'A1' in both NG and HS conditions; B) 7 Q-peptides had sub-optimal features with low confidence in a correct detection, were classified 'A2', in both NG and HS conditions; C) 16 peptides analyte counterparts were below the limit of detection in an LC-ESI-MS experiment, classified 'B1', in both NG and HS conditions; D) 8 Q-peptides did not pass a 1% FDR threshold when processing with mProphet, classified 'B2', in both NG and HS conditions; E) 54 Q-peptides were not observable as either standard or analyte, classified 'C,' in both NG and HS conditions.

For each classification type, there was a disparity in the number of peptides observed between NG and HS conditions. For 'A1' class Q-peptides, the disparity was a direct result of the abundance level. For example, Zuo1 had an absolute abundance of 21,700 cpc in NG, however fell below the limit of detection under conditions of HS and so was classified 'B'. This was the same for the reverse situation as observed with Hsc82 with a high abundance in HS (35,900 cpc), but was below the limit of detection in NG (for protein quantifications, see Appendix 12).

'B1' peptides were expected to differ between conditions. These peptides were below the limit of detection concerning their analyte counterparts. As discussed above, protein and consequently peptide levels would undoubtedly change in response to each condition and thus the proteins that lie below the limit of detection (~700 cpc) would differ. I found that for the chaperones Hsp32_Sno4_Hsp33, a single non-unique Q-peptide was below the limit of detection, whilst three non-unique peptides are classified 'C' in both conditions with a final non-unique peptide unable to pass a 1 % FDR after mProphet processing. This suggested these chaperones are consistently of low abundance in *S. cerevisiae*. This was also observed to be the case in previous attempts (Brownridge et al., 2013, Lawless et al., 2016).

The difference in abundance of the proteomic background as a result of unique experimental conditions coupled to the modest reproducibility of peptide chromatography is likely responsible for the disparity between 'A2', 'B2' and 'C' type Q-peptides. 'C' peptides were classed so due to their inability to be observed in an SRM experiment. Due to competing peptides at the chromatographic column, the retention time of the peptide will differ between conditions. To ensure I was able to capture every Q-peptide to the best of my ability, an unscheduled experiment was run prior to every quantification experiment in both conditions, to ensure the correct three-minute window was assigned. I observed retention times under HS deviating from those observed under NG conditions, and in at least 20 cases, identified peptide light and heavy counterparts in one condition without being able to observe the equivalent in the second condition. By running quality control experiments before and after the runs, I was able to eliminate signal detection and chromatographic issues. Of the six peptides classified as 'A1' under NG conditions, but classified 'C' under HS conditions, only one (IGLQIVQFINEPSAALLAHAEQFPFEK) contributed to a significant fold change observed in Ssz1. To observe its effect on the absolute quantification of Ssz1, removal of this peptide from Ssz1 data resulted in a slightly higher NG cpc value of 16,300 cpc, a fold change of 1.58 and an adjusted p value of 0.05. I observed no reason for full removal of this peptide as it behaved well

under NG conditions with its lack of signal in HS likely due to chromatography or reproducibility issues as discussed above rather than a direct result of peptide suitability for an LC-MS experiment, so was retained for use in my final cpc calculations.

Despite these disparities in the classification numbers, chaperones were quantified under HS conditions using the same technique as described in Chapter 3. The median copy per cell value was determined on a per chaperone basis across all 'A1' peptides targeting the respective chaperone across all biological replicates. Furthermore, I determined fold change as the ratio of the median HS cpc to the median NG cpc. I used these median values to determine a fold change error, as depicted in Figure 6.3. Full results are supplied in Appendix 12.

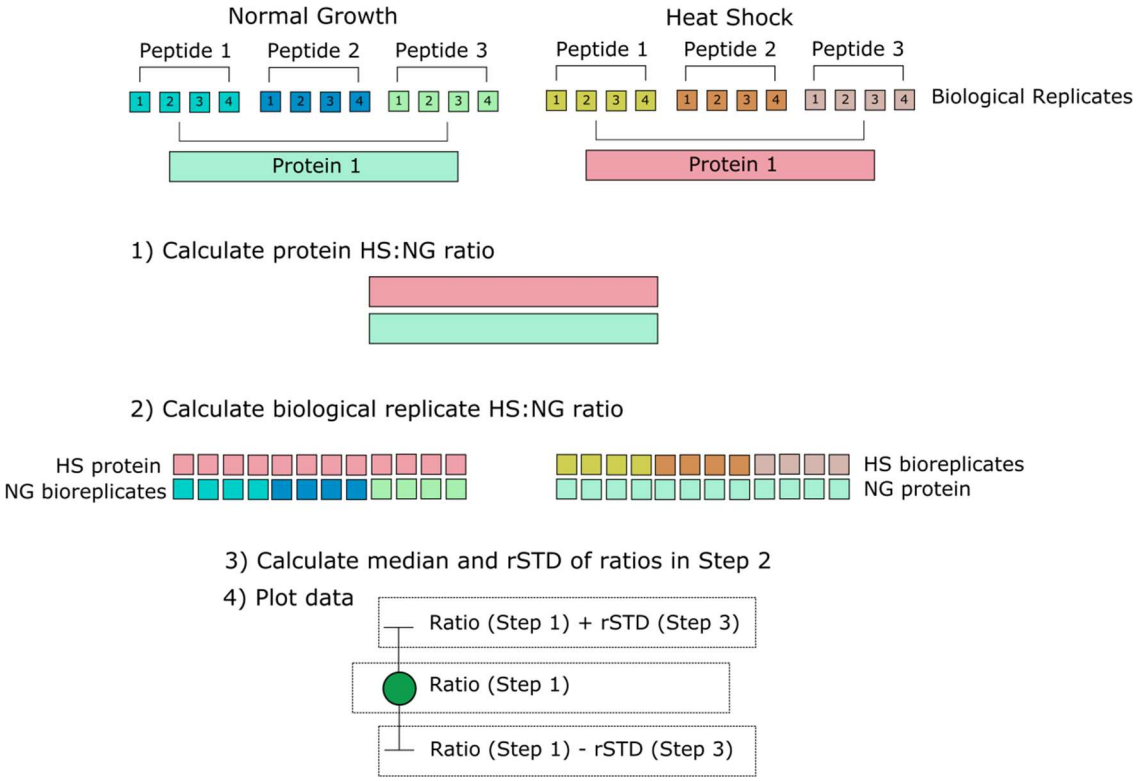



**Figure 6.3) Calculation of error for fold changes in response to HS**

Errors were calculated in 4 steps; firstly, the HS:NG ratio between median cpc values for a protein of interest were determined to give the protein fold change. Secondly, HS:NG ratios were calculated with respect to either the median HS cpc or the median NG cpc to produce an array of ratios depicting biological replicate fold changes. The median and robust standard deviation across these ratios gave an indicator of error. Finally, the upper and lower limits for the absolute cpc fold change in response to heat shock (as calculated in step 1) were determined using the robust standard deviation of the ratios.

## 6.2 Absolute quantification of chaperones in response to HS

A major benefit I anticipated from selecting more than two Q-peptides where possible during the design phase was the improvement of rCV and the increased likelihood of being able to detect protein. I determined the rCV across biological replicates for all 'A1' peptides on a per protein basis. It was predicted that increasing the number of peptides used for quantification could increase the confidence in the protein cpc value (Brownridge et al., 2013). However, I found that higher numbers of 'A1' Q-peptides used towards quantification generally led to a small increase in the median rCV in both conditions, and there was a lack of a clear material gain in precision by attempting to increase the number of Q-peptides from two (as is typically used in QconCAT design strategy (Brownridge et al., 2013, Lawless et al., 2016)) to up to five (Figure 6.4).

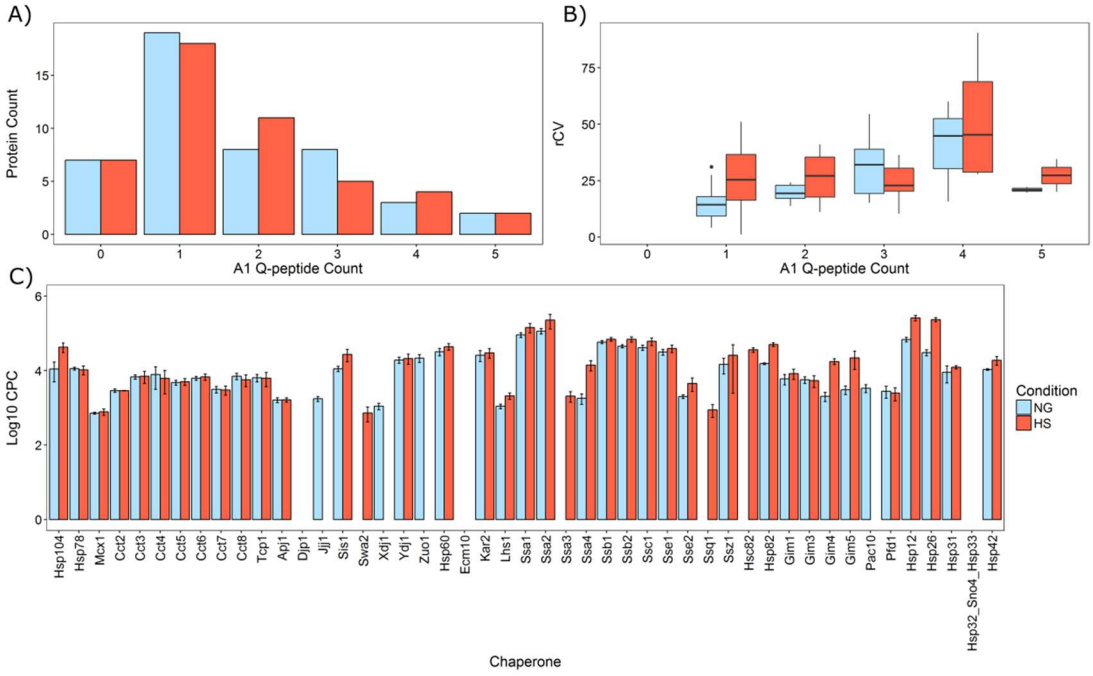



**Figure 6.4) Absolute quantification of chaperones in NG and HS conditions.**

A) The majority of chaperones quantifications were performed using 1 or 2 'A1' class Q-peptides. B) I observed a small increase in the median rCV for chaperone quantification as the number of 'A1' Q-peptides used for quantification increased. C) I observed upregulation for the majority of chaperones in response to HS.

Even with stringent quality control steps and design consideration for the selection of up to five Q-peptides per chaperone (described in Chapter 3, 4 and 5), the majority of chaperones were quantified by two or fewer 'A1'-type Q-peptides, illustrating one of the inherent challenges associated with peptide-based targeted proteomics. There are often few suitable quantotypic peptides for use in an absolute quantification experiment. Quantifications

performed under conditions of HS had an increased median rCV, 28.0 compared to 18.9 under NG conditions. To investigate, I examined the distribution of cpc values across biological replicates for Q-peptides that were classed as 'A1' in both NG and HS (Figure 6.5); no clear systematic trend was observed. However, as discussed in Chapter 3, I observed significant difference in the cpc values determined in ChapCAT 6 HS biological replicates 1 and 3 to those determined in replicates 2 and 4, causing a higher rCV for these Q-peptides in HS conditions. ChapCAT 6 targeted chaperones of the CCT complex, which despite these larger rCVs were not significantly upregulated in response to heat shock. Despite this, my data demonstrated good agreement between peptides common to a parent protein, with abundance shifts matched between conditions.

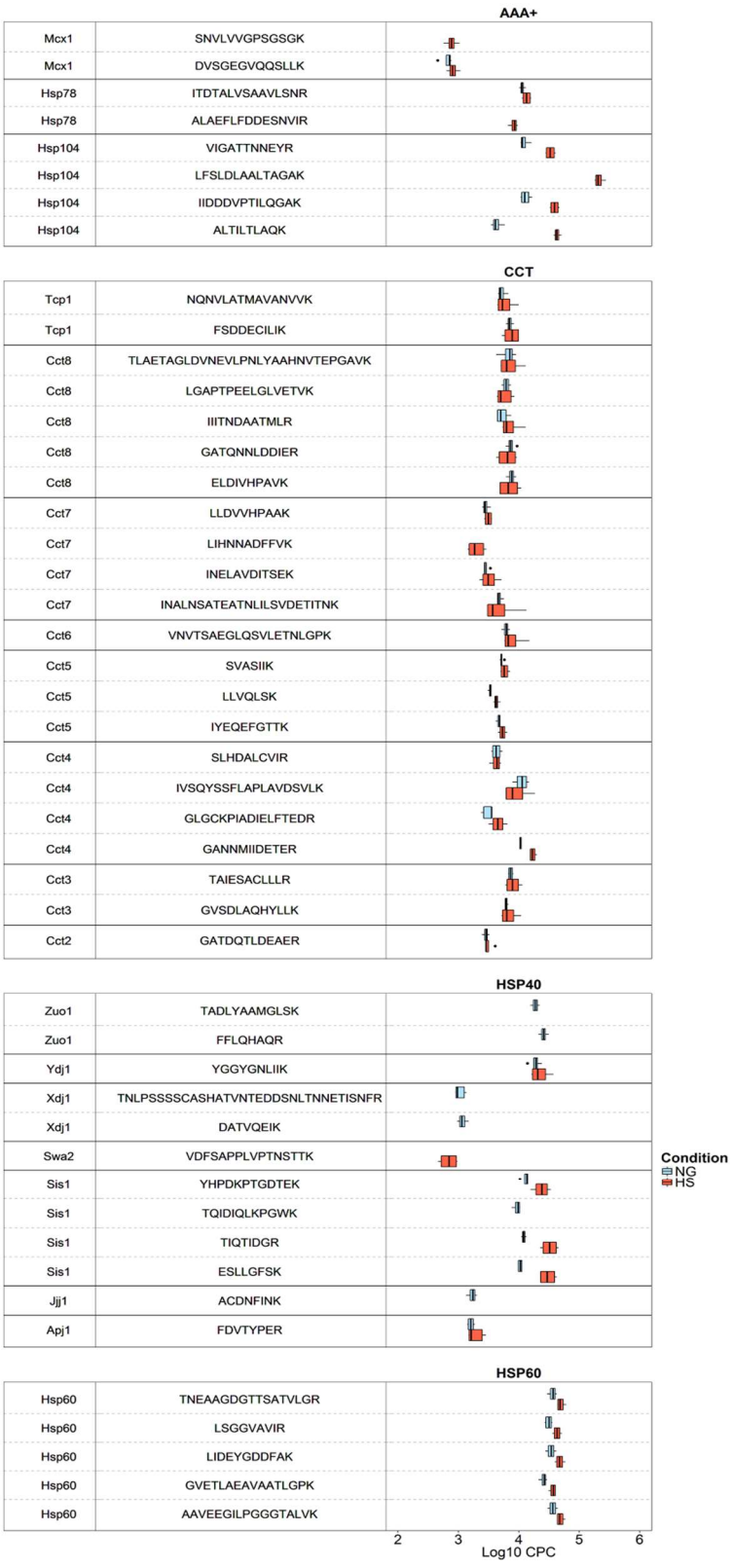
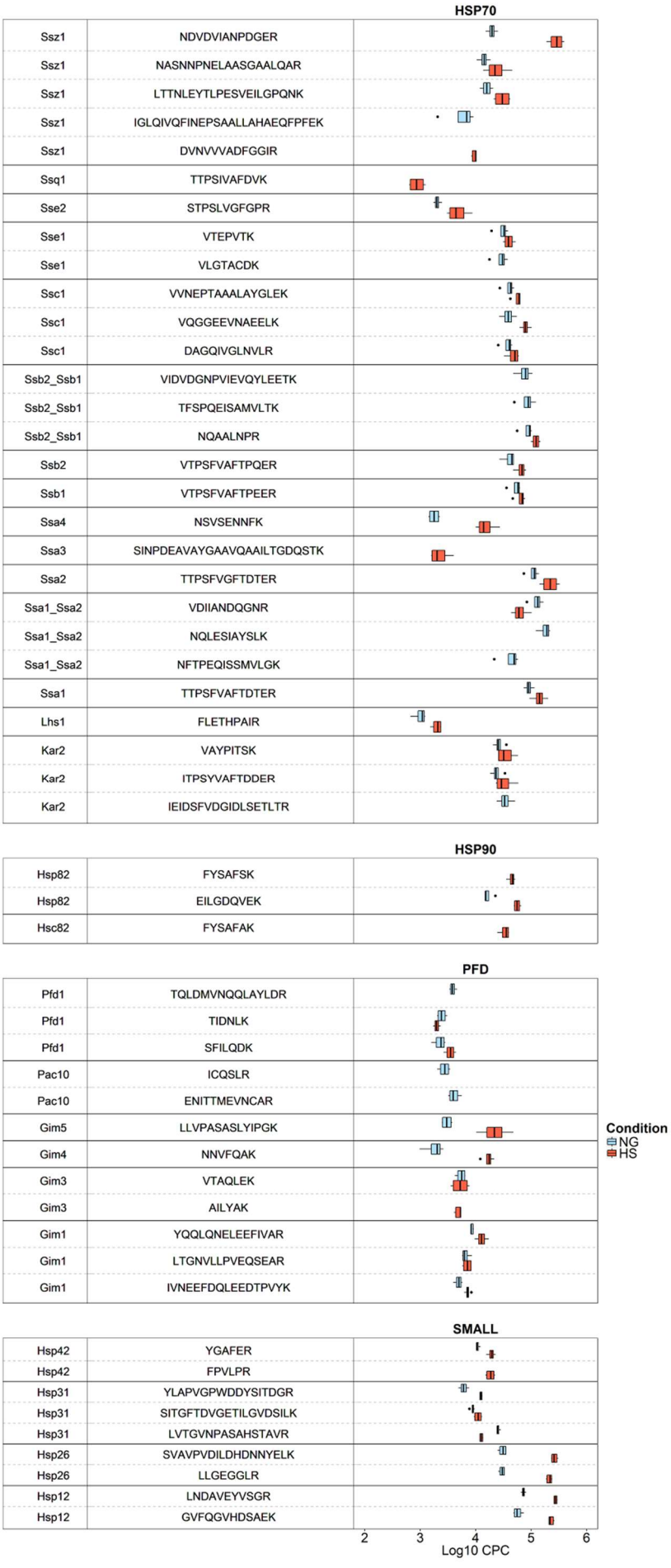



**Figure 6.5) Observation of the spread of data, reflecting rCV, for each 'A1' Q-peptide under conditions of NG and HS – continued overleaf**

To investigate the cpc values on a per peptide basis the spread of biological replicate data points unique to each condition is observed. There is no clear systematic trend in the spread of data for a condition, with common peptides to a parent protein sharing good agreement in cpc value.


HSP70
Ssz1 NDVDVIANPDGER
Ssz1 NASNNPNELAASGAALQAR
Ssz1 LTTNLEYTLPESVEILGPQNK
Ssz1 IGLQIVQFINEPSAALLAHAEQFPFEK
Ssz1 DVNVVVADFGGIR
Ssq1 TTPSIVAFDVK
Sse2 STPSLVGFGPR
Sse1 VTEPVTK
Sse1 VLGTACDK
Ssc1 VVNEPTAAALAYGLEK
Ssc1 VQGGEEVNAEELK
Ssc1 DAGQIVGLNVLR
Ssb2_Ssb1 VIDVDGNPVIEVQYLEETK
Ssb2_Ssb1 TFSPQEISAMVLTK
Ssb2_Ssb1 NQAALNPR
Ssb2 VTPSFVAFTPQER
Ssb1 VTPSFVAFTPEER
Ssa4 NSVSENNFK
Ssa3 SINPDEAVAYGAAVQAAILTGDQSTK
Ssa2 TTPSFVGFTDTER
Ssa1_Ssa2 VDIIANDQGNR
Ssa1_Ssa2 NQLESIAYSLK
Ssa1_Ssa2 NFTPEQISSMVLGK
Ssa1 TTPSFVAFTDTER
Lhs1 FLETHPAIR
Kar2 VAYPITSK
Kar2 ITPSYVAFTDDER
Kar2 IEIDSFVDGIDLSETLTR
HSP90
Hsp82 FYSAFSK
Hsp82 EILGDQVEK
Hsc82 FYSAFAK
PFD
Pfd1 TQLDMVNQQLAYLDR
Pfd1 TIDNLK
Pfd1 SFILQDK
Pac10 ICQSLR
Pac10 ENITTMEVNCAR
Gim5 LLVPASASLYIPGK
Gim4 NNVFQAK
Gim3 VTAQLEK
Gim3 AILYAK
Gim1 YQQLQNELEEFIVAR
Gim1 LTGNVLLPVEQSEAR
Gim1 IVNEEFDQLEEDTPVYK
SMALL
Hsp42 YGAFER
Hsp42 FPVLPR
Hsp31 YLAPVGPWDDYSITDGR
Hsp31 SITGFTDVGETILGVDSILK
Hsp31 LVTGVNPASAHSTAVR
Hsp26 SVAVPVDILDHDNNYELK
Hsp26 LLGEGGLR
Hsp12 LNDAVEYVSGR
Hsp12 GVFQGVHDSAEK
2
3
4
5
6
Log10 CPC
Condition
NG
HS

Under NG conditions, absolute protein levels were defined for 40 chaperone proteins, ranging from 700 to 114,000 cpc (Figure 6.6). A slight increase in the median cpc level (from 7,500 to 13,100 cpc) was observed under HS conditions, with values ranging from 700 cpc to 260,000 cpc, indicative of an average increase in chaperone abundance in response to HS. Although 40 chaperones were quantified under both conditions, there were differences in the proteins for which cpc values were determined. Specifically, Pac10, Zuo1, Xdj1 and Jjj1 were quantified by 'A1' peptides in NG, but had 'B1' and 'B2' peptides under HS conditions. Additionally, Ssa3, Swa2, Ssaq1 and Hsc82 were quantified by 'A1' peptides in HS but had 'B1' and 'B2' peptides under NG conditions. The transition from 'A1' class to 'B1' class was indicative of decreasing (to below the limit of detection) absolute protein abundance, suggesting that the chaperones were down regulated under the respective condition. Only five chaperones from the 49 proteins targeted (Ecm10, Djp1 and those belonging to the protein group Hsp32_Sno4_Hsp33) failed to yield any absolute quantitative information under both conditions. I estimated from the 'B' peptide data that Ecm10 and Djp1 lie below 2,200 cpc in both conditions, whilst the combined abundance for Hsp32, Sno4 and Hsp33 is below 700 cpc in both conditions. Successful quantification was achieved for 36 proteins under both growth conditions (for protein abundances and fold changes, see Appendix 12).

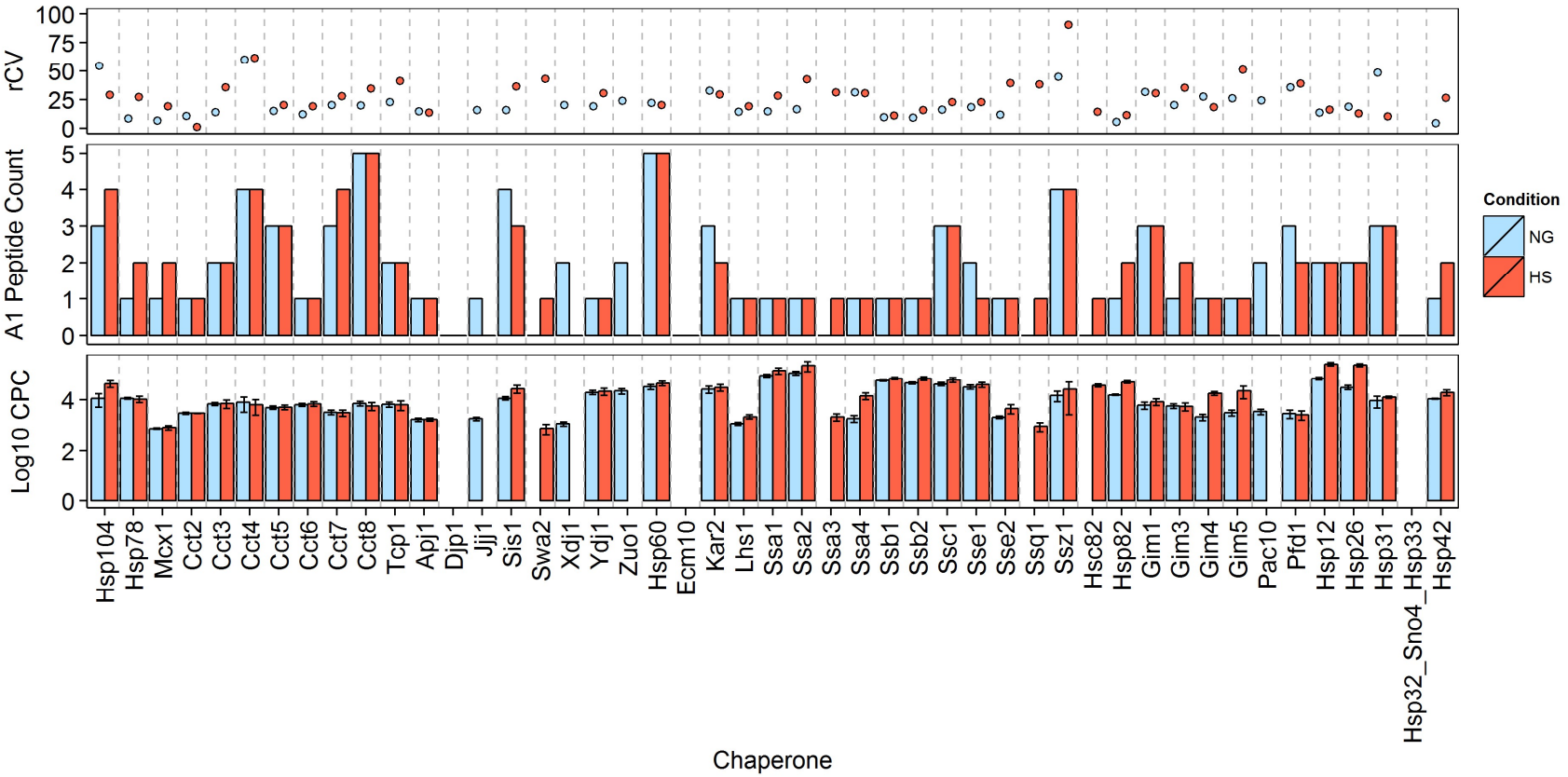



**Figure 6.6) Absolute protein quantification performed using 'A1' peptides**

Cpc values were obtained for 36 chaperones under both conditions of NG and HS. rCV values are below 40, with minor exceptions.

As indicated in previous genomic studies, I anticipated the Hsps were of higher absolute abundance under HS conditions. In order to determine the HSR-induced fold change in response to HS, the ratio of cpc under NG and HS were used. To calculate significance, an unpaired t test ($p < 0.05$) was performed between all logged NG biological replicates and all logged HS biological replicates on a per protein basis for all 'A1' peptides, correcting p-values for multiple testing using the Benjamini-Hochberg FDR approach (Figure 6.7). As discussed in Chapter 3, for protein groups that shared non-unique Q-peptides (Hsp32_Sno4_Hsp33), protein quantification was hypothesised to be the sum of the cpc values determined via any unique Q-peptides.

Thirteen proteins were observed as significantly expressed in response to HS (adjusted p value < 0.05), with a median fold change of 3.3: Hsp104, Sis1, Hsp60, Lhs1, Ssa4, Ssc1, Ssz1, Hsp82, Gim4, Gim5, Hsp12, Hsp26 and Hsp42 (Figure 6.7). All of these significantly changing chaperones, except Lhs1, Ssc1, Ssz1, Gim4 and Gim5, are known direct targets of the HSR modulator, Hsf1 (discussed in Section 1.4) according to genome wide studies performed by Hahn and colleagues (Hahn et al., 2004). According to Hahn and colleague's dataset, 16 of my 49 chaperones quantified are direct targets of Hsf1, having a median fold change of 1.9. These 16 chaperones fall into six of the chaperone subclasses: two AAA+ chaperones (Hsp78, Hsp104), three Hsp40 chaperones (Apj1, Sis1, Ypj1), one Hsp60 chaperone (Hsp60), six Hsp70 chaperones (Kar2, Ssa1, Ssa2, Ssa4, Sse1 and Sse2), one Hsp90 chaperone (Hsp82) and three small chaperones (Hsp12, Hsp26, Hsp42). As reported by Albanese and colleagues (Albanese et al., 2006) after genomics studies, there are two distinct chaperone networks in the cytosol of *S. cerevisiae*: the chaperones linked to protein synthesis (CLIPs) – repressed in response to heat shock in terms of their genomic expression levels, and the stress induced chaperones (Hsps) - induced following heat shock. In terms of absolute abundance, I observed no chaperone to be significantly down-regulated following heat shock, and surprisingly also observed up-regulation of a selection of cytosolic CLIP-type chaperones as classified by Albanese and colleagues (Figure 6.8) (Albanese et al., 2006). I discuss the biological significance of the significantly folding chaperones in the Discussion section of this thesis.

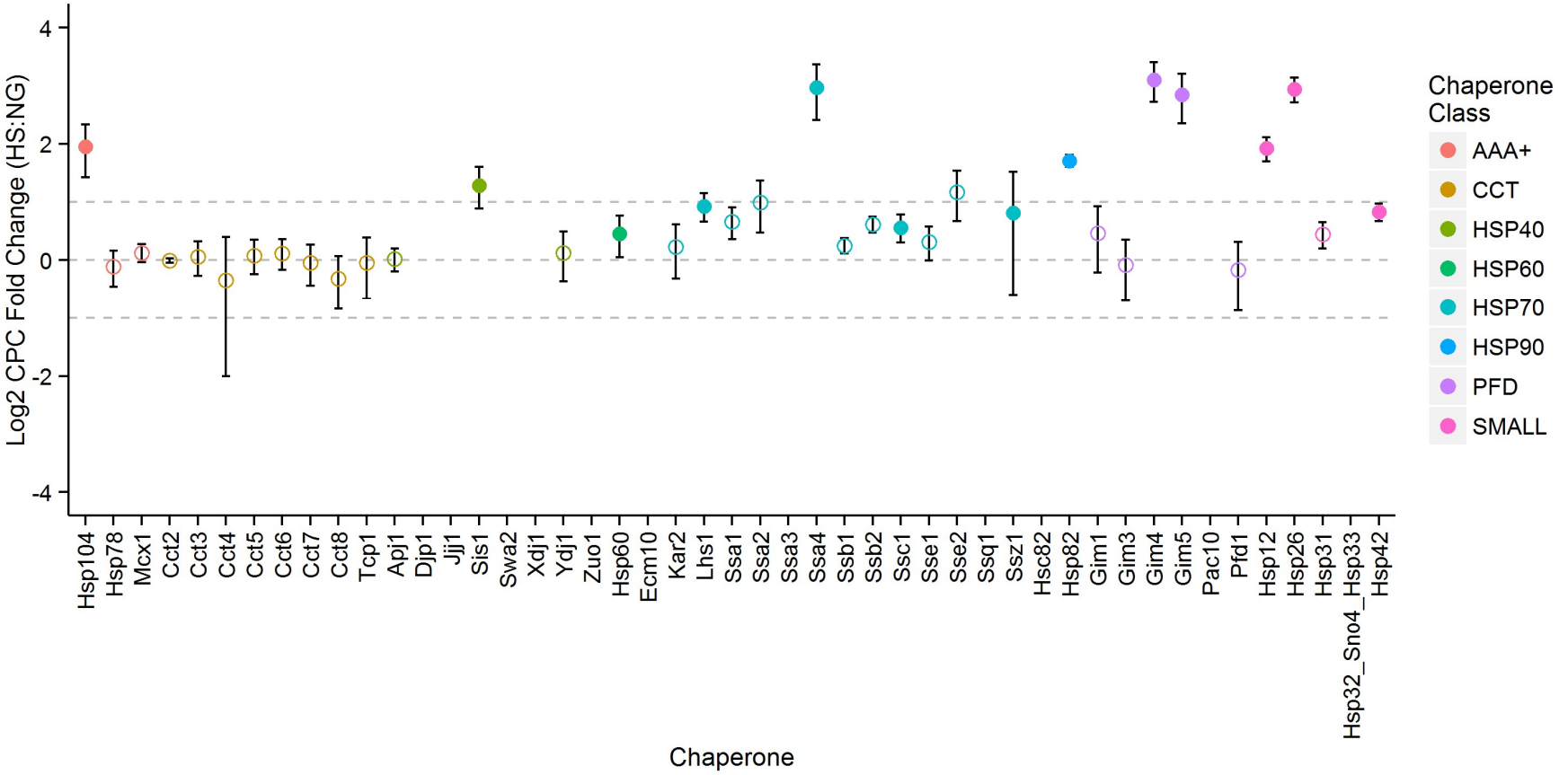



**Figure 6.7) Upregulation of known HSF1 targets.**

By performing an unpaired t-test between all biological replicates to determine final absolute protein abundance in NG and HS, the corresponding p-values and thus significant changing proteins (filled points) were determined.

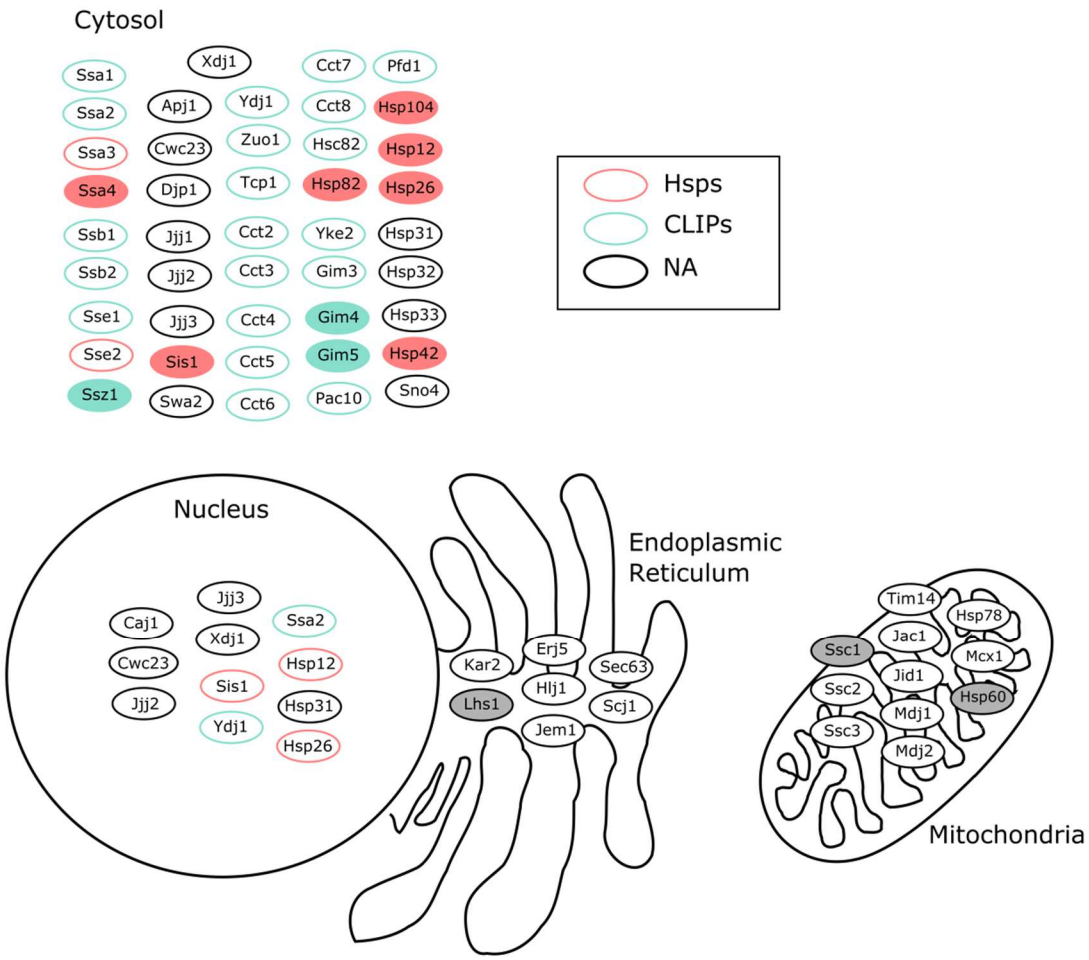



**Figure 6.8) The CLIPs and Hsps subsets of cytosolic chaperones**

In an alternative view of Figure 1.2, using clustering techniques on genomic data, (Albanese et al., 2006) identified two distinct cytosolic chaperone networks that were either repressed (CLIPs – outlined blue) or induced in terms of their gene expression levels in response to heat shock (Hsps – outlined red). The remaining chaperones in *S. cerevisiae* did not meet the grouping criteria due to their subcellular locations, function or genomic response to stress (here they are given the class 'NA'). Despite this, these chaperones may still play a role in the cellular response to stress. Chaperones are localised according to those depicted by Gong and colleagues (Gong et al., 2009). I found a subset of both Hsp and CLIP chaperones to be significantly upregulated in response to heat shock in terms of their absolute copy per cell values (chaperones are filled in).

## 6.3 SRM-correlated label free quantification

After performing a label-free study of the whole proteome of NG and HS *S. cerevisiae* samples as described in Methods, I quantified 1671 and 1816 proteins respectively, with 1644 yeast proteins in common between the two conditions according to MaxQuant (with a maximum Q-value for protein identification of 0.0091). In this set, 37 chaperones were also quantified in a relative manner, one of which was Sec63. Sec63 had not been targeted in my SRM experiments as it was included in the construct for ChapCAT010, which failed to express at sufficient levels.

Chaperone protein abundances determined using label free quantification demonstrated good agreement with respect to SRM-based QconCAT quantification, comparing the chaperone cpc values with their corresponding median MaxLFQ intensities reported by MaxQuant (Cox et al., 2014). A logged comparison of the 32 chaperones identified under NG conditions (set A) produced a Spearman rank correlation coefficient of 0.898, with an $R^2$ value of 0.762 (Figure 6.9a). For the 31 chaperones in common in HS (set B), the Spearman rank correlation coefficient was 0.848 whilst the $R^2$ value for the linear regression was 0.734 (Figure 6.9b). As these studies agreed well, I was able to calibrate the label-free data using a MaxLFQ SRM-normalisation approach, similar to that published previously (Rosenberger et al., 2014).

To convert the label-free quantification data to absolute values, condition-dependent linear regression was performed using the R package 'aLFQ' (Rosenberger et al., 2014) to predict 'mod-cpc' values from the median MaxLFQ intensities modelled on the cpc values for chaperones identified in both the SRM and unfractionated label-free datasets. Using leave-one-out cross-validation, this yielded a mean fold error of 1.8 for set A and 2.0 for set B. This regression approach normalised the label-free MaxLFQ values to compute SRM-corrected label free values (herein termed 'mod-cpc') for all 1644 proteins that were identified under both NG ($R^2$ = 0.762, Slope = 1.21, Intercept = 4.18, F = 96.16, p = 7.21 x $10^{-11}$) and HS ($R^2$ = 0.734, Slope = 1.05, Intercept = 4.85, F = 79.92, p = 7.87 x $10^{-10}$). A similar approach towards absolute quantification of the proteome has been performed in *E. coli* by Schmidt and colleagues (Schmidt et al., 2016).

To determine validity of SRM normalisation of the label-free quantification data in this manner, I compared the mod-cpc fold change in 30 ChapCAT-quantified proteins in response to HS. As mod-cpc calculated under either NG or HS conditions had a mean fold error of around two, I expected a greater error when calculating fold changes using mod-cpc values (Figure 6.9c). This proved to be the case, with a lower but still reasonable Spearman's rank correlation coefficient of 0.631 and an $R^2$ of 0.677. Finally, I assessed the ability of unfractionated label-free experiments to observe accurate differences in chaperone protein abundance between NG and HS. To do so, I compared the fold change of the median MaxLFQ intensity (HS/NG) for 30 chaperones identified in both conditions to their fold change counterparts. This produced a result with an $R^2$ of 0.649 (Spearman's rank correlation coefficient was 0.601), again showing reasonable agreement but slightly worse than the mod-cpc *versus* cpc acquired fold changes (Figure 6.9d). The aLFQ-based normalisation was therefore slightly superior in estimating protein fold changes than those determined by a standard label free experiment. When comparing the MaxLFQ fold changes and the mod-cpc fold changes using an unpaired Wilcoxon test (U test, $p < 0.05$), I observed no significant difference in the ranks of the fold changes of the chaperones ($p = 0.93$). I also assessed whether significantly upregulated chaperones in the cpc dataset were present in the top 10 equivalent set in the mod-cpc and MaxLFQ datasets, with both mod-cpc and MaxLFQ datasets ordered by decreasing fold changes (Table 6.1 and 6.2, respectively). I observed 7 significantly upregulated chaperones (according to my SRM dataset) in both the top 10 for the mod-cpc and MaxLFQ datasets. By performing an unpaired t test across the NG and HS biological replicate values for MaxLFQ intensity, I was able to observe only three significantly up-regulated chaperones in response to HS according to their adjusted p-value (adjusted using the Benjamini-Hochberg FDR approach). Due to my modelling approach I could not determine p-values for the mod-cpc dataset. However, according to the top 10 approach, using MaxLFQ SRM-normalisation improved my chances of identifying significant changes as determined by the gold standard SRM approach compared to using a purely label-free approach based on MaxLFQ intensities.

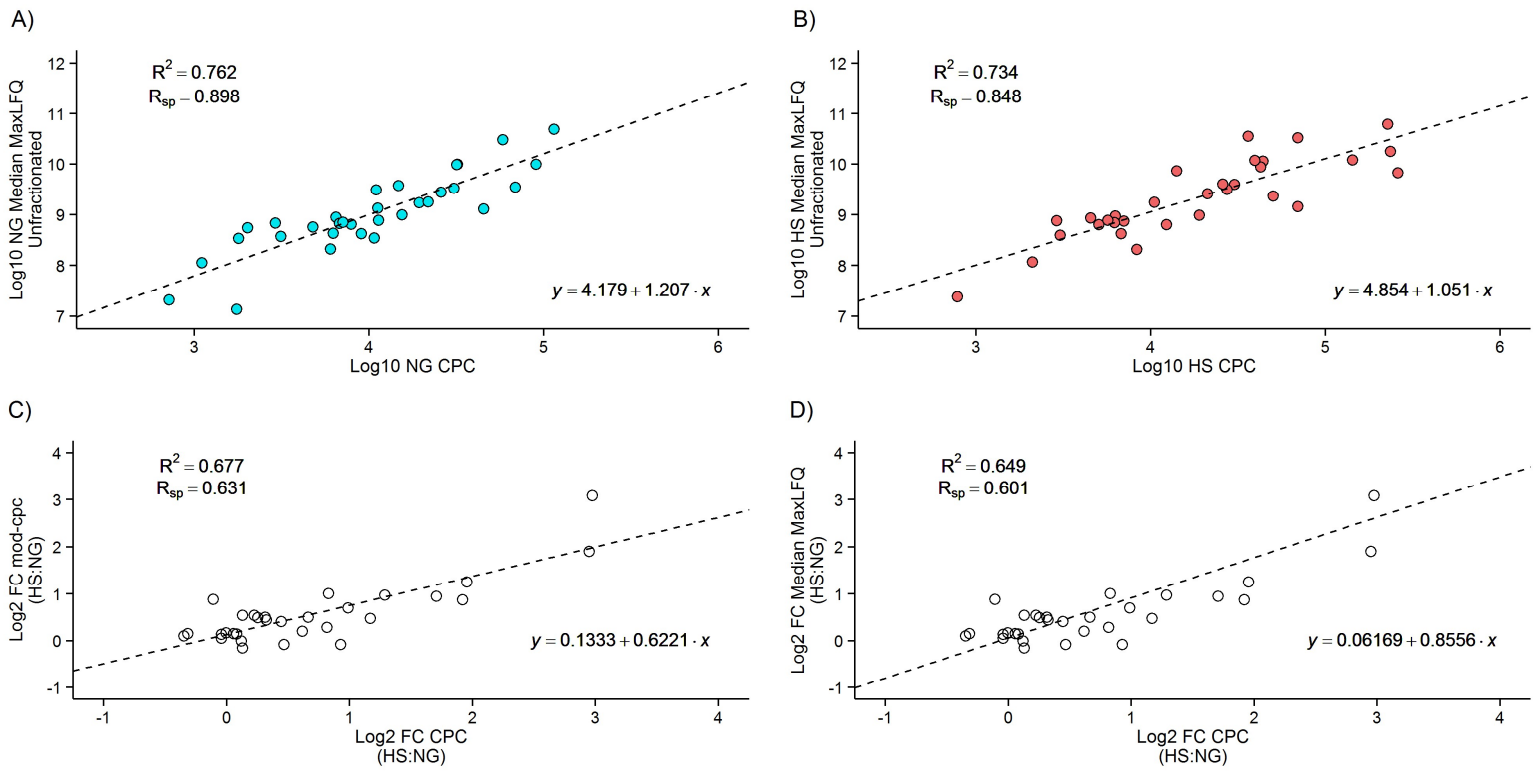



**Figure 6.9) Assessment of the abilities of relative quantification and MaxLFQ SRM-normalisation.**

After performing an unfractionated label free experiment, I compared the relative quantification of chaperones observed in NG conditions (A) and HS conditions (B). I performed MaxLFQ SRM-normalisation to obtain mod-cpc values ofr chaperones, and determined their fold changes according to their mod-cpc values. C) Upon comparison of these fold changes with absolute fold changes; I observed decreased agreement as a magnitude of the error in the model. D) I assessed the ability of relative quantification to accurately define fold errors, finding that the agreement between relative fold changes and the absolute fold changes is less than the agreement between the fold changes obtained following MaxLFQ SRM-normalisation.

**Table 6.1) Chaperones with the top ten mod-cpc fold changes**
Chaperones are ordered according to their mod-cpc (HS/NG) fold changes in decreasing order.

| | | | Absolute Dataset | | Single Shot Label Free |
|---|---|---|---|---|---|
| **Ensembl** | **Protein** | **Gene** | **Fold Change (cpc)** | **Adjusted P value** | **Fold Change (mod-cpc)** |
| P22202 | Ssa4 | YER103W | 7.85 | 2.194E-03 | 8.48 |
| P15992 | Hsp26 | YBR072W | 7.71 | 5.060E-07 | 3.75 |
| P31539 | Hsp104 | YLL026W | 3.87 | 8.940E-07 | 2.39 |
| Q12329 | Hsp42 | YDR171W | 1.78 | 3.040E-04 | 2.02 |
| P25294 | Sis1 | YNL007C | 2.44 | 2.000E-06 | 1.97 |
| P02829 | Hsp82 | YPL240C | 3.26 | 8.773E-04 | 1.93 |
| P33416 | Hsp78 | YDR258C | 0.93 | 6.235E-01 | 1.85 |
| P22943 | Hsp12 | YFL014W | 3.78 | 1.000E-07 | 1.84 |
| P15108 | Hsc82 | YMR186W | NA | NA | 1.77 |
| P10592 | Ssa2 | YLL024C | 1.98 | 6.224E-02 | 1.63 |

**Table 6.2) Chaperones with the top ten MaxLFQ intensity fold changes**

Chaperones are ordered according to decreasing MaxLFQ (HS/NG) fold change. Adjusted p values are determined according to an unpaired t test across the NG and HS biological replicate values for MaxLFQ intensity and adjusted according to a Benjamini-Hochberg FDR.

| | | | **Absolute Dataset** | | **Single Shot Label Free** | |
|---|---|---|---|---|---|---|
| **Ensembl** | **Protein** | **Gene** | **Fold Change (cpc)** | **Adjusted P value** | **Fold Change (MaxLFQ)** | **Adjusted P value** |
| P22202 | Ssa4 | YER103W | 7.85 | 2.194E-03 | 21.67 | 6.900E-02 |
| P15992 | Hsp26 | YBR072W | 7.71 | 5.060E-07 | 5.43 | 1.162E-01 |
| P31539 | Hsp104 | YLL026W | 3.87 | 8.940E-07 | 2.87 | 2.435E-01 |
| Q12329 | Hsp42 | YDR171W | 1.78 | 3.040E-04 | 2.77 | 2.369E-01 |
| P02829 | Hsp82 | YPL240C | 3.26 | 8.773E-04 | 2.35 | 2.412E-02 |
| P25294 | Sis1 | YNL007C | 2.44 | 2.000E-06 | 2.35 | 1.310E-02 |
| P33416 | Hsp78 | YDR258C | 0.93 | 6.235E-01 | 2.27 | 2.443E-01 |
| P22943 | Hsp12 | YFL014W | 3.78 | 1.000E-07 | 1.94 | 3.295E-01 |
| P32590 | Sse2 | YBR169C | 2.25 | 7.086E-02 | 1.56 | 1.520E-01 |
| P15108 | Hsc82 | YMR186W | NA | NA | 1.53 | 1.913E-01 |

In terms of absolute modcpc values, protein abundances under NG conditions ranged from 400 to 338,000 modcpc; whilst under HS conditions, protein abundances ranged from 300 to 480,000 modcpc (Figure 6.10).

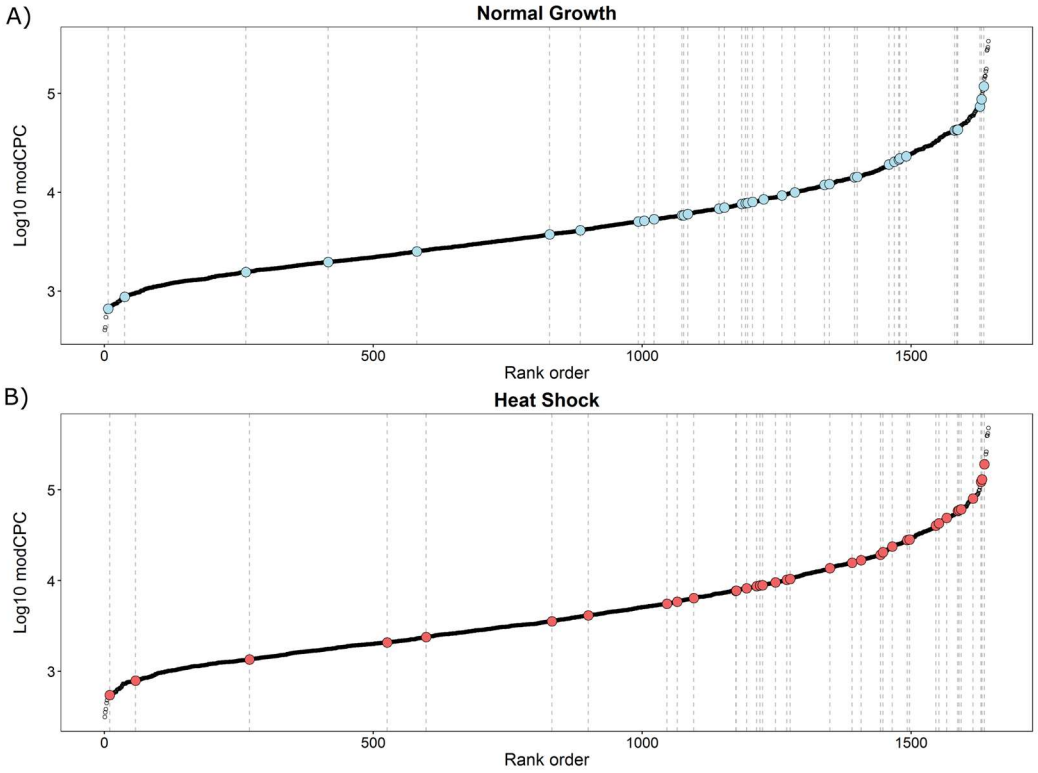



**Figure 6.10) The modcpc values of 1644 proteins ranges from $10^3$ to $10^5$ under conditions of NG and HS**

A) Chaperones (blue points) used to perform linear regression between NG cpc values and NG MaxLFQ intensities in order to obtain modcpc values. B) The same 30 chaperones (red points) were also used to perform linear regression between their HS cpc values and respective HS MaxLFQ intensities.

In a recent study performed by Picotti and colleagues (Picotti et al., 2009), one hundred proteins spanning all levels of cellular abundance (41 to 1,000,000 cpc) in *S. cerevisiae* were targeted and detected in an SRM strategy, using five proteotypic peptides per targeted protein. Indeed, using SRM I quantified chaperones from 700 cpc to 114,000 cpc under NG conditions (700 cpc to 259,000 cpc under HS conditions), with a smaller range likely the result of only targeting chaperone proteins. Picotti and colleagues separated these one hundred proteins into classes dependent upon their abundance, with the uppermost class (524,288 cpc to 1,255,722 cpc) containing the proteins YGL008C, YKL060C, YLR355C, YLR249W and YDR382W. In my label-free experiment, I expected to detect these high abundant proteins; however, only YGL008C was identified at 97,500 and 137,800 mod-cpc under NG and HS conditions respectively. In a DDA experiment, proteins of higher abundance are more routinely identified, therefore it is surprising that previously reported highly abundant proteins are missing from the dataset;

however, discrepancies in quantification can arise due to a difference in growth conditions. Whilst cultures in both experiments were grown under normal conditions at 30 °C, samples grown by Picotti and colleagues (Picotti et al., 2009) were grown to log-phase whilst my samples were grown to late-log/early-stationary phase. Picotti and colleagues did not determine abundances under heat shock, and so the data for this condition was not comparable. Picotti and colleagues quantified 21 proteins using stable isotope reference peptides that spanned the cpc range, twelve of which were also quantified via the mod-cpc approach. Both the $R^2$ value and the spearman's rank correlation coefficient was high when comparing the logged values, demonstrating good agreement between the quantification for the twelve proteins spanning the range of $10^2$ to $10^6$ cpc (Figure 6.11).

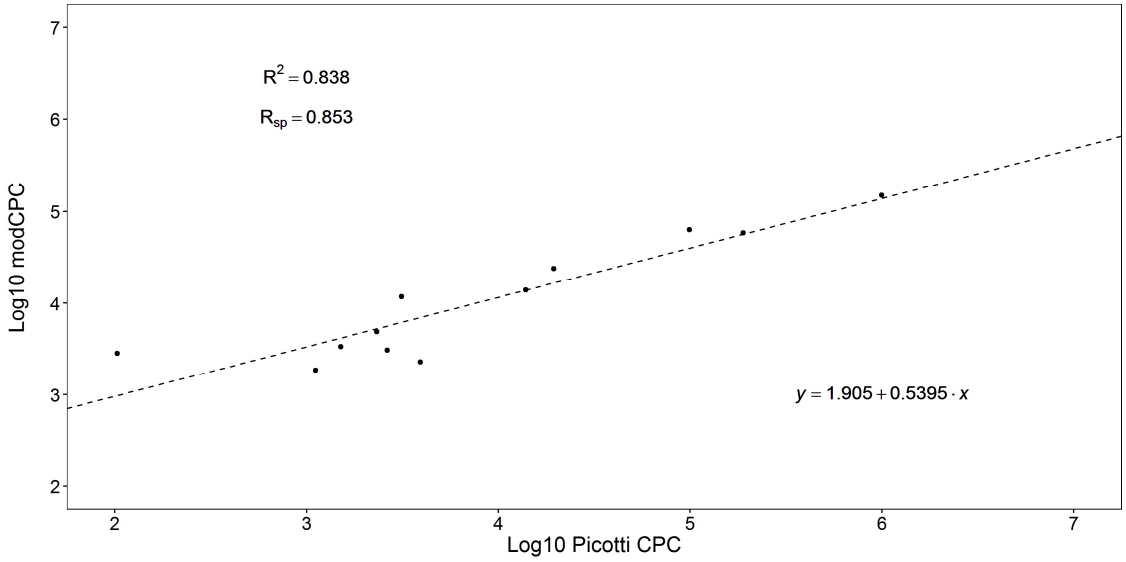



**Figure 6.11) Comparison of SRM-determined cpc values by Picotti and colleagues with mod-cpc values for twelve proteins**

Although mod-cpc values derived from a linear modelling approach using chaperone cpc values and their MaxLFQ intensites spanned a slightly smaller magnitude range, agreement between twelve proteins for which SRM cpc values and mod-cpc values were known is high, with an $R^2$ of 0.838 and a spearman's rank correlation coefficient of 0.853.

In summary, according to the MaxLFQ SRM-normalised model for 1644 proteins in NG and HS conditions (Appendix 13), the median mod-cpc under NG and HS was 3700 and 3500 respectively, with a median fold change of 0.97. Performing an unpaired Wilcoxon test (U test, $p < 0.05$) on protein concentration across biological replicates determined via a Bradford Assay indicated no significant change in response to HS ($p = 0.80$). No change in cell size was observed during cell counting, with the average size of a cell under NG and HS 4.18 µm. However, total protein abundance of the 1644 proteins increased in response to HS to 18,482,888 mod-cpc, a 1.18 fold-change. I discuss this data in biological context in the Discussion section of this thesis.

I used these 1644 mod-cpc values as measurements of abundance towards modelling the 'chaperome' response to heat shock, as is discussed in the next chapter.

# Chapter 7 - Informatics approaches to modelling the *S. cerevisiae* chaperome

To understand the wider role of each chaperone, one must also investigate its client proteins. By monitoring the changes in the clientele for a chaperone in response to various conditions, we can elucidate the role of the chaperone in a particular condition or pathway. For instance, the PFD complex is believed to interact predominantly with actin and tubulin (Vainberg et al., 1998), and is therefore a crucial player in the biosynthetic pathway for these cytoskeletal proteins. It is a co-chaperone of the CCT complex (often referred to as chaperonins), and acts by transferring its substrate proteins to this complex for refolding. It is therefore expected to have a distinct client protein profile, which should be manifest in proteomics experiments. Such experiments require the chaperone network to be characterised, and the abundance of the chaperone and each client to be known.

To investigate the workload of each chaperone in terms of its substrates, its interacting proteins (herein termed 'clients') should first be identified; the so-called 'chaperome'. Here I describe the chaperome network as all the chaperone-client interactions in the proteome. To identify the chaperome interactions, Gong and colleagues (Gong et al., 2009) have carried out an extensive proteome-wide affinity purification experiment in which each of 63 known chaperones were tagged and their interactors isolated through tandem affinity purification pull downs and MS used to identify the interacting partner. Other approaches have also added to this data set, including the use of yeast two-hybrid experiments (Ito et al., 2001) and genomic studies to identify co-regulated genes and thus likely to participate in the same functional process (Albanese et al., 2006). Tandem affinity experiments, whilst able to identify interaction partners, are subject to high numbers of false positives due to nonspecific binding or common contaminants, towards which informatics approaches have been generated to curate data as well as dedicated databases for use towards filtering data (Choi et al., 2011, Mellacheruvu et al., 2013, Lawless and Hubbard, 2014). Various protein-protein interaction database repositories are available for download online, consisting of curated datasets of protein-protein interactions (PPI) submitted to the repository. Such repositories often contain large numbers of interactions validated across multiple experiments, and may be scored within the database to give an indicator of the likelihood of the interaction being a 'true positive'. Like Brownridge and colleagues (Lawless and Hubbard, 2014, Brownridge et al., 2013), I created a high quality chaperome interaction dataset by downloading chaperone-proteins interactions reported in

the PPI repositories BioGrid, String and IntAct (Stark et al., 2006, Szklarczyk et al., 2015, Hermjakob et al., 2004).

Although informative, an interaction network alone does not indicate the response of the network to various conditions. One method for introducing abundance data is to weight interactions according to the relative abundances of each protein pair in an interaction. Weighting interactions by the respective abundances of the protein pair can result in a reorganisation of the network topology and modular subnetworks may be identified. For instance, Mihalik and colleagues weighted interactions by the average mRNA abundance of the two interacting proteins and the median interaction weight decreased in response to heat shock, indicative of a less intensive 'resource-sparing' interactome (Mihalik and Csermely, 2011). Alternatively, and the approach that I used in this study, is to use the abundances of all interacting proteins to inform conclusions about the protein of interest. This work has been carried out previously under NG conditions using absolute abundances determined via the QconCAT strategy by Brownridge and colleagues (Brownridge et al., 2013). By determining the volume (total abundance) of all interacting proteins to a single chaperone and the workload (the total abundance of client proteins mediated by all copies of the chaperone per minute), Brownridge and colleagues were able to deduce the major chaperone players in the chaperome network under normal conditions, supporting the general hypothesis that chaperones responsible for mediating the most folding of the cell are generally highly abundant.

I extended this analysis to cover the heat shock response. Firstly, I created a novel high quality dataset, searching for reciprocal interactions reported in at least two of three up-to-date dataset repositories available online, and was able to use absolute copy per cell values for chaperones and clients by virtue of the SRM-normalisation approach as described in Chapter 6.

## 7.1 Chaperone-client interactions

I created a high quality chaperone:non-chaperone protein interaction dataset using *S. cerevisiae* interactions reported within the IntAct, String and BioGrid databases. This dataset contained 2761 interactions, for which all non-chaperone proteins were considered client proteins of the chaperone it was interacting with. A single client protein is able to interact with multiple chaperone proteins, as the network of chaperones is known to pass client proteins between classes for further folding or degradation. My dataset containing 2761 interactions covered 60 chaperones and 1412 client proteins.

In order to determine the protein volume encountered by each chaperone, I used the mod-cpc values determined for both NG and HS in Chapter 6 as the abundance of each client protein. However, only 1644 proteins were present in both NG and HS conditions following MaxQuant processing. Of the 1412 client proteins in my interaction dataset, only 635 had a mod-cpc value reported in my SRM-normalised label free dataset, covering 55 chaperones. Despite the missing data, I was able to ensure that the relationship between the number of clients and the number of clients with abundance data per chaperone was comparable (Figure 7.1). Both the Spearman's correlation coefficient and adjusted $R^2$ for a linear model was 0.95. Therefore, I was able to base my model on the 635 clients with abundance data as a representative of the entire interaction dataset.

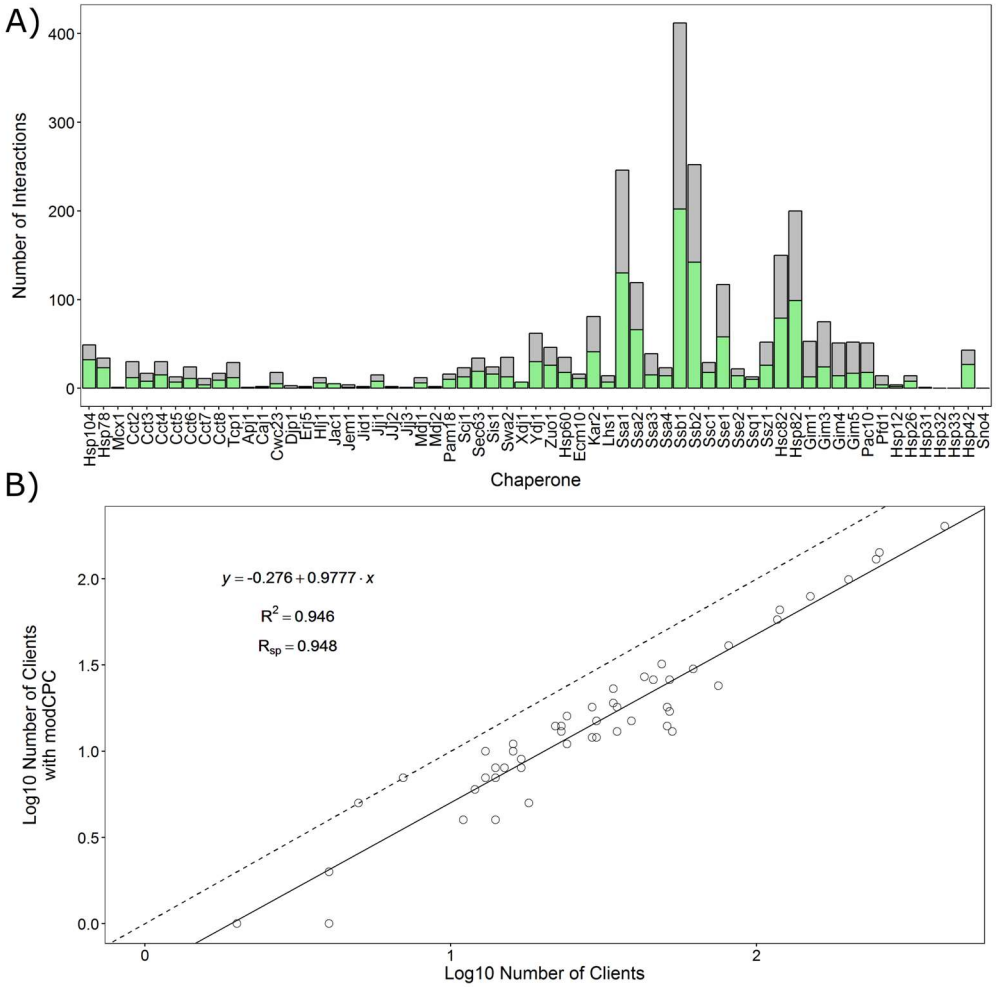



**Figure 7.1) Mod-cpc values were available for 635 nonchaperone protein clients**

A) I obtained mod-cpc values for 635 (green) out of 1412 clients across all 60 chaperones in the interaction dataset. Interactions shown in grey involved client proteins that were not detected in the label free quantification dataset and so quantification data was missing. B) Despite missing data, agreement between the number of clients with a mod-cpc value and the number of clients per chaperone was high. The linear regression line is presented as a solid black line.

Assuming that each interactor identified for each chaperone has its folding mediated by the chaperone, I tested whether the abundance of the chaperone is related to the number of its client proteins. Under conditions of NG and HS, I observed significant correlation between the copies of chaperone and the number of client proteins the chaperone mediated folding of (Spearman's rank correlation coefficients of 0.620 and 0.576 respectively) (Figure 7.2). In response to HS, I observed a lower correlation and regression, suggesting chaperone abundance did not agree as well with the number of interacting clients. To note, I assumed identical chaperone-client interactions under both conditions.

In my high quality dataset, I observed general trends for the number of interactors across different chaperone classes. Hsp70 and small chaperones are reported to be promiscuous, able to interact with a wide range of proteins. I found Hsp70 proteins typically interacting with the largest number of client proteins in my dataset, with Ssb1 and Ssb2 interacting with 202 and 142 clients respectively. Chaperones of Hsp40 family with specialised roles, for example Mdj2 as a mitochondrial import motor and Erj5, a Hsp40 chaperone within the ER, interacted with the fewest number of clients, with only one client reported for Jjj2, Jid1, Caj1, Jem1, Erj5 and Mdj2 (see Appendix 14 for full data table). Interestingly, Hsp40 chaperones mediate the client delivery to, and the ATPase activity of, Hsp70 chaperones, but appear to be lower in abundance and interacting with fewer clients than its Hsp70 chaperone partner. For instance, Sis1 interacts with the Hsp70 chaperone Ssa1, but interacts with only 16 client proteins whilst Ssa1 is reported to interact with 130, with almost a 9-fold chaperone abundance difference (11,000 and 90,000 cpc for Sis1 and Ssa1 respectively).

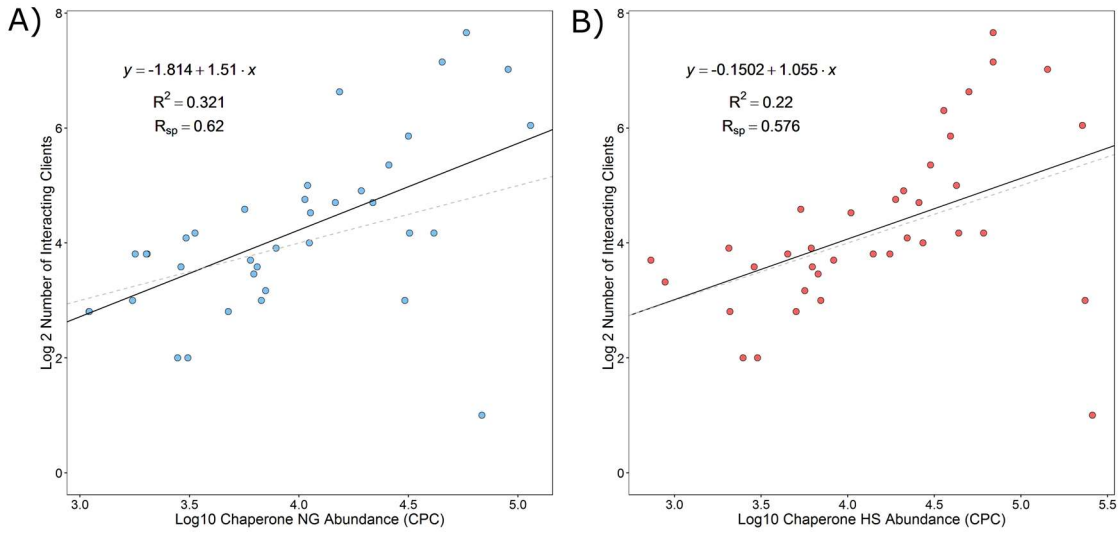



**Figure 7.2) Comparison of chaperone abundance to number of interacting clients**

A) Comparison under conditions of NG; B) Comparison under conditions of HS; I assumed identical interaction networks for both conditions.

## 7.2 Protein volume mediated by chaperones

Using the mod-cpc values for all 635 client proteins I could deduce the client protein volume (defined as the total number of client protein copies a chaperone interacts with) interacting with each chaperone under conditions of NG and HS. As each client protein could interact with multiple chaperones during its protein folding process, I assumed that each chaperone would interact with equal copy numbers of client protein, such that each chaperone 'shares the load'. Therefore, if a client of 300 copies interacted with three chaperone proteins, each chaperone protein would interact with 100 copies of the client protein. To calculate the total client volume mediated by each chaperone, I summed the client mod-cpc values (with any 'share the load' adjustments taken into consideration) (Figure 7.3).

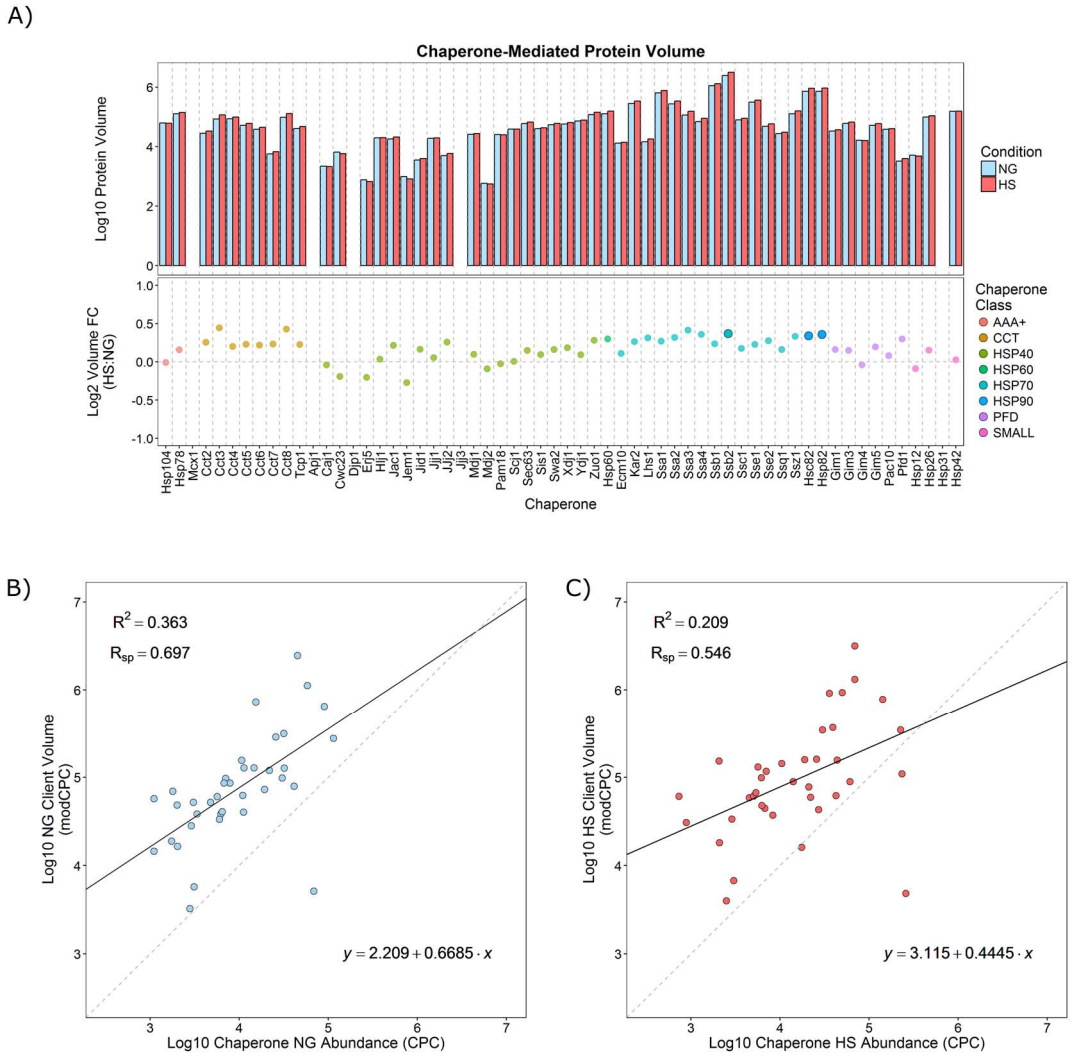



**Figure 7.3) Chaperone-mediated protein volume correlates well with abundance between conditions**

A) Upper panel: total mod-cpc of clients on a per chaperone basis (the total protein volume mediated by the chaperone), under NG and HS. Lower panel: fold change in protein volume mediated by the chaperone in response to HS conditions. Significant changes are enlarged and outlined in black (Ssb2, Hsc82 and Hsp82, $p < 0.05$); B) Comparison between the protein volume mediated and chaperone abundance under NG conditions; C) Comparison between the protein volume mediated and chaperone abundance under HS conditions.

I found the chaperone Ssb2 interacted with the most copies of client proteins under both conditions of NG and HS at 2,449,378 cpc and 3,165,577 cpc respectively. Ssb2 (and its paralog Ssb1) is a cytoplasmic chaperone that is able to interact with the ribosome to function in nascent chain co-translational protein folding and so is able to interact with a wide range of client proteins. In comparison, under both NG and HS conditions, the Hsp40 chaperone Mdj2 interacted with the least copies of client protein (588 cpc and 551 cpc respectively), only interacted with one client protein with a mod-cpc value (Tim44, a component of the TIM23 complex – the mitochondrial import motor) according to my high quality dataset. I was unable to determine client protein volume mediated by the chaperones Mcx1, Apj1, Hsp31, Djp1, Jjj3,

Sno4, Hsp33 and Hsp32 due to either their lack of interactions in the dataset or lack of mod-cpc values associated with any client proteins.

To determine any significance between NG and HS client volumes, I performed a paired Wilcoxon test between all NG and all HS client mod-cpc values for each chaperone (*U* test, $p < 0.05$). I found the client volumes for Ssb2, Hsc82 and Hsp82 to be significantly different ($p = 0.019$, 0.04 and 0.04 respectively), with total protein volume mediated by these chaperones as upregulated (1.29, 1.26 and 1.28 -fold respectively). Of these chaperones, only Hsp82 was deemed significantly upregulated in terms of its absolute cpc value in response to HS, whilst Ssb2 and Hsc82 remained unchanged (Chapter 6). Agreement between the chaperone cpc and the protein volume is higher under NG conditions in comparison to HS, with a Spearman's rank correlation coefficient of 0.70 under NG (Figure 7.3b) and 0.55 under HS (Figure 7.3c). To observe correlation between the protein volume fold change and the chaperone cpc fold change in response to HS, I performed linear regression and Spearman's rank correlation on the logged fold changes. There was no relationship between the chaperone cpc fold change and the change in protein volume mediated by the chaperone (Spearman's rank correlation coefficient was -0.17), whilst data did not well suit the linear model ($r^2$ was 0.11). Chaperone cpc fold changes appeared to be slightly higher than the client volume fold changes in response to HS, although this was not the case for every chaperone (Figure 7.4).

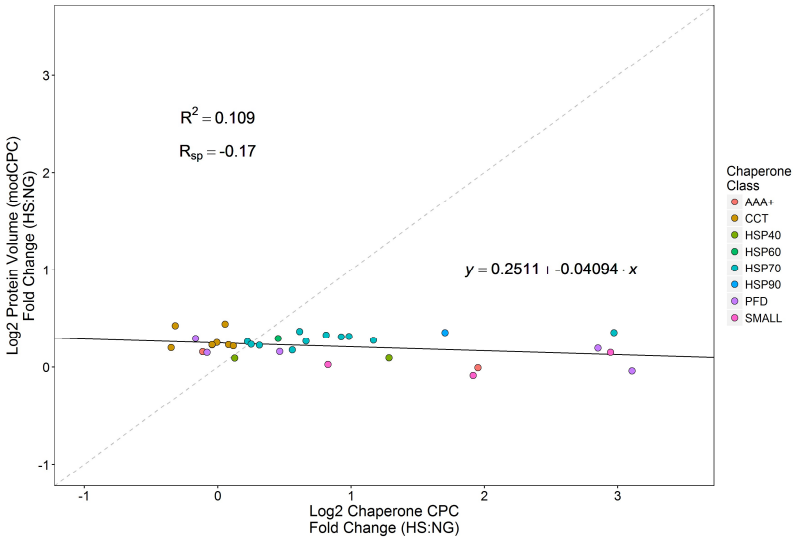



**Figure 7.4) Chaperone response to HS versus client volume response to HS**

I compared the absolute fold change of a chaperone in response to HS to its client protein volume fold change in response to HS.

## 7.3 Chaperone workload

I extended this analysis further by calculating the chaperone workload per minute, i.e. the number of copies per client all chaperone copies interact with per minute. Using the same calculation as that used by Brownridge and colleagues (Brownridge et al., 2013) (see Methods for full calculation) I was able to determine synthesis rates for each client protein. To do so, I used degradation rates determined by Christiano and colleagues (Christiano et al., 2014) via MS, assuming that the cell was at steady state and that flux in molecules per unit time was the responsibility of individual chaperones. I also assume degradation rates are identical between conditions. Any missing values were replaced by the geometric mean across the entire degradation rate dataset, such that the mean $k_{deg}$ was unchanged. Cell growth rates were not formally taken into account in this model as abundances and turnover rates are taken from different yeast studies. Growth rates vary during batch growth of cells, and so error constants could be added to represent the growth rate. However, given the constraints of the model (see the Discussion chapter of this thesis) I kept my equation identical to that used in previous literature.

I observed total flux per minute ('workload') to be highest for the Hsp70 chaperone Ssb2 under conditions of both NG and HS (~2900 mod-cpc $min^{-1}$ and ~3750 mod-cpc $min^{-1}$ respectively), with a median fold change in response to HS of 1.15 (Figure 7.5). Again, I observed an apparent trend for the chaperone classes. Hsp40 chaperones that resided in subcellular compartments such as the ER and mitochondria had low total workloads per minute, indicative of a slow folding time. Specifically, the lowest was Mdj2 with a total workload under NG and HS conditions of 0.70 and 0.66 mod-cpc $min^{-1}$ respectively. This data suggests that promiscuous chaperones with higher number of interactors are able to mediate the folding of more copies of client protein faster than those that only interact with a few clients. To determine any significance between NG and HS client volumes, I performed a paired Wilcoxon test between all NG and all HS $k_{syn}$ client values for each chaperone (*U* test, $p < 0.05$). The average workload fold change across all chaperones was 1.13, however I found only the workload for Ssb2 significantly changing in response to HS ($p = 1 \times 10^{-14}$). This data suggests that whilst chaperone abundance and client volume does increase significantly for particular chaperones, chaperones (except Ssb2) do not alter their speed of protein folding. Due to the lack of available turnover datasets for stress response, an assumption of my model is that $k_{deg}$ values do not change in response to HS. In reality however, the unfolded protein response initiated as a result of stress

causes increased degradation of proteins (Verghese et al., 2012), and further experiments are required to determine the resulting changes in protein turnover.

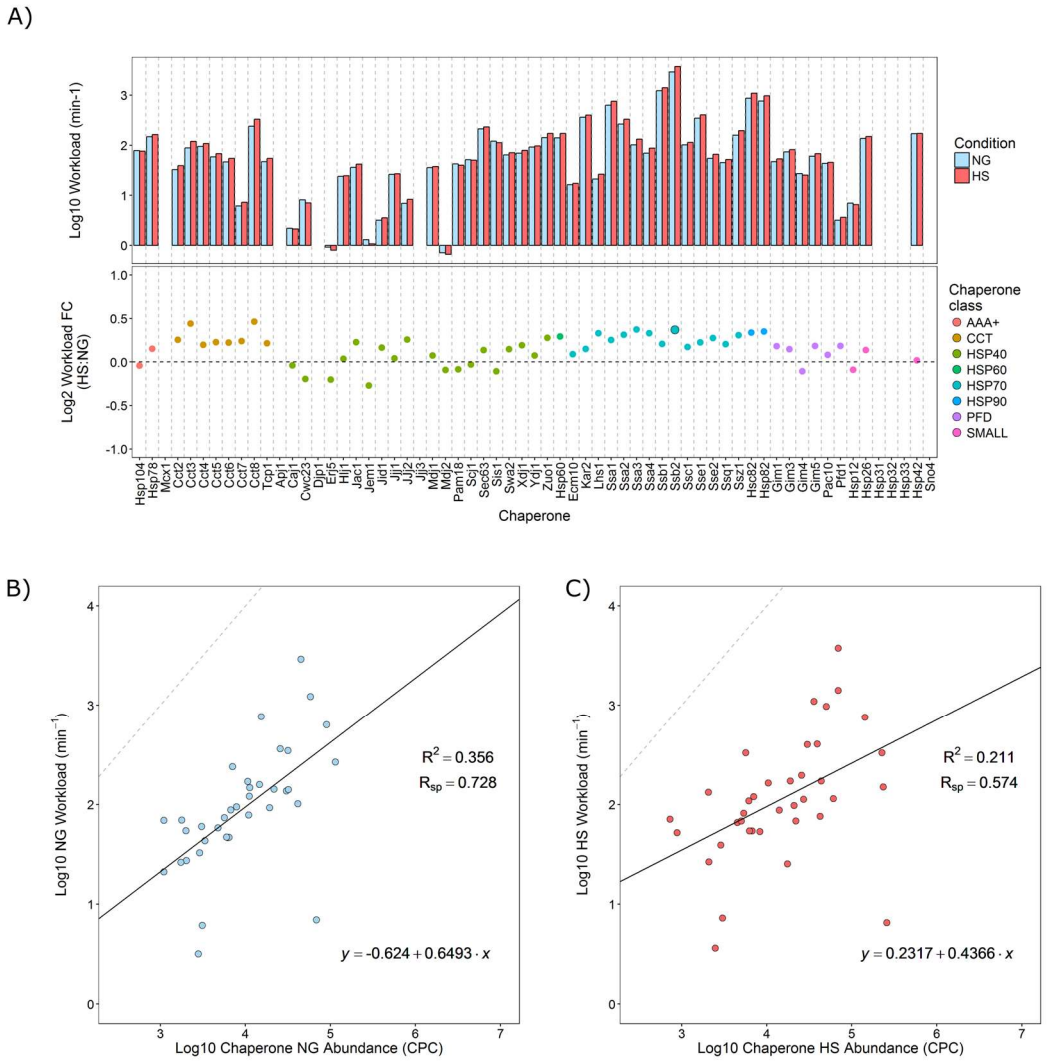



**Figure 7.5) Chaperone workload and abundance correlates well between conditions**

A) Upper panel: workload on a per chaperone basis (the total $K_{syn}$ of all client proteins), under NG and HS. Lower panel: fold change in chaperone workload in response to HS conditions. Significant changes are enlarged and outlined in black (Ssb2, $p < 0.05$); B) Comparison between the protein volume mediated and chaperone abundance under NG conditions; C) Comparison between the chaperone workload and chaperone abundance under HS conditions.

Furthermore, I compared the chaperone cpc abundances to their respective total workloads. Under NG conditions, I observe good agreement between the ranks of the chaperone abundance and the total workload (Spearman's rank correlation coefficient was 0.728) whilst under HS, this agreement was only moderate (Spearman's rank correlation coefficient was 0.574) (Figures 7.5b and c).

Finally, I compared the chaperone response to HS to the change in its workload in response to HS (Figure 7.6). I found no linear relationship between the two variables, suggesting the chaperone abundance response is not a direct response to changing workload. Brownridge and colleagues extend this modelling further to take into account the chaperone efficiency, that is the number of client molecules mediated per chaperone molecule per minute, but due to the limitations of my model and missing data in terms of client and chaperone abundance I did not deem it suitable to extend this analysis further.

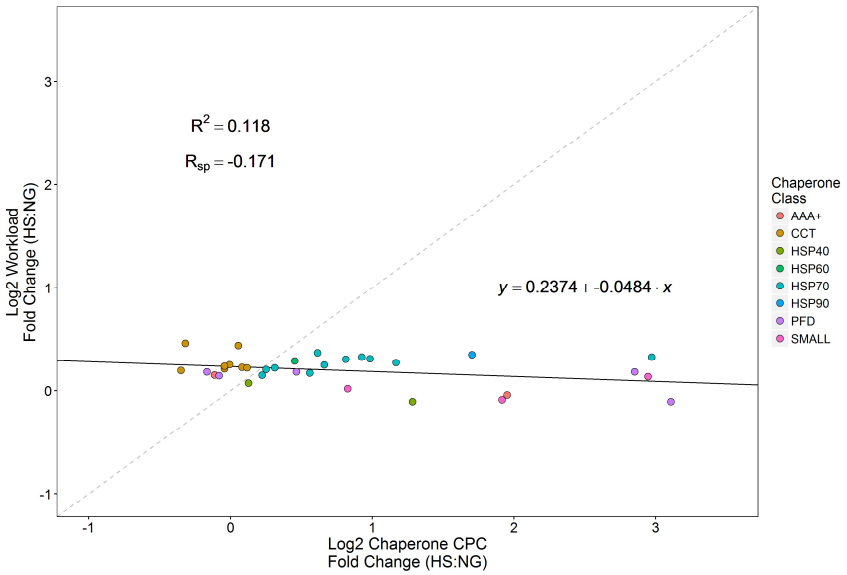



**Figure 7.6) Chaperone response to HS versus workload response to HS**

I compared the absolute fold change of a chaperone in response to HS to its workload fold change in response to HS.

This modelled data suggests that chaperones that mediate the folding of a large number of client proteins (particularly Ssb2) are significantly upregulated in terms of their protein volume mediated and workload. In addition, of the three chaperones identified as significantly upregulated in terms of their protein volume, we observed significant upregulation of two of chaperones in terms of their absolute abundance. This indicates chaperone upregulation in order to mediate the additional protein volume. However there are limitations to this model, further discussed in the next chapter.

# Chapter 8 - Discussion and Further Work

In the work presented in this thesis, a variety of quantitative proteomics methods and accompanying bioinformatics strategies have been used to study the ability of a model eukaryotic cell system to deploy its protein chaperones in response to an environmental stress in order to maintain proteostatsis. A modified QconCAT strategy, using up to five Q-peptides per chaperone to perform absolute quantification of 40 chaperones under conditions of NG and HS was used, to elucidate the chaperone response to heat shock. Furthermore, using an extended classification technique, additional quality control checks were developed. Coupled to an approach that determines the level of the completion of each tryptic digest, only the highest quality peptide data were used for quantification. Using a high quality interaction dataset, gathered from three data repositories, the chaperome changes in response to heat shock were subsequently modelled. The overall findings from these different studies and their implications for further work suggestions are discussed in this chapter.

## 8.1 The QconCAT strategy using up to five Q-peptides per chaperone

Using the QconCAT strategy, in which up to five Q-peptides were selected per target chaperone I have quantified 40 chaperones under conditions of NG and HS. Previous attempts to had only used two peptides per target and had taken a more “hand-crafted” approach to peptide selection. Here, the MC:Pred and CONSeQuence online prediction tools were also used directly to select Q-peptides for each protein of interest; specifically they were used, where possible, to aid in selection of the top five Q-peptide candidates for each chaperone. A major advantage of the use of Q-peptides for absolute quantification is that there is an internal standard for every single peptide targeted. The standard and analyte peptide are identical, although there is a mass difference corresponding to the label; both counterparts behave in a mass spectrometer in the same way, and share the same chromatographic properties, providing a method of quality control for every peptide used for quantification.

In previous attempts (Brownridge et al., 2013, Lawless et al., 2016), the SRM tool mProphet was used to elucidate the true peak corresponding to each peptide. However, during manual verification, we observed various issues with this selection tool, particularly for peptides with low intensity peak groups and changing transition profiles between the standard and analyte counterpart. Low abundance peptides were often assigned to background noise rather than a true peak group or, in the latter case of changing transition profiles, despite the

inconsistent transition profile the peak groups passed the 1 % mProphet threshold. In this instance, without manual intervention, it would likely have resulted in erroneous absolute quantification. The pipeline used here processed the data first with Skyline to identify the true peak group for each peptide according to the co-elution between the standard and analyte counterparts. This meant that *in silico* retention time alteration was possible via manual adjustment, such that subsequently mProphet could only select a true peak within a 1 minute retention time window about the peak group selected by Skyline. Even so, in a few examples, individual peptide biological replicates failed to pass the 1 % mProphet FDR threshold, and were subject to removal prior to absolute quantification of the peptide and parent protein. Although mProphet provides an excellent automatable tool for prediction of the true peak and a means of ensuring each peak group is validated according to an FDR threshold, thorough manual verification is arguably still required to ensure quantification is performed using the 'true peaks' rather than background noise. This would ensure that any peak groups that pass a 1 % mProphet FDR threshold have the same product ion profile between the standard and analyte peptide groups, and that background noise has not been selected as the true peptide signal. Although mProphet uses a sophisticated scoring system based on multiple profile features to ensure the quality of peaks selected it appears it can still be "fooled", this highlights the trade-off between fully automatable software pipelines and the superior (but more time-consuming) results that can be obtained by manual intervention by an expert. However, such a manual approach is not practical for a very large number of attempted QconCAT quantifications such as performed on the CoPY project (Lawless et al., 2016).

For the reasons outlined above, an updated classification scheme was implemented in which 'B' peptides were categorised into 'B1' and 'B2', and 'A' peptides were categorised into 'A1' and 'A2' peptides, with 'A2' peptides classified according to manual verification. A major advantage to using multiple peptides to quantify a single protein is that the peptide-level abundance should be equal amongst these peptides. This notion can be used as an indicator of peptides that could result in an erroneous quantification, particularly in cases where only two 'A' type peptides are available for quantification of the parent protein with an abundance difference of over two-fold. This rule has been used in previous approaches (Lawless et al., 2016). Here, using up to five candidate peptides, where there are more than two 'A' peptides, if all have passed the 1 % mProphet FDR threshold and behave in the same way in response to heat shock whilst passing all other quality control checks, quantification was performed using all the 'A' peptides.

Additional SRM processing tools have been published that attempt to improve quantification using SRM experiments, but still require a degree of manual verification. MSstats is available through the Bioconductor package in R, (available at http://www.msstats.org) (Choi et al., 2014). Analysis is performed by modelling all the available data to the protein level and accounting for stochastic variation using a model that accounts for individual peptide variance. MSstats utilises mProphet as its first step in analysis to identify high scoring peak groups according to the FDR threshold, before performing statistical examination and processing of the peak groups. The peak areas are log transformed and normalised via linear mixed effect models such that variation between runs is reduced. MSstats also has the capability to identify transition peak groups that were incorrectly selected by mProphet and replaced by the correct peak group, but this requires manual verification (Surinova et al., 2013). This automated processing tool is similar to the manual techniques performed here, however we do not normalise peak areas between biological replicates prior to absolute quantification and instead perform various manual verification checks to determine the peptides most suitable for quantification. Due to time constraints, the data generated here were not reanalysed using MSstats, although doing so may improve the number of 'A1' peptides identified in each condition. In instances where multiple peaks can occur within a single retention time window, a second tool SRM collider (available at http://www.srmcollider.org/srmcollider/srmcollider.py) may be used to determine potentially interfering transitions in a given proteomic background, and can be used to elucidate the true peptide peak (Rost et al., 2012). In my own experiments, multiple peaks in a single window did not often occur, and in such instances the standard peptide could be used to identify the true peak by means of coelution.

In absolute quantification strategies, it is crucial to ensure that both standard and analyte peptides are digested to completion. If the surrogate peptides generated from the protein lead to signal split into multiple overlapping peptides due to missed cleavage, the signal will be attenuated and quantification underestimated. Literature and experiments have proven that the presence of dibasic sequences and/or the presence of a glutamine or aspartic acid residues downstream of the cleavage site is common in peptides with missed cleavages. During the selection phase of the Q-peptides, peptides with dibasic sequences were removed as potential surrogates, however peptides *N*-terminal glutamic acid and aspartic acid residues could be selected – in hindsight this should have been a selection feature. Despite this, I found very few peptides (6 %) not to be digested to completion. To avoid this issue, a relatively recent development in QconCAT technology includes the use of spacer peptides that contain

sequences equivalent to those surrounding the terminal peptide cleavage sites in the analyte protein, such that the efficiency of proteolysis between the standard and analyte is expected to be equivalent and spectral data comparable (Cheung et al., 2015). However, Cheung and colleagues (Cheung et al., 2015) have noted that flanking sequences shorter than six residues can cause quantitative error due to the random appearance of other amino acid residues in close proximity to cleavage sites, resulting in unpredictable consequences for the digestion rates of QconCATs. The flanking sequences can also occupy a large proportion of the QconCAT in addition to the Q-peptides, resulting in an increase in cost as more QconCATs have to be synthesised to target all the proteins of interest. In the approach used here, spacer peptides were not included due to cost and consistency issues; we wanted the data to be consistent and comparable with the previous QconCAT experiments. Despite the best efforts of prediction tools, a small number of peptides can still be miscleaved as standard or analyte, and so should be identified and removed prior to performing absolute quantification. Whilst missed cleavages were not seen to be a major issue, it should become common practice in absolute quantification strategies for the digestion process to be monitored; in this way, the completion of digestion of both the standard and analyte peptide is known and quantifications made are robust, lending increased confidence to the data.

A previous study using QconCATs to target chaperones resulted in quantification of 51 of the 63 chaperones (using CopyCATs) under NG conditions. However, due to expression issues with the ChapCATs, it was only possible to determine absolute values (or upper limits) for 49 chaperones supporting comparisons with values in the previous study, discussed in the following section.

## 8.2 Comparison to a previous chaperone QconCAT study

Whilst both the CopyCAT study and this study used the QconCAT strategy to determine absolute copy numbers per cell for the known chaperones in *S. cerevisiae*, a small number of significant differences in the methodology were present. These could contribute to changes in the absolute values under NG conditions obtained for any Q-peptides that were used in both datasets and for the absolute abundances of the chaperones common to both datasets. Firstly, in the current approach, the *S. cerevisiae* were grown under batch conditions rather than as steady-state (chemostat) cultures. Batch grown cultures encounter a variable growth rate due to changing environment, whilst chemostat cultures remain steadily controlled at a predetermined single growth rate. Changes in growth rates are known to affect protein and/or

transcript levels, with proteins involved in the stress response reported to be down-regulated following an upshift in growth rate (Castrillo et al., 2007, Airoldi et al., 2016). It has been reported previously that cells grown under batch conditions encounter an environmental stress response upon entering the stationary phase due to exhaustion of nutrients (Saldanha et al., 2004, Gasch and Werner-Washburne, 2002), which does not occur in chemostat. Some of the differences observed are thus likely to due to differences in growth conditions. Secondly, the original CopyCAT study used a maximum of two Q-peptides per chaperone protein to determine cpc, whilst up to five Q-peptides were considered in this approach. This likely resulted in changes to observed median cpc values and attendant rCV values. Where protein abundances do not agree well, the majority have a higher cpc value under chemostat conditions which is likely a result of carbon-limitation in the chemostat condition as well as the slower (but consistent) growth rate (Castrillo et al., 2007, Stone et al., 1990). As an example, Hsp12, known to become upregulated in nutrient limitation and slow growth (Stone et al., 1990), is quantified as 364,319 and 68,598 cpc under chemostat and batch NG conditions respectively, quantified by 2 Q-peptides in both instances. Of these two Q-peptides, LNDAVEYVSGR was used in both chemostat and batch datasets, with a cpc value of 319,003 and 72,998 cpc respectively, demonstrating at the peptide level this difference in abundance.

Using the ChapCATs and following the quality control steps as discussed, it was possible to determine the absolute abundances of 40 chaperones under conditions of NG and HS, discussed further in the next section.

## 8.3 Quantification of chaperones under conditions of NG

When grouped according to chaperone functional class (Gong et al., 2009), Hsp90 chaperones have the lowest total cpc value (15,000 cpc, 2.1 % of the total cpc determined), whilst Hsp70 chaperones are the most abundant and occupy the largest total cpc value (411,000 cpc, 56.5 % of the total cpc determined). An important caveat to note is that the two unsuccessful ChapCAT constructs contained Hsp40 class chaperones; therefore, only six of the 24 Hsp40 chaperones were quantified here. Despite this, in agreement with epitope tagging studies (Ghaemmaghami et al., 2003) Zuo1, Sis1 and Ydj1 were determined to be the most abundant of the Hsp40 chaperones quantified (21,000; 11,000 and 19,000 respectively), with the quantified Hsp40 chaperones making up 8.1 % of the total chaperone cpc determined in NG conditions. A major benefit of absolute quantification is the ability to define the relative levels of total intracellular protein for members of interacting proteins, especially in complexes. For

example, Zuo1 is known to interact with the Hsp70 chaperone Ssz1 as part of the RAC complex and is able to recruit Ssb1 and Ssb2 to mediate co-translational folding of nascent chains. This data suggests a stoichiometry 12:8:33:25:1 (Zuo1:Ssz1:Ssb1:Ssb2:Jjj1). This is unsurprising; Zuo1 is required for the ATPase activity of more than one chaperone, whilst Ssz1 only requires Zuo1. Ssb1 and Ssb2 on the other hand are also able to utilise additional Hsp40 co-chaperones (including Jjj1) and so are much higher in abundance, therefore a 1:1 stoichiometry between Zuo1 and its interacting Hsp70 chaperones is not required.

Chaperones that make up the subunits of the PFD and CCT complex have been reported as having 1:1 stoichiometry in the native complex (with the structures solved using electron microscopy) (Siegert et al., 2000). According to their cpc values under NG conditions, PFD chaperones had the approximate cellular stoichiometry 3:3:1:2:2:1 (Gim1:Gim3:Gim4:Gim5:Pac10:Pfd1). However, these chaperones had rCVs above 20 %, with a maximum of 35 % for Pfd1. This suggested sub-optimal agreement between the peptides from the same parent chaperone. A similar observation was made with chaperones of the CCT complex, with Cct2 being the least abundant and having the cellular stoichiometry 1:2:3:2:2:1:2:2 (Cct2:Cct3:Cct4:Cct5:Cct6:Cct7:Cct8:Tcp1). Cct4, the highest abundant CCT chaperone (7800 cpc), also had the highest rCV of 60 %, determined using 4 'A1' peptides. A closer look at these peptides demonstrated a disagreement in cpc. Two peptides (GANNMIIDETER and IVSQYSSFLAPLAVDSVLK) had a median of approximately 10,000 cpc whilst the remaining Q-peptides (GLGCKPIADIELFTEDR and SLHDALCVIR) had a median of 5000 cpc. Despite my stringent quality control steps, no reason was found to remove GANNMIIDETER and IVSQYSSFLAPLAVDSVLK as 'A1' peptides, and subsequently resulted in a higher Cct4 cpc value with a higher rCV. In comparison, Cct8 was targeted by 5 'A1' peptides, having an rCV of 19 % with all Q-peptides having a median cpc of ~6000. Due to the nature of these high rCVs it was difficult to ascertain whether these results better reflect the true cellular stiochiometry, but disparities between the stoichiometry of chaperones within the fully formed complexes and the stoichiometry of chaperone abundance within the cell may be a result of additional levels of chaperone being present within the cytosol that is not yet part of the CCT/PFD complexes. Even so, there is little suggestion for other roles of these chaperones and so to synthesise or degrade the chaperones in unequal amounts would be sub-optimal for the cell.

For the Hsp90 chaperones, only Hsp82 was quantifiable via 'A1' peptides under NG conditions, whilst Hsc82 was below the limit of detection. This is surprising as Hsc82 is reported as constitutively expressed and higher in abundance than Hsp82 under NG conditions, whilst

Hsp82 was deemed highly inducible following stress (Borkovich et al., 1989). In the previous experiment, Hsc82 was detected by western blotting using a polyclonal antibody specific to Hsc82 and Hsp82, with yeast samples grown at 25 °C. In my approach, yeast samples were grown at 30 °C with absolute cpc determined using a standard identical to the unique peptides in Hsc82 and Hsp82. The difference in temperature may be enough to permit higher levels of Hsp82 when compared to Hsc82, whilst lower temperatures have this abundance relationship inverted – Hsc82 is more abundant than Hsp82. In addition, my cultures are batch grown and would have inevitably encountered fluctuating growth rates and a potential entry into stationary phase due to the sampling methods, a higher Hsp82 abundance may be attributed to these changes. At 30 °C, I was able to define the upper limit for Hsc82 using the 'B1' peptide, HSEFVAYPIQLLVTK, as ~700 cpc, whilst Hsp82 is highly abundant at 15,000 cpc.

Of the AAA+ chaperones, I observed cytosolic Hsp104 and mitochondrial Hsp78 to be present in almost equal abundances (11,000 cpc). I found mitochondrial Mcx1 to be the lowest abundant AAA+ chaperone, with a cpc of ~700. Although both Hsp78 and Mcx1 are localised to the mitochondrial matrix, Mcx1 does not have proteolytic function and does not fulfil unique function under NG conditions, indicating that other mitochondrial chaperone proteins (including Hsp78 and Ssc1) are able to substitute for a loss of Mcx1 (van Dyck et al., 1998).

With regards to the small class chaperones, I was unable to quantify the chaperones Hsp32, Sno4 and Hsp33, with their summed cpc values having an upper limit of approximately 700 cpc. Under NG conditions, Hsp26 is reported as an inactive oligomer unable mediate the folding of substrate proteins. However, I found Hsp26 to be present at 30,000 cpc, making it the second most abundant small class chaperone in *S. cerevisiae*, however I am unable to comment on its activity. Interestingly, Hsp42, reported as the main sHsp in *S. cerevisiae* under NG conditions (Haslbeck et al., 2004) was present almost three times lower than Hsp26 at 11,000 cpc. Instead I found the chaperone Hsp12 as the most abundant sHsp in NG conditions, with its reported induction in response to entry into the stationary phase (Welker et al., 2010). In this approach, *S. cerevisiae* were grown under batch conditions, and therefore likely encountered variable growth rates due to changing environment. It has been previously reported that cells grown under batch conditions encounter an environmental stress response upon entering the stationary phase due to exhaustion of nutrients (Saldanha et al., 2004, Gasch and Werner-Washburne, 2002). Although I harvested cells 30 minutes after reaching an $OD_{600}$ of 2.0 in the late-log phase, the environmental stress response may have already been triggered at this point, allowing for an increase in the abundance of Hsp12 and even Hsp26. This does not

affect the conclusions made on the chaperone response to heat shock as both NG and HS samples were grown in batch and harvested at the same time point. I discuss the chaperone, proteome and chaperome responses to heat shock in the following section.

## 8.4 The response to HS

The response to HS can be characterised in three ways from studies performed here. Firstly, by determining absolute quantification of the chaperones, I have been able to characterise the chaperone response to heat shock. Secondly, through my SRM-normalisation approaches for which 1644 protein modcpc values could be calculated under conditions of NG and HS I have been able to assess the proteome response, albiet this does not cover all proteins in *S. cerevisiae*. Finally, using a combination of a high quality chaperone-client interaction dataset and modcpc values for client proteins, I have been able to model the chaperome response to HS. The conclusions and considerations for each are discussed here.

### 8.4.1 The chaperone response to HS

*Sections of this discussion were originally published in the journal 'Proteomics': Mackenzie, R. J., Lawless, C., Holman, S. W., Lanthaler, K., Beynon, R. J., Grant, C. M., Hubbard, S. J. & Eyers, C. E. 2016. Absolute protein quantification of the yeast chaperome under conditions of heat shock. Proteomics DOI: 10.1002/pmic.201500503.*

One would expect the majority of chaperones to become significantly upregulated in response to HS as the level of unfolded protein and thus demand on the chaperone increases as a result of stress. Somewhat surprisingly, only thirteen chaperones of the 49 targeted were found to become significantly upregulated following HS, with a median fold change of 3.3. The proteins found to be significantly upregulated are further discussed here.

In particular, Ssz1 of the RAC complex was significantly upregulated (1.75 fold change, $p = 0.01$) whilst its Hsp40 co-chaperone for the formation of the RAC complex is not detectable under HS. In addition, the CLIPs chaperones Gim4 and Gim5 abundance levels were significantly upregulated at the protein level though not at the mRNA expression levels (Gasch et al., 2000). As previously discussed in this thesis, mRNA levels and protein abundance levels tend to correlate but often only modestly or incompletely; post-translational and post-transcriptional processing pathways are expected to be responsible. As such, if mRNA levels were to become

significantly decreased, for the same change to be observed at the protein level extended periods of time and additional processing routes may be required.

It was previously indicated that upon stress, cells act to down-regulate protein translation at the ribosome in order to reduce nascent chain folding demand on the chaperone machinery (Morano et al., 2012). Indeed, Zuo1, a major player of the RAC, was not quantifiable under HS conditions with only a single 'B2' peptide observed that failed to pass a 1 % FDR. Within the RAC, Zuo1 is in complex with the Hsp70 Ssz1, with the latter observed under HS at 25,700 cpc. The relative lack of Zuo1 indicates a reduced ability of RAC to form and promote co-translational folding at the ribosome. It also suggests Zuo1 may be function-limiting and acting as the regulatory factor. However, with Ssz1 levels still significant, one may speculate Ssz1 participates predominantly in post-translational folding of misfolded protein rather than directed towards nascent chain folding after heat shock.

Gim4 and Gim5 are both significantly upregulated in response to HS (8.6-fold and 7.2 fold respectively). Both are subunits of the PFD complex, crucial for mediating the folding of actin and tubulin by CCT in the cytosol. The six PFD complex subunits are in a one:one ratio, with two subunits being α-like and four subunits being β-like. Therefore, it is surprising to observe significant upregulation of only two subunits in response to HS, neither of which are known stress-induced targets of Hsf1.

The ability for the cell to survive heat shock and ensure protection against further environmental stress is largely dependent on its ability to maintain proteostasis. Failure to remove misfolded and/or aggregated protein via refolding mediated by chaperones may lead to their sequestration in designated protein quality control (PQC) foci or inclusion bodies to prevent cytotoxicity prior to their degradation. As anticipated, I successfully observed significant upregulation of chaperones known to be involved in the HSR. Of the Hsp70s, I observed very low levels of Ssa4 (1,800 cpc) in NG conditions, with significant upregulation of ~7-fold (to 14,000 cpc) observed in HS. The related heat-inducible homolog Ssa3 was quantified only in HS (at 2,000 cpc), with the native peptide undetected in NG (considered as a class 'B' Q-peptide), consistent with up-regulation under HS. With regards to the Hsp70 chaperones of the ER, Lhs1 increased (1.9 fold, $p = 0.02$) whilst levels of Kar2, a direct target of Hsf1, was not significantly upregulated. Ssa2, previously reported to be non-heat inducible (Ellwood and Craig, 1984, Lindquist and Craig, 1988), was observed at 226,000 cpc under conditions of HS, almost 2-fold higher than the levels found under NG. However, due to issues with incomplete

proteolysis (as discussed in Chapter 5) and the necessary removal of some Q-peptides as quantification standards rendered this apparent fold-change non-significant.

The Hsp40 chaperone, Sis1, a known co-chaperone of the Hsp70 chaperone homologs Ssa1 and Ssa2, was significantly up-regulated over 2-fold to 27,000 cpc. Whilst I observe Sis1 and Ydj1 to be the most abundant HSP40 chaperones (of those targeted) excepting Zuo1 under NG conditions, their cpc values were much lower (11,000 and 19,000 cpc respectively) in comparison to values reported in the literature determined via TAP-tagging and quantitative western blotting approaches (20,300 and 119,000 cpc respectively), albeit for a different yeast strain (BY4741) (Ghaemmaghami et al., 2003). Sis1 is able to stimulate the ATPase activity of Hsp70 chaperones, shuttling substrates between the cytosol and nucleus. Sis1 has also been linked to targeting of misfolded substrates to the PQC-degradation system and a role of protection from prion toxicity alongside the AAA+ chaperone Hsp104 (Shiber et al., 2013, Glover and Lindquist, 1998, Shorter and Lindquist, 2008, Summers et al., 2013). In agreement with previous findings, the AAA+ chaperone Hsp104 was low under NG conditions (11,000 cpc) but increased significantly to 42,000 cpc (4-fold, $p = 8.94 \times 10\text{-}7$) following exposure to HS. As discussed in Chapter 1, Hsp104 functions in a complementary role to the water-displacing molecule trehalose, stabilising proteins at physiological concentrations (Lindquist and Kim, 1996, Verghese et al., 2012). Hsp104 is able to co-operate with Ydj1 and Ssa1 to refold previously denatured proteins that have become aggregated (Bosl et al., 2006). Unlike conventional chaperones, Hsp104 functions specifically to dissociate aggregates that have formed due to overloaded cellular chaperone capacity, freely localising to, and removing those proteins that are terminally misfolded and contained within the perivacuolar insoluble protein deposit and juxtanuclear compartments (the PQC foci) (Bosl et al., 2006, Verghese et al., 2012).

Literature suggests marked down-regulation of the Hsp70 class ribosome-associated chaperones Ssb1 and Ssb2 in response to heat stress, inferred from mRNA abundances, albeit at prolonged times and varying temperatures (Craig and Jacobsen, 1985, Lopez et al., 1999). However, at the protein level, no significant difference for these chaperones was observed; both Ssb1 and Ssb2 were present in HS at 69,000 cpc, compared to 58,000 and 49,000 cpc in NG for Ssb1 and Ssb2 respectively. Given that the correlation observed between mRNA levels and protein abundances is generally modest, particularly under transitions associated with stress (Liu et al., 2016), these results suggest post-transcriptional regulation is in play. mRNA half-lives are typically shorter than those of proteins, so a decrease in mRNA abundance may not be reflected immediately at the protein level. Conceivably, a reduction in Ssb1 and Ssb2

protein levels may be observed upon prolonged heat shock conditions, and a study comparing the absolute copy per cell numbers of chaperones in response to heat shock across various *S. cerevisiae* strains would give further insight, however, this is not the focus of this thesis.

Of the Hsp90 family, Hsp82, which functions in the final stages of protein folding and protein complex assembly receiving client proteins from Hsp70 chaperones via the co-chaperone Sti1, (Verghese et al., 2012, O'Connell et al., 2014) was significantly upregulated three-fold to 50,000 cpc in response to HS. Similarly, Hsc82 cpc levels were also elevated in HS, being quantified only under these conditions (akin to Ssa3), at 36,000 cpc. This supports previous studies that reported modest increases in Hsc82 levels in response to heat shock compared to larger Hsp82 changes (Borkovich et al., 1989). Although Hsp82 is thought to work with Ssa1/Ssa2 to deactivate Hsf1 following cellular recovery (Bonner et al., 2000), there was no significant heat stress-induced change in these co-regulators, in contrast to the observations for Hsp82.

Three of the four 'small' class chaperones quantified under both conditions (Hsp12, Hsp26 and Hsp42) were observed to be significantly upregulated. Hsp26 exhibited the greatest significant change following HS, with its absolute abundance increasing to 235,000 cpc, almost an 8-fold increase over that under NG conditions. Although significant, the ~2-fold up-regulation of Hsp42 was notably lower than that of Hsp12 and Hsp26, whose levels increased 3.8-fold and 7.7-fold in response to HS respectively. These observations were consistent with previous studies, which suggest that the majority of small heat shock protein family members, Hsp42 aside, are functionally inactive under NG conditions (Haslbeck et al., 2004). Small class chaperones are able to form large oligomeric complexes containing unfolded protein within their hollow structures upon stress, thus protecting from aggregation until folding can occur.

Using SRM-normalisation, label free MaxLFQ intensities were converted into a measure of absolute abundance for 1644 proteins under conditions of NG and HS. Although in *S. cerevisiae* there exists in excess of 4000 proteins present in the cell, data for only 1644 were obtained via our protocol, presumably representing most of the more abundant proteins. The next section discusses the proteome in context of these 1644 proteins. This number could be improved through MS analysis of the samples using fractionation to improve the chances of identifying lower abundance proteins. As data dependent acquisitions were used, ions from peptides that are more abundant are more routinely identified and observed above background noise than those belonging to low abundant proteins. Therefore, the proteins that were

determined have absolute mod-cpc values that are likely to represent the majority of the total proteome in response to HS. Interestingly, more proteins were quantified under HS (quantifying 1671 and 1816 proteins in NG and HS respectively). In the consideration of time, this was not further investigated during my studies. I discuss the 1644 proteins common to NG and HS conditions further here.

### 8.4.2 The proteome response to heat shock

Whilst I observed no change in cell size or cellular protein concentration, I observed a 1.18 fold change in terms of total mod-cpc protein abundance in response to heat shock. On a local scale I observed few proteins with significant differences that might contribute to this change. Due to the nature of my modelling approach, I defined significance as a fold change +/- 2-fold. I identified five proteins (Hsp26, Ssa4, Hsp104, Rgi1 and Hsp42) with a fold change over two in response to HS, whilst only Elo2 becomes significantly downregulated under the same conditions, exhibiting a fold change of less than 0.5. In addition to the previously identified Hsp26, Hsp42, Hsp104, and Ssa4, Rgi1 (Respiratory Growth Induced protein 1) also becomes upregulated (2.2–fold) upon HS. Although the precise function of Rgi1 has yet to be elucidated, it appears to have a role in regulating energy metabolism and drug resistance, with high expression levels reported under a wide range of conditions, inclusive of high temperatures, cold stress and the unfolded protein response (Domitrovic et al., 2010, Gasch et al., 2000, Travers et al., 2000). In contrast, Elo2, a fatty acid elongase localised to the endoplasmic reticulum, becomes downregulated 0.5-fold. Elo2 is involved in sphingolipid biosynthesis (essential components of membranes and thus important for cellular integrity) and transport from the late endosome to the vacuole as part of the secretory pathway (David et al., 1998). Its paralog Elo1, was not identified in the label free analysis.

The observation that the proteome does not exhibit significant changes globally is not surprising. During the HSR, the cell would attempt to maintain homeostasis such that the global protein abundance would remain constant. As such, one would expect very few significant changes. However, this mod-cpc analysis does highlight a limitation with normalisation in large scale proteomics; normalising both NG and HS median MaxLFQ intensities may result in under-representation of true significant changes that are occurring. Here, I was able to identify only those that are on the extreme ends of regulation. Despite this, I was able to provide protein-level evidence, in terms of absolute copies per cell quantification, of the yeast cell's ability to maintain overall proteostasis and the ability to adapt to heat shock conditions.

I was able to further investigate the role of the chaperones in response to HS by modelling their protein volume and workload using client proteins modcpc values as determined by SRM-normalisation of label free MaxLFQ intensities. I discuss these findings further in the next section.

### 8.4.3 The chaperome response to heat shock

Using modelling approaches, I was able to identify the protein volume of Hsc82, Hsp82 and Ssb2 significantly upregulated in response to HS. In addition, the workload of Ssb2 was significantly upregulated in response. Of these chaperones, Hsc82 was not quantifiable under NG conditions via the QconCAT strategy, and so signficance could not be assessed. However, all proteins were in abundance in HS conditions, with Hsp82 significantly upregulated in terms of their abundance, and with Hsc82 quantified, upregulation can be suggested. Interestingly, Ssb2 was not deemed significantly upregulated in terms of its absolute abundance in response to heat shock. I do not discuss the biological relevance of Ssb2 and Hsp82 here as they have been discussed in the previous 'chaperone response to heat shock' section. However, there were many assumptions made during this modelling phase that should be taken into consideration that I wish to discuss further here.

Firstly, although a wide range of interaction data exists, collected via epitope-tagging strategies; such data is subject to a high number of false positives through non-specific protein and contaminants binding to the epitope tag. Therefore, I created a high quality dataset using interactions listed in three online interaction repositories containing the interaction data from multiple published experiments. This interaction dataset consisted of 60 chaperones and 1412 client proteins, a stark contrast to the 4340 client proteins identified via TAP tagging by Gong and colleagues (Gong et al., 2009). A major difference here is that for my high quality datasets, only interactions that also had a reciprocal interaction reported in one of the repositories is considered a true interaction. Often in published experiments, the interactions are identified through tagging of a protein of interest, such as is done in TAP; if the client protein is not a targeted protein and is not abundant, it is unlikely to be identified in many experiments as an interactor. In addition, many interactions are reported if the gene for the protein is co-expressed with another. The STRING repository provides a means of assessing the likelihood of the interaction being true by using a normalised scoring techinque, with each score assigned according to the experimental method used – whether the interaction is defined by association

or indirect interaction, and the number of times the interaction is reported, corrected for the probability of randomly observing an interaction (von Mering et al., 2005). To this end, I accepted only interactions with a STRING score over 0.7 as potential interactions in keeping with previous high quality interaction filters (Brownridge et al., 2013).

A major assumption made during these studies is that the client network for a chaperone is identical irrespective of conditions. Chaperones are able to play multiple roles within the life cycle of a protein, from co/post-translational folding to mediation of transportation of a protein across a membrane to degradation and removal. If a chaperone plays a transportation role under NG conditions, but plays a role in recovery from misfolding or degradation of proteins in response to stress, it is highly likely that its clientele will differ between the two conditions. Although few studies have identified interactions between a chaperone and a client protein (Romanova and Chernoff, 2009) under stressed conditions, there exists no comprehensive interaction dataset unique to a particular condition. A second consideration is the missing data observed in combining the SRM-normalised modcpc values for 1644 proteins and the 1412 clients. As such, only 635 client proteins had a modcpc value assigned. Therefore, it is assumed that the remaining 777 clients are low in abundance and any changes to these proteins in response to a condition would have little significance. Despite this, as the 635 clients had a modcpc value under both NG and HS conditions, I was able to consider the changes in chaperone protein volume based on these 635 clients only; however this may not be truly reflective of the entire chaperome and further work is required to increase the number of clients for which quantitiative data is present.

Furthermore, I encountered a similar issue with missing data for degradation rates (and thus synthesis rates) for each of the client proteins. Again, it was assumed that the degradation rate of a client is the equal under conditions of NG and HS. It has been previously reported that at the mRNA-level, HS results in a change in the mRNA stability and transcription rate (Castells-Roca et al., 2011). However, there appears to have been no study towards defining the synthesis or degradation rates of proteins under HS. Therefore, although I determined a significant increase in the modelled workload for the chaperone Ssb2 in response to HS, further work is required to prove this notion. I have summarised the further work perspective of these studies in the next section.

## 8.5 Further work perspectives

In order to create a condition-unique comprehensive interaction network complete with quantitative data, I propose TAP experiments to be performed on each of the chaperones of interest and the isolated clients quantified by MS analysis either by relative or absolute techniques. Although TAP experiments have been previously performed by Gong and colleagues (Gong et al., 2009), quantitative data was lacking and was subject to samples from non stressed conditions only. TAP experimentation does bear an issue with contaminants and false positives, however bioinformatics tools are available to curate this data.

In addition to this, studies regarding determination of condition-dependent protein degradation rates are required before being able to formally assess the workload of a chaperone under different conditions. Previous experiments by Christiano and colleagues compared the degradation rates of proteins between yeast strains, however were focussed upon NG conditions only.

Finally, an experiment that has not been performed here but would highlight the role of a chaperone in response to HS in terms of misfolding, is one that assesses the extent of known client misfolding in particular conditions. Therefore, a chaperome network may be divided into functionally distinct nodes: those that are folding protein for recovery from the HS, and those that are removing protein to allow for cellular recovery. Feng and colleagues (Feng et al., 2014) have developed the limited proteolysis SRM workflow to this end. Firstly, proteins are digested under nondenaturing conditions, such that the sites of initial proteolysis are dictated by the structure properties of the substrate protein resulting in large protein fragments. The sample is then subject to denaturing conditions resulting in shorter peptide fragments but with unique termini. The same original sample is subject to denaturing proteolysis, and both digests are run on LC-MS/MS platforms to identify protein structural differences by virtue of the peptide sequences identified. Further experiments involving the cross-linking of chaperones and their client proteins will also allow for the identification of a true chaperone-client interaction during the specified condition.

## References


ACAMPORA, G. & HERMANS, J., JR. 1967. Reversible denaturation of sperm whale myoglobin. I. Dependence on temperature, pH, and composition. *J Am Chem Soc,* 89**,** 1543-7.

AIROLDI, E. M., MILLER, D., ATHANASIADOU, R., BRANDT, N., ABDUL-RAHMAN, F., NEYMOTIN, B., HASHIMOTO, T., BAHMANI, T. & GRESHAM, D. 2016. Steady-state and dynamic gene expression programs in Saccharomyces cerevisiae in response to variation in environmental nitrogen. *Mol Biol Cell,* 27**,** 1383-96.

AKERFELT, M., MORIMOTO, R. I. & SISTONEN, L. 2010. Heat shock factors: integrators of cell stress, development and lifespan. *Nat Rev Mol Cell Biol,* 11**,** 545-55.

ALBAN, A., DAVID, S. O., BJORKESTEN, L., ANDERSSON, C., SLOGE, E., LEWIS, S. & CURRIE, I. 2003. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics,* 3**,** 36-44.

ALBANESE, V., YAM, A. Y., BAUGHMAN, J., PARNOT, C. & FRYDMAN, J. 2006. Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. *Cell,* 124**,** 75-88.

ANCKAR, J. & SISTONEN, L. 2011. Regulation of HSF1 function in the heat stress response: implications in aging and disease. *Annu Rev Biochem,* 80**,** 1089-115.

ANDRECHT, S. & VON HAGEN, J. 2008. General Aspects of Sample Preparation for Comprehensive Proteome Analysis. *In:* VON HAGEN, J. (ed.) *Proteomics Sample Preparation.* Weinheim: WILEY-VCH.

ANFINSEN, C. B. 1973. Principles that govern the folding of protein chains. *Science,* 181**,** 223-30.

ARNDT, V., ROGON, C. & HOHFELD, J. 2007. To be, or not to be--molecular chaperones in protein degradation. *Cell Mol Life Sci,* 64**,** 2525-41.

BALLINGER, C. A., CONNELL, P., WU, Y., HU, Z., THOMPSON, L. J., YIN, L. Y. & PATTERSON, C. 1999. Identification of CHIP, a novel tetratricopeptide repeat-containing protein that interacts with heat shock proteins and negatively regulates chaperone functions. *Mol Cell Biol,* 19**,** 4535-45.

BALOGH, G., HORVATH, I., NAGY, E., HOYK, Z., BENKO, S., BENSAUDE, O. & VIGH, L. 2005. The hyperfluidization of mammalian cell membranes acts as a signal to initiate the heat shock protein response. *FEBS J,* 272**,** 6077-86.

BARBOUR, J., WEISE, S., MEYER, H., E. & WARSCHEID, B. 2008. Mass Spectrometry. *In:* VON HAGEN, J. (ed.) *Proteomics Samples Preparation.* Weinheim: WILEY-VCH.

BECK, F., BURKHART, J. M., GEIGER, J., ZAHEDI, R. P. & SICKMANN, A. 2012. Robust workflow for iTRAQ-based peptide and protein quantification. *Methods Mol Biol,* 893**,** 101-13.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57**,** 289-300.

BEYNON, R. J., DOHERTY, M. K., PRATT, J. M. & GASKELL, S. J. 2005. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods,* 2**,** 587-9.

BOCHTLER, M., HARTMANN, C., SONG, H. K., BOURENKOV, G. P., BARTUNIK, H. D. & HUBER, R. 2000. The structures of HsIU and the ATP-dependent protease HsIU-HsIV. *Nature,* 403**,** 800-5.

BONNER, J. J., CARLSON, T., FACKENTHAL, D. L., PADDOCK, D., STOREY, K. & LEA, K. 2000. Complex regulation of the yeast heat shock transcription factor. *Mol Biol Cell,* 11**,** 1739-51.

BOOTH, C. R., MEYER, A. S., CONG, Y., TOPF, M., SALI, A., LUDTKE, S. J., CHIU, W. & FRYDMAN, J. 2008. Mechanism of lid closure in the eukaryotic chaperonin TRiC/CCT. *Nat Struct Mol Biol,* 15**,** 746-53.

BORKOVICH, K. A., FARRELLY, F. W., FINKELSTEIN, D. B., TAULIEN, J. & LINDQUIST, S. 1989. hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol,* 9**,** 3919-30.

BOSL, B., GRIMMINGER, V. & WALTER, S. 2006. The molecular chaperone Hsp104--a molecular machine for protein disaggregation. *J Struct Biol,* 156**,** 139-48.

BROMBERG, Z., GOLOUBINOFF, P., SAIDI, Y. & WEISS, Y. G. 2013. The membrane-associated transient receptor potential vanilloid channel is the central heat shock receptor controlling the cellular heat shock response in epithelial cells. *PLoS One,* 8**,** e57149.

BROWNRIDGE, P. & BEYNON, R. J. 2011. The importance of the digest: proteolysis and absolute quantification in proteomics. *Methods,* 54**,** 351-60.

BROWNRIDGE, P., LAWLESS, C., PAYAPILLY, A. B., LANTHALER, K., HOLMAN, S. W., HARMAN, V. M., GRANT, C. M., BEYNON, R. J. & HUBBARD, S. J. 2013. Quantitative analysis of chaperone network throughput in budding yeast. *Proteomics,* 13**,** 1276-91.

BROWNRIDGE, P. J., HARMAN, V. M., SIMPSON, D. M. & BEYNON, R. J. 2012. Absolute multiplexed protein quantification using QconCAT technology. *Methods Mol Biol,* 893**,** 267-93.

BRUN, V., DUPUIS, A., ADRAIT, A., MARCELLIN, M., THOMAS, D., COURT, M., VANDENESCH, F. & GARIN, J. 2007. Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics,* 6**,** 2139-49.

BUKAU, B. & HORWICH, A. L. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell,* 92**,** 351-66.

CABRITA, L. D., DOBSON, C. M. & CHRISTODOULOU, J. 2010. Protein folding on the ribosome. *Curr Opin Struct Biol,* 20**,** 33-45.

CALLONI, G., CHEN, T., SCHERMANN, S. M., CHANG, H. C., GENEVAUX, P., AGOSTINI, F., TARTAGLIA, G. G., HAYER-HARTL, M. & HARTL, F. U. 2012. DnaK functions as a central hub in the E. coli chaperone network. *Cell Rep,* 1**,** 251-64.

CAPLAN, A. J., TSAI, J., CASEY, P. J. & DOUGLAS, M. G. 1992. Farnesylation of YDJ1p is required for function at elevated growth temperatures in Saccharomyces cerevisiae. *J Biol Chem,* 267**,** 18890-5.

CARROLL, K. M., SIMPSON, D. M., EYERS, C. E., KNIGHT, C. G., BROWNRIDGE, P., DUNN, W. B., WINDER, C. L., LANTHALER, K., PIR, P., MALYS, N., KELL, D. B., OLIVER, S. G., GASKELL, S. J. & BEYNON, R. J. 2011. Absolute quantification of the glycolytic pathway in yeast: deployment of a complete QconCAT approach. *Mol Cell Proteomics,* 10**,** M111 007633.

CASTELLS-ROCA, L., GARCIA-MARTINEZ, J., MORENO, J., HERRERO, E., BELLI, G. & PEREZ-ORTIN, J. E. 2011. Heat shock response in yeast involves changes in both transcription rates and mRNA stabilities. *PLoS One,* 6**,** e17272.

CASTRILLO, J. I., ZEEF, L. A., HOYLE, D. C., ZHANG, N., HAYES, A., GARDNER, D. C., CORNELL, M. J., PETTY, J., HAKES, L., WARDLEWORTH, L., RASH, B., BROWN, M., DUNN, W. B., BROADHURST, D., O'DONOGHUE, K., HESTER, S. S., DUNKLEY, T. P., HART, S. R., SWAINSTON, N., LI, P., GASKELL, S. J., PATON, N. W., LILLEY, K. S., KELL, D. B. & OLIVER, S. G. 2007. Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol,* 6**,** 4.

CAUSTON, H. C., REN, B., KOH, S. S., HARBISON, C. T., KANIN, E., JENNINGS, E. G., LEE, T. I., TRUE, H. L., LANDER, E. S. & YOUNG, R. A. 2001. Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell,* 12**,** 323-37.

CHENG, M. Y., HARTL, F. U., MARTIN, J., POLLOCK, R. A., KALOUSEK, F., NEUPERT, W., HALLBERG, E. M., HALLBERG, R. L. & HORWICH, A. L. 1989. Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature,* 337**,** 620-5.

CHERNOFF, Y. O., LINDQUIST, S. L., ONO, B., INGE-VECHTOMOV, S. G. & LIEBMAN, S. W. 1995. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. *Science,* 268**,** 880-4.

CHEUNG, C. S., ANDERSON, K. W., WANG, M. & TURKO, I. V. 2015. Natural flanking sequences for peptides included in a quantification concatamer internal standard. *Anal Chem,* 87**,** 1097-102.

CHOE, L., D'ASCENZO, M., RELKIN, N. R., PAPPIN, D., ROSS, P., WILLIAMSON, B., GUERTIN, S., PRIBIL, P. & LEE, K. H. 2007. 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics,* 7**,** 3651-60.

CHOI, H., LARSEN, B., LIN, Z. Y., BREITKREUTZ, A., MELLACHERUVU, D., FERMIN, D., QIN, Z. S., TYERS, M., GINGRAS, A. C. & NESVIZHSKII, A. I. 2011. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods,* 8**,** 70-3.

CHOI, M., CHANG, C. Y., CLOUGH, T., BROUDY, D., KILLEEN, T., MACLEAN, B. & VITEK, O. 2014. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics,* 30**,** 2524-6.

CHRISTIANO, R., NAGARAJ, N., FROHLICH, F. & WALTHER, T. C. 2014. Global proteome turnover analyses of the Yeasts S. cerevisiae and S. pombe. *Cell Rep,* 9**,** 1959-65.

CONFALONIERI, F. & DUGUET, M. 1995. A 200-amino acid ATPase module in search of a basic function. *Bioessays,* 17**,** 639-50.

CONNELL, P., BALLINGER, C. A., JIANG, J., WU, Y., THOMPSON, L. J., HOHFELD, J. & PATTERSON, C. 2001. The co-chaperone CHIP regulates protein triage decisions mediated by heat-shock proteins. *Nat Cell Biol,* 3**,** 93-6.

COX, J., HEIN, M. Y., LUBER, C. A., PARON, I., NAGARAJ, N. & MANN, M. 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics,* 13**,** 2513-26.

COX, J. & MANN, M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol,* 26**,** 1367-72.

COX, J. & MANN, M. 2009. Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap. *J Am Soc Mass Spectrom,* 20**,** 1477-85.

COX, J., NEUHAUSER, N., MICHALSKI, A., SCHELTEMA, R. A., OLSEN, J. V. & MANN, M. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res,* 10**,** 1794-805.

CRAIG, E. A. & JACOBSEN, K. 1985. Mutations in cognate genes of Saccharomyces cerevisiae hsp70 result in reduced growth rates at low temperatures. *Mol Cell Biol,* 5**,** 3517-24.

CROWE, J. H. 2007. Trehalose as a "chemical chaperone": fact and fantasy. *Adv Exp Med Biol,* 594**,** 143-58.

CUELLAR, J., MARTIN-BENITO, J., SCHERES, S. H., SOUSA, R., MORO, F., LOPEZ-VINAS, E., GOMEZ-PUERTAS, P., MUGA, A., CARRASCOSA, J. L. & VALPUESTA, J. M. 2008. The structure of CCT-Hsc70 NBD suggests a mechanism for Hsp70 delivery of substrates to the chaperonin. *Nat Struct Mol Biol,* 15**,** 858-64.

DAVID, D., SUNDARABABU, S. & GERST, J. E. 1998. Involvement of long chain fatty acid elongation in the trafficking of secretory vesicles in yeast. *J Cell Biol,* 143**,** 1167-82.

DE GODOY, L. M., OLSEN, J. V., COX, J., NIELSEN, M. L., HUBNER, N. C., FROHLICH, F., WALTHER, T. C. & MANN, M. 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature,* 455**,** 1251-4.

DE MARCO, A., VIGH, L., DIAMANT, S. & GOLOUBINOFF, P. 2005. Native folding of aggregation-prone recombinant proteins in Escherichia coli by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress Chaperones,* 10**,** 329-39.

DEKKER, C., STIRLING, P. C., MCCORMACK, E. A., FILMORE, H., PAUL, A., BROST, R. L., COSTANZO, M., BOONE, C., LEROUX, M. R. & WILLISON, K. R. 2008. The interaction network of the chaperonin CCT. *EMBO J,* 27**,** 1827-39.

DEL ALAMO, M., HOGAN, D. J., PECHMANN, S., ALBANESE, V., BROWN, P. O. & FRYDMAN, J. 2011. Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. *PLoS Biol,* 9**,** e1001100.

DEMAND, J., ALBERTI, S., PATTERSON, C. & HOHFELD, J. 2001. Cooperation of a ubiquitin domain protein and an E3 ubiquitin ligase during chaperone/proteasome coupling. *Curr Biol,* 11**,** 1569-77.

DESIERE, F., DEUTSCH, E. W., KING, N. L., NESVIZHSKII, A. I., MALLICK, P., ENG, J., CHEN, S., EDDES, J., LOEVENICH, S. N. & AEBERSOLD, R. 2006. The PeptideAtlas project. *Nucleic Acids Res,* 34**,** D655-8.

DEUTSCH, E. W. 2010. The PeptideAtlas Project. *Methods Mol Biol,* 604**,** 285-96.

DEUTSCH, E. W., LAM, H. & AEBERSOLD, R. 2008. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep,* 9**,** 429-34.

DOBSON, C. M. 2001. Protein folding and its links with human disease. *Biochem Soc Symp***,** 1-26.

DOBSON, C. M. 2003. Protein folding and misfolding. *Nature,* 426**,** 884-90.

DOMITROVIC, T., KOZLOV, G., FREIRE, J. C., MASUDA, C. A., DA SILVA ALMEIDA, M., MONTERO-LOMELI, M., ATELLA, G. C., MATTA-CAMACHO, E., GEHRING, K. & KURTENBACH, E. 2010. Structural and functional study of YER067W, a new protein involved in yeast metabolism control and drug resistance. *PLoS One,* 5**,** e11163.

DOYLE, S. M., SHORTER, J., ZOLKIEWSKI, M., HOSKINS, J. R., LINDQUIST, S. & WICKNER, S. 2007. Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity. *Nat Struct Mol Biol,* 14**,** 114-22.

ELIAS, J. E. & GYGI, S. P. 2010. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol Biol,* 604**,** 55-71.

ELLIS, R. J. 2001a. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol,* 11**,** 114-9.

ELLIS, R. J. 2001b. Molecular chaperones: inside and outside the Anfinsen cage. *Curr Biol,* 11**,** R1038-40.

ELLIS, R. J. & MINTON, A. P. 2006. Protein aggregation in crowded environments. *Biol Chem,* 387**,** 485-97.

ELLWOOD, M. S. & CRAIG, E. A. 1984. Differential regulation of the 70K heat shock gene and related genes in Saccharomyces cerevisiae. *Mol Cell Biol,* 4**,** 1454-9.

ENGLANDER, S. W. & MAYNE, L. 2014. The nature of protein folding pathways. *Proc Natl Acad Sci U S A,* 111**,** 15873-80.

ETCHELLS, S. A., MEYER, A. S., YAM, A. Y., ROOBOL, A., MIAO, Y., SHAO, Y., CARDEN, M. J., SKACH, W. R., FRYDMAN, J. & JOHNSON, A. E. 2005. The cotranslational contacts between ribosome-bound nascent polypeptides and the subunits of the hetero-oligomeric chaperonin TRiC probed by photocross-linking. *J Biol Chem,* 280**,** 28118-26.

EYERS, C. E., LAWLESS, C., WEDGE, D. C., LAU, K. W., GASKELL, S. J. & HUBBARD, S. J. 2011. CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. *Mol Cell Proteomics,* 10**,** M110 003384.

FALSONE, S. F., KUNGL, A. J., REK, A., CAPPAI, R. & ZANGGER, K. 2009. The molecular chaperone Hsp90 modulates intermediate steps of amyloid assembly of the Parkinson-related protein alpha-synuclein. *J Biol Chem,* 284**,** 31190-9.

FENG, Y., DE FRANCESCHI, G., KAHRAMAN, A., SOSTE, M., MELNIK, A., BOERSEMA, P. J., DE LAURETO, P. P., NIKOLAEV, Y., OLIVEIRA, A. P. & PICOTTI, P. 2014. Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol,* 32**,** 1036-44.

FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 246**,** 64-71.

FINKA, A., SOOD, V., QUADRONI, M., RIOS PDE, L. & GOLOUBINOFF, P. 2015. Quantitative proteomics of heat-treated human cells show an across-the-board mild depletion of housekeeping proteins to massively accumulate few HSPs. *Cell Stress Chaperones,* 20**,** 605-20.

FORSYTHE, H. L., JARVIS, J. L., TURNER, J. W., ELMORE, L. W. & HOLT, S. E. 2001. Stable association of hsp90 and p23, but Not hsp70, with active human telomerase. *J Biol Chem,* 276**,** 15571-4.

FRANZMANN, T. M., WUHR, M., RICHTER, K., WALTER, S. & BUCHNER, J. 2005. The activation mechanism of Hsp26 does not require dissociation of the oligomer. *J Mol Biol,* 350**,** 1083-93.

FRYDMAN, J., NIMMESGERN, E., OHTSUKA, K. & HARTL, F. U. 1994. Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature,* 370**,** 111-7.

FUJIWARA, K., ISHIHAMA, Y., NAKAHIGASHI, K., SOGA, T. & TAGUCHI, H. 2010. A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J,* 29**,** 1552-64.

GALL, W. E., HIGGINBOTHAM, M. A., CHEN, C., INGRAM, M. F., CYR, D. M. & GRAHAM, T. R. 2000. The auxilin-like phosphoprotein Swa2p is required for clathrin function in yeast. *Curr Biol,* 10**,** 1349-58.

GASCH, A. P., HUANG, M., METZNER, S., BOTSTEIN, D., ELLEDGE, S. J. & BROWN, P. O. 2001. Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol Biol Cell,* 12**,** 2987-3003.

GASCH, A. P., SPELLMAN, P. T., KAO, C. M., CARMEL-HAREL, O., EISEN, M. B., STORZ, G., BOTSTEIN, D. & BROWN, P. O. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell,* 11**,** 4241-57.

GASCH, A. P. & WERNER-WASHBURNE, M. 2002. The genomics of yeast responses to environmental stress and starvation. *Funct Integr Genomics,* 2**,** 181-92.

GHAEMMAGHAMI, S., HUH, W. K., BOWER, K., HOWSON, R. W., BELLE, A., DEPHOURE, N., O'SHEA, E. K. & WEISSMAN, J. S. 2003. Global analysis of protein expression in yeast. *Nature,* 425**,** 737-41.

GLOVER, J. R. & LINDQUIST, S. 1998. Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell,* 94**,** 73-82.

GLOVER, J. R. & TKACH, J. M. 2001. Crowbars and ratchets: hsp100 chaperones as tools in reversing protein aggregation. *Biochem Cell Biol,* 79**,** 557-68.

GONG, Y., KAKIHARA, Y., KROGAN, N., GREENBLATT, J., EMILI, A., ZHANG, Z. & HOURY, W. A. 2009. An atlas of chaperone-protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways in the cell. *Mol Syst Biol,* 5**,** 275.

GREENE, M. K., MASKOS, K. & LANDRY, S. J. 1998. Role of the J-domain in the cooperation of Hsp40 with Hsp70. *Proc Natl Acad Sci U S A,* 95**,** 6108-13.

GROUSL, T., IVANOV, P., FRYDLOVA, I., VASICOVA, P., JANDA, F., VOJTOVA, J., MALINSKA, K., MALCOVA, I., NOVAKOVA, L., JANOSKOVA, D., VALASEK, L. & HASEK, J. 2009. Robust heat shock induces eIF2alpha-phosphorylation-independent assembly of stress granules containing eIF3 and 40S ribosomal subunits in budding yeast, Saccharomyces cerevisiae. *J Cell Sci,* 122**,** 2078-88.

GUILHAUS, M. 1995. Principles and Instrumentation in Time-of-Flight Mass-Spectrometry - Physical and Instrumental Concepts. *Journal of Mass Spectrometry,* 30**,** 1519-1532.

HAHN, J. S., HU, Z., THIELE, D. J. & IYER, V. R. 2004. Genome-wide analysis of the biology of stress responses through heat shock transcription factor. *Mol Cell Biol,* 24**,** 5249-56.

HARTL, F. U. 1996. Molecular chaperones in cellular protein folding. *Nature,* 381**,** 571-9.

HARTL, F. U. & HAYER-HARTL, M. 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science,* 295**,** 1852-8.

HASHIKAWA, N., MIZUKAMI, Y., IMAZU, H. & SAKURAI, H. 2006. Mutated yeast heat shock transcription factor activates transcription independently of hyperphosphorylation. *J Biol Chem,* 281**,** 3936-42.

HASLBECK, M., BRAUN, N., STROMER, T., RICHTER, B., MODEL, N., WEINKAUF, S. & BUCHNER, J. 2004. Hsp42 is the general small heat shock protein in the cytosol of Saccharomyces cerevisiae. *EMBO J,* 23**,** 638-49.

HASLBECK, M., FRANZMANN, T., WEINFURTNER, D. & BUCHNER, J. 2005. Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol,* 12**,** 842-6.

HASLBECK, M., WALKE, S., STROMER, T., EHRNSPERGER, M., WHITE, H. E., CHEN, S., SAIBIL, H. R. & BUCHNER, J. 1999. Hsp26: a temperature-regulated chaperone. *EMBO J,* 18**,** 6744-51.

HATTENDORF, D. A. & LINDQUIST, S. L. 2002. Cooperative kinetics of both Hsp104 ATPase domains and interdomain communication revealed by AAA sensor-1 mutants. *EMBO J,* 21**,** 12-21.

HAYES, A., ZHANG, N., WU, J., BUTLER, P. R., HAUSER, N. C., HOHEISEL, J. D., LIM, F. L., SHARROCKS, A. D. & OLIVER, S. G. 2002. Hybridization array technology coupled with chemostat culture: Tools to interrogate gene expression in Saccharomyces cerevisiae. *Methods,* 26**,** 281-90.

HEBERT, A. S., MERRILL, A. E., BAILEY, D. J., STILL, A. J., WESTPHALL, M. S., STRIETER, E. R., PAGLIARINI, D. J. & COON, J. J. 2013. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Methods,* 10**,** 332-4.

HERMJAKOB, H., MONTECCHI-PALAZZI, L., LEWINGTON, C., MUDALI, S., KERRIEN, S., ORCHARD, S., VINGRON, M., ROECHERT, B., ROEPSTORFF, P., VALENCIA, A., MARGALIT, H., ARMSTRONG, J., BAIROCH, A., CESARENI, G., SHERMAN, D. & APWEILER, R. 2004. IntAct: an open source molecular interaction database. *Nucleic Acids Res,* 32**,** D452-5.

HOLMAN, S. W., SIMS, P. F. & EYERS, C. E. 2012. The use of selected reaction monitoring in quantitative proteomics. *Bioanalysis,* 4**,** 1763-86.

HORWICH, A. L., FENTON, W. A., CHAPMAN, E. & FARR, G. W. 2007. Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol,* 23**,** 115-45.

HUANG, P., GAUTSCHI, M., WALTER, W., ROSPERT, S. & CRAIG, E. A. 2005. The Hsp70 Ssz1 modulates the function of the ribosome-associated J-protein Zuo1. *Nat Struct Mol Biol,* 12**,** 497-504.

ISHIHAMA, Y., ODA, Y., TABATA, T., SATO, T., NAGASU, T., RAPPSILBER, J. & MANN, M. 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics,* 4**,** 1265-72.

ITO, T., CHIBA, T., OZAWA, R., YOSHIDA, M., HATTORI, M. & SAKAKI, Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A,* 98**,** 4569-74.

JAKOB, U., GAESTEL, M., ENGEL, K. & BUCHNER, J. 1993. Small heat shock proteins are molecular chaperones. *J Biol Chem,* 268**,** 1517-20.

JARNUCZAK, A. F., EYERS, C. E., SCHWARTZ, J. M., GRANT, C. M. & HUBBARD, S. J. 2015. Quantitative proteomics and network analysis of SSA1 and SSB1 deletion mutants reveals robustness of chaperone HSP70 network in Saccharomyces cerevisiae. *Proteomics,* 15**,** 3126-39.

JIANG, J., MAES, E. G., TAYLOR, A. B., WANG, L., HINCK, A. P., LAFER, E. M. & SOUSA, R. 2007. Structural basis of J cochaperone binding and regulation of Hsp70. *Mol Cell,* 28**,** 422-33.

JOHNSON, J. L. 2012. Evolution and function of diverse Hsp90 homologs and cochaperone proteins. *Biochim Biophys Acta,* 1823**,** 607-13.

JOHNSON, J. L. & CRAIG, E. A. 2001. An essential role for the substrate-binding region of Hsp40s in Saccharomyces cerevisiae. *J Cell Biol,* 152**,** 851-6.

JONES, A. R., SIEPEN, J. A., HUBBARD, S. J. & PATON, N. W. 2009. Improving sensitivity in proteome studies by analysis of false discovery rates for multiple search engines. *Proteomics,* 9**,** 1220-9.

JUN, H., KIESELBACH, T. & JONSSON, L. J. 2012. Comparative proteome analysis of Saccharomyces cerevisiae: a global overview of in vivo targets of the yeast activator protein 1. *BMC Genomics,* 13**,** 230.

KARP, N. A., HUBER, W., SADOWSKI, P. G., CHARLES, P. D., HESTER, S. V. & LILLEY, K. S. 2010. Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics,* 9**,** 1885-97.

KEICH, U., KERTESZ-FARKAS, A. & NOBLE, W. S. 2015. Improved False Discovery Rate Estimation Procedure for Shotgun Proteomics. *J Proteome Res,* 14**,** 3148-61.

KENNAWAY, C. K., BENESCH, J. L., GOHLKE, U., WANG, L., ROBINSON, C. V., ORLOVA, E. V., SAIBIL, H. R. & KEEP, N. H. 2005. Dodecameric structure of the small heat shock protein Acr1 from Mycobacterium tuberculosis. *J Biol Chem,* 280**,** 33419-25.

KERNER, M. J., NAYLOR, D. J., ISHIHAMA, Y., MAIER, T., CHANG, H. C., STINES, A. P., GEORGOPOULOS, C., FRISHMAN, D., HAYER-HARTL, M., MANN, M. & HARTL, F. U. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell,* 122**,** 209-20.

KIM, K. K., KIM, R. & KIM, S. H. 1998. Crystal structure of a small heat-shock protein. *Nature,* 394**,** 595-9.

KIM, M. S. & PANDEY, A. 2012. Electron transfer dissociation mass spectrometry in proteomics. *Proteomics,* 12**,** 530-42.

KIM, Y. E., HIPP, M. S., BRACHER, A., HAYER-HARTL, M. & HARTL, F. U. 2013. Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem,* 82**,** 323-55.

KIRKPATRICK, D. S., GERBER, S. A. & GYGI, S. P. 2005. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods,* 35**,** 265-73.

KOLL, H., GUIARD, B., RASSOW, J., OSTERMANN, J., HORWICH, A. L., NEUPERT, W. & HARTL, F. U. 1992. Antifolding activity of hsp60 couples protein import into the mitochondrial matrix with export to the intermembrane space. *Cell,* 68**,** 1163-75.

LANGER, T., LU, C., ECHOLS, H., FLANAGAN, J., HAYER, M. K. & HARTL, F. U. 1992. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature,* 356**,** 683-9.

LAWLESS, C., HOLMAN, S. W., BROWNRIDGE, P., LANTHALER, K., HARMAN, V. M., WATKINS, R., HAMMOND, D. E., MILLER, R. L., SIMS, P. F., GRANT, C. M., EYERS, C. E., BEYNON, R. J. & HUBBARD, S. J. 2016. Direct and Absolute Quantification of over 1800 Yeast Proteins via Selected Reaction Monitoring. *Mol Cell Proteomics,* 15**,** 1309-22.

LAWLESS, C. & HUBBARD, J. S. 2014. Analysis of Chaperone Network Throughput. *In:* HOURY, A. W. (ed.) *The Molecular Chaperones Interaction Networks in Protein Folding and Degradation.* New York, NY: Springer New York.

LAWLESS, C. & HUBBARD, S. J. 2012. Prediction of missed proteolytic cleavages for the selection of surrogate peptides for quantitative proteomics. *OMICS,* 16**,** 449-56.

LEE, T. Y., HUANG, H. D., HUNG, J. H., HUANG, H. Y., YANG, Y. S. & WANG, T. H. 2006. dbPTM: an information repository of protein post-translational modification. *Nucleic Acids Res,* 34**,** D622-7.

LEHTINEN, S., MARSELLACH, F. X., CODLIN, S., SCHMIDT, A., CLEMENT-ZIZA, M., BEYER, A., BAHLER, J., ORENGO, C. & PANCALDI, V. 2013. Stress induces remodelling of yeast interaction and co-expression networks. *Mol Biosyst,* 9**,** 1697-707.

LEITNER, A., JOACHIMIAK, L. A., BRACHER, A., MONKEMEYER, L., WALZTHOENI, T., CHEN, B., PECHMANN, S., HOLMES, S., CONG, Y., MA, B., LUDTKE, S., CHIU, W., HARTL, F. U., AEBERSOLD, R. & FRYDMAN, J. 2012. The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure,* 20**,** 814-25.

LEONHARDT, S. A., FEARSON, K., DANESE, P. N. & MASON, T. L. 1993. HSP78 encodes a yeast mitochondrial heat shock protein in the Clp family of ATP-dependent proteases. *Mol Cell Biol,* 13**,** 6304-13.

LEROUX, M. R., FANDRICH, M., KLUNKER, D., SIEGERS, K., LUPAS, A. N., BROWN, J. R., SCHIEBEL, E., DOBSON, C. M. & HARTL, F. U. 1999. MtGimC, a novel archaeal chaperone related to the eukaryotic chaperonin cofactor GimC/prefoldin. *EMBO J,* 18**,** 6730-43.

LI, J., SOROKA, J. & BUCHNER, J. 2012. The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim Biophys Acta,* 1823**,** 624-35.

LINDQUIST, S. & CRAIG, E. A. 1988. The heat-shock proteins. *Annu Rev Genet,* 22**,** 631-77.

LINDQUIST, S. & KIM, G. 1996. Heat-shock protein 104 expression is sufficient for thermotolerance in yeast. *Proc Natl Acad Sci U S A,* 93**,** 5301-6.

LIU, H., SADYGOV, R. G. & YATES, J. R., 3RD 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem,* 76**,** 4193-201.

LIU, Y., BEYER, A. & AEBERSOLD, R. 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell,* 165**,** 535-50.

LLORCA, O., MARTIN-BENITO, J., RITCO-VONSOVICI, M., GRANTHAM, J., HYNES, G. M., WILLISON, K. R., CARRASCOSA, J. L. & VALPUESTA, J. M. 2000. Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO J,* 19**,** 5971-9.

LLORCA, O., MCCORMACK, E. A., HYNES, G., GRANTHAM, J., CORDELL, J., CARRASCOSA, J. L., WILLISON, K. R., FERNANDEZ, J. J. & VALPUESTA, J. M. 1999. Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature,* 402**,** 693-6.

LOPEZ, N., HALLADAY, J., WALTER, W. & CRAIG, E. A. 1999. SSB, encoding a ribosome-associated chaperone, is coordinately regulated with ribosomal protein genes. *J Bacteriol,* 181**,** 3136-43.

LUM, R., TKACH, J. M., VIERLING, E. & GLOVER, J. R. 2004. Evidence for an unfolding/threading mechanism for protein disaggregation by Saccharomyces cerevisiae Hsp104. *J Biol Chem,* 279**,** 29139-46.

MACLEAN, B., TOMAZELA, D. M., SHULMAN, N., CHAMBERS, M., FINNEY, G. L., FREWEN, B., KERN, R., TABB, D. L., LIEBLER, D. C. & MACCOSS, M. J. 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics,* 26**,** 966-8.

MARCH, R. E. & TODD, J. F. J. 2005. Theory of Quadrupole Instruments. *In:* J., H. N. (ed.) *Quadrupole Ion Trap Mass Spectrometry.* Second ed. USA: John Wiley & Sons.

MARTIN-BENITO, J., BOSKOVIC, J., GOMEZ-PUERTAS, P., CARRASCOSA, J. L., SIMONS, C. T., LEWIS, S. A., BARTOLINI, F., COWAN, N. J. & VALPUESTA, J. M. 2002. Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO J,* 21**,** 6377-86.

MARTINEZ-PASTOR, M. T., MARCHLER, G., SCHULLER, C., MARCHLER-BAUER, A., RUIS, H. & ESTRUCH, F. 1996. The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO J,* 15**,** 2227-35.

MCALISTER, G. C., NUSINOW, D. P., JEDRYCHOWSKI, M. P., WUHR, M., HUTTLIN, E. L., ERICKSON, B. K., RAD, R., HAAS, W. & GYGI, S. P. 2014. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem,* 86**,** 7150-8.

MCCLELLAN, A. J., XIA, Y., DEUTSCHBAUER, A. M., DAVIS, R. W., GERSTEIN, M. & FRYDMAN, J. 2007. Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell,* 131**,** 121-35.

MELLACHERUVU, D., WRIGHT, Z., COUZENS, A. L., LAMBERT, J. P., ST-DENIS, N. A., LI, T., MITEVA, Y. V., HAURI, S., SARDIU, M. E., LOW, T. Y., HALIM, V. A., BAGSHAW, R. D., HUBNER, N. C., AL-HAKIM, A., BOUCHARD, A., FAUBERT, D., FERMIN, D., DUNHAM, W. H., GOUDREAULT, M., LIN, Z. Y., BADILLO, B. G., PAWSON, T., DUROCHER, D., COULOMBE, B., AEBERSOLD, R., SUPERTI-FURGA, G., COLINGE, J., HECK, A. J., CHOI, H., GSTAIGER, M., MOHAMMED, S., CRISTEA, I. M., BENNETT, K. L., WASHBURN, M. P., RAUGHT, B., EWING, R. M., GINGRAS, A. C. & NESVIZHSKII, A. I. 2013. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods,* 10**,** 730-6.

MENSONIDES, F. I., BRUL, S., KLIS, F. M., HELLINGWERF, K. J. & TEIXEIRA DE MATTOS, M. J. 2005. Activation of the protein kinase C1 pathway upon continuous heat stress in Saccharomyces cerevisiae is triggered by an intracellular increase in osmolarity due to trehalose accumulation. *Appl Environ Microbiol,* 71**,** 4531-8.

MERRILL, A. E., HEBERT, A. S., MACGILVRAY, M. E., ROSE, C. M., BAILEY, D. J., BRADLEY, J. C., WOOD, W. W., EL MASRI, M., WESTPHALL, M. S., GASCH, A. P. & COON, J. J. 2014. NeuCode labels for relative protein quantification. *Mol Cell Proteomics,* 13**,** 2503-12.

MEYER, A. E., HUNG, N. J., YANG, P., JOHNSON, A. W. & CRAIG, E. A. 2007. The specialized cytosolic J-protein, Jjj1, functions in 60S ribosomal subunit biogenesis. *Proc Natl Acad Sci U S A,* 104**,** 1558-63.

MIHALIK, A. & CSERMELY, P. 2011. Heat shock partially dissociates the overlapping modules of the yeast protein-protein interaction network: a systems level model of adaptation. *PLoS Comput Biol,* 7**,** e1002187.

MILLER-FLEMING, L., ANTAS, P., PAIS, T. F., SMALLEY, J. L., GIORGINI, F. & OUTEIRO, T. F. 2014. Yeast DJ-1 superfamily members are required for diauxic-shift reprogramming and cell survival in stationary phase. *Proc Natl Acad Sci U S A,* 111**,** 7012-7.

MILLER, P. E. & DENTON, M. B. 1986. The Quadrupole Mass Filter - Basic Operating Concepts. *Journal of Chemical Education,* 63**,** 617-622.

MORANO, K. A. 2007. New tricks for an old dog: the evolving world of Hsp70. *Ann N Y Acad Sci,* 1113**,** 1-14.

MORANO, K. A., GRANT, C. M. & MOYE-ROWLEY, W. S. 2012. The response to heat shock and oxidative stress in Saccharomyces cerevisiae. *Genetics,* 190**,** 1157-95.

MULLER, P., RUCKOVA, E., HALADA, P., COATES, P. J., HRSTKA, R., LANE, D. P. & VOJTESEK, B. 2013. C-terminal phosphorylation of Hsp70 and Hsp90 regulates alternate binding to co-chaperones CHIP and HOP to determine cellular protein folding/degradation balances. *Oncogene,* 32**,** 3101-10.

MURATA, S., MINAMI, Y., MINAMI, M., CHIBA, T. & TANAKA, K. 2001. CHIP is a chaperone-dependent E3 ligase that ubiquitylates unfolded protein. *EMBO Rep,* 2**,** 1133-8.

MURPHY, M. P. 2009. How mitochondria produce reactive oxygen species. *Biochem J,* 417**,** 1-13.

NARBERHAUS, F. 2002. Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol Mol Biol Rev,* 66**,** 64-93; table of contents.

NATHAN, D. F., VOS, M. H. & LINDQUIST, S. 1997. In vivo functions of the Saccharomyces cerevisiae Hsp90 chaperone. *Proc Natl Acad Sci U S A,* 94**,** 12949-56.

NEUWALD, A. F., ARAVIND, L., SPOUGE, J. L. & KOONIN, E. V. 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res,* 9**,** 27-43.

NIKOLOV, M., SCHMIDT, C. & URLAUB, H. 2012. Quantitative mass spectrometry-based proteomics: an overview. *Methods Mol Biol,* 893**,** 85-100.

O'CONNELL, J. D., TSECHANSKY, M., ROYALL, A., BOUTZ, D. R., ELLINGTON, A. D. & MARCOTTE, E. M. 2014. A proteomic survey of widespread protein aggregation in yeast. *Mol Biosyst,* 10**,** 851-61.

OH, E., BECKER, A. H., SANDIKCI, A., HUBER, D., CHABA, R., GLOGE, F., NICHOLS, R. J., TYPAS, A., GROSS, C. A., KRAMER, G., WEISSMAN, J. S. & BUKAU, B. 2011. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell,* 147**,** 1295-308.

ONG, S. E., BLAGOEV, B., KRATCHMAROVA, I., KRISTENSEN, D. B., STEEN, H., PANDEY, A. & MANN, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics,* 1**,** 376-86.

PARKER, R. & SHETH, U. 2007. P bodies and the control of mRNA translation and degradation. *Mol Cell,* 25**,** 635-46.

PARSELL, D. A., KOWAL, A. S., SINGER, M. A. & LINDQUIST, S. 1994. Protein disaggregation mediated by heat-shock protein Hsp104. *Nature,* 372**,** 475-8.

PATEL, S. & LATTERICH, M. 1998. The AAA team: related ATPases with diverse functions. *Trends Cell Biol,* 8**,** 65-71.

PERKINS, D. N., PAPPIN, D. J., CREASY, D. M. & COTTRELL, J. S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis,* 20**,** 3551-67.

PERRY, R. H., COOKS, R. G. & NOLL, R. J. 2008. Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications. *Mass Spectrometry Reviews,* 27**,** 661-699.

PFUND, C., LOPEZ-HOYO, N., ZIEGELHOFFER, T., SCHILKE, B. A., LOPEZ-BUESA, P., WALTER, W. A., WIEDMANN, M. & CRAIG, E. A. 1998. The molecular chaperone Ssb from Saccharomyces cerevisiae is a component of the ribosome-nascent chain complex. *EMBO J,* 17**,** 3981-9.

PICOTTI, P., BODENMILLER, B., MUELLER, L. N., DOMON, B. & AEBERSOLD, R. 2009. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell,* 138**,** 795-806.

PIPER, P. W., TALREJA, K., PANARETOU, B., MORADAS-FERREIRA, P., BYRNE, K., PRAEKELT, U. M., MEACOCK, P., RECNACQ, M. & BOUCHERIE, H. 1994. Induction of major heat-shock proteins of Saccharomyces cerevisiae, including plasma membrane Hsp30, by ethanol levels above a critical threshold. *Microbiology,* 140 ( Pt 11)**,** 3031-8.

PRODROMOU, C., SILIGARDI, G., O'BRIEN, R., WOOLFSON, D. N., REGAN, L., PANARETOU, B., LADBURY, J. E., PIPER, P. W. & PEARL, L. H. 1999. Regulation of Hsp90 ATPase activity by tetratricopeptide repeat (TPR)-domain co-chaperones. *EMBO J,* 18**,** 754-62.

QIAN, S. B., MCDONOUGH, H., BOELLMANN, F., CYR, D. M. & PATTERSON, C. 2006. CHIP-mediated stress recovery by sequential ubiquitination of substrates and Hsp70. *Nature,* 440**,** 551-5.

RABILLOUD, T. 2012. The whereabouts of 2D gels in quantitative proteomics. *Methods Mol Biol,* 893**,** 25-35.

REITER, L., RINNER, O., PICOTTI, P., HUTTENHAIN, R., BECK, M., BRUSNIAK, M. Y., HENGARTNER, M. O. & AEBERSOLD, R. 2011. mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods,* 8**,** 430-5.

RIVERS, J., SIMPSON, D. M., ROBERTSON, D. H., GASKELL, S. J. & BEYNON, R. J. 2007. Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol Cell Proteomics,* 6**,** 1416-27.

ROMANOVA, N. V. & CHERNOFF, Y. O. 2009. Hsp104 and prion propagation. *Protein Pept Lett,* 16**,** 598-605.

RONSEIN, G. E., PAMIR, N., VON HALLER, P. D., KIM, D. S., ODA, M. N., JARVIK, G. P., VAISAR, T. & HEINECKE, J. W. 2015. Parallel reaction monitoring (PRM) and selected reaction

monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J Proteomics,* 113**,** 388-99.

ROSENBERGER, G., LUDWIG, C., ROST, H. L., AEBERSOLD, R. & MALMSTROM, L. 2014. aLFQ: an R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. *Bioinformatics,* 30**,** 2511-3.

ROSS, P. L., HUANG, Y. N., MARCHESE, J. N., WILLIAMSON, B., PARKER, K., HATTAN, S., KHAINOVSKI, N., PILLAI, S., DEY, S., DANIELS, S., PURKAYASTHA, S., JUHASZ, P., MARTIN, S., BARTLET-JONES, M., HE, F., JACOBSON, A. & PAPPIN, D. J. 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics,* 3**,** 1154-69.

ROST, H., MALMSTROM, L. & AEBERSOLD, R. 2012. A computational tool to detect and avoid redundancy in selected reaction monitoring. *Mol Cell Proteomics,* 11**,** 540-9.

ROTTGERS, K., ZUFALL, N., GUIARD, B. & VOOS, W. 2002. The ClpB homolog Hsp78 is required for the efficient degradation of proteins in the mitochondrial matrix. *J Biol Chem,* 277**,** 45829-37.

ROWLEY, A., JOHNSTON, G. C., BUTLER, B., WERNER-WASHBURNE, M. & SINGER, R. A. 1993. Heat shock-mediated cell cycle blockage and G1 cyclin expression in the yeast Saccharomyces cerevisiae. *Mol Cell Biol,* 13**,** 1034-41.

SAAVEDRA, C., TUNG, K. S., AMBERG, D. C., HOPPER, A. K. & COLE, C. N. 1996. Regulation of mRNA export in response to stress in Saccharomyces cerevisiae. *Genes Dev,* 10**,** 1608-20.

SALDANHA, A. J., BRAUER, M. J. & BOTSTEIN, D. 2004. Nutritional homeostasis in batch and steady-state culture of yeast. *Mol Biol Cell,* 15**,** 4089-104.

SANCHEZ, Y. & LINDQUIST, S. L. 1990. HSP104 required for induced thermotolerance. *Science,* 248**,** 1112-5.

SCHAUPP, A., MARCINOWSKI, M., GRIMMINGER, V., BOSL, B. & WALTER, S. 2007. Processing of proteins by the molecular chaperone Hsp104. *J Mol Biol,* 370**,** 674-86.

SCHECHTER, I. & BERGER, A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun,* 27**,** 157-62.

SCHELTEMA, R. A., HAUSCHILD, J. P., LANGE, O., HORNBURG, D., DENISOV, E., DAMOC, E., KUEHN, A., MAKAROV, A. & MANN, M. 2014. The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol Cell Proteomics,* 13**,** 3698-708.

SCHIRMER, E. C., QUEITSCH, C., KOWAL, A. S., PARSELL, D. A. & LINDQUIST, S. 1998. The ATPase activity of Hsp104, effects of environmental conditions and mutations. *J Biol Chem,* 273**,** 15546-52.

SCHMIDT, A., KOCHANOWSKI, K., VEDELAAR, S., AHRNE, E., VOLKMER, B., CALLIPO, L., KNOOPS, K., BAUER, M., AEBERSOLD, R. & HEINEMANN, M. 2016. The quantitative and condition-dependent Escherichia coli proteome. *Nat Biotechnol,* 34**,** 104-10.

SCHWANHAUSSER, B., BUSSE, D., LI, N., DITTMAR, G., SCHUCHHARDT, J., WOLF, J., CHEN, W. & SELBACH, M. 2011. Global quantification of mammalian gene expression control. *Nature,* 473**,** 337-42.

SHANER, L., WEGELE, H., BUCHNER, J. & MORANO, K. A. 2005. The yeast Hsp110 Sse1 functionally interacts with the Hsp70 chaperones Ssa and Ssb. *J Biol Chem,* 280**,** 41262-9.

SHI, Y., MOSSER, D. D. & MORIMOTO, R. I. 1998. Molecular chaperones as HSF1-specific transcriptional repressors. *Genes Dev,* 12**,** 654-66.

SHIBER, A., BREUER, W., BRANDEIS, M. & RAVID, T. 2013. Ubiquitin conjugation triggers misfolded protein sequestration into quality control foci when Hsp70 chaperone levels are limiting. *Mol Biol Cell,* 24**,** 2076-87.

SHORTER, J. & LINDQUIST, S. 2008. Hsp104, Hsp70 and Hsp40 interplay regulates formation, growth and elimination of Sup35 prions. *EMBO J,* 27**,** 2712-24.

SHUI, W., XIONG, Y., XIAO, W., QI, X., ZHANG, Y., LIN, Y., GUO, Y., ZHANG, Z., WANG, Q. & MA, Y. 2015. Understanding the Mechanism of Thermotolerance Distinct From Heat Shock Response Through Proteomic Analysis of Industrial Strains of Saccharomyces cerevisiae. *Mol Cell Proteomics,* 14**,** 1885-97.

SIEGERS, K., WALDMANN, T., LEROUX, M. R., GREIN, K., SHEVCHENKO, A., SCHIEBEL, E. & HARTL, F. U. 1999. Compartmentation of protein folding in vivo: sequestration of non-native polypeptide by the chaperonin-GimC system. *EMBO J,* 18**,** 75-84.

SIEGERT, R., LEROUX, M. R., SCHEUFLER, C., HARTL, F. U. & MOAREFI, I. 2000. Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell,* 103**,** 621-32.

SIEPEN, J. A., KEEVIL, E. J., KNIGHT, D. & HUBBARD, S. J. 2007. Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J Proteome Res,* 6**,** 399-408.

SILVA, J. C., DENNY, R., DORSCHEL, C., GORENSTEIN, M. V., LI, G. Z., RICHARDSON, K., WALL, D. & GEROMANOS, S. J. 2006a. Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale. *Mol Cell Proteomics,* 5**,** 589-607.

SILVA, J. C., GORENSTEIN, M. V., LI, G. Z., VISSERS, J. P. & GEROMANOS, S. J. 2006b. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics,* 5**,** 144-56.

SINGER, M. A. & LINDQUIST, S. 1998. Multiple effects of trehalose on protein folding in vitro and in vivo. *Mol Cell,* 1**,** 639-48.

SINGH, S. M. & PANDA, A. K. 2005. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng,* 99**,** 303-10.

SOBOTT, F., WATT, S. J., SMITH, J., EDELMANN, M. J., KRAMER, H. B. & KESSLER, B. M. 2009. Comparison of CID versus ETD based MS/MS fragmentation for the analysis of protein ubiquitination. *J Am Soc Mass Spectrom,* 20**,** 1652-9.

SOUSA, M. C., TRAME, C. B., TSURUTA, H., WILBANKS, S. M., REDDY, V. S. & MCKAY, D. B. 2000. Crystal and solution structures of an HslUV protease-chaperone complex. *Cell,* 103**,** 633-43.

STARK, C., BREITKREUTZ, B. J., REGULY, T., BOUCHER, L., BREITKREUTZ, A. & TYERS, M. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res,* 34**,** D535-9.

STEAD, D. A., PREECE, A. & BROWN, A. J. 2006. Universal metrics for quality assessment of protein identifications by mass spectrometry. *Mol Cell Proteomics,* 5**,** 1205-11.

STONE, R. L., MATARESE, V., MAGEE, B. B., MAGEE, P. T. & BERNLOHR, D. A. 1990. Cloning, sequencing and chromosomal assignment of a gene from Saccharomyces cerevisiae which is negatively regulated by glucose and positively by lipids. *Gene,* 96**,** 171-6.

STREET, T. O., LAVERY, L. A. & AGARD, D. A. 2011. Substrate binding drives large-scale conformational changes in the Hsp90 molecular chaperone. *Mol Cell,* 42**,** 96-105.

STROMER, T., EHRNSPERGER, M., GAESTEL, M. & BUCHNER, J. 2003. Analysis of the interaction of small heat shock proteins with unfolding proteins. *J Biol Chem,* 278**,** 18015-21.

SUMMERS, D. W., WOLFE, K. J., REN, H. Y. & CYR, D. M. 2013. The Type II Hsp40 Sis1 cooperates with Hsp70 and the E3 ligase Ubr1 to promote degradation of terminally misfolded cytosolic protein. *PLoS One,* 8**,** e52099.

SURINOVA, S., HUTTENHAIN, R., CHANG, C. Y., ESPONA, L., VITEK, O. & AEBERSOLD, R. 2013. Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. *Nat Protoc,* 8**,** 1602-19.

SYKA, J. E., COON, J. J., SCHROEDER, M. J., SHABANOWITZ, J. & HUNT, D. F. 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A,* 101**,** 9528-33.

SZKLARCZYK, D., FRANCESCHINI, A., WYDER, S., FORSLUND, K., HELLER, D., HUERTA-CEPAS, J., SIMONOVIC, M., ROTH, A., SANTOS, A., TSAFOU, K. P., KUHN, M., BORK, P., JENSEN, L. J. & VON MERING, C. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res,* 43**,** D447-52.

TAKEMORI, Y., ENOKI, Y., YAMAMOTO, N., FUKAI, Y., ADACHI, K. & SAKURAI, H. 2009. Mutational analysis of human heat-shock transcription factor 1 reveals a regulatory role for oligomerization in DNA-binding specificity. *Biochem J,* 424**,** 253-61.

TIAN, G., LEWIS, S. A., FEIERBACH, B., STEARNS, T., ROMMELAERE, H., AMPE, C. & COWAN, N. J. 1997. Tubulin subunits exist in an activated conformational state generated and maintained by protein cofactors. *J Cell Biol,* 138**,** 821-32.

TING, L., RAD, R., GYGI, S. P. & HAAS, W. 2011. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods,* 8**,** 937-40.

TRAVERS, K. J., PATIL, C. K., WODICKA, L., LOCKHART, D. J., WEISSMAN, J. S. & WALTER, P. 2000. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell,* 101**,** 249-58.

TROTTER, E. W., BERENFELD, L., KRAUSE, S. A., PETSKO, G. A. & GRAY, J. V. 2001. Protein misfolding and temperature up-shift cause G1 arrest via a common mechanism dependent on heat shock factor in Saccharomycescerevisiae. *Proc Natl Acad Sci U S A,* 98**,** 7313-8.

TROTTER, E. W., KAO, C. M., BERENFELD, L., BOTSTEIN, D., PETSKO, G. A. & GRAY, J. V. 2002. Misfolded proteins are competent to mediate a subset of the responses to heat shock in Saccharomyces cerevisiae. *J Biol Chem,* 277**,** 44817-25.

TU, B. P. & WEISSMAN, J. S. 2004. Oxidative protein folding in eukaryotes: mechanisms and consequences. *J Cell Biol,* 164**,** 341-6.

UNLU, M., MORGAN, M. E. & MINDEN, J. S. 1997. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis,* 18**,** 2071-7.

VABULAS, R. M., RAYCHAUDHURI, S., HAYER-HARTL, M. & HARTL, F. U. 2010. Protein folding in the cytoplasm and the heat shock response. *Cold Spring Harb Perspect Biol,* 2**,** a004390.

VAINBERG, I. E., LEWIS, S. A., ROMMELAERE, H., AMPE, C., VANDEKERCKHOVE, J., KLEIN, H. L. & COWAN, N. J. 1998. Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell,* 93**,** 863-73.

VAN DYCK, L., DEMBOWSKI, M., NEUPERT, W. & LANGER, T. 1998. Mcx1p, a ClpX homologue in mitochondria of Saccharomyces cerevisiae. *FEBS Lett,* 438**,** 250-4.

VAUDEL, M., BARSNES, H., BERVEN, F. S., SICKMANN, A. & MARTENS, L. 2011. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics,* 11**,** 996-9.

VERGHESE, J., ABRAMS, J., WANG, Y. & MORANO, K. A. 2012. Biology of the heat shock response and protein chaperones: budding yeast (Saccharomyces cerevisiae) as a model system. *Microbiol Mol Biol Rev,* 76**,** 115-58.

VOGEL, C. & MARCOTTE, E. M. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet,* 13**,** 227-32.

VOISINE, C., CRAIG, E. A., ZUFALL, N., VON AHSEN, O., PFANNER, N. & VOOS, W. 1999. The protein import motor of mitochondria: unfolding and trapping of preproteins are distinct and separable functions of matrix Hsp70. *Cell,* 97**,** 565-74.

VON MERING, C., JENSEN, L. J., SNEL, B., HOOPER, S. D., KRUPP, M., FOGLIERINI, M., JOUFFRE, N., HUYNEN, M. A. & BORK, P. 2005. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res,* 33**,** D433-7.

VOOS, W. & ROTTGERS, K. 2002. Molecular chaperones as essential mediators of mitochondrial biogenesis. *Biochim Biophys Acta,* 1592**,** 51-62.

WANG, J., SONG, J. J., FRANKLIN, M. C., KAMTEKAR, S., IM, Y. J., RHO, S. H., SEONG, I. S., LEE, C. S., CHUNG, C. H. & EOM, S. H. 2001a. Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure,* 9**,** 177-84.

WANG, J., SONG, J. J., SEONG, I. S., FRANKLIN, M. C., KAMTEKAR, S., EOM, S. H. & CHUNG, C. H. 2001b. Nucleotide-dependent conformational changes in a protease-associated ATPase HslU. *Structure,* 9**,** 1107-16.

WANG, M., WEISS, M., SIMONOVIC, M., HAERTINGER, G., SCHRIMPF, S. P., HENGARTNER, M. O. & VON MERING, C. 2012. PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics,* 11**,** 492-500.

WASHBURN, M. P., WOLTERS, D. & YATES, J. R., 3RD 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol,* 19**,** 242-7.

WELKER, S., RUDOLPH, B., FRENZEL, E., HAGN, F., LIEBISCH, G., SCHMITZ, G., SCHEURING, J., KERTH, A., BLUME, A., WEINKAUF, S., HASLBECK, M., KESSLER, H. & BUCHNER, J. 2010. Hsp12 is an intrinsically unstructured stress protein that folds upon membrane association and modulates membrane function. *Mol Cell,* 39**,** 507-20.

WENDLER, P., SHORTER, J., SNEAD, D., PLISSON, C., CLARE, D. K., LINDQUIST, S. & SAIBIL, H. R. 2009. Motor mechanism for protein threading through Hsp104. *Mol Cell,* 34**,** 81-92.

WENGER, C. D., LEE, M. V., HEBERT, A. S., MCALISTER, G. C., PHANSTIEL, D. H., WESTPHALL, M. S. & COON, J. J. 2011. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Methods,* 8**,** 933-5.

WIESER, R., ADAM, G., WAGNER, A., SCHULLER, C., MARCHLER, G., RUIS, H., KRAWIEC, Z. & BILINSKI, T. 1991. Heat shock factor-independent heat control of transcription of the CTT1 gene encoding the cytosolic catalase T of Saccharomyces cerevisiae. *J Biol Chem,* 266**,** 12406-11.

WILM, M. 2009. Quantitative proteomics in biological research. *Proteomics,* 9**,** 4590-605.

WISNIEWSKI, J. R., HEIN, M. Y., COX, J. & MANN, M. 2014. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol Cell Proteomics,* 13**,** 3497-506.

WISNIEWSKI, J. R., OSTASIEWICZ, P., DUS, K., ZIELINSKA, D. F., GNAD, F. & MANN, M. 2012. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol Syst Biol,* 8**,** 611.

WOTTON, D., FREEMAN, K. & SHORE, D. 1996. Multimerization of Hsp42p, a novel heat shock protein of Saccharomyces cerevisiae, is dependent on a conserved carboxyl-terminal sequence. *J Biol Chem,* 271**,** 2717-23.

XU, Z., HORWICH, A. L. & SIGLER, P. B. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature,* 388**,** 741-50.

YAM, A. Y., XIA, Y., LIN, H. T., BURLINGAME, A., GERSTEIN, M. & FRYDMAN, J. 2008. Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat Struct Mol Biol,* 15**,** 1255-62.

YAMAMOTO, A., MIZUKAMI, Y. & SAKURAI, H. 2005. Identification of a novel class of target genes and a novel type of binding sequence of heat shock transcription factor in Saccharomyces cerevisiae. *J Biol Chem,* 280**,** 11911-9.

YAMAMOTO, N., MAEDA, Y., IKEDA, A. & SAKURAI, H. 2008. Regulation of thermotolerance by stress-induced transcription factors in Saccharomyces cerevisiae. *Eukaryot Cell,* 7**,** 783-90.

YANG, W. Y. & GRUEBELE, M. 2003. Folding at the speed limit. *Nature,* 423**,** 193-7.

YOKOM, A. L., GATES, S. N., JACKREL, M. E., MACK, K. L., SU, M., SHORTER, J. & SOUTHWORTH, D. R. 2016. Spiral architecture of the Hsp104 disaggregase reveals the basis for polypeptide translocation. *Nat Struct Mol Biol,* 23**,** 830-7.

ZOU, J., GUO, Y., GUETTOUCHE, T., SMITH, D. F. & VOELLMY, R. 1998. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell,* 94**,** 471-80.

ZUEHLKE, A. & JOHNSON, J. L. 2010. Hsp90 and co-chaperones twist the functions of diverse client proteins. *Biopolymers,* 93**,** 211-7.

# Appendix 1 - Buffer solutions

Concentration, Substance/Solution, Company (if required)

Luria Broth (LB) pH 7.0 - autoclaved

- 10 $gL^{-1}$ Bacto Tryptone
- 5 $gL^{-1}$ Bacto Yeast Extract
- 5 $gL^{-1}$ NaCl

Luria Agar (LA) pH 7.0 - autoclaved

- 10 $gL^{-1}$ Bacto Tryptone
- 5 $gL^{-1}$ Bacto Yeast Extract
- 5 $gL^{-1}$ NaCl
- 15 $gL^{-1}$ Bacto Agar

Starting Minimal Media – filter sterilised

- 48 mM $Na_2HPO_4$ (anhydrous)
- 22 mM $KH_2PO_4$
- 2.2 mM NaCl
- 18.8 mM $NH_4Cl$
- 1 M $MgSO_4$
    - M $CaCl_2$
- 20 % (w/v) Glucose
- 0.5 % (w/v) Thiamine

M9 Minimal Media – filter sterilised

- 48 mM $Na_2HPO_4$ (anhydrous)
- 22 mM $KH_2PO_4$
- 2.2 mM NaCl
- 18.8 mM $NH_4Cl$
- 1 M $MgSO_4$
- 0.1 M $CaCl_2$
- 20 % (w/v) Glucose
- 0.5 % (w/v) Thiamine

- 10 mg $mL^{-1}$ Proline
- 10 mg $mL^{-1}$ Histidine (monohydrochloride monohydrate)
- 10 mg $mL^{-1}$ Glycine
- 10 mg $mL^{-1}$ Alanine
- 10 mg $mL^{-1}$ Valine
- 10 mg $mL^{-1}$ Leucine
- 10 mg $mL^{-1}$ Isoleucine
- 10 mg $mL^{-1}$ Phenylalanine
- 10 mg $mL^{-1}$ Tryptophan
- 10 mg $mL^{-1}$ Serine
- 10 mg $mL^{-1}$ Threonine
- 10 mg $mL^{-1}$ Methionine
- 10 mg $mL^{-1}$ Aspartic Acid
- 10 mg $mL^{-1}$ Glutamic Acid (Monosodium)
- 10 mg $mL^{-1}$ Glutamine
- 10 mg $mL^{-1}$ L-Arginine ($^{13}C_6$ HCl, 98 ATOM%, $^{13}C$, 95 % CP, Sigma)
- 10 mg $mL^{-1}$ L-Lysine ($^{13}C_6$ HCl, 98 ATOM%, $^{13}C$, 95 % CHE, Sigma)

5X SDS Sample Buffer

- 5 % SDS (w/v) pH 7.2
- 12 % (v/v) Glycerol
- 0.125 M Tris HCl pH 6.8
- Bromoblue
- 5 % (v/v) β-mercaptoethanol

Binding Buffer

- 0.3 M NaCl
- 10 mM Imidazole
- 50 mM $NaH_2PO_4$
- 1 per 10 mL EDTA-free protease inhibitor tablet
- 10 mg Lysozyme

Guanidine Buffer

- 6 M Guanidine Hydrochloride

- 0.3 M NaCl
- 50 mM $NaH_2PO_4$
- 1 per 10 mL EDTA-free protease inhibitor tablet
- 10 mg Lysozyme
- For wash of Ni-NTA columns: 10 mM Imidazole

Elution Buffer

- 0.3 M NaCl
- 250 mM Imidazole
- 50 mM $NaH_2PO_4$
- 1 per 10 mL EDTA-free protease inhibitor tablet

Transfer Buffer

- 29 mM Glycine
- 0.0375% (w/v) SDS
- 58 mM Tris-HCl pH 7.5
- 20 % (v/v) MeOH

1X PBS-Tween

- 1 mM $NaHPO_4$ pH 7.4
- 0.09 % NaCl
- 0.01% Tween20

## Appendix 2 - In-house Perl script for *in silico* scheduling

Raw files (final_mrml.xml) from mProphet are processed alongside report files (.csv) exported by Skyline are processed using the Perl script 'In_silico_window.pl' available on the attached Appendix CD. To run via command line:

*perl In_silico_window.pl Exported_report.csv final_mrml.xml*

## Appendix 3 - mProphet search parameters

Data processed with mProphet is searched against parameters contained within the file 'copy.params'. The parameters are as follows:

copy.params

```
use_reference 1
allow_pgpair_wo_reference_partner 0
allow_pgpair_wo_target_partner    1
make_dummy_peakgroup       1
fill_pgs_with_noise 1
light_label   light
heavy_label   heavy
reference_isoform   heavy
minSN_target  5
minSN_reference 5
select_nbest_peakgroups_target    4
select_nbest_peakgroups_reference 2
use_decoy     1
decoy_schema  AQUA
main_vars
       log10_total_xic|intensity_correlation_with_assay|xcorr_coelution_score|
xcorr_shape_score|light_heavy_correlation|light_heavy_shape_score|light_heavy_
coelution_score|abs_Tr_deviation
main_score    light_heavy_shape_score
min_peak_width      7
select_nbest_peaks  5
max_Tr_difference   5
denoise_parameter1  5
```

## Appendix 4 - Selection of Q-peptides for ChapCAT design

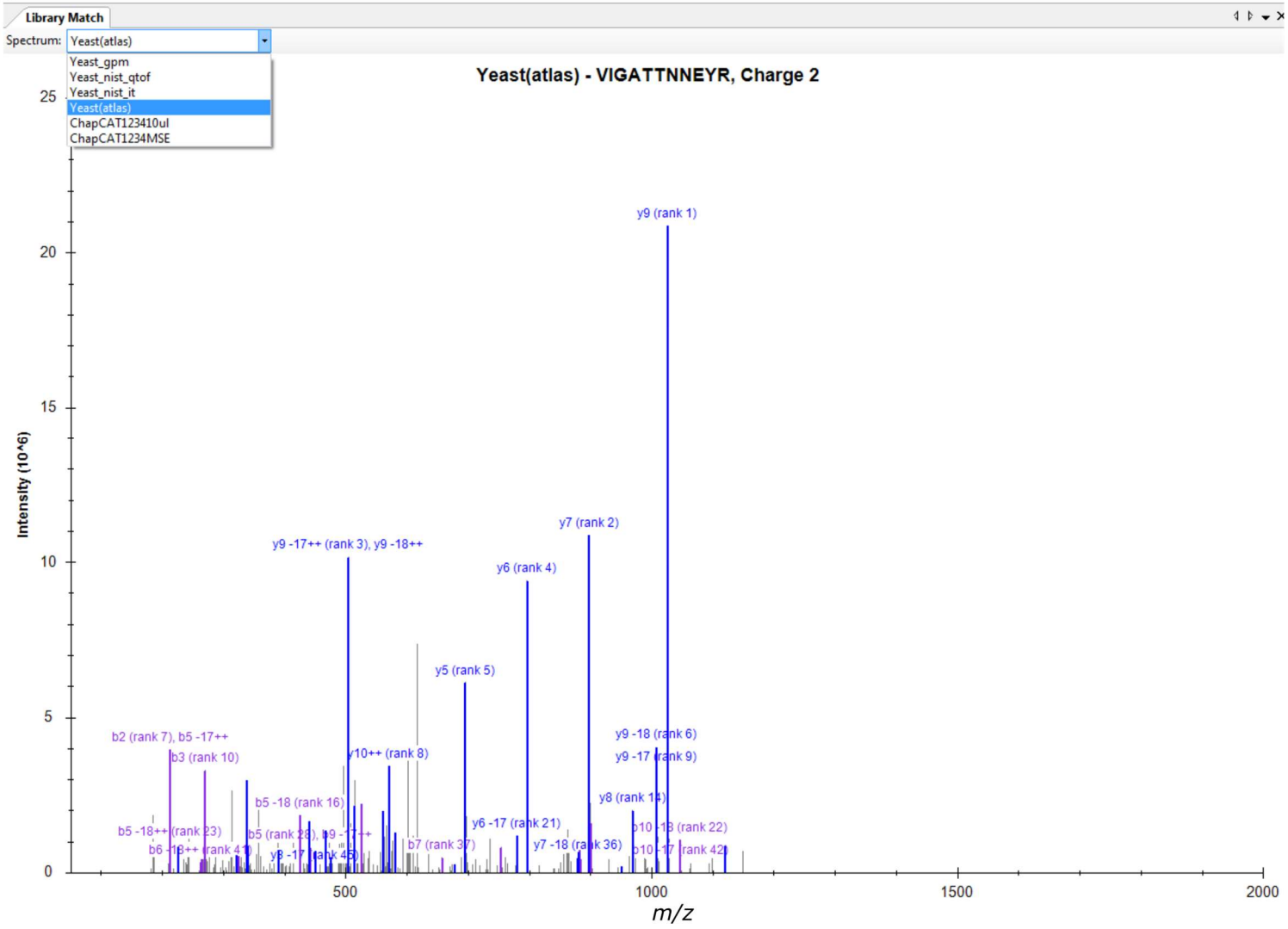
Up to five Q-peptides were selected as surrogates for quantification of 63 chaperone proteins. Table available on the attached cd, in the file Appendices.xlsx.

## Appendix 5 - Full size figures for Figure 3.16

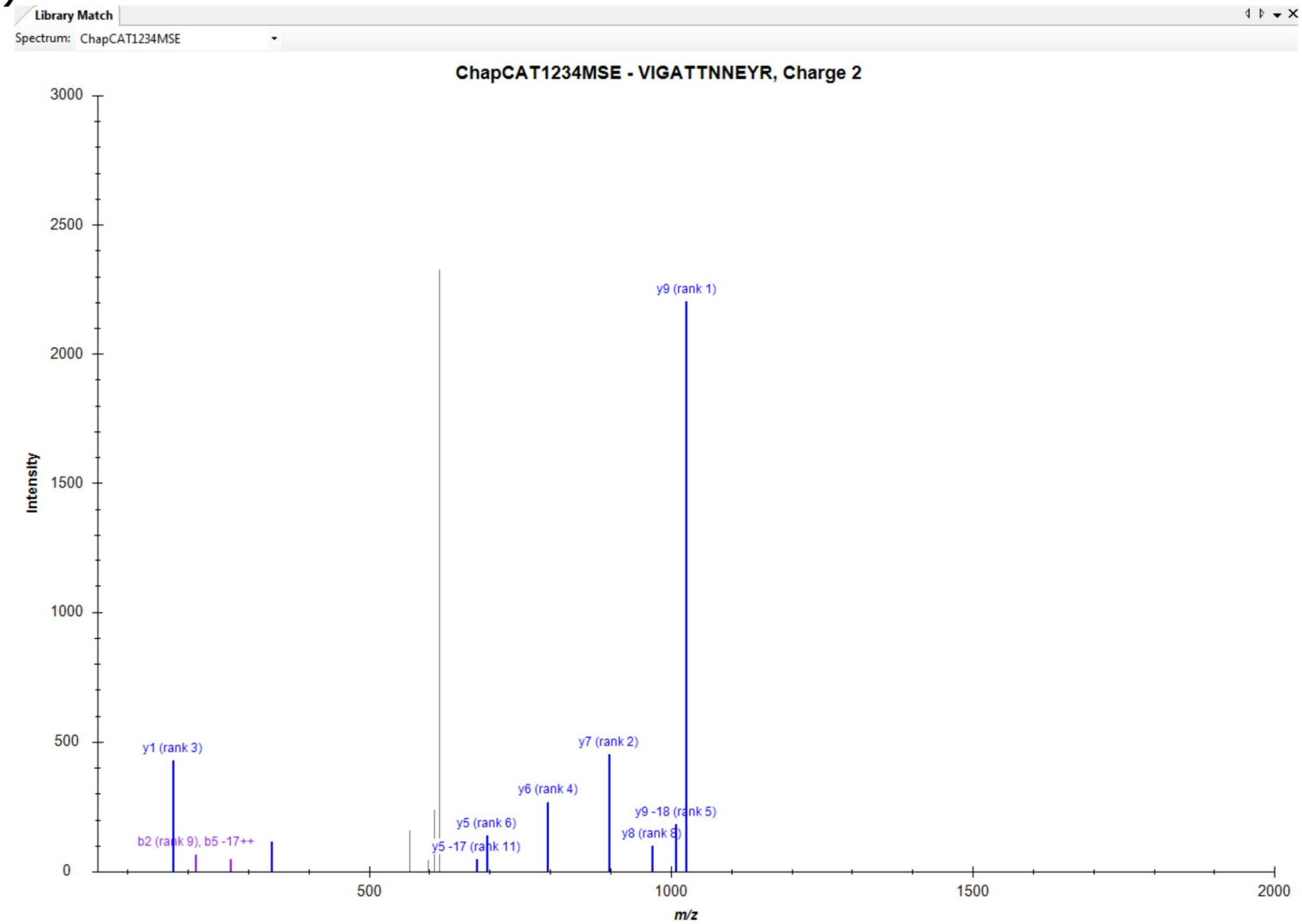
As an example using the peptide VIGATTNNEYR from ChapCAT 1 targeting Hsp104. Peptide product ions are searched against known spectral databases, including SRM atlas

(shown) (A). Following an $MS^E$ experiment, the top seven most intense ions are selected (B). An unscheduled SRM experiment is performed on a digest containing only ChapCAT, targeting these seven ions, with the top three most intense ions selected as the final transitions for use towards absolute quantification (C). A ChapCAT and yeast digest is subject to an unscheduled SRM experiment targeting the top three transitions, allowing for determination of the the retention time in the respective condition (D). Yeast digests containing 10 fmol, 1 fmol and 250 amoles of ChapCAT are subject to scheduled SRM using the retention times determine in D, with both heavy and light isotope variants targeted (E). The concentration matching closest to 1:10 (light:heavy or heavy:light) is used towards absolute quantification of the light analyte peptide (F – heavy labelled variant is shown).

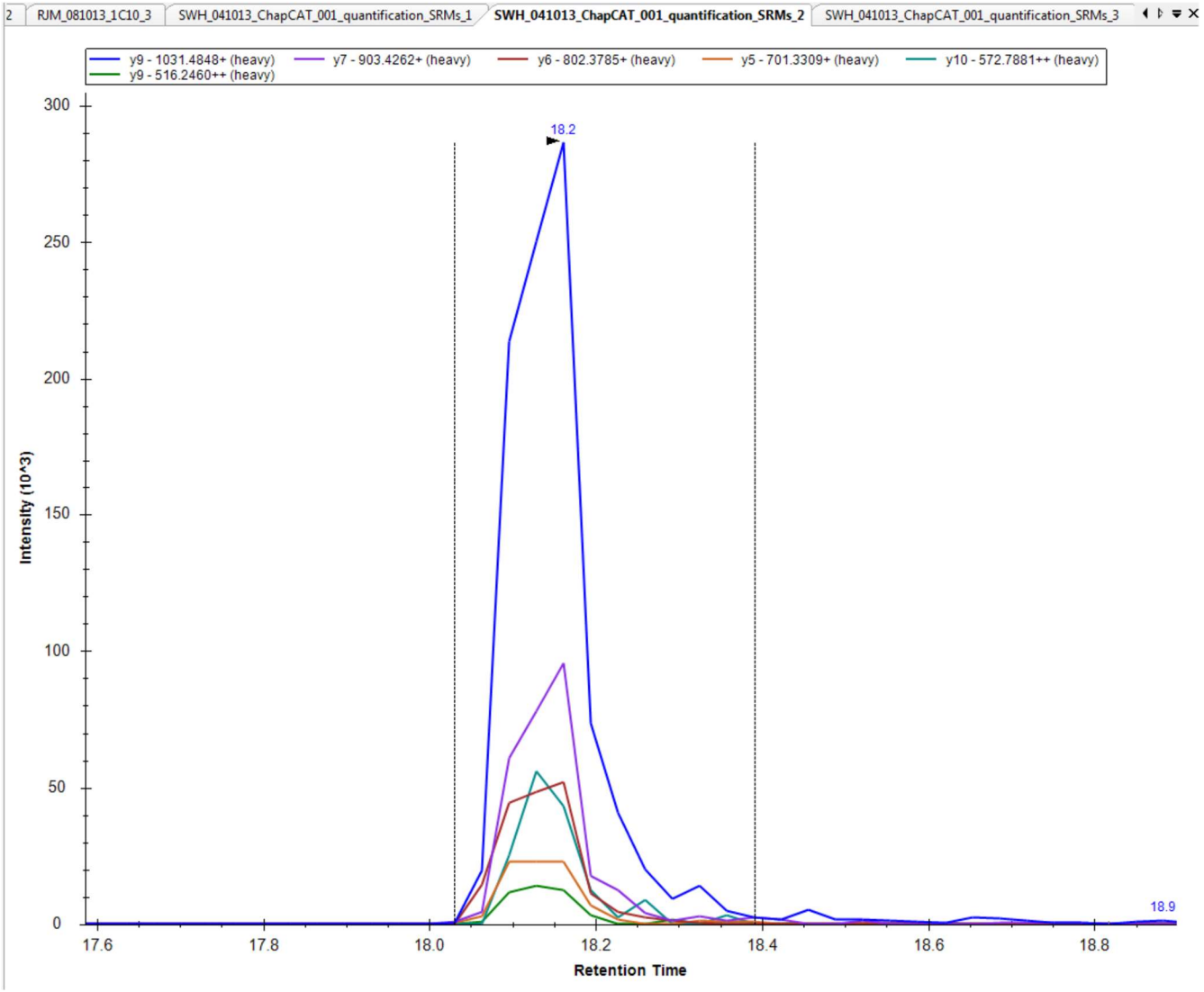
A)


Library Match
Spectrum:
Yeast(atlas)
Yeast_gpm
Yeast_nist_qtof
Yeast_nist_it
Yeast(atlas)
ChapCAT123410ul
ChapCAT1234MSE
Yeast(atlas) - VIGATTNNEYR, Charge 2
Intensity (10^6)
m/z
y9 (rank 1)
y7 (rank 2)
y9 -17++ (rank 3), y9 -18++
y6 (rank 4)
y5 (rank 5)
b2 (rank 7), b5 -17++
b3 (rank 10)
y10++ (rank 8)
y9 -18 (rank 6)
y9 -17 (rank 9)
b5 -18 (rank 16)
y8 (rank 14)
y6 -17 (rank 21)
b5 -18++ (rank 23)
y10 -18 (rank 22)
y7 -18 (rank 36)
b7 (rank 37)
y10 -17 (rank 42)

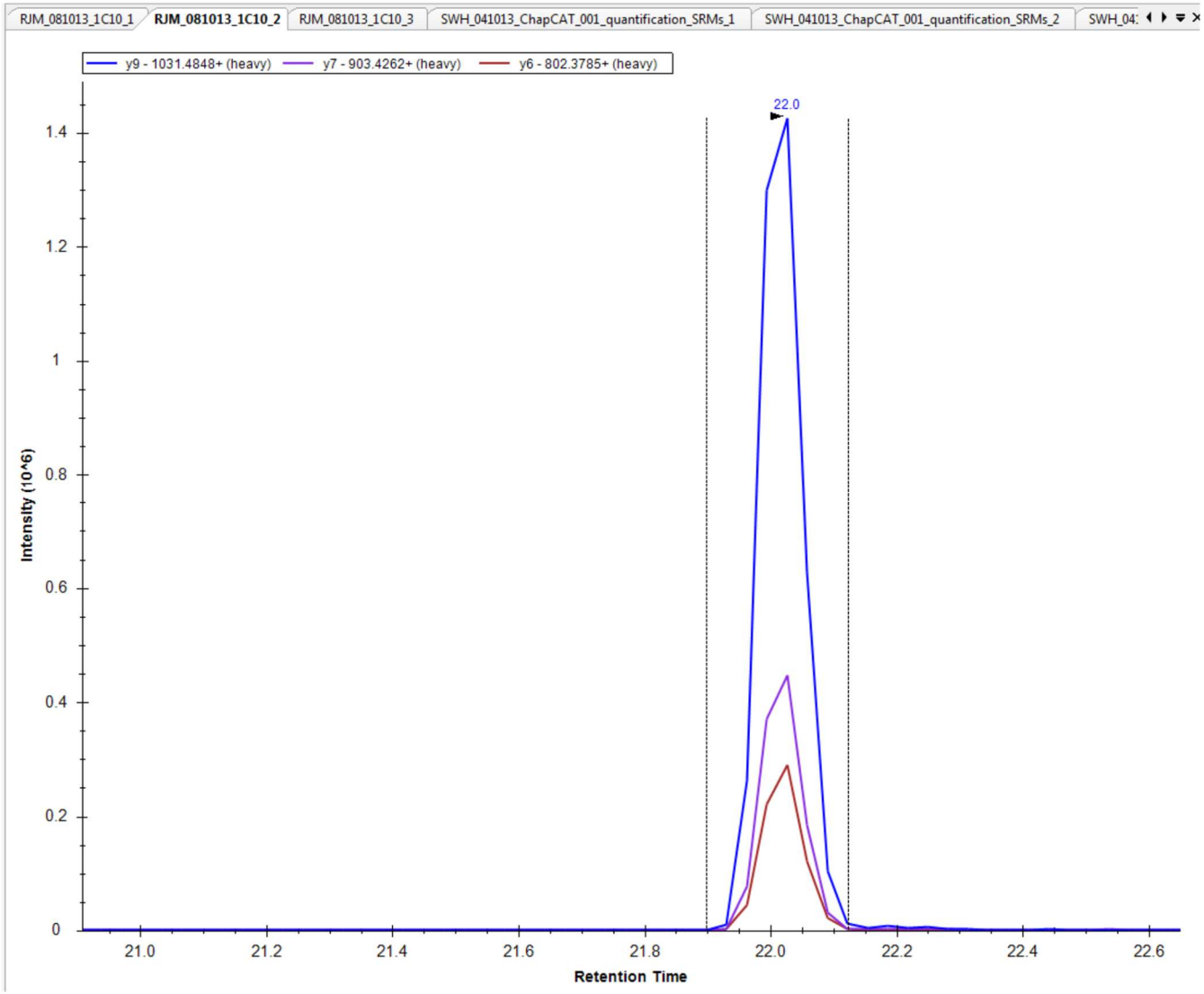
B)


Library Match
Spectrum: ChapCAT1234MSE
ChapCAT1234MSE - VIGATTNNEYR, Charge 2
Intensity
m/z
3000
2500
2000
1500
1000
500
0
500
1000
1500
2000
y9 (rank 1)
y1 (rank 3)
y7 (rank 2)
y6 (rank 4)
y5 (rank 6)
y9 -18 (rank 5)
y8 (rank 8)
b2 (rank 9), b5 -17++
y5 -17 (rank 11)

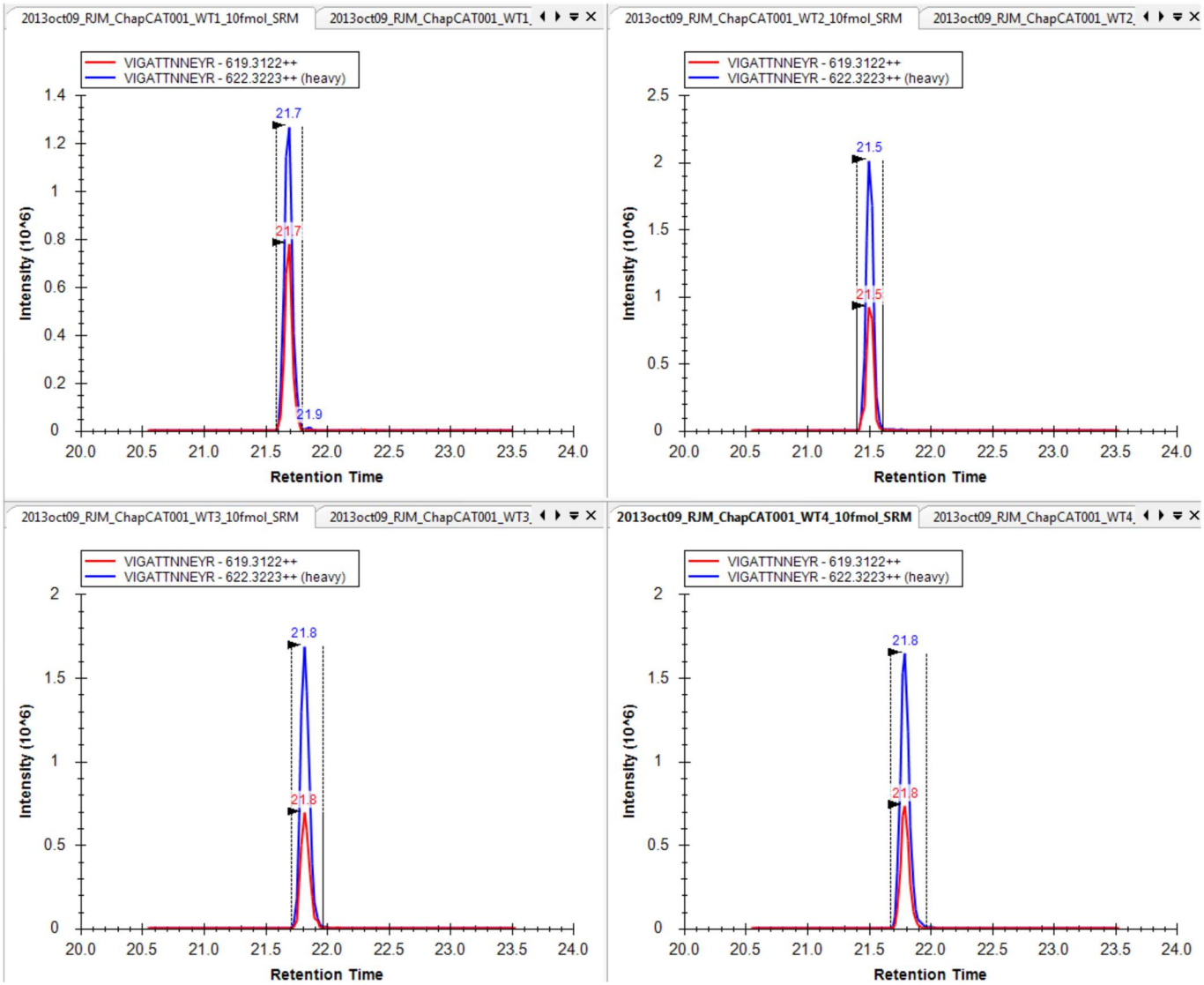
C)


2
RJM_081013_1C10_3
SWH_041013_ChapCAT_001_quantification_SRMs_1
SWH_041013_ChapCAT_001_quantification_SRMs_2
SWH_041013_ChapCAT_001_quantification_SRMs_3
y9 - 1031.4848+ (heavy)
y7 - 903.4262+ (heavy)
y6 - 802.3785+ (heavy)
y5 - 701.3309+ (heavy)
y10 - 572.7881++ (heavy)
y9 - 516.2460++ (heavy)
Intensity (10^3)
300
250
200
150
100
50
0
18.2
18.9
17.6
17.8
18.0
18.2
18.4
18.6
18.8
Retention Time

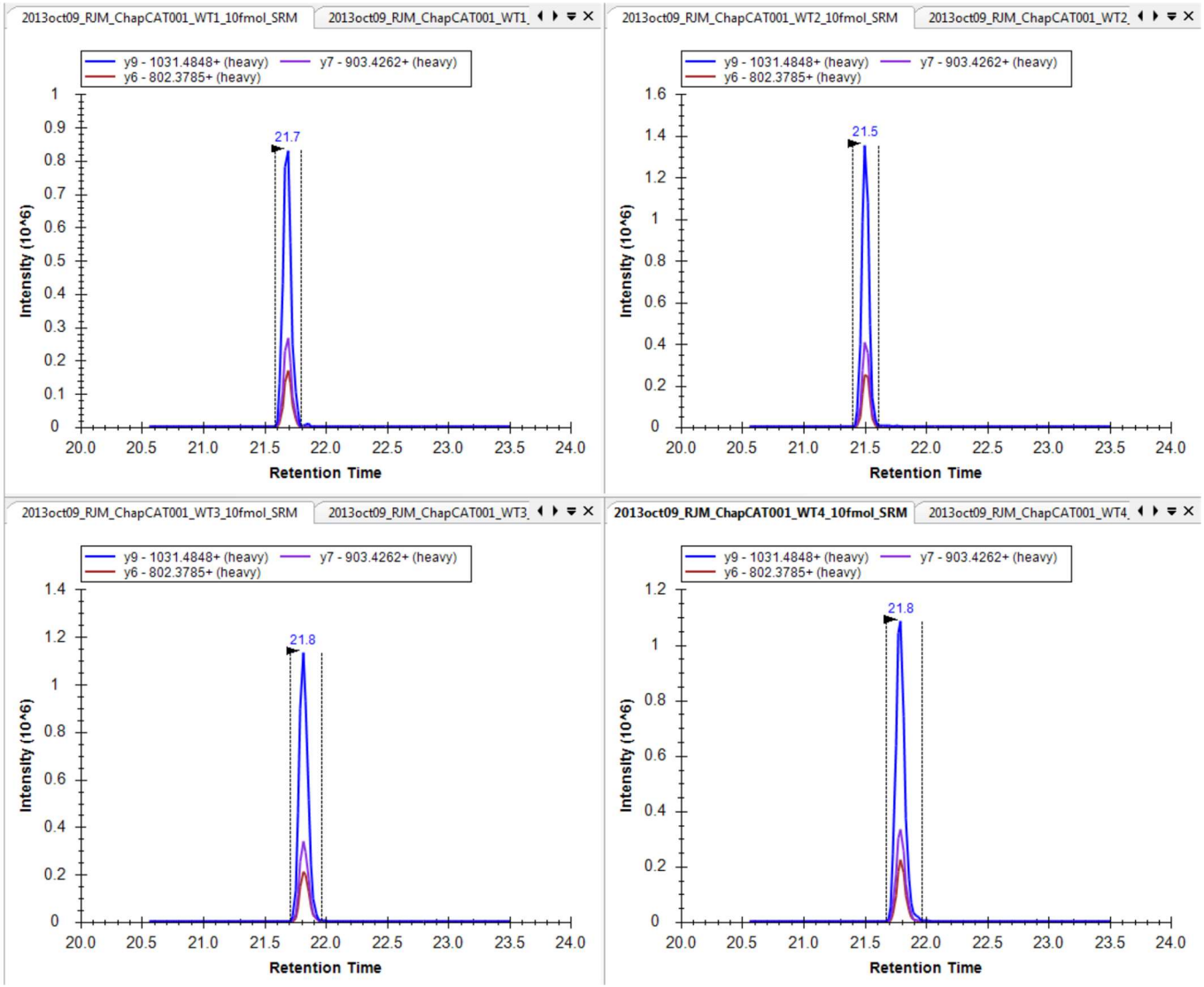
D)


RJM_081013_1C10_1
RJM_081013_1C10_2
RJM_081013_1C10_3
SWH_041013_ChapCAT_001_quantification_SRMs_1
SWH_041013_ChapCAT_001_quantification_SRMs_2
y9 - 1031.4848+ (heavy)
y7 - 903.4262+ (heavy)
y6 - 802.3785+ (heavy)
22.0
Intensity (10^6)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
21.0
21.2
21.4
21.6
21.8
22.0
22.2
22.4
22.6
Retention Time

E)


2013oct09_RJM_ChapCAT001_WT1_10fmol_SRM
VIGATTNNEYR - 619.3122++
VIGATTNNEYR - 622.3223++ (heavy)
21.7
21.7
21.9
Intensity (10^6)
Retention Time
2013oct09_RJM_ChapCAT001_WT2_10fmol_SRM
VIGATTNNEYR - 619.3122++
VIGATTNNEYR - 622.3223++ (heavy)
21.5
21.5
Intensity (10^6)
Retention Time
2013oct09_RJM_ChapCAT001_WT3_10fmol_SRM
VIGATTNNEYR - 619.3122++
VIGATTNNEYR - 622.3223++ (heavy)
21.8
21.8
Intensity (10^6)
Retention Time
2013oct09_RJM_ChapCAT001_WT4_10fmol_SRM
VIGATTNNEYR - 619.3122++
VIGATTNNEYR - 622.3223++ (heavy)
21.8
21.8
Intensity (10^6)
Retention Time

F)


2013oct09_RJM_ChapCAT001_WT1_10fmol_SRM
2013oct09_RJM_ChapCAT001_WT2_10fmol_SRM
2013oct09_RJM_ChapCAT001_WT3_10fmol_SRM
2013oct09_RJM_ChapCAT001_WT4_10fmol_SRM
y9 - 1031.4848+ (heavy)
y7 - 903.4262+ (heavy)
y6 - 802.3785+ (heavy)
21.7
21.5
21.8
21.8
Intensity (10^6)
Retention Time

## Appendix 6 - Unpaired Wilcoxon tests between biological replicates for absolute quantification

To observe reproducibility between ChapCAT biological replicates for each condition an unpaired Wilcoxon test was performed between the cpc values for each ChapCAT biological replicate (A1 and A2 class Q-peptides); with all cpc values for each of the other biological replicates for that ChapCAT under the particular condition. Table available on the attached cd, in the file Appendices.xlsx.

## Appendix 7 - Proteolysis assays for ChapCAT 4 and ChapCAT 5

In order to investigate the effects of any proteolytic products on the classification, absolute abundance and raw intensities of the standard Q-peptide I observed these with respect to the Q-peptides position in the QconCAT (depicted as order). Data for proteolysed ChapCAT 4 and 5 are presented in Figure A.1.

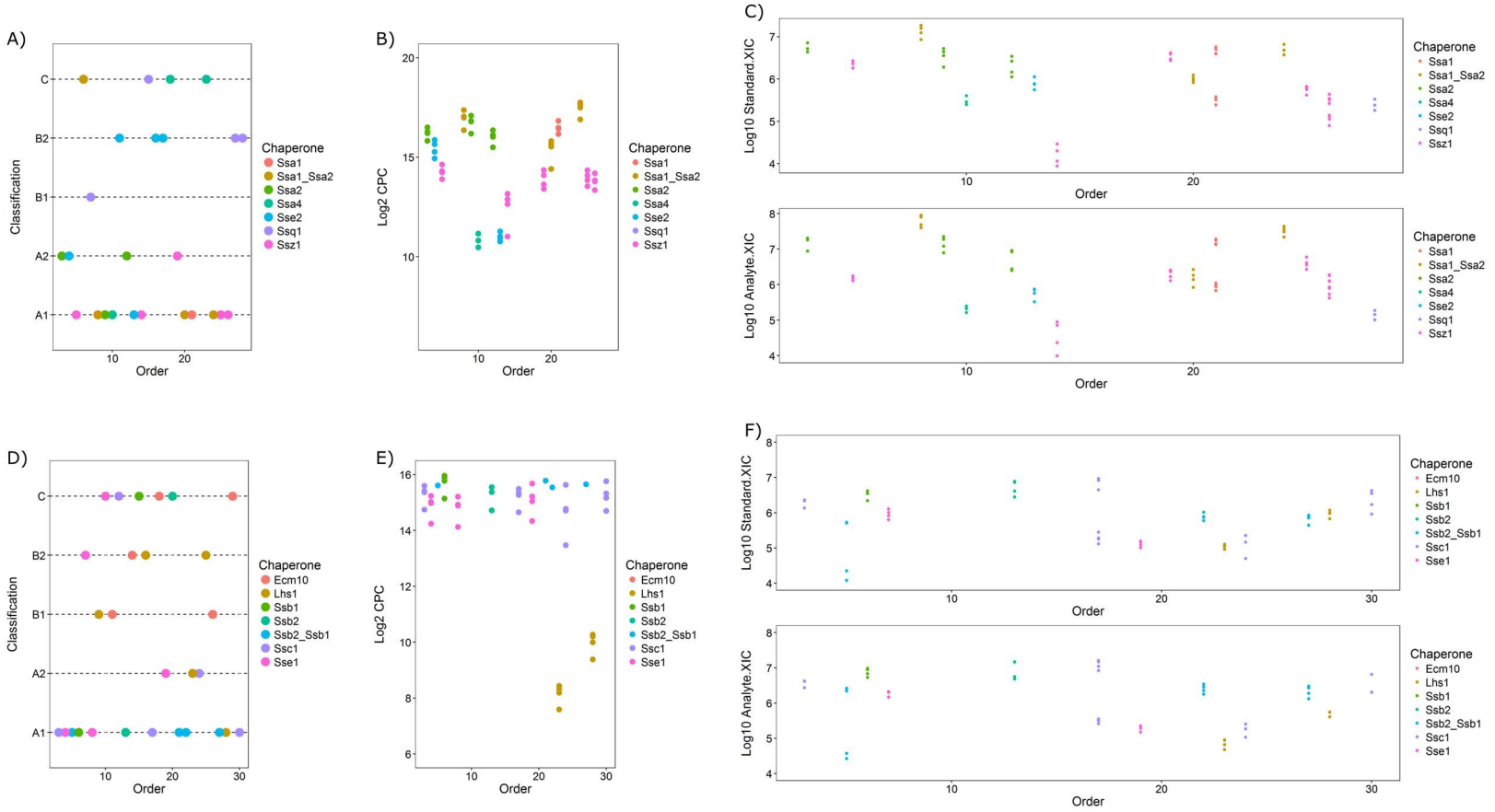



**Figure A.1) Proteolysed ChapCAT Q-peptide classifications, copy per cell values and XICs with respect to Q-peptide position**

A – C) Proteolysed ChapCAT 4 classifications, absolute Q-peptide abundances and XICs respectively with respect to the Q-peptide order within the construct. D – F) Proteolysed ChapCAT 5 classifications, absolute Q-peptide abundances and XICs respectively with respect to the Q-peptide order within the construct.

## Appendix 8 - Transitions used for absolute quantification

The transitions used during SRM for absolute quantification for each peptide was identical for standard and analyte counterparts. Table available on the attached cd, in the file Appendices.xlsx.

## Appendix 9 - Digestion rate constants for standard and analyte peptides

To determine the extent a peptide is digested to completion, non-linear modelling was used to determine *k* values and subsequent completion time (in minutes) for each Q-peptide standard and analyte counterpart. Table available on the attached cd, in the file Appendices.xlsx.

## Appendix 10 - Peptide classification and absolute quantification under condition of NG and HS

The successful expression and purification of eight ChapCATs targeting 49 chaperone proteins resulted in a total of 222 Q-peptides subject to targeted analysis under conditions of NG and HS. Each peptide was classified according to suitability for quantification. Table available on the attached cd, in the file Appendices.xlsx.

## Appendix 11 - Differences in peptide classification between NG and HS conditions

Differences in the peptide classification were observed between NG and HS, resulting in differing numbers of 'A1' peptides used towards absolute quantification under each condition. Table available on the attached cd, in the file Appendices.xlsx.

## Appendix 12 - Absolute Protein Quantification

Using 'A1' Q-peptides, the copy per cell value for each chaperone was determined as the median copy per cell value across all biological replicates for all 'A1' Q-peptides under the particular condition. If a chaperone is targeted by an 'A1' Q-peptide, it was given the classification 'A'. If no Q-peptides are available but the chaperone is targeted by 'B' class Q-peptides, it was given the classification 'B'. If 'B' chaperones are targeted by 'B1' Q-peptides, an upper cpc limit may be defined and is presented as the cpc value under the according condition. Table available on the attached cd, in the file Appendices.xlsx.

## Appendix 13 - Mod-cpc values for 1644 proteins

Using SRM-normalisation, the median MaxLFQ label free intensity for 1644 proteins identified in both NG and HS conditions could be transformed into a mod-cpc value. Table available on the attached cd, in the file Appendices.xlsx.

## Appendix 14 - Chaperone volume and workload modelling

Using the mod-cpc values for clients of each chaperone, the respective protein volume mediated and workload of each chaperone could be calculated. Table available on the attached cd, in the file Appendices.xlsx.