



UPPSALA  
UNIVERSITET

UPTEC X 18 018

Examensarbete 30 hp  
Juni 2018

# Investigating the use of isotope-labeled standards as calibrants in label-free quantification

---

Linda Breimark





UPPSALA  
UNIVERSITET

**Teknisk- naturvetenskaplig fakultet  
UTH-enheten**

Besöksadress:  
Ångströmlaboratoriet  
Lägerhyddsvägen 1  
Hus 4, Plan 0

Postadress:  
Box 536  
751 21 Uppsala

Telefon:  
018 – 471 30 03

Telefax:  
018 – 471 30 00

Hemsida:  
<http://www.teknat.uu.se/student>

## Abstract

### **Investigating the use of isotope-labeled standards as calibrants in label-free quantification**

*Linda Breimark*

The ability to accurately identify and quantify proteins in complex samples is of great importance in the field of proteomics. Using mass spectrometry, samples can be analysed and quantified either by the incorporation of a labelled standard of known concentration, or by label free quantification. Label free quantification has many benefits, including time, cost, and ease of use, but is not as accurate as the use of isotope label standards. In this project, the possibility of increasing accuracy in quantification results from LFQ using a set of isotope labelled standards, QPrESTs, is investigated. The standards were produced by metabolic incorporation of heavy Lysine and Arginine during expression in *E. coli*. They were then quality controlled using SDS-PAGE for purity analysis, and LC-MS/MS for quantification and confirmation of MW. Human cell lysate samples spiked with a set of 21 QPrEST standards were analysed by LC-MS/MS and quantified by QPrEST-H/L intensity ratios and intensity based LFQ. In the LFQ protein quantification indices obtained from MaxQuant were combined with BCA results, or with calibration curves obtained from spiked in QPrEST standards. The LFQ results that best matched those obtained from QPrEST-H/L were those that used the calibration curves for quantification, which were found in a ~3-fold range, with a correlation coefficient varying from 0.67 to 1. Assuming that QPrEST-H/L is the most accurate quantification method used, this indicates that the use of QPrEST standards as calibrants can be beneficial when it comes to increasing the accuracy in LFQ.

Handledare: Tove Boström  
Ämnesgranskare: Jonas Bergquist  
Examinator: Jan Andersson  
ISSN: 1401-2138, UPTec X 18 018



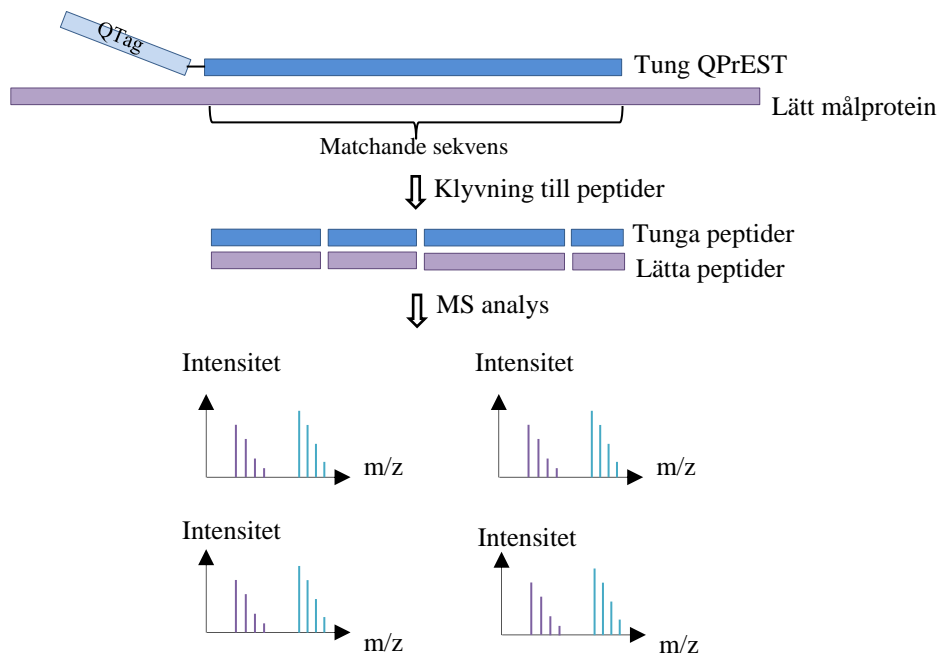
## Sammanfattning

Ett protein består av en sekvens aminosyror, som sitter ihop en efter en i en viss sekvens, som är unik för varje protein. Proteomik är en gren inom biologi där man undersöker proteiner och deras roll i celler. Man forskar kring vilka proteiner som uttrycks i en viss vävnad eller celltyp, vad dom gör och hur mycket som finns av dom. Genom att studera många olika prover från tex människor med olika hälsostatus så kan man hitta proteiner som spelar roll i olika sjukdomar, och utifrån den informationen hitta sätt att lättare ställa diagnos, eller hitta nya behandlingsmetoder. Proteomik är därför ett viktigt forskningsområde.

Traditionellt har antikroppar använts för att identifiera och kvantifiera proteiner i ett prov. Antikroppar är biologiska molekyler som binder till ett visst protein på ett visst ställe, och som därför kan användas för att urskilja proteinerna i ett prov. Men antikroppar är svåra och dyra att tillverka, och kräver användning av försöksdjur, vilket gör att man gärna vill använda andra metoder. Ett vanligt alternativ är då att använda masspektrometri, MS, som är en teknik där laddade molekyler separeras med hjälp av magnetfält, beroende på förhållandet mellan deras massa,  $m$ , och laddning,  $z$ , vilket ger kvoten  $m/z$ . Proteiner kan analyseras hela eller klyvda i kortare bitar, peptider. Joniserade proteiner eller peptider som analyseras med masspektrometri kommer att ge ett masspektrum, där intensiteten för olika joner med olika  $m/z$  ses som staplar. Beroende på mönstret av staplar i spektrumet så kan man räkna ut vilka proteiner som finns i det analyserade provet. För att kunna kvantifiera proteinerna, kan man tillsätta märkta standarder till sitt prov. Standarderna är proteiner eller peptider, med en aminosyrasekvens som matchar den hos proteinet man vill analysera, målproteinet, men med skillnaden att dom är märkta så att dom kan urskiljas från de andra proteinerna i masspektrumet. Standarderna kan vara märkta på många olika sätt, men en vanlig metod är att använda aminosyror med tunga isotoper av kväve och kol. Det gör att standarden kommer att ha samma aminosyrasekvens som målproteinet men lite högre massa, och därför får den ett lite högre  $m/z$  värde. Genom att man vet koncentrationen av standarden som används, så kan man jämföra intensiteterna för målproteinet med intensiteten för standarden, och på så sätt kvantifiera målproteinet. Man kan också kvantifiera målproteinet utan att använda märkta standarder. Då gör man först en analys av hur mycket protein som finns totalt i provet, och sedan jämför man intensiteterna för varje protein från masspektrumet för att avgöra hur stor del av den totala proteinkoncentrationen som hör till ett visst protein. Fördelen med att inte använda standarder är att det är billigare och lättare, men nackdelen är att man inte får lika exakta resultat som när man använder märkta

standarder. I denna studie testas det om ett litet set med isotopmärkta standarder mot ett fåtal målproteiner i ett prov, kan användas för att få ett bättre resultat för kvantifieringen av övriga proteiner i provet utan att behöva använda märkta standarder för varenda protein som ska kvantifieras. Standarderna som används heter QPrEST, och tillverkas i *Escherichia coli*. Koncentrationen av standarden bestäms noggrant med MS analys, med hjälp av Q-taggen, som

är en del av QPrESTens sekvens. Lätt Q-tag tillsätts till standarden, och kvoten mellan intensiteten av tung Q-tag jämfört med lätt Q-tag gör att man kan bestämma koncentrationen väldigt exakt. När koncentrationen för varje QPrEST-standard är känd, så tillsätts lite av varje standard till de prover som ska analyseras. Proverna i den här studien är cellysat från olika humana cellinjer. Cellysat spetsade med QPrESTar analyseras sedan med MS (se figur 1), och målproteinerna kvantifieras utifrån intensitetskvoten mellan tung och lätt peptid, H/L. Genom att koncentrationer som fås för de olika målproteinerna plottas mot deras intensitet så får man en linjär kurva, vars ekvation kan användas för att beräkna koncentrationen av alla proteiner i provet utifrån deras intensitet. Man skapar alltså en kalibreringskurva utifrån dom proteiner som kvantifieras med hjälp av QPrEST- standarder. Därefter kvantifieras målproteinerna helt utan standarder, genom att ta deras respektive intensitet och dela med den totala intensiteten för alla proteiner i provet, och sedan multiplicera den kvoten med den uppmätta totala proteinkoncentrationen i provet. Den metod som antas fungera bäst och ge mest korrekt kvantifiering är kvantifieringen som görs direkt utifrån H/L-kvoten. Om man jämför de andra två kvantifieringsmetoderna mot denna, så ser man att metoden med en kallibringskurva ger väldigt lika resultat, inom 3ggr skillnad, medan resultaten från kvantifiering helt utan standarder avviker mycket från resultaten med H/L, med upp till 10ggr skillnad. Detta indikerar att metoden med kallibreringskurva ger en noggrannare kvantifiering än metoden helt utan standarder. Med mer studier och optimering av metoden så kan detta kanske leda till en kvantifieringsmetod där koncentrationer för alla proteiner i provet kan bestämmas med god noggrannhet, med hjälp av en QPrEST-mix med enbart ett fåtal tunga standarder, vilket skulle vara både ekonomiskt och tidsbesparande jämfört med att tillverka standarder för alla proteiner i provet.



**Figur 1: Schematisk struktur av en QPrEST, och användning av en QPrEST i MS-analys. Tung QprEST blandas med målproteiner, klyvs och analyseras med MS. Intensitetskvoten Tung/Lätt som fås för varje peptid används för att beräkna koncentrationen av målproteinet.**

# Table of contents

<b>Abbreviations.....</b>	<b>1</b>
<b>1 Introduction .....</b>	<b>3</b>
1.1 Mass spectrometry .....	3
1.2 Protein Quantification by MS .....	5
1.2.1 Labelling for quantification .....	5
1.2.2 Label-free quantification .....	7
1.3 Atlas Antibodies .....	8
1.3.1 QPrEST .....	9
1.4 Aims and objectives .....	10
<b>2 Materials and methods .....</b>	<b>11</b>
2.1 Production of QPrESTs.....	11
2.2 QC of produced QPrESTs.....	12
2.2.1 Molecular weight determination of QPrEST .....	12
2.2.2 QPrEST Quantification .....	12
2.2.3 QPrEST Purity analysis .....	13
2.3 Preparation of cell lysate for MS .....	13
2.3.1 Cell lysis.....	13
2.3.2 Bicinchoninic assay .....	14
2.4 Identification of quantotypic peptides .....	14
2.4.1 Sample preparation .....	14
2.4.2 Data analysis .....	15
2.5 Preparation of adjusted QPrEST-LFQ mix and samples for quantification .....	15
2.6 Quantification of proteins in human cell lines.....	16
<b>3 Results .....</b>	<b>17</b>
3.1 Production and QC of QPrEST proteins .....	17
3.2 Identification of Quantotypic peptides .....	20
3.3 LFQ .....	25
<b>4 Discussion.....</b>	<b>29</b>
<b>5 Acknowledgements .....</b>	<b>34</b>
<b>References .....</b>	<b>35</b>
<b>Appendix .....</b>	<b>38</b>





## Abbreviations

Arg	Arginine
BCA	Bicinchoninic assay
CM	Chloramphenicol
IMAC	ion metal affinity chromatography
KM	Kanamycin
LC	Liquid chromatography
LFQ	Label free quantification
Lys	Lysine
MS	Mass spectrometry
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Tpm	transcripts per million



# 1 Introduction

The discovery of protein ionization techniques allowing the use of MS analysis, has led to a new way of studying proteomics, enabling a system- approach rather than the traditional way of investigating one protein at a time (Bantscheff *et al.* 2012). The more classical techniques used in proteomics are largely based on immuno affinity. The multiplexing ability is therefore limited, and development of assays towards new targets is very much dependent on the ability to produce good antibodies. Assay development is therefore costly as well as time consuming. There is also an ethical aspect of antibody production, as it requires the use of animals for immunisation. MS is thus a complementing method with many advantages, e.g. allowing multiplexing and not being dependent on immuno affinity. The ability to quantify proteins is of great importance in the field of proteomics, to increase knowledge about biological function and impact of different proteins. For instance, there is much to gain by measuring protein concentrations directly when investigating gene expression, rather than measuring only the mRNA levels, which will not account for the process of translation (Cutillas & Timms 2010) and might therefore not always be a good indicator of protein abundance. Another example of where MS is useful is in studying organelle proteomes (Cutillas & Timms 2010). In this type of study, an important feature is the ability to quantify the proteins in a sample, as one need to be able to distinguish proteins that are expressed inside the organelle and part of its proteome, from proteins that are not. The contaminant proteins may only be present in small amounts, but using qualitative MS means that even proteins in small amounts are detected. By quantifying the ingoing proteins, contaminants can be identified based on their low abundance in the sample (Cutillas & Timms 2010). Quantification of proteins is also important in the search for biomarkers, which rely on comparisons of protein levels between different cells, tissues etc. (Cutillas & Timms 2010). In conclusion, the use of LC-MS is widely applicable in the field of proteomics, and the ability to quantify the proteins in a sample is many times crucial.

## 1.1 Mass spectrometry

Mass spectrometry is an analytical method where ionized compounds are separated depending on their mass ( $m$ ) to charge ( $z$ ) ratio, also known as  $m/z$ . Based on the  $m/z$  values detected, molecular weight, analyte composition etc. can be inferred (Price & Nairn 2009). The instrumentation consists of an ionization source, a mass analyser and a detector. The mass analyser measures the  $m/z$  values of the generated ions, while the detector counts the number of ions at each  $m/z$  value (Aebersold & Mann 2003). MS is often performed in tandem (MS/MS) meaning that the ionized analyte is subject to two rounds of MS, the second one after further fragmentation of the analyte (Aebersold & Mann 2003). Two types of ion sources often used are electrospray ionization (ESI) and matrix- assisted laser desorption ionization (MALDI). In ESI the analyte is dissolved in a volatile solvent, which is ionised by passing through a thin needle under high voltage, giving rise to an ion-spray, with ions that can enter the mass spectrometer. In MALDI the analyte is instead dried onto a matrix,

and laser beams directed at the surface ionize the analyte (Price & Nairn 2009). When it comes to mass analysers and detectors, there are several alternatives, and also different ways to combine them. Some examples of mass analysers are the time of flight (TOF), ion trap, quadrupole and Fourier transform ion cyclotron (Aebersold & Mann 2003). By applying magnetic fields, the ions are guided inside the mass spectrometer, and can therefore be separated and detected depending on their  $m/z$  values. Liquid chromatography is often coupled online to MS (LC-MS). This allows a separation of sample components prior to MS. LC-MS is used in various applications of biomolecular research, for example exploring gene expression levels by quantification of protein levels in cells (Cutillas & Timms 2010).

As previously mentioned, MS is often coupled to LC, to achieve a separation of analyte before entering the mass spectrometer. The LC-systems used are high pressure LC (HPLC)(Bantscheff *et al.* 2012) or ultra-high pressure LC (UPLC)(Cutillas & Timms 2010). The ion source of choice for LC-MS is often ESI since it ionises liquid sample, which enables online coupling to the LC. If MALDI is to be used, fractions from LC must be gathered and then analysed offline (Cutillas & Timms 2010, Boström 2014). To obtain small elution volume and high concentration of eluted proteins from the LC system, the flowrate used is low (Cutillas & Timms 2010) and miniaturization of the column is applied. The chromatographic dilution of the sample scale with the column diameter and length, and decrease as dimensions are slimmed down (Rieux *et al.*). Downscaling the column will therefore lead to smaller elution volumes for the proteins, and thus higher resolution. The downscaled columns go by the name nano LC columns, and have an inner diameter of about  $\sim 75\mu\text{m}$ . A commonly used stationary phase for online coupling to the MS, is reverse phase (Aebersold & Mann 2003, Cutillas & Timms 2010, Bantscheff *et al.* 2012). The reverse phase matrix consists of silica beads covered in C8-C18 alkyl chains. As a result, the hydrophilic molecules are eluted first, while the hydrophobic molecules interact with the stationary phase and need to be eluted by altering the mobile phase polarity. The three main types of MS used in proteomics are Quadrupole, orbitrap TOFs, and ion traps, often in combination with each other (Steen & Mann 2004).

When it comes to data acquisition, there are different modes to choose between. Data dependent acquisition, DDA, is when only the ions of highest intensity are chosen for further fragmentation and a second analysis. The combined information from the precursor ion spectrum with its fragment ions spectra is used for identification and quantification of analyte components. This mode is often used for proteome-wide analysis (Cutillas & Timms 2010, Boström 2014). The alternative to data dependent, is data independent acquisition, DIA. In DIA the scanning is instead performed in a stepwise manner, where ions from a certain  $m/z$  are chosen for fragmentation regardless of their intensities. Data analysis of DIA data is not as straight forward as for DDA data, as the relation between precursor and fragment ions is lost as many different ions are fragmented simultaneously, which makes identification, and thus quantification, a challenging task. The  $m/z$  range can be chosen to

include all ions, or only a small range. The choice of range will also affect the level of difficulty of data analysis.

Typical sample preparation of a protein sample includes reduction and alkylation of cysteines, and might also include digestion of the protein into peptides, and if necessary desalting of samples (Aebersold & Mann 2003). If internal standards are to be used, these can be added to the sample as well, before or after digestion depending on what standard is used (Villanueva *et al.* 2014).

## 1.2 Protein Quantification by MS

Quantification of proteins and peptides is then performed based on the spectra obtained from the MS. Quantification can be done in a relative or absolute manner (Steen & Mann 2004, Bantscheff *et al.* 2012). Relative quantification will give the relative amounts (e.g. percent) of a protein when comparing multiple samples, or the relative amount of the proteins in one sample. Absolute quantification will give an absolute value of the amount of protein present in the sample (e.g. copies/cell) (Steen & Mann 2004, Bantscheff *et al.* 2012). For some applications, it is enough to study the relative amounts of protein while others require absolute amounts. The task of obtaining accurate absolute amounts for multiple proteins in complex samples is not straightforward, and several strategies towards this goal exist. These strategies can be divided into two main approaches, namely labeled quantification and label-free quantification (Cutillas & Timms 2010, Bantscheff *et al.* 2012).

### 1.2.1 Labelling for quantification

One way of quantifying proteins by mass spectrometry is to use labeled proteins/peptides. The idea is to incorporate a labeled version of the protein/peptide you want to quantify, in the MS analysis. Either mixed in with the analyte, or run separately, as a sample with only the labeled protein/peptides. The resulting MS spectra for the labeled and unlabeled peptides can then be compared, and quantities inferred. There are many different alternatives when it comes to labelling, but the quantification is still based on the comparison between differentially labeled proteins and peptides (Cutillas & Timms 2010). Two examples of strategies for labelling are chemical- and metabolic labelling. Regardless of what labelling method is used, it is always best to introduce the labels or labeled proteins/peptides as early in the workflow as possible, to minimize variations introduced during sample preparation (Cutillas & Timms 2010, Bantscheff *et al.* 2012, Boström 2014, Villanueva *et al.* 2014).

In chemical labelling, the proteins are first produced and harvested, and then labeled by chemical derivatization of the protein, adding isobaric tags or stable isotopes. This will give rise to differences between the labeled and unlabeled protein. For example mass differences between peptides of the same sequence, or different ion fragments in MS/MS from peptides with the same mass (Cutillas & Timms 2010, Bantscheff *et al.* 2012, Boström 2014). Based on these differences, comparisons of MS results from peptides in the analyte are made, and quantities inferred for the proteins in the sample. The downside to using chemical labelling

is that not all methods are suitable for all samples, for example labelling by isotope coded affinity tags (ICAT) is done by alkylation of cysteines, which will only be possible if the peptides that are to be labeled contain cysteines. Other downsides are that many of the chemical labels are added to the sample at a late stage of sample preparation, and also the ability to multiplex is limited to the number of labels available (Bantscheff *et al.* 2012).

Metabolic labelling is another alternative. The basic principle is to metabolically incorporate stable isotopes during protein expression, which will give the labeled proteins a mass difference compared to their unlabeled counterparts (Cutillas & Timms 2010, Boström 2014), without altering other properties of the protein, such as sequence, digestibility or ionization patterns. The labeled proteins/peptides are mixed with the sample that is to be analysed prior to MS analysis, sometimes as early as before sample digestion, which is a major difference compared to the chemical labelling methods. Often heavy isotopes are used for the labelling, and the MS analysis will result in a spectrum with peaks from heavy (labeled) and light (unlabeled) versions of the proteins/peptides, separated by their mass, *m*. The intensities of peaks from light and heavy peptides can then be compared, and quantities inferred. To quantify a certain protein in a complex mixture, for example in a cell lysate, a known amount of heavy isotope labeled protein/peptide can be spiked in to the sample, enabling absolute quantification (Villanueva *et al.* 2014). This method is called “spike in standard”. There are many different methods of producing labeled proteins and peptides, and thus many different spike in standards are available, all with pros and cons. It is therefore important to carefully consider what kind of label is most suitable for the application at hand. All samples cannot be labeled with all techniques, for example clinical samples of tissue etc. cannot, for obvious reasons, be labeled *in vivo*, only *in vitro*. One example of a metabolic labelling technique is stable isotope labelling of amino acids in cell culture (SILAC)(Geiger *et al.* 2011). Here, cells are grown in medium enriched with light- or heavy-isotope versions of amino acids. The expressed proteins will thus have the same sequence, posttranslational modifications etc. but slightly different mass compared to their target peptide. Another example of a metabolically labeled protein is QconCAT (Bantscheff *et al.* 2012). The principle is to select tryptic peptides from the target, and then produce a concatenated synthetic gene corresponding to the chosen peptides. The synthetic gene is then cloned into a vector that can be expressed in for example *Escherichia.coli*, in a heavy amino acid enriched medium. Isotopically labeled peptides and proteins can also be produced synthetically and used in a similar manner to the metabolically labeled, which as discussed have many advantages compared to chemically labeled ones. However, there is a drawback when it comes to synthetic proteins. Namely the lack of posttranslational modifications, structural motives etc. found in the metabolically synthesised proteins and peptides. This can lead to differences in digestion efficiency between endogenous and synthetic peptides, which in turn will affect the accuracy of results (Cutillas & Timms 2010, Bantscheff *et al.* 2012, Boström 2014). One example of a commercially available synthetic peptide standard is AQUA peptides. They are short peptides, up to ~15 amino acids in length, with a sequence

matching that of a tryptic target peptide (Bantscheff *et al.* 2012, Villanueva *et al.* 2014). The standard is thus added to the sample after digestion.

Labelling of peptides and proteins for MS-Quantification is a widely used strategy, with the ability to achieve very precise and accurate quantification results, both absolute and relative. The process of producing the labeled proteins and peptides is however time consuming and can be very costly (Arike *et al.* 2012). In proteomic studies when large sets of proteins need to be quantified, it would mean that labeled versions of all the proteins of interest must be produced. Therefore, the labeled approach is not always a viable option, and thus alternative label-free methods for quantification have been developed.

### **1.2.2 Label-free quantification**

In label-free MS quantification, proteins are quantified based on comparison of results from several separate runs. Multiple samples are prepared and analysed, and then quantification can be done by comparing the measured peptide intensities, or by comparing spectral counts between the different runs (Zhu *et al.* 2010). Spectral counting relies on the fact that protein abundance normally corresponds to the number of tryptic peptides, i.e. a highly abundant protein will give rise to more tryptic peptides than a low abundant one. For peptide intensities, it has been shown that they increase as the concentration of protein in the analyte increase, thus enabling intensity-based quantification (Zhu *et al.* 2010). To obtain absolute amounts for the analysed proteins, one can add an internal standard, i.e. a protein of known amount, to the sample. Based on MS results for the standard, quantities for the other proteins in the sample can be inferred.

It is also possible to obtain absolute amounts without using a standard. This is done by comparing the fraction of different proteins in the mixture to the total amount of protein in the sample, and from that calculate the abundance of each (Arike *et al.* 2012). The accuracy of label free quantification is not as good as for the labeled methods (Arike *et al.* 2012), since variations between runs, differences in sample handling and in the analysis itself, will affect the results. One of the benefits of label-free quantification is the low cost compared to labeled methods, and also the fact that quantification can be performed simultaneously for all the identified proteins in a sample.

To be able to quantify the proteins though, they must first be identified. Often the MS is run in data dependent mode, with only ions of the highest intensity picked for further fragmentation and MS/MS scan. Low abundant proteins will not give rise to as many peptides as the high abundant ones and might therefore not trigger as many MS/MS scans resulting in fewer peptide identifications for low abundant proteins. However, if an identification has been made in one sample in a run, the MS1 spectra from that sample can be compared and matched to that from the other samples in the run, which might lead to identifications in the other samples as well, even though MS2 data is absent for the peptide in those samples (Higgs *et al.* 2013).

When performing MS analysis of a protein sample, the number of identified peptides from a protein correspond to the size of the protein (Bantscheff *et al.* 2012). This fact can be used to calculate a protein abundance index (PAI) for each protein, which reflect the relative amount of different proteins in the sample. The PAI is defined as number of observed peptides per protein divided by the number of theoretically observable peptides. To achieve an absolute quantification, the experimentally modified PAI (emPAI) can be used. The emPAI is based on the observation that the number of peptides identified is proportional to the total amount of protein in the sample (Ishihama *et al.* 2005). The emPAI for a protein is defined as  $10^{\text{PAI}-1}$ , and the molar fraction of a protein is calculated by dividing its emPAI with the total sum of all emPAIs of the sample.

A method similar to emPAI is called APEX (short for absolute protein expression). In APEX, machine learning is used to calculate peptide detection probabilities, in order to predict the number of spectral counts arising from a molecule of a certain protein (Ahrné *et al.* 2013). This theoretical estimation is compared to the experimental data, to give an estimation of the absolute amount of the protein in the sample. Note that both methods are based on spectral counting. Another approach for performing label free quantification is by the extracted ion chromatograms (XIC), and peptide intensities (Ahrné *et al.* 2013). The intensities of a proteins peptides correspond to the amount of protein present in the sample (Bantscheff *et al.* 2012). Intensity based absolute quantification, iBAQ, is a method in which the ion intensities for all peptides from one protein are summed up, and then that sum is normalized by the number of theoretically observable peptides. This will result in what is called an “iBAQ value” which can be used as a quantification index (Schwanhäusser *et al.* 2011). MaxLFQ is another intensity based LFQ method, available in the software package MaxQuant, which calculate an LFQ-intensity for each protein by normalisation of intensities across samples (Cox *et al.* 2014). Another method is the Top3, where only the top three peptide intensities for each protein is used to estimate the protein abundance (Ahrné *et al.* 2013).

An example of the use of LFQ, is the study by Zeiler *et al.* (2014), where copy numbers for the entire murine platelet proteome are determined. The motivation for the study is to gain insight and quantitative knowledge of the platelet proteome. This in turn can aid the search for possible drug targets, involved in cardiovascular disease. The LFQ- method used is iBAQ, with a set of protein standards (SILAC-PrESTs) as calibrants for estimation of absolute amounts for all remaining platelet proteins. They were able to accurately determine copy numbers of all platelet proteins.

### 1.3 Atlas Antibodies

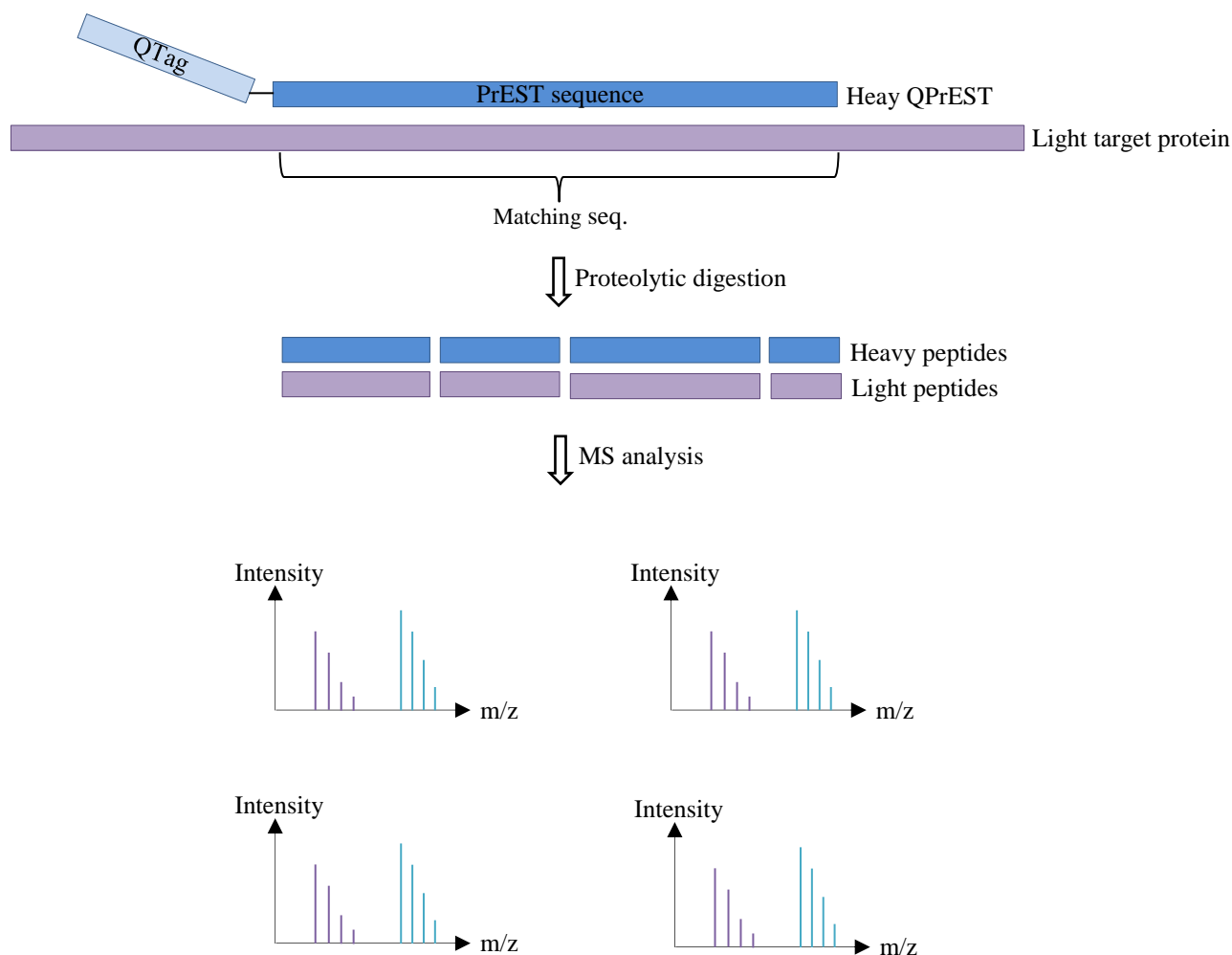
Atlas Antibodies is a rapidly growing company, with products being distributed worldwide. The company is located in Bromma, and was founded in 2006 by researchers working within the Human Protein Atlas (HPA)(Atlas Antibodies 2018a). HPA is a project aimed at mapping the expression and localization of all proteins in the human body, by using



immunohistochemistry. Within the HPA project a process for development and quality control of antibodies was developed, in order to supply high quality antibodies for the project. When a demand for these antibodies appeared from researchers outside the HPA project, Atlas Antibodies was founded to commercialize the antibodies. The first product was the Triple A Polyclonals, which are rabbit polyclonals developed within the HPA project. Triple A Polyclonals are available towards more than 75% of the human protein coding genes (Atlas Antibodies 2018b). In 2012 a new product was developed, namely PrecisA Monoclonals which are mouse monoclonals (Atlas Antibodies 2018c). Currently almost 400 PrecisA Monoclonals are available. The most recently launched products are the PrEST antigens and QPrEST standards, which were both launched in 2014. PrESTs (Protein Epitope Signature Tag) are antigens for generation of antibodies. Each PrEST is a recombinant protein fragment, designed to have lowest possible sequence similarity to any other human protein, in order to give highly specific recognition from its corresponding antibody (Atlas Antibodies 2018d). The QPrESTs are heavy-isotope labeled PrESTs. These heavy labeled protein fragments are used as standards for MS-based quantification of proteins.

### **1.3.1 QPrEST**

QPrEST standards (see Figure 1) are isotope labeled recombinant proteins consisting of ~50-150 amino acids (Boström 2016). The sequence consists of a protein epitope signature tag (PrEST) and a HisABP quantification tag (Q tag) used for purification and quantification. The PrEST part matches a fragment of the target protein, carefully chosen to show as low sequence similarity as possible to other proteins. Each PrEST sequence contains multiple tryptic peptides and may therefore give several peptide ratios (heavy/light) for each target protein which is beneficial for the subsequent quantification. The QPrESTs are spiked into the sample at an early stage, prior to digestion, which reduce the errors that can occur during sample preparation. This is beneficial compared to many of the other labeled standards discussed in section 1.2. that are added later in the workflow (Boström 2016). The QPrESTs are introduced metabolically, and the PrEST sequence is relatively long and matching the target protein, with the same sequences flanking each digestion site. It can therefore be assumed, and has been shown (Atlas Antibodies 2016), that the digestion efficiency of this recombinant protein will be quite similar to that of the target protein. When synthetically produced concatenated peptides are used, the digestion efficiency of the standard may not reflect that of its target protein equally well.



**Figure 1: Schematic structure of a QPrEST, and the use of QPrEST standards in MS analysis.** Heavy QPrEST is mixed with light target protein, digested and analysed by MS. The intensity ratio H/L obtained for each peptide is used for absolute quantification of the light target protein.

## 1.4 Aims and objectives

This project is a pilot study investigating the use of QPrESTs as calibration standards in label-free absolute quantification, and if that use can be beneficial compared to other LFQ methods in terms of quantification accuracy and specificity. A part of this will be to try and find a suitable set of QPrESTs, for future development of a “QPrEST-LFQ kit”, intended for use in human cell lines. Once a set of QPrESTs have been selected, they will be used for quantification and the results evaluated and compared to results achieved without QPrESTs as calibrants. The development of a standard-kit of QPrESTs that can be easily used to achieve better results from absolute LFQ, would be beneficial in many aspects. For Atlas Antibodies, it can lead to a new product in their catalogue. But for customers and researchers, it could lead to better results and a wider applicability of LFQ, which is both a cheaper and less time-consuming method for protein quantification compared to labeled approaches. Progress and successful research is in turn socially beneficial.

## 2 Materials and methods

### 2.1 Production of QPrESTs

A total of 33 QPrESTs were produced according to a protocol developed at Atlas Antibodies. For most QPrESTs the *E.coli* used to express the different QPrESTs (Rosetta *E.coli* cells auxotrophic for Lys and Arg) were already transformed with QPrEST plasmid, and available as glycerol stocks. For those not available as glycerol stock, transformations were made.

Glycerol stocks with *E. coli* cells auxotrophic for Lys and Arg, containing QPrEST plasmid, were streaked out onto Kanamycin/Chloramphenicol, Km/Cm, agar plates (40mg/ml LB Agar, distilled water, 50 µg/ml Km, 10 µg/ml Cm) and incubated o/n, at 37°C. For those QPrESTs not available as glycerol stocks, transformations were made from plasmid preps, already prepared from the different QPrESTs. The plasmid preps were transformed into competent *E. coli* cells thawed on ice. 4µl of plasmid prep was added to an aliquot of competent cells and mixed gently. It was then incubated on ice for five minutes. After incubation, cells were heat shocked for 30 sec in 42°C water bath followed by two minutes of incubation on ice. After the two-minute incubation, 80 µl of room temperature SOC media was added to the cells, which were then incubated in 37°C, at 250 rpm for 1 hour. The cells were then streaked out onto Km/Cm agar plates (40 mg/ml LB Agar, distilled water, 50 µg/ml Km, 10 µg/ml Cm) and incubated over night at 37°C.

The following day a single colony was picked and inoculated in 5 ml of TSB+Y (30 mg/ml Tryptic Soy Broth, 5 mg/ml Yeast Extract, distilled water) with 50 µg/ml Km, 34 µg/ml Cm, in a 50-ml falcon tube. This was done for each plate. The cultures were grown shaking at 180 rpm, 37°C o/n. The next day, QPrEST medium (500 mM Na<sub>2</sub>HPO<sub>4</sub>, 500 mM KH<sub>2</sub>PO<sub>4</sub>, 250 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5% Glycerol, 0.5% Glucose, 2% Lactose, 200 mM MgSO<sub>4</sub>, 50 mM FeCl<sub>3</sub>, 20 mM CaCl<sub>2</sub>, 10 mM MnCl<sub>2</sub>, 10 mM ZnSO<sub>4</sub>, 2 mM CoCl<sub>2</sub>, 2 mM CuSO<sub>4</sub>, 2 mM NiSO<sub>4</sub>, 200 µg/ml of each heavy amino acid (Lys and Arg), 200 µg/ml of the 18 remaining amino acids (light), 50 mg/ml Km and 34 mg/ml Cm), which is the medium used for inducing overexpression, was prepared. For each of the o/n cultures 10 µl of culture was mixed with 10 ml of QPrEST medium in a 100-ml E-flask. The E-flasks were put on shaking at 180 rpm, 37°C for 24 h. OD<sub>600</sub> was measured for all cultures. The cells were harvested after the 24 hours of incubation in the QPrEST medium. The cell cultures were transferred into centrifuge tubes and centrifuged (using a fixed angle rotor) at 2700 x g for 10 min at 4°C. The supernatant was discarded, and the pellet resuspended in 5 ml IMAC lysis buffer pH 8.0 (7 M Guanidinium chloride, 47 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.65 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM Tris-HCl pH 8.0, 100 mM NaCl, distilled water) with 20 mM β-mercaptoethanol. Tubes were put on shaking at 150 rpm, 37°C for 2 h. After lysis the tubes were centrifuged for 40 min, 4°C at 17 100 x g. The crude lysate (supernatant) was transferred to new tubes. The lysate obtained after harvesting and concentration was purified using IMAC on an ASPEC GX-274 robot. 2 ml of HisPur<sup>TM</sup> Cobalt Resin (Thermo scientific) was used for each column. The columns

were washed with 2 column volumes (CV) IMAC wash buffer pH 8.0-8.2 (6 M Guanidinium chloride, 46.6 mM Na<sub>2</sub>HPO<sub>4</sub>, 3.4 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl). The crude lysates were added to the columns, followed by 80-150 CV of IMAC wash buffer. QPrESTs were eluted by adding 3 CV IMAC elution buffer (6 M Urea, 46.6 mM Na<sub>2</sub>HPO<sub>4</sub>, 3.4 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM Imidazole). The collected eluate from the purification was diluted to a final concentration of 1 M Urea by adding 15 ml PBS (2 mM NaH<sub>2</sub>PO<sub>4</sub>, 8 mM Na<sub>2</sub>HPO<sub>4</sub>, 150 mM NaCl). The QPrEST eluates were then concentrated with Pierce concentrators (9 K MWCO) at 4200 rpm, 20°C for approximately 25 min to a final volume of ~2.5-3.5 ml. The concentrated QPrESTs were transferred to 15-ml falcon tubes and centrifuged at 4000 rpm for 5 min at 20°C. The QPrESTs were then transferred to screw cap tubes, except for the last ~500 µl at the bottom of the falcon tubes, which was discarded. The absorbance at 280 nm was measured using Nanodrop (in triplicates), and the concentrations calculated using Lambert-Beers law.

## 2.2 QC of produced QPrESTs

Quality control of the produced QPrESTs was performed in three separate steps, namely molecular weight determination, quantification and purity analysis. Both molecular weight determination and quantification were performed using mass spectrometry. The purity analysis was performed using sodium dodecyl sulfate polyacrylamide gel electrophoresis, SDS-PAGE and image analysis.

### 2.2.1 Molecular weight determination of QPrEST

QPrESTs were thawed and vortexed for 30 s, then centrifuged at 15000 x g for 1 min. 10 µl of QPrEST was added to a well on a 96 well plate. The QPrEST was reduced by adding 0.4 µl 250 mM dithiothreitol (DTT) to the well. The plate was vortexed and spun down, then incubated for 1 h in room temperature. The sample was alkylated by addition of 0.4 µl of 400 mM Iodoacetamide (IAA), and the plate vortexed and spun down then incubated for 30 min at room temperature in the dark. The sample was then diluted by addition of 90 µL 0.1% Formic acid (FA). The samples were then analysed by LC-MS on a Dionex Ultimate 3000 LC system coupled to a Bruker impact II (Q-TOF). The flowrate of the LC was set to 200 µl/min. The buffers used were A (ACN 0.1%FA) and B (0.1%FA) in a 6 min gradient from 4-90% B. The column was a 5cm C4 column (RP-H4). For ionization the standard ESI source is used. The mass range was set to 300-3000 m/z, and the spectra rate at 1Hz. Data acquisition and analysis was performed using the software BioPharma Compass.

### 2.2.2 QPrEST Quantification

QPrESTs were thawed and vortexed for 30 s followed by centrifugation at 15000 x g for 1 min. 20 µl of 1M DTT was diluted with 1180 µl 0.1 M NH<sub>4</sub>HCO<sub>3</sub>, to a final DTT concentration of 17 mM. For all QPrESTs (in triplicates) 30 µl of the diluted DTT was added to a well on a 96 well plate followed by 3.9 µl of light Q-tag and then 5 µl of QPrEST. Three wells with a control, with heavy Q-tag instead of QPrEST, were also prepared as described above. The plate was vortexed and spun down, then incubated in room temperature for 1 h.

After incubation, 2  $\mu$ l 400 mM IAA was added to each well. The plate was vortexed and spun down, then incubated at room temperature in the dark for 30 min. After the 30 min incubation 1  $\mu$ l of 100 ng/ $\mu$ l trypsin was added to each well. The plate was vortexed and spun down, then incubated for 16 h in room temperature. To a new plate, 120  $\mu$ l of 0.1% FA was added to the wells. The samples from the incubated plate were then moved to the new plate and mixed with the FA. The plate was vortexed and spun down and kept in fridge at +8°C until MS analysis. The samples were then analysed by LC-MS/MS on the same LC-MS setup as in 2.2.1. The flowrate of the LC was set to 180  $\mu$ l/min. The buffers used was A (ACN 0.1%FA) and B (0.1%FA) in a 6 min gradient from 15-35% B. The columns used were a C18 trap column (PepMap™ C18, nanoViper™) followed by a 15cm C18 column (PepMap™ C18, nanoViper™). A top three method was used, with mass range set to 150-2200 m/z, and the spectra rate 2 Hz for MS1 and 6 Hz for MS2. Data acquisition and analysis was performed using the software BioPharma Compass. Data acquisition was performed by Hystar, and data analysis by Proteinscape. A Q-tag peptide was used for the quantification, where the median of three replicates was used. The variation of the replicates was calculated.

### 2.2.3 QPrEST Purity analysis

QPrESTs were thawed and vortexed for 30 s, then centrifuged at 15000 x g for 1 min. Samples were prepared in 1.5-ml Eppendorf tubes by mixing of 1  $\mu$ g QPrEST, 7.5  $\mu$ l 4x Laemmli Sample Buffer (BioRad), 1  $\mu$ l 1 M DTT and milliQ water up to 30  $\mu$ l. Samples were heated in a heat block at 95°C for five minutes. After heating they were centrifuged for 1 minute at 12000 rpm. They were then loaded onto a precast gel (4-20% Criterion™ TGX precast gel, BioRad), assembled in a gel tank filled with cold 1xTGS (tris, glycerol, SDS) running buffer (0.5l 10x TGS BioRad, 4.5l milliQ water). In an empty lane, 10  $\mu$ l of molecular weight marker (PageRuler™ Plus Prestained Ladder, Thermo scientific) was loaded. The gel was then run at 200 V for 40 min. After running, the gel was disassembled and put in a plastic tray. The gel was covered in with milliQ and put on gentle shaking for 15 min. The milliQ water was replaced every 5 min. After rinsing, the water was removed, and the gel was stained with ~20 ml of SimplyBlue SafeStain (life technologies) for 60 min on gentle shaking. The gel was then washed with milliQ water, for 2 h at gentle shaking. The water was replaced after 1 h.

The gel was then photographed by a UV camera, and the band-intensities of each lane on the gel estimated using the software ImageLab. Image presentation and lane- and band-detection was manually adjusted for an optimized image analysis (e.g. removing falsely detected bands like air bubbles and adjusting lane width etc.).

## 2.3 Preparation of cell lysate for MS

### 2.3.1 Cell lysis

Cell pellets from HeLa and A549 cells were thawed on ice. To each pellet, 2 ml of lysis buffer (PBS, 1% SDC) was added for every  $10^7$  cells. The pellet was resuspended using a 10 ml syringe with a 23G  $\frac{3}{4}$  needle. The cells were passed through the needle 10 times. After

resuspension, the cells were sonicated (40% amplitude, 1s pulse and 1s rest) for 10 to 60 s, until they foam. The lysates were then centrifuged at 17100 x g for 30 min at 4°C. The supernatants were transferred into new tubes and stored at -80°C as aliquots of 10<sup>5</sup> cells.

### **2.3.2 Bicinchoninic assay**

Bovine serum albumin (BSA) standard was diluted, in triplicates, to 1.0, 0.80, 0.60, 0.40 and 0.20 mg/ml. Cell lysates of HeLa and A549 were diluted, in triplicates, to 1:2, 1:5 and 1:10 in PBS. 25 µl from each of the standard and lysate samples were added to separate wells on a 96 well plate. BCA reagents A and B were mixed 50 parts A to 1 part B. 200 µl of BCA reagent mixture was added to each well. The plate was sealed and incubated at 37°C for 30 min. The absorbance of each well at 570 nm was measured using a Tecan sunrise instrument. A standard curve was created based on the absorbance of the BSA standard samples. The concentration of the cell lysates was calculated using the linear equation obtained from the standard curve.

## **2.4 Identification of quantotypic peptides**

### **2.4.1 Sample preparation**

QPrESTs were thawed and vortexed for 30 s, then centrifuged at 15000 x g for 1 min. A mastermix of QPrESTs with PBS, 1 M Urea, was prepared to a final concentration of 0.5 pmol/µl for each QPrEST. Two dilutions of 1:10 and 1:100 were prepared from the Qmix (dilution in 25mM Ambic). Qmix, and lysate samples (HeLa and A549 cell lysates) with spiked in Qmix in three different concentrations, were prepared in duplicates (Qmix samples) or triplicates (lysate samples) for MS analysis by Sodium deoxycholate, SDC, digestion. For Qmix samples, 10 µl of undiluted Qmix were used, and for lysate samples a lysate volume corresponding to 10<sup>5</sup> cells were used. The samples were prepared in 1.5-ml Eppendorf tubes. To each sample 10% SDC in 25 ml Ambic was added in the volume needed to give a final SDC percentage of five in each sample. To lysate samples, Qmix was spiked in to a final concentration of 1, 0.1 or 0.01 pmol/QPrEST. To all samples, 2 µl of 100 mM DTT in 25 mM Ambic was added, and the samples incubated at 95°C for 10 min. After incubation, 2 µl of 600 mM IAA in 25 mM Ambic was added to each sample. The samples were then incubated in darkness at room temperature for 30 min. After incubation the samples were diluted with 25 mM Ambic to a final volume of 200 µl, and a final SDC concentration of 1%. 15 µl of Trypsin in 25 mM Ambic was added to give a 1:20 or 1:50 enzyme to substrate ratio for the Qmix and lysate samples respectively. The samples were incubated overnight (~16 h) at 37°C. The following day, 115 µl of 0.1% TFA was added to each sample, and then incubated at room temperature for 30 min. The samples were then centrifuged at 13,000 x g for 10 min. The supernatants were transferred to new tubes, and acidity checked for pH < 3, by pH indicator paper. The samples were then desalted using inhouse made stage tips with three layers of C18 material. The stage tips were conditioned by adding 150 µl of MeOH, followed by 150 µl 80% ACN 0.5% HAc and finally 150 µl 0.5%. After each addition the tips were centrifuged for 1.5 min at 3000 x g, before adding

the next component. Following conditioning, the collected sample-supernatants were added to the stage tips, one for each sample, and centrifuged at 3000 x g for 1.5 min. The tips were then washed by addition of 150 µl 0.5% HAc and centrifugation at 3000 x g for 1.5 min. The tips were then moved to new collection tubes, and the sample eluted by two rounds of adding 20 µl 80% ACN 0.5% HAc and centrifuging at 3000 x g for 1 min. The eluates were then dried in a speed vac for 15 min. Dried desalted samples were dissolved in 0.1% FA to give a final concentration of 50 fmol per QPrEST per microliter for Qmix samples, and 0.25 µg protein per microliter for lysate samples. The dissolved samples were moved to a 96 well plate and analysed by LC-MS/MS on a Dionex Ultimate 3000 coupled to a Thermo QE orbitrap instrument. The flowrate of the LC was set to 300 µl/min. The buffers used was A (ACN 0.1% FA) and B (0.1% FA) in a 90 min gradient from 4-35% B. The columns used was a C18 trap column (PepMap™ C18, nanoViper™) followed by a 50 cm C18 column (Acclaim PepMap™ RSLC nanoViper). An EASY-spray source was used. For the MS the resolution was set to 70000, the AGC target 3e6, maximum IT 100 ms and scan range 400-1600 m/z. For the MS/MS a top 10 method was used, with resolution set to 17500, AGC target 5e4, maximum IT 100ms and 1.6 m/z. Data acquisition and analysis was performed using the software Xcalibur.

#### **2.4.2 Data analysis**

Raw data acquired by Xcalibur was imported into MaxQuant (version 1.5.3.30) and searched against the human proteome (UniprotKB). On Group specific parameters, multiplicity was set to 2, and Heavy labels Arg10 and Lys8 chosen. Enzyme was set to Trypsin/P. On Global parameters Min. peptide length for unspecific search was set to 6. All other settings were left at default.

Results were also imported into skyline and analysed. A library was built from the msms file obtained by MaxQuant, and FASTA files for all QPrEST sequences were imported, to create a protein list with all tryptic peptides of the QPrESTs, including those with one missed cleavage. After that, the acquired raw data files were imported as results and matched to the library. All peptides were manually evaluated, and only those with good peaks and high signal for both heavy and light peptide were kept. For the peptides that remained, heavy to light ratios obtained by skyline were compared to those obtained by MaxQuant, and only peptides that were identified by both softwares, and showed similar results were deemed as “proteotypic”.

### **2.5 Preparation of adjusted QPrEST-LFQ mix and samples for quantification**

Once a list of proteotypic peptides had been obtained, the heavy to light ratios for peptides from each QPrEST were used to create a new mastermix of QPrESTs. The average of H/L ratios for the proteotypic peptides were calculated for each QPrEST, and the amount of QPrEST needed to achieve a ratio close to one for both tested cell lines (Hela and A549) determined. For the preparation of the QPrEST mastermix, QPrESTs were thawed and

vortexed for 30 s, then centrifuged at 15000 x g for 1 min. The mastermix (QPrESTs in PBS 1 M Urea) was then prepared with adjusted amounts for all QPrESTs. New lysate samples were prepared for MS, according to the same protocol described in 2.4, spiked with the amount of adjusted QPrEST mastermix calculated to give the appropriate H/L ratio. Lysate from five human cell lines were used, namely A549, Caco-2, Hek293, HeLa and U2OS. All cell lines were spiked with the same amount of QPrEST mastermix, and the samples prepared in triplicates. The samples were then subjected to LC-MS/MS by the same method as described in 2.4. Once again, the data acquired was analysed using MaxQuant and Skyline. This time though, only the proteotypic peptides were considered. The results in skyline were manually evaluated and compared to the results obtained by MaxQuant. Peptides that were readily identified in all cell lines by both methods, with H/L ratios that concurred between both methods were listed as “quantotypic” for use in quantification.

## 2.6 Quantification of proteins in human cell lines

Quantification of proteins in the tested cell lines was performed using H/L ratio for the target proteins of all the QPrESTs used in the adjusted mastermix. LFQ was then performed for six of the QPrEST target proteins, by different LFQ methods. Quantification was thus performed in six different ways for each cell lysate. Firstly, by labeled quantification of all QPrEST target proteins, then for six of the targets, by five different LFQ approaches, hereafter referred to as maxLFQ-BCA, iBAQ-BCA, QPrEST-maxLFQ, QPrEST-iBAQ and QPrEST intensity.

For the labeled quantification the amount of spiked QPrEST was compared to the measured H/L ratio of its quantotypic peptides, to calculate the amount of its light target protein, resulting in an absolute quantification of the target. For the LFQ methods, protein quantification indices, PQIs, for all identified proteins in the lysates were obtained by MaxQuants built in functions “maxLFQ” and “iBAQ”. The raw data files for each cell lysate were grouped together into separate parameter groups prior to the search.

For the first two LFQ methods, maxLFQ-BCA and iBAQ-BCA, a “PQI total” was calculated for each cell line, by summing all protein PQIs obtained in the cell line, excluding PQIs of proteins marked as common contaminants. Protein quantities were then estimated by dividing the individual PQI for each target protein by the total PQI, and then multiplying the ratio obtained by the total amount of protein,  $C_{p,tot}$ , in the cell lysate that was used to prepare the sample (measured by BCA), see eq. 1. The calculated amount of protein was then multiplied by its molar weight to give an amount in pmol.

$$\frac{PQI_{target}}{PQI_{tot}} \cdot C_{p,tot} = C_{p,target} \quad (\text{eq.1})$$

For the other three LFQ approaches, QPrEST-maxLFQ, -iBAQ and -intensity, all but six QPrESTs were used to create a calibration curve from which absolute protein amounts could



be inferred based on each proteins PQI. The common logarithm of the PQIs (or the measured intensity) for QPrEST target proteins were plotted against the  $\log_{10}$  of the absolute concentration (pmol) of the target protein (earlier determined by H/L ratios of the QPrESTs and respective targets). A linear curve was fitted to the data, and the linear equation obtained was used to quantify the remaining six QPrEST target proteins that had not been used to create the calibration curve.

Finally, the absolute quantities obtained by the different LFQ methods were compared to those obtained by the labeled quantification. The quantification results were also compared to RNA data (tpm) for the proteins (Uhlen *et al.* 2015).

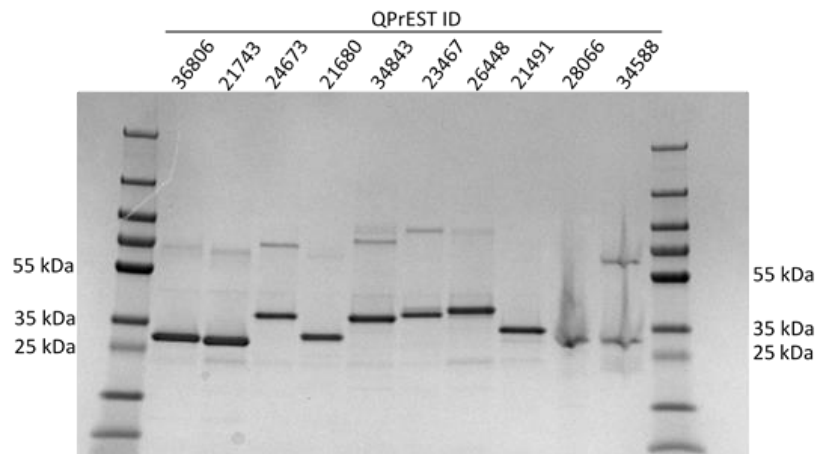
## 3 Results

### 3.1 Production and QC of QPrEST proteins

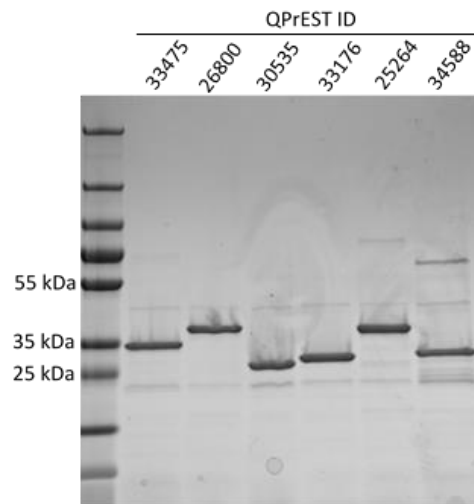
A total of 33 QPrESTs were produced in an *E. coli* strain auxotrophic for Lysine and Arginine (Matic *et al.* 2011) as described by (Studier 2005), and purified using IMAC. The produced QPrESTs then went through three stages of quality control, purification, molecular weight determination and quantification. The purity of purified QPrESTs was evaluated by SDS-PAGE analysis and comparing band intensity for the different protein bands for each QPrEST. The limit of approval was set to 85% purity, i.e. 85% of the intensity must come from QPrEST main band, dimers and trimers. In table 1 the estimated purity for the produced QPrESTs are listed. Figure 2-5 show the gels for all QPrESTs that met the purity criteria. The main band of each QPrEST was found at a weight corresponding well to its theoretical weight, for all but QPrEST37513, which is found at a higher weight than expected. In total, 25 of the 33 produced QPrESTs met the purity criteria. The theoretical weight for each QPrEST is listed in table 1. Determination of molecular weight was performed by MS, in order to verify the molecular weight of the produced QPrESTs when compared to the theoretical weight. The experimentally determined weights are listed in table 1. QPrEST20919 was not identified in the sample, and thus experimental weight data is absent. For QPrEST25076 the difference between the theoretical weight and the experimentally determined is ~59 Da. This difference corresponds well to the weight of iodoacetamide, at ~57 Da. In figure 6, an example of a deconvoluted MS spectra (for QPrEST21743) can be seen. There is a peak at 27830 with intensity  $10^4$ , which correspond well to the theoretical weight of 27832 Da. A known amount of light Q-tag was spiked into the produced QPrEST and concentration determined by MS and analysis of the heavy to light ratio of the Q-tag *ISEATDGLSDFLK*. The average concentration from three replicates, and the coefficient of variation(CV) are listed in Table 1. An example of the obtained MS spectrum for QPrEST26210 is presented in figure 7. In total, 25 QPrESTs met all the established criteria.

**Table 1: Results from quality control of produced QPrESTs. The theoretical molecular weights are for the alkylated proteins. \*Approved QPrESTs that meet all established QC-criteria.**

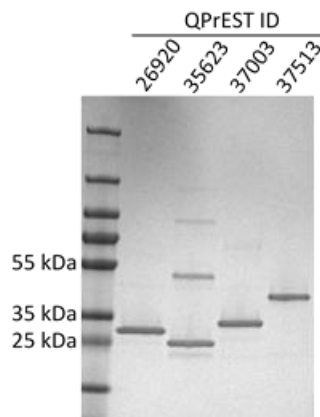
QPrEST ID	Purity %	MW theoretical	MW experimental	Conc(μM)	Conc CV(%)
* <i>QPrEST 33176</i>	90.03	27256.20	27254.14	28.7	0.4
* <i>QPrEST 34588</i>	89.04	30033.33	30031.00	10.4	0.9
* <i>QPrEST 26800</i>	94.66	31513.41	31510.97	22.1	1.7
* <i>QPrEST 26448</i>	90.75	30927.49	30925.94	9.2	4.1
<i>QPrEST 38329</i>	60.98	32073.66	32070.85	8.5	14.3
* <i>QPrEST 30535</i>	85.21	27985.24	27983.89	18.0	4.7
* <i>QPrEST 33475</i>	86.65	32680.14	32677.62	12.8	2.5
* <i>QPrEST 25264</i>	87.76	34846.06	34843.06	18.6	1.4
* <i>QPrEST 28066</i>	100	29062.88	29060.56	7.5	2.4
<i>QPrEST 20919</i>	55.75	34858.79	-	6.0	6.7
<i>QPrEST 29980</i>	53.24	27588.20	27589.11	15.6	0.6
* <i>QPrEST 37003</i>	100	30942.00	30939.56	16.6	0.5
* <i>QPrEST 23467</i>	98.20	32302.90	32300.18	20.6	1.4
* <i>QPrEST 21680</i>	100	28765.24	28762.88	31.3	6.2
* <i>QPrEST 37513</i>	96.12	30370.51	30368.16	20.8	9.1
<i>QPrEST 26516</i>	82.00	32374.79	32372.49	6.4	2.1
<i>QPrEST 37494</i>	83.84	34984.30	34982.31	5.0	3.4
* <i>QPrEST 21743</i>	89.74	27832.01	27830.02	15.9	1.1
* <i>QPrEST 36806</i>	98.76	28567.06	28564.64	16.6	2.0
* <i>QPrEST 38984</i>	86.31	29575.10	29572.35	17.8	1.2
* <i>QPrEST 25096</i>	100	28791.05	8788.917	21.6	4.9
* <i>QPrEST 26920</i>	100	29598.49	29596.10	21.0	6.7
* <i>QPrEST 24673</i>	97.02	34567.02	34564.59	20.4	2.7
* <i>QPrEST 21491</i>	90.20	27645.52	27643.47	15.9	1.9
* <i>QPrEST 35623</i>	91.06	25870.92	25869.79	19.1	2.0
* <i>QPrEST 36934</i>	100	31659.46	31661.58	19.6	5.6
* <i>QPrEST 23845</i>	100	32256.69	32258.92	23.2	7.0
* <i>QPrEST 26210</i>	88.57	29068.38	29070.34	19.2	1.8
* <i>QPrEST 25076</i>	83.68	28783.77	28724.72	29.2	1.5
<i>QPrEST 20566</i>	100	33939.93	33937.58	4.6	9.7
<i>QPrEST 25400</i>	73.94	33473.40	33471.28	16.7	2.4
* <i>QPrEST 34843</i>	92.42	32623.00	32620.96	7.7	2.5
<i>QPrEST 24883</i>	69.34	31433.46	31431.29	7.4	2.7



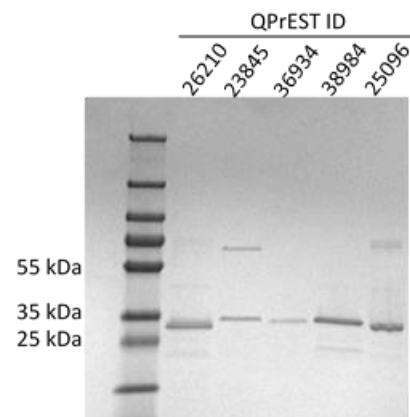
**Figure 2:** Gel from SDS PAGE analysis of produced QPrESTs. From left to right: QPrEST 36806, 21743, 24673, 21680, 34843, 23467, 26448, 21491, 28066 and 34588.



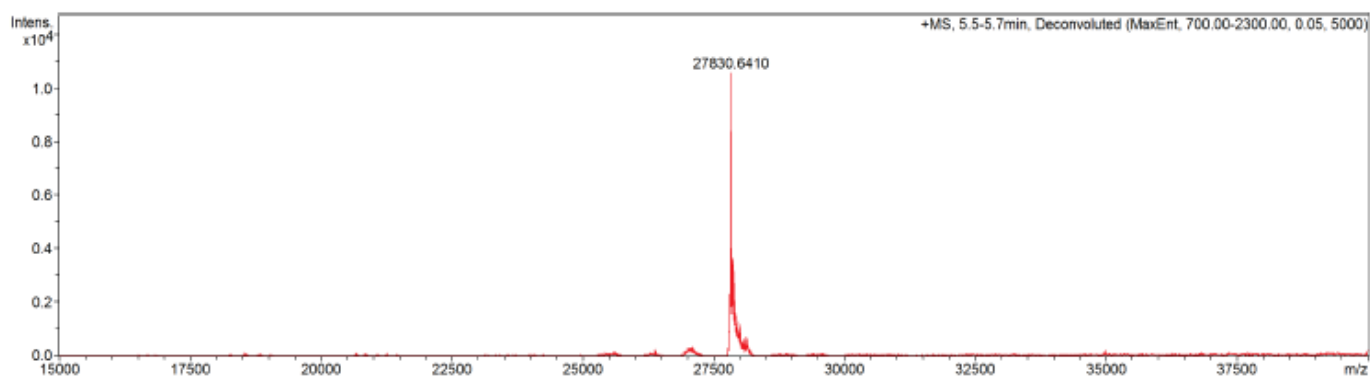
**Figure 3:** Gel from SDS PAGE analysis of produced QPrESTs. From left to right: QPrEST 33475, 26800, 30535, 33176, 25264 and 34588.



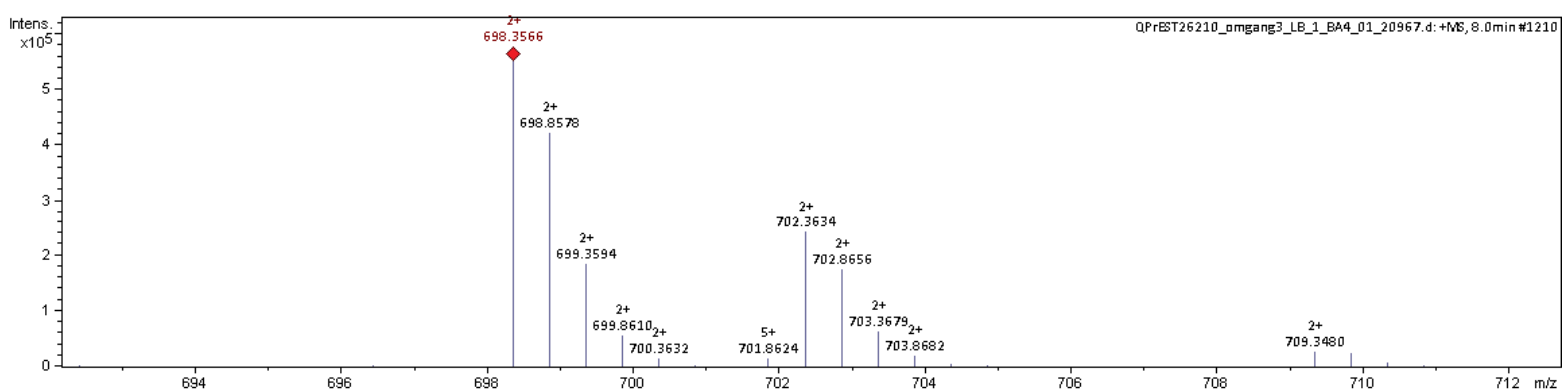
**Figure 4:** Gel from SDS PAGE analysis of produced QPrESTs. From left to right: QPrEST 26920, 35623, 37003 and 37513.



**Figure 5:** Gel from SDS PAGE analysis of produced QPrESTs. From left to right: QPrEST 26210, 23845, 36934, 38984 and 25096.



**Figure 6: Deconvoluted MS Spectra for QPrEST 21743**



**Figure 7: MS spectrum for heavy and light Q-tag peptide ISEATDGLSDFLK, from quantification of QPrEST26210.**

### 3.2 Identification of Quantotypic peptides

Cell lysate samples, with concentrations determined by BCA, were spiked with different concentrations of QPrEST mastermix and analysed by LC-MS/MS. The acquired data was analysed using the softwares MaxQuant and Skyline. A QPrEST mastermix was prepared, containing 25 of the QPrESTs produced in this project, that met the QC criteria (see table 1). The mix was prepared to contain an amount of 100 pmol of each QPrEST, and two dilutions of 1:10 and 1:100 were made. The mastermix or diluted mastermix was spiked into lysate samples to give a final concentration of 1, 0.1 or 0.01 pmol per QPrEST in each sample (10  $\mu$ l lysate, corresponding to 10<sup>5</sup> cells). Analysis of the MS data for these samples revealed that only a few peptides received a H/L ratio by MaxQuant. For most peptides only the heavy peptide was identified, or no peptide was identified at all. In skyline only a few peptides showed good peaks for both heavy and light peptides, while most light peptide peaks were undistinguishable from the background. The only QPrEST that had a good peak for both heavy and light peptides, was QPrEST26210, targeting AIMP1 gene product. The criteria for a peptide to be deemed as proteotypic was that a H/L ratio should be identified in both cell lines. It should also show the same H/L ratio in skyline as in MaxQuant. Since the

majority of produced QPrESTs did not give a good signal, the only proteotypic peptides that could be selected from these 25 QPrESTs were those of QPrEST26210.

A new set of QPrESTs to use for the remainder of the project was therefore chosen, from QPrESTs available in the company stock. From the QPrESTs in stock, those that showed the least variation in tpm data for different human cell lines were chosen. To further ensure that the light target peptides of these QPrESTs could be identified in the lysate samples, skyline was used. FASTA files for all the new QPrEST sequences were imported, to create a new protein list with tryptic peptides, and raw data files from the previous LC-MS experiment were used to identify proteotypic peptides within the new set of QPrEST standards. The data was manually evaluated, and only QPrESTs that showed more than one light target peptide with good signal were chosen from the stock. The chosen new QPrESTs were used for a new mastermix, spiked into lysate samples in different concentrations as before, and analysed in the same manner. An example from skyline of a proteotypic peptide is shown in figure 8. The estimated heavy to light ratios (H/L) for the peptides deemed as proteotypic are presented in table 2. The H/L ratios for the samples spiked with 1 pmol QPrEST varied between 0.02-49.00 with the majority between 1-10. For the samples spiked with lower concentration of QPrEST only a few obtained a H/L ratio due to the heavy QPrEST not being identified in the sample. Proteotypic peptides could be identified in all but one of the chosen QPrESTs.

From the proteotypic peptides identified in the new QPrEST mastermix, H/L ratios were used to determine what amount of each QPrEST to spike into the lysate samples to achieve a H/L ratio close to one for both cell lines. From that, a new QPrEST mastermix with adjusted amounts of QPrESTs was prepared (see Table A1 in appendix). Lysate samples of five different human cell lines were then spiked with the adjusted QPrEST mastermix, and analysed by LC-MSMS, followed by data analysis of quantotypic peptides in MaxQuant and skyline. The peptides were evaluated once more, and only peptides giving a good signal and concurring H/L ratios from both skyline and MaxQuant were selected as quantotypic and used for later quantification of the target proteins. The final list of QPrESTs and their quantotypic peptides, with corresponding H/L ratios is displayed in table 3. There were 13 QPrESTs that had more than one proteotypic peptide, of which 9 were used to create the QPrEST-LFQ calibration curves.

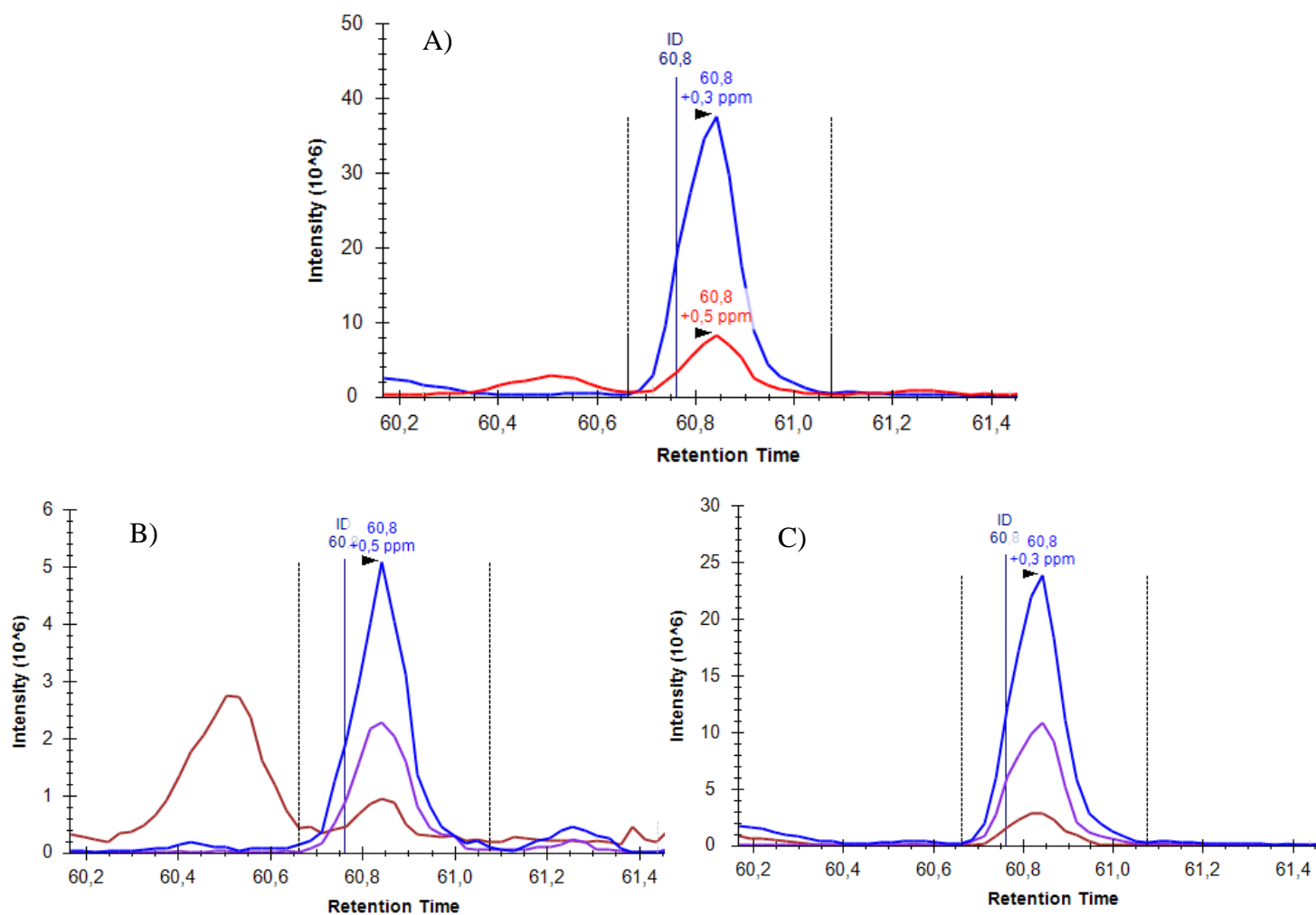


Figure 8: Chromatograms from Skyline of proteotypic peptide IPLNDLFR of QPrEST 23147. A) Heavy (blue) and Light (red) peptides. B) Precursors of Light Peptide. C) Precursors of Heavy peptide.

**Table 2: H/L ratios of proteotypic peptides.**

QPrEST ID	Gene	Peptide seq.	pmol spiked	H/L HeLa	H/L A549
QPrEST20480	PDLIM1	GHFFVEDQIYCEK	1	18.8	5.1
		CGTGIVGVFVK	1	28.8	4.9
QPrEST22716	GSN	AQPVQVAEGSEPDGFWEALGGK	1	7.4	6.9
QPrEST25474	RPS28	EGDVLTLLESEREAR	1	1.0	0.7
		GPVREGDVLTLLESEREAR	1	1.0	0.6
		VEFMDDTSR	1	1.0	0.7
		EGDVLTLLESER	1	1.0	0.6
QPrEST20520	ATOX1	LGGVKYDIDLPNKK	1	10.2	5.4
QPrEST25396	RAB10	AFLTAEILR	1	5.6	1.6
			0.1	0.1	-
		KTPVKEPNSENVDISSGGGVGTGWK	1	6.9	1.7
QPrEST21629	GGCT	NPSAAFFCVAR	1	19.1	23.6
QPrEST23174	LAMP2	IPLNDLFR	1	4.3	2.7
QPrEST33555	AKT1S1	SLPVSVPVWGFK	1	49.0	30.6
QPrEST35130	TXNDC17	SWCPDCVQAEPVVR	1	5.57	2.44
			0.1	0.08	-
		YEEVSVSGFEEFHR	1	7.71	3.30
		TIFAYFTGSK	1	7.96	3.42
QPrEST26315	TARDBP	FGGNPGGFGNQGGFGNSR	1	8.83	6.35
			0.1	0.12	-
		GISVHISNAEPK	1	8.74	7.25
QPrEST24019	PEBP1	APVAGTCYQAEWDDYVPK	1	0.37	0.29
		GNDISSGTVLSDYVSGGPPK	1	0.61	0.44
QPrEST33915	OTUB1	EYAEDDNIYQQK	1	6.27	2.81
		IQQEIAVQNPLVSR	1	6.63	3.08
		LELSVLYK	1	6.68	3.28
QPrEST25436	YWHAG	YLAEVATGEKR	1	2.26	1.62
			0.1	-	0.06
		YLAEVATGEK	1	2.44	1.50
QPrEST23611	VDAC1	EHINLGCDMDFDIAGPSIR	1	1.79	1.22
		LTFDSSFSPNTGKK	1	1.72	1.44
			0.1	-	0.02
		VTQSNFAVGKYK	1	1.73	1.43
			0.1	0.06	0.05
		WNTDNTLGTEITVEDQLAR	1	2.27	1.87
			0.1	0.06	0.04
		LTFDSSFSPNTGK	1	2.28	1.87
QPrEST23714	PIIB	DKPLKDVIADCGK	1	0.91	0.69
		DTNGSQFFITTVK	1	1.17	1.02
		IEVEKPFAlAKE	1	1.22	1.06
		VLEGMEVVR	1	1.57	1.32
QPrEST22627	GAPDH	WGDAGAeyVVESTGVFTTMEK	1	0.11	0.10
		LVINGNPITIFQERDPSK	1	0.14	0.10

**Table 3: Final list of QPrESTs H/L ratios for quantotypic peptides, in different human cell lines. \*QPrESTs used QPrEST-LFQ calibration curves.**

QPrEST ID	Peptide seq.	H/L HeLa	H/L A549	H/L Caco-2	H/L Hek293	H/L U2OS
<i>QPrEST26315*</i>	FGGNPGGFGNQGGFGNSR	6.10	4.49	3.32	3.96	7.21
	GISVHISNAEPK	5.76	4.08	2.77	3.44	8.53
<i>QPrEST33555</i>	SLPVSVPVWGFK	3.35	1.77	2.00	-	4.34
<i>QPrEST22881</i>	DQQEAAALVDMVNDGVEDLR	0.89	0.66	0.19	1.43	1.52
<i>QPrEST25436*</i>	YLAEVATGEK	10.06	6.69	6.70	11.56	20.17
	YLAEVATGEKR	9.59	6.66	5.93	12.76	23.24
<i>QPrEST23174</i>	IPLNDLFR	3.98	2.40	1.82	3.11	6.80
<i>QPrEST22716</i>	AQPVQVAEGSEPDGFWEALGGK	5.75	5.61	2.22	8.17	5.76
<i>QPrEST23611*</i>	LTFDSSFSPNTGKK	2.01	1.61	0.69	1.68	3.77
	VTQSNFAVGKY	1.88	1.58	0.66	1.57	4.56
<i>QPrEST24019</i>	APVAGTCYQAEWDDYVPK	1.09	0.86	0.61	0.73	2.70
<i>QPrEST25474</i>	EGDVLTLLESER	1.59	1.11	0.93	1.08	2.93
	EGDVLTLLESEREAR	1.70	1.20	0.93	1.15	3.18
	GPVREGDVLTLLESER	1.69	1.10	1.03	1.18	3.19
	VEFMDDTSR	1.84	1.23	0.92	1.80	3.08
<i>QPrEST23714*</i>	DKPLKDVIIADCGK	1.32	1.19	0.62	1.72	1.87
	DTNGSQFFITTVK	1.73	1.60	0.87	-	2.36
	IEVEKPFAIAKE	2.13	1.78	1.02	2.63	2.25
<i>QPrEST22627*</i>	LVINGNPITIFQERDPSK	1.94	1.35	0.96	3.20	2.32
	WGDAGAEYVVESTGVFTTMEK	1.79	1.31	0.86	2.63	2.07
<i>QPrEST21629</i>	LQDFKLDFGNSQGK	4.33	5.65	2.22	2.74	8.76
	NPSAAFFCVAR	4.57	5.34	2.70	2.89	8.83
<i>QPrEST33915*</i>	EYAEDDNIYQQK	7.61	3.42	2.97	5.24	10.10
	IQQEIAVQNPLVSR	7.72	3.46	3.08	4.60	14.44
	LELSVLYK	7.75	3.79	3.48	4.81	16.14
<i>QPrEST25396</i>	AFLTLAEDILR	5.65	1.67	1.75	4.73	6.35
	KTPVKEPNSENVDISSGGGVGTGWK	7.42	2.00	2.20	5.72	8.57
<i>QPrEST20520</i>	LGGVKYDIDLPNKK	6.74	3.72	1.47	4.60	4.51
	VCIESEHSMDTLLATLKK	5.40	2.95	1.34	4.02	4.43
<i>QPrEST20480</i>	GHFFVEDQIYCEK	6.27	1.84	0.87	2.50	-
<i>QPrEST22812</i>	GLVLGPIHK	0.20	0.17	0.11	0.97	0.89
<i>QPrEST35130*</i>	SWCPDCVQAEPVVR	6.49	2.67	5.97	4.66	11.90
	TIFAYFTGSK	9.10	3.80	5.89	6.57	15.57
	YEEVSVSGFEEFHR	9.25	3.91	7.39	6.02	14.20
<i>QPrEST26210</i>	IWEQIQPDLTNDECVATYK	5.93	4.29	2.38	4.50	8.66
<i>QPrESTH3*</i>	EIAQDFKTDLR	1.23	4.91	0.89	0.40	1.26
	YRPGTVALR	1.08	5.39	0.80	0.36	1.13
<i>QPrESTH4*</i>	KTVTAMDVVYALKR	2.01	9.60	1.47	0.67	1.69
	TVTAMDVVYALK	1.81	9.30	1.46	0.57	1.75
	TVTAMDVVYALKR	1.81	8.80	1.39	0.71	1.70
	VFLENVIR	1.74	8.44	1.38	0.56	1.68



### 3.3 LFQ

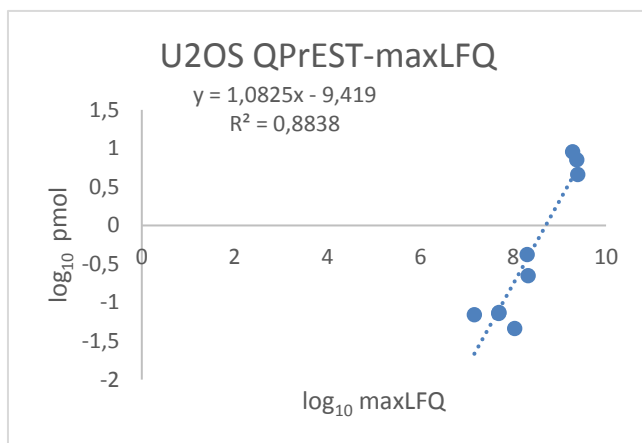
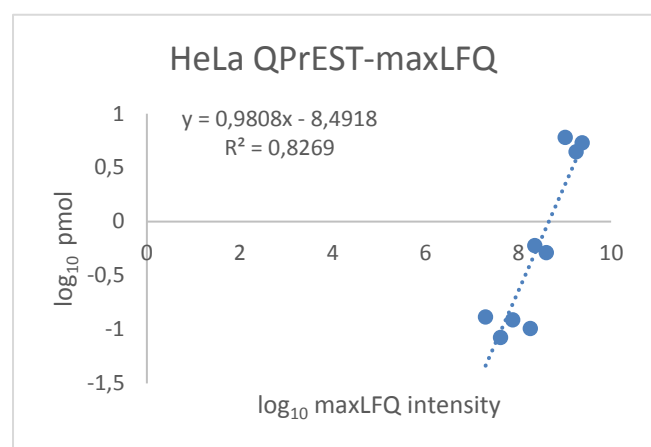
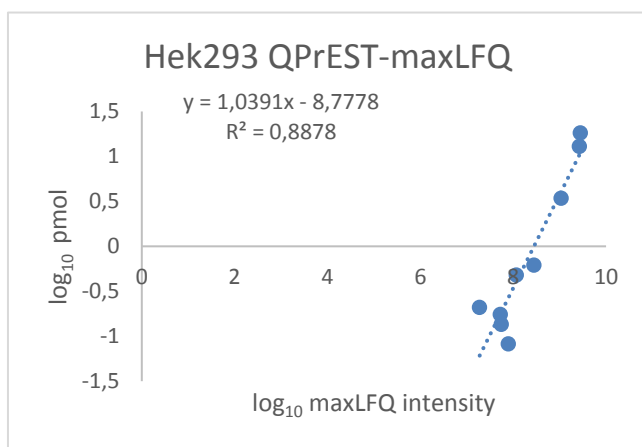
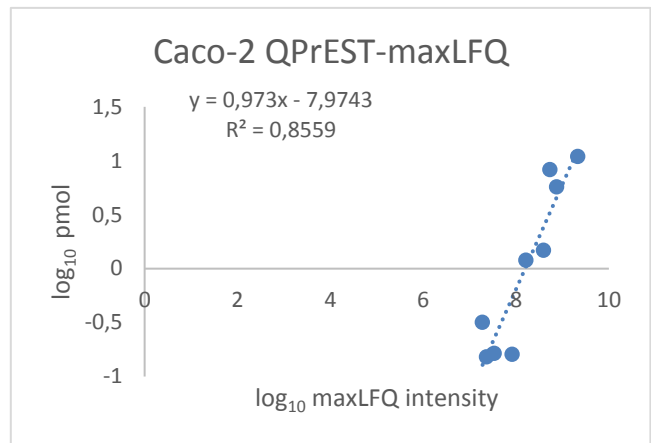
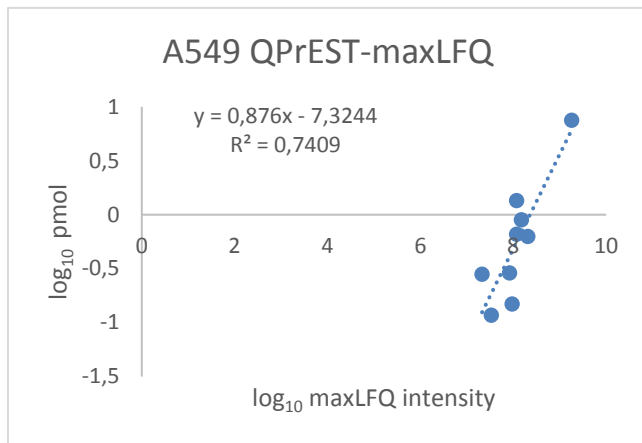
Quantification of all QPrEST target proteins was performed by H/L ratios and by five different LFQ-methods. Two methods used the PQI obtained by maxLFQ and iBAQ (built into MaxQuant) to estimate the absolute amount of protein based on measured protein concentration in the cell lysates. The other three LFQ methods used the concentrations of target proteins, estimated by a method with high accuracy (QPrEST-H/L) as calibration standards to normalize the intensities of remaining proteins, and by that achieve better quantification of them, compared to LFQ without using calibration standards (see Figure 9 - 11). Only target protein concentrations obtained by H/L quantification of QPrESTs with quantotypic peptides were used to create the curve. From the linear equation, the absolute quantity (pmol) was calculated for six QPrESTs not used to create the regression model. The linear curves obtained by different methods show  $R^2$  and linear equations. The result obtained for the six QPrEST targets quantified by all methods, are summarised in table 5. The dynamic range of the HeLa-, A549-, Caco-2-, Hek293- and U2OS- curves were ~0.05-7.62 pmol, ~0.04-7.50 pmol, ~0.08-15.58 pmol, ~0.06-5.50 pmol and ~0.02-4.55 pmol respectively. The quantification results for all QPrEST targets from labeled quantification are presented in table 4, along with reported RNA sequencing (tpm) data for each protein. The proteins with a low tpm often show a low H/L-estimated concentration, and vice versa. The estimated concentrations from the three QPrEST-LFQ methods used are more in agreement with those estimated by H/L-quantification, than the maxLFQ-BCA and iBAQ-BCA are.

**Table 4: QPrEST-H/L quantification results, in pmol, for target proteins in five human cell lines along with RNA sequencing data (tpm) for each target.**

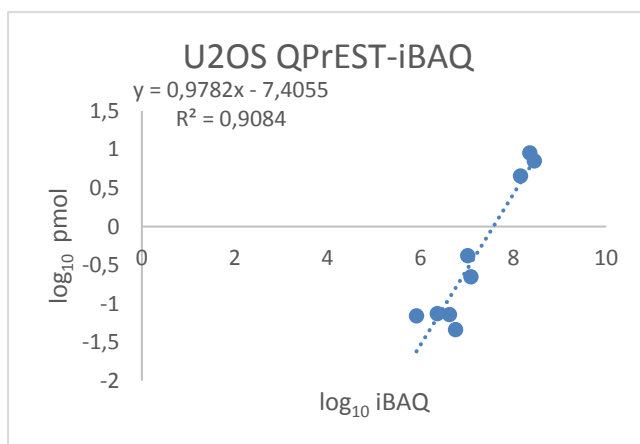
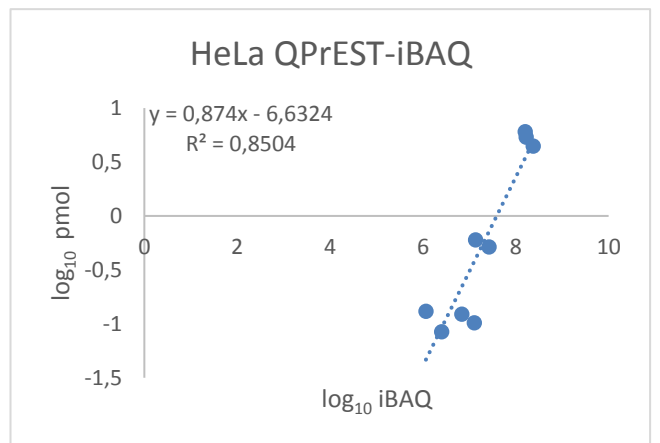
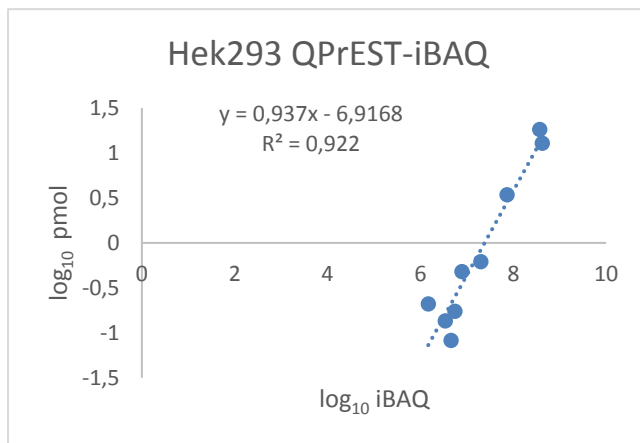
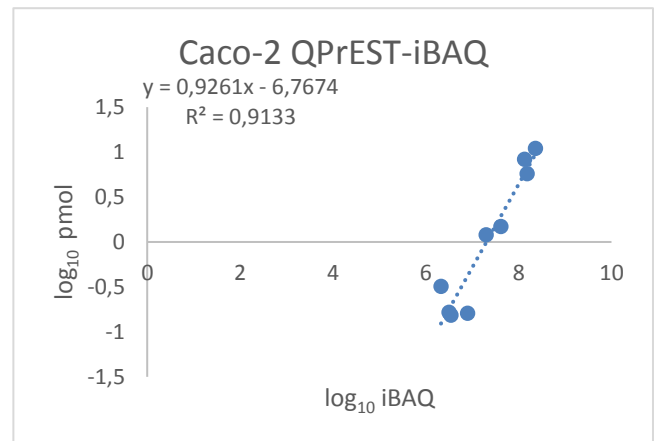
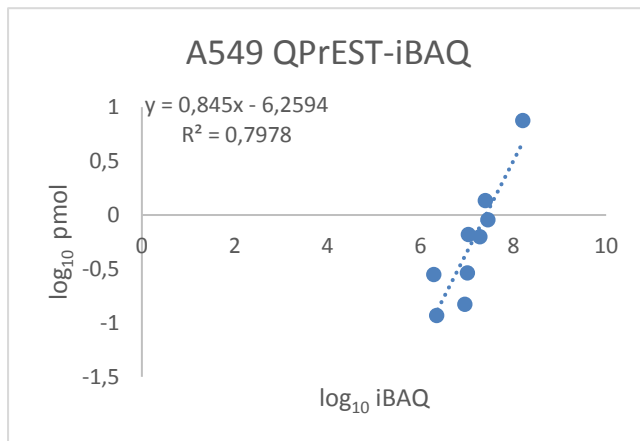
QPrEST ID	Gene	tpm HeLa	H/L HeLa	tpm A549	H/L A549	tmp Caco-2	H/L Caco-2	tpm Hek293	H/L Hek293	tpm U2OS	H/L U2OS
22881	GSTP1	536	3.39	1002	4.57	1673	15.58	488	2.10	670	1.97
22716	GSN	44	0.09	64	0.09	60	0.22	9	0.06	43	0.09
25474	RPS28	71	0.39	88	0.58	110	0.70	708	0.51	131	0.22
25396	RAB10	39	0.15	121	0.54	65	0.51	49	0.19	58	0.13
20520	ATOX1	75	0.08	156	0.15	194	0.36	190	0.12	159	0.11
20480	PDLIM1	36	0.05	161	0.18	314	0.38	127	0.13	21	-
26315	TARDBP	184	0.08	150	0.12	163	0.16	151	0.14	217	0.07
33555	AKT1S1	32	0.06	76	0.11	38	0.10	108	-	65	0.05
25436	YWHAG	104	0.10	206	0.15	119	0.16	100	0.08	150	0.05
23174	LAMP2	56	0.17	68	0.28	52	0.37	52	0.21	90	0.10
23611	VDAC1	225	0.51	285	0.63	535	1.48	304	0.61	252	0.22
24019	PEBP1	141	3.66	169	4.68	93	6.54	340	5.50	122	1.48
23714	PPIB	488	0.60	600	0.66	795	1.20	374	0.48	833	0.42
22627	GAPDH	4494	5.36	9714	7.51	6313	10.97	2834	3.43	6438	4.55
21629	GGCT	65	0.05	44	0.04	54	0.08	83	0.07	76	0.02
33915	OTUB1	79	0.13	179	0.28	110	0.32	99	0.21	77	0.07
22812	SLC3A2	337	7.62	627	0.81	551	14.22	111	1.55	231	1.69
35130	TXNDC17	92	0.12	140	0.29	72	0.15	100	0.17	65	0.07
26210	AIMP1	-	0.08	-	0.12	-	0.21	-	0.11	-	0.06

**Table 5 : Quantification results by LFQ of target proteins in five human cell lines, using five different LFQ-methods.**

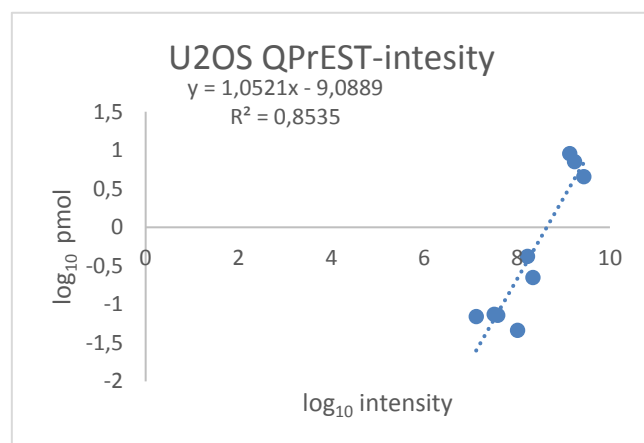
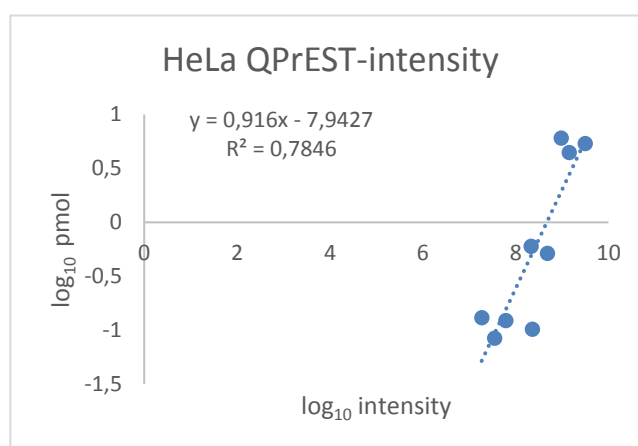
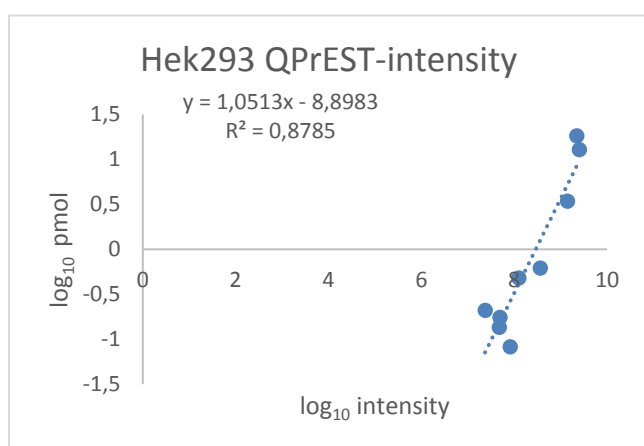
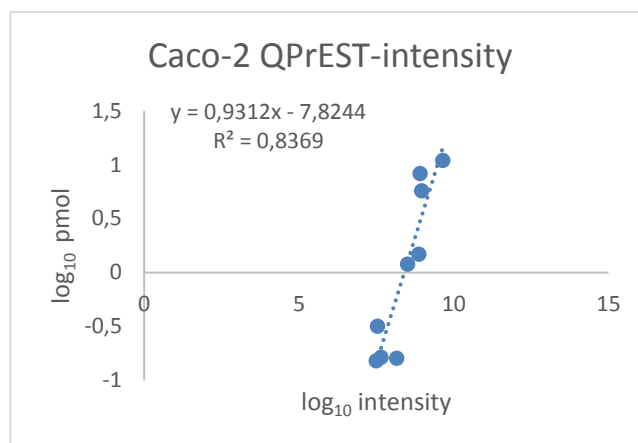
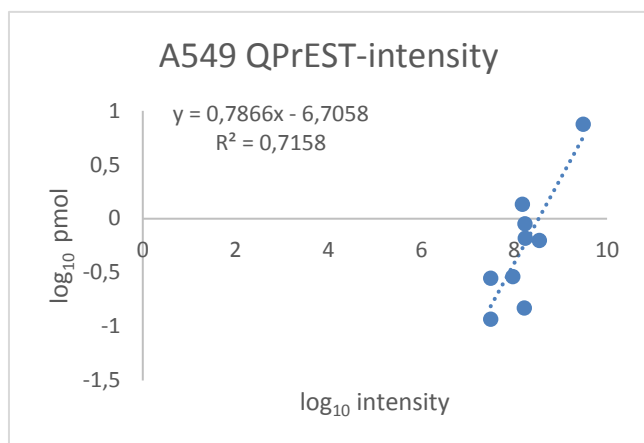
<i>QPrEST ID</i>	<i>Cell line</i>	<i>maxLFQ- BCA</i>	<i>iBAQ- BCA</i>	<i>QPrEST- maxLFQ</i>	<i>QPrEST- iBAQ</i>	<i>QPrEST- intensity</i>	<i>QPrEST H/L quant.</i>
<i>QPrEST 22881</i>	A549	4.80	5.72	1.22	1.39	1.18	<b>4.57</b>
	Caco-2	87.98	42.15	6.73	8.79	7.93	<b>15.58</b>
	Hek293	4.18	5.61	0.62	1.02	0.93	<b>2.10</b>
	HeLa	2.92	2.97	0.75	0.83	0.76	<b>3.39</b>
	U2OS	7.00	6.92	0.78	0.86	0.80	<b>1.97</b>
<i>QPrEST 22716</i>	A549	0.03	0.01	0.04	0.02	0.05	<b>0.09</b>
	Caco-2	0.38	0.04	0.12	0.05	0.13	<b>0.22</b>
	Hek293	0.01	0.00	0.00	0.00	0.01	<b>0.06</b>
	HeLa	0.03	0.01	0.03	0.01	0.03	<b>0.09</b>
	U2OS	0.11	0.03	0.03	0.01	0.04	<b>0.09</b>
<i>QPrEST 25474</i>	A549	4.44	12.30	0.44	1.05	0.33	<b>0.58</b>
	Caco-2	15.13	15.89	0.42	1.30	0.34	<b>0.70</b>
	Hek293	7.26	19.54	0.35	1.18	0.28	<b>0.51</b>
	HeLa	3.12	9.36	0.28	0.87	0.24	<b>0.39</b>
	U2OS	3.61	11.23	0.12	0.48	0.11	<b>0.22</b>
<i>QPrEST 25396</i>	A549	0.69	0.50	0.22	0.17	0.24	<b>0.54</b>
	Caco-2	2.84	0.62	0.23	0.17	0.23	<b>0.51</b>
	Hek293	0.52	0.32	0.07	0.07	0.07	<b>0.19</b>
	HeLa	0.19	0.10	0.05	0.04	0.05	<b>0.15</b>
	U2OS	0.42	0.25	0.04	0.03	0.04	<b>0.13</b>
<i>QPrEST 20520</i>	A549	2.69	1.00	0.28	0.13	0.07	<b>0.15</b>
	Caco-2	5.19	3.48	0.15	0.32	0.14	<b>0.36</b>
	Hek293	0.87	1.32	0.04	0.09	0.03	<b>0.12</b>
	HeLa	0.32	0.54	0.03	0.07	0.03	<b>0.08</b>
	U2OS	1.00	1.76	0.03	0.08	0.03	<b>0.11</b>
<i>QPrEST 20480</i>	A549	0.75	0.58	0.35	0.29	0.42	<b>0.18</b>
	Caco-2	5.25	1.22	0.66	0.50	0.73	<b>0.38</b>
	Hek293	0.84	0.51	0.18	0.16	0.21	<b>0.13</b>
	HeLa	0.16	0.09	0.07	0.06	0.08	<b>0.05</b>
	U2OS	0.02	0.00	0.00	0.00	0.00	<b>-</b>



**Figure 9: Calibration curves for quantification using QPrEST-maxLFQ, in five human cell lines, A549, Caco-2, Hek293, HeLa and U2OS. The amounts of protein (pmol) used were obtained by QPrEST-H/L. The maxLFQ intensities were calculated by MaxQuant.**



**Figure 10: Calibration curves for quantification using QPrEST-iBAQ, in five human cell lines, A549, Caco-2, Hek293, HeLa and U2OS. The amounts of protein (pmol) used were obtained by QPrEST-H/L. The iBAQ values were calculated by MaxQuant.**



**Figure 11: Calibration curves for quantification using QPrEST-intensities, in five human cell lines, A549, Caco-2, Hek293, HeLa and U2OS. The amounts of protein (pmol) used were obtained by QPrEST-H/L. The protein- intensities were obtained in MaxQuant.**

## 4 Discussion

Label free quantification has many benefits compared to label-based techniques, when it comes to ease of use, cost of assay development and more. However, the accuracy in

quantification is low compared to labeled quantification, and therefore the usefulness and applicability is limited. Development of more accurate LFQ-methods is therefore of great interest, and also the motivation behind this study, where the use of isotope labeled standards, QPrESTs, in LFQ is investigated.

QPrEST-LFQ in a range of human cell lines was performed, by creating a calibration curve from PQIs and QPrEST H/L ratios obtained from MS-analysed lysate samples spiked with a mix of QPrEST-standards. The selection of which QPrESTs to use was made with the criteria that the tpm value for the QPrEST gene target had a CV<45% in the ten cell lines A-431, A549, Caco-2, Hek293, HeLa, MCF7, PC3, SK-MEL-30, U-2 OS and U-251. The assumption was that a low variance in tpm indicated a target protein that would be found in similar amounts in all cell lines. Additionally, only QPrESTs with five or more theoretical tryptic peptides were chosen, to maximize the chance of finding several proteotypic peptides from each. The QPrESTs were also chosen to cover target proteins with varying abundance. The criteria described above resulted in a list of 34 QPrESTs, of which 33 were produced in this project. The produced QPrESTs were subjected to exhaustive quality control, including purity analysis by SDS PAGE, molecular weight determination and quantifications by mass spectrometry. Only QPrESTs that met the quality control criteria were used in the later stages of the project. In total, the number of approved QPrESTs summed up to 25. Of those that were not approved, most failed because of impurities. In purity analysis of almost all QPrESTs, bands can be seen at ~25 and ~40 kDa. Since these bands are present at the same weight in all QPrESTs, one possible explanation is that they correspond to *E. coli* proteins, however this has not been investigated. The purification of QPrESTs was done with IMAC (HisTag), but apparently there are some other proteins in the lysate than QPrEST that also have affinity for the matrix used. Low concentration samples more often seem to have problems with impurities, which would also make sense since there would be more unoccupied binding sites available on the matrix, if the concentration of QPrEST is lower. That would mean lower competition for binding, and more of the proteins with some affinity for the gel could remain bound. However, this theory does not explain why there are also examples of very low concentration QPrESTs that have no problem with impurities, like QPrEST20566 at 100% purity. There is a need to further investigate what those often-seen impurities are, to provide an answer to why the purification fails for some of the QPrESTs. Another issue with some of the QPrESTs was the presence of QPrEST fragments, i.e. degradation products or truncated protein variants. Since the application of the QPrESTs are as internal standards these degraded proteins are a problem. The determination of QPrEST concentration will be affected, since the truncated proteins will contribute to the final concentration. Also, if there are truncated versions of QPrEST in the sample, the product peptides from QPrESTs compared to target protein might not be the same, neither the digestion efficiency, which in turn will introduce an unknown error in the quantification. Therefore, QPrESTs showing signs of degradation in the purity analysis were excluded regardless if they met the purity criteria or not. Another phenomena observed for some of the QPrESTs was that their estimated weight from the SDS PAGE gel differed slightly from

the theoretically and experimentally determined. SDS PAGE is not a reliable method for determining molecular weight, for example, proteins that were not fully denatured during sample prep might show different migration pattern than expected. In this project, all molecular weights were also determined by MS, and therefore deviations from the expected weight that appeared in the SDS-PAGE analysis were ignored.

There was one QPrEST, 20919, that was not identified at all during molecular weight determination by MS. In the SDS-PAGE analysis, a band was visible at ~35 kDa, which correspond well to the theoretical weight. It was also possible to quantify the QPrEST as its tryptic peptides were identified. However, not all proteins ionize well, and if it is not ionized it is not going to fly and cannot be detected. QPrEST 20919 was therefore excluded.

All QPrESTs that were approved, were combined into a mastermix and spiked into lysate samples, that were analysed by LC-MS/MS and subsequent data analysis by MaxQuant and skyline. The idea was to run a few lysate samples spiked with different amounts of the mastermix, to obtain H/L ratios, from which the composition of the mastermix could be altered in order to obtain H/L ratios closer to one. The reason for aiming at a 1:1 ratio of QPrEST and its target protein is that the range of linearity for the ratios are not known. Thus, very high or very low ratios might not reflect the true value but could be due to saturation effects. Unfortunately, the results from the lysate samples spiked with the 25QPrEST-mastermix were inconclusive. For lysate samples spiked with 1:10 and 1:100 dilutions of QPrEST-mix, very few peptides had been identified and received a H/L ratio by MaxQuant. For the lysate sample spiked with undiluted QPrEST mix, the heavy peptides were found, but very few of the light target peptides. The ratios obtained were high, the majority~50-70. To investigate this further, the data was imported into skyline, where the chromatograms revealed that the light target peptides found had a very noisy signal with no clear peak, and low intensity. Since the heavy peptides had such high intensities, this resulted in very high H/L ratios. Only one of the QPrESTs gave a good signal for both heavy and light peptides. The majority of the produced QPrESTs were thus not suitable for quantification in the untargeted manner needed for the intended application.

A set of new QPrESTs were therefore chosen, from those available in stock at Atlas Antibodies AB. The criteria this time was lowered, with no demand on the number of tryptic peptides. When it comes to variation between cell lines the limited number of available QPrESTs, meant that the criteria of a CV<45% could not be kept. Instead it was a matter of choosing QPrESTs that showed the least variation in tpm, while still covering a broad range of target concentrations. To avoid ending up with a new set of QPrESTs with poor light peptide targets, a skyline evaluation of the chosen QPrESTs were made in advance, before any samples were prepared. The raw data from earlier MS-runs did not contain any of the new heavy QPrESTs, but since it was from whole cell lysate, it contained all the light target peptides. The light target peptides could thus be located, and QPrEST were chosen according to which targets generated high intensities and clear peaks. From the first pick of new QPrESTs, more than half were ruled out because of poor signal from the light target, in the

skyline evaluation. Had this type of investigation been performed before the original list of QPrESTs were assembled, it would surely have resulted in a different pick.

After skyline evaluation, a new QPrEST mastermix was prepared from the newly selected QPrEST set. It was spiked into lysate samples as before, in three different concentrations. The results from the highest concentration of spike in (undiluted mastemix) show that most peptides have been identified by MaxQuant, and the majority of ratios is found at 0.1-10. Looking at the data in skyline about half of the peptides showed a good signal for both light and heavy peptide and were able to confirm the ratios obtained in MaxQuant. However, for samples prepared with dilutions of QPrEST-mastermix, results were still lacking. A closer inspection of the MaxQuant results revealed that the light peptides were often identified with good signal, whereas the heavy QPrEST peptides were not. The suspected cause of the missing identification is the dilution of the QPrEST mastermix. The mastermix itself consist of QPrESTs in PBS with 1 M Urea. The dilutions however were prepared in 25 mM Ambic, which might have affected the stability of the QPrESTs.

The results from the highest spiked concentration of QPrEST mastermix were all in a range relatively close to the desired. Adjusted amounts to reach a ratio closer to one were calculated, and a mix with the adjusted amounts prepared. This mix was spiked into lysate samples from five different cell lines and analysed as before. The resulting H/L ratios were all in the range of 0.3-10, with a few exceptions. The results were then quantified in six different ways, five of which are LFQ methods (see 2.6). The most accurate quantification method of the ones used is assumed to be the one with QPrEST H/L ratios used directly to estimate the concentration of the target protein. To evaluate the performance of the LFQ methods, the quantification results for each protein are thus compared to those obtained by the QPrEST quantification. In the three QPrEST-LFQ methods where a calibration curve is used for the quantification (see 2.6), the model is built based on calculated PQIs, and quantification results from QPrEST H/L quantification. To ensure unbiased evaluation across all LFQ methods, six QPrESTs that were not used to build the linear models were quantified. The evaluation of model performance is thus based on the quantification results of these six proteins. One strategy towards optimizing the performance of the QPrEST-LFQ methods, was to only use proteins that had more than one quantotypic peptide to create the calibration curve. The presence of multiple peptides allows for them to verify each other, since they should both display the same H/L ratio.

From looking at the quantification results, see Table 5, it seems that the two methods whose results differed most from the QPrEST-H/L results were maxLFQ-BCA and the iBAQ-BCA. The results from maxLFQ-BCA was found on average in a 6.7-fold range from the QPrEST-H/L results. For iBAQ-BCA they were found in a 9.9- fold range. The correlation coefficient for the results vary between cell lines, from 0.5- 0.99 for maxLFQ-BCA, and from 0.19 – 0.94 for iBAQ-BCA. For the other three LFQ methods, where QPrESTs were used for linear regression, the results look more like those from QPrEST-H/L. The results from QPrEST-maxLFQ compared to QPrEST-H/L is found on average within a 3- fold range. For QPrEST-



iBAQ and QPrEST-intensity the results are found within an average of 3.2-fold and 2.8-fold range respectively. Looking at the correlation coefficient reveals that there is a high correlation with QPrEST-H/L results for the three QPrEST-LFQ methods compared to the correlation obtained by maxLFQ-BCA and iBAQ-BCA. The correlation coefficient for QPrEST-maxLFQ vary between 0.92-1.00 for different cell lines, and for QPrEST-iBAQ and QPrEST-intensity it varies between 0.67-0.99 and 0.98-1.00 respectively. The fact that the methods using labeled standards seems more accurate, when comparing to the QPrEST-H/L results, than the ones relying on total protein measurement was expected and in agreement with other research (Ahrné *et al.* 2013). The obtained deviation and correlation for the QPrEST-LFQ results compared to the QPrEST-H/L results, also correspond well to earlier reported results by similar approaches (Zeiler *et al.* 2014).

The linearity of the regression curve varied a lot between cell lines and could perhaps be improved by choosing QPrESTs with less variation in tpm value between cell lines. In this project the list of QPrESTs to choose from was limited to those in stock, which resulted in a suboptimal pick. It would also be beneficial to find suitable QPrESTs with more than two quantotypic peptides, to further increase the robustness of H/L quantification. Another interesting thing would be to investigate the range of linearity for the H/L ratio for each QPrEST, to ensure that the ratio obtained from the QPrEST mastermix is in a reliable range.

Regarding the tpm data (see table 4), it seems to correspond well to the QPrEST-H/L estimated quantities of the different proteins in their respective cell line. Overall, tpm therefore seem to be a good predictor of relative protein abundances. There are however a few examples where this is not true, for example for PEBP1 gene product, where the reported tpm for each cell line is relatively low compared to the protein abundance estimated by QPrEST H/L ratio. The reason behind this inconsistency might be due to error in quantification, or in tpm measurement, or it might actually reflect biological functions, like translation rate or lifetime and degradation rate. Further investigation is needed to decide on an explanation, but this example serves to illustrate the need for accurate protein quantification methods, as RNA data alone is not necessarily a good reflection of the proteome.

In this project, isotope labeled QPrESTs were produced, and used to quantify a set of proteins in different cell lines by LFQ. It has been shown that the use of QPrESTs as calibrants in LFQ is beneficial, compared to methods based on measured total protein abundance. There is potential in the idea, and further development and optimization of the QPrEST mastermix and subsequent data analysis could eventually lead to an easy to use QPrEST-LFQ method, with one mastermix that is ready to go for quantification in a large range of human cell lines.

## 5 Acknowledgements

Ett stort tack till min handledare på Atlas Antibodies, Tove Boström, och även Marie Utterbäck som bidragit med mycket handledning. Det har varit otroligt roligt och lärorikt att få jobba med er, tack för allt! Ännu ett stort tack till Jonas Bergquist, professor vid Uppsala Universitet och ämnesgranskare för projektet, som både bidragit med stort kunnande inom området samt uppmuntran och peppning när projektet strulade. Tack till Jan Andersson och Lena Henriksson, projektexaminator och koordinator, som funnits med och stöttat, dels under projektets gång, men även under hela utbildningen.

Sist men inte minst, ett stort tack till Nicole Yacoub, Caroline Jansson, Matilda Jonsson och Bonoshree Dey. Ni vet att utan er hade ingenting av detta varit möjligt. Tack!

## References

- Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. *Nature* 422: 198–207.
- Ahrné E, Molzahn L, Glatter T, Schmidt A. 2013. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *PROTEOMICS* 13: 2567–2578.
- Arike L, Valgepea K, Peil L, Nahku R, Adamberg K, Vilu R. 2012. Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*. *Journal of Proteomics* 75: 5437–5448.
- Atlas Antibodies. 2018a. Our Mission and History | Atlas Antibodies. online 2018: <https://atlasantibodies.com/about/mission-and-history>. Accessed May 23, 2018.
- Atlas Antibodies. 2018b. About Triple A Polyclonals | Atlas Antibodies. online 2018: <https://atlasantibodies.com/learn-more/about-triple-a-polyclonals>. Accessed May 23, 2018.
- Atlas Antibodies. 2018c. About PreciA Monoclonals | Atlas Antibodies. online 2018: <https://atlasantibodies.com/learn-more/about-precisa-monoclonals>. Accessed May 23, 2018.
- Atlas Antibodies. 2018d. About PrEST Antigens | Atlas Antibodies. online 2018: <https://atlasantibodies.com/learn-more/about-prest-antigens>. Accessed May 23, 2018.
- Atlas Antibodies. 2016. QPrEST standards and endogenous proteins show similar digestion efficiency.
- Bantscheff M, Lemeer S, Savitski MM, Kuster B. 2012. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and Bioanalytical Chemistry* 404: 939–965.
- Boström T. 2014. High-throughput protein analysis using mass spectrometry-based methods. *Biotechnology*, KTH Royal Institute of Technology
- Boström T. 2016. QPrEST<sup>TM</sup>—isotope-labeled multipепptide standards for quantitative mass spectrometry-based proteomics. *Nature Methods* 13: iv–vi.
- Cox J, Hein MY, Lubеr CA, Paron I, Nagaraj N, Mann M. 2014. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics* 13: 2513–2526.
- Cutillas PR, Timms JF (eds). 2010. LC-MS/MS in proteomics: methods and applications. Humana Press, New York, NY.

- Geiger T, Wisniewski JR, Cox J, Zanivan S, Kruger M, Ishihama Y, Mann M. 2011. Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature Protocols* 6: 147–157.
- Higgs RE, Butler JP, Han B, Knierman MD. 2013. Quantitative Proteomics via High Resolution MS Quantification: Capabilities and Limitations. *International Journal of Proteomics* 2013: 1–10.
- 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. *Molecular & Cellular Proteomics* 4: 1265–1272.
- Matic I, Jaffray EG, Oxenham SK, Groves MJ, Barratt CLR, Tauro S, Stanley-Wall NR, Hay RT. 2011. Absolute SILAC-Compatible Expression Strain Allows Sumo-2 Copy Number Determination in Clinical Samples. *Journal of Proteome Research* 10: 4869–4875.
- Price NC, Nairn J. 2009. *Exploring proteins: a student's guide to experimental skills and methods*. Oxford University Press, Oxford ; New York.
- Rieux L, Sneekes E-J, Swart R. Nano LC: Principles, Evolution, and State-of-the-Art of the Technique | LCGC. online: <http://www.chromatographyonline.com/nano-lc-principles-evolution-and-state-art-technique>. Accessed February 13, 2018.
- Schwanhäusser 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: 337–342.
- Steen H, Mann M. 2004. The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology* 5: 699–711.
- Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* 41: 207–234.
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigartyo CA-K, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F. 2015. Tissue-based map of the human proteome. *Science* 347: 1260419–1260419.
- Villanueva J, Carrascal M, Abian J. 2014. Isotope dilution mass spectrometry for absolute quantification in proteomics: Concepts and strategies. *Journal of Proteomics* 96: 184–199.

Zeiler M, Moser M, Mann M. 2014. Copy Number Analysis of the Murine Platelet Proteome Spanning the Complete Abundance Range. *Molecular & Cellular Proteomics* 13: 3435–3445.

Zhu W, Smith JW, Huang C-M. 2010. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology* 2010: 1–6.

## Appendix

**Table A1: Concentration of different QPrESTs in QPrEST mastermix with adjusted amounts, and the amount of each spiked into lysate samples.**

QPrEST ID	Gene	Conc in mastermix	pmol spiked
<i>QPrEST26315</i>	TARDBP	50	0.5
<i>QPrEST33555</i>	AKT1S1	20	0.2
<i>QPrEST22881</i>	GSTP1	300	3.0
<i>QPrEST25436</i>	YWHAG	100	1.0
<i>QPrEST23174</i>	LAMP2	67	0.67
<i>QPrEST22716</i>	GSN	50	0.5
<i>QPrEST23611</i>	VDAC1	100	1.0
<i>QPrEST24019</i>	PEBP1	400	4.0
<i>QPrEST25474</i>	RPS28	67	0.67
<i>QPrEST23714</i>	PPIB	100	1.0
<i>QPrEST22627</i>	GAPDH	1000	10.0
<i>QPrEST21629</i>	GGCT	20	0.2
<i>QPrEST33915</i>	OTUB1	100	1.0
<i>QPrEST25396</i>	RAB10	100	1.0
<i>QPrEST20520</i>	ATOX1	50	0.5
<i>QPrEST20480</i>	PDLIM1	33	0.33
<i>QPrEST22812</i>	SLC3A2	150	1.5
<i>QPrEST35130</i>	TXNDC17	100	1.0
<i>QPrEST26210</i>	AIMP1	50	0.5
<i>QPrESTH3</i>	H3	697	6.97
<i>QPrESTH4</i>	H4	816	8.16