ANALYSIS OF THE REGULATION OF BIOLOGICAL NETWORKS USING QUANTITATIVE PROTEOMICS

by

Anders Riis Kristensen

B.Sc., University of Southern Denmark, 2006

M.Sc., University of Southern Denmark, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2013

Abstract

A protein network can be thought of as a graph with nodes and edges, where nodes represent proteins and edges represent protein-protein interactions. Neither proteins nor their interactions are stable constituents in the cell; they are constantly changing in response to external stimulation or internal programming. Changes in protein expression are regulated by transcription, translation and protein degradation, whereas protein interaction changes have been shown in focused studies to be regulated by post translational modification. To investigate the processes influencing the regulation of protein expression and interaction changes, mass spectrometry based proteomics was applied because it has two key advantages for the study of protein networks: 1) it directly detects peptides from the proteins, and so does not rely on antibodies or the generation of fusion proteins; 2) by combining mass spectrometry based proteomics with quantitative techniques, such as stable isotope labeling by amino acids in cell culture (SILAC), it is possible to quantify thousands of proteins in a single experiment.

Here a systems biology approach was applied to investigate protein expression change, synthesis and degradation of proteins during cellular differentiation in two different cell lines. This allowed observing that protein expression during cellular differentiation is largely controlled by changes in the relative synthesis rate, whereas the relative degradation rate of the majority of proteins changed little. By comparing the data with previously published data of mRNA levels, there could be provide strong evidence that the generally poor correlation observed between transcript and protein levels can be explained once the protein synthesis and degradation rates are taken into account.

To study how protein interactions change in response to perturbation, a novel approach was developed combining size exclusion chromatography (SEC) and protein correlation profiling (PCP)-SILAC. Stimulation with epidermal growth factor (EGF) caused 351 proteins to alter their interactions with other proteins and, interestingly, when compared to previously published phosphorylation data, these proteins tended to also have altered phosphorylation under similar experimental conditions. This approach allowed identification of protein interactions in numbers comparable to other high throughput techniques, but also enabled quantification of protein stoichiometry between proteins participating in multiple complexes.

Preface

Part of Chapter 1.4 has been submitted for publication. Kristensen A.R. and Foster L.J. High throughput strategies for probing the different organization levels of protein interaction networks. This literature review was written primarily by me, with input and editing by LJF.

A version of chapter 2 has been submitted for publication. Kristensen A.R., Gsponer J. and Foster L.J. Protein synthesis rate is the predominant regulator of protein expression during differentiation. I was the primary researcher involved in this project, where I conceived and performed the experiments and analyzed the data, with inputs from JG and LJF. The manuscript was primarily written by me, with editing from JG and LJF.

A shorter version of chapter 3 has been published. Kristensen A.R., Gsponer J. and Foster L.J. A high-throughput approach for measuring temporal changes in the interactome, Nature Methods 9(9) 907-909. I was the primary researcher involved in this project, where I conceived and performed the experiments and analyzed the data, with inputs from JG and LJF The manuscript was primarily written by me, with editing from LJF and JG.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMPK	AMP activated protein kinase
AP	Affinity purification
AP-2	Adapter-related protein complex 2
AP-3	Adapter-related protein complex 3
APOBEC3G	DNA dC->dU-editing enzyme APOBEC-3G
AQUA	Absolute quantitation
BAC	Bacterial artificial chromosome
BAD	Bcl2 antagonist of cell death
CDC42	Cell division control protein 42
CFP	Cyan fluorescence protein
CID	Collision induced dissociation
CLN2	G1/S-specific cyclin CLN2
CORUM	Comprehensive Resource of Mammalian protein complexes
CXMS	Cross linking combined with mass spectrometry
DEPTOR	DEP domain-containing mTOR-interacting protein
DUB	Deubiquitinating enzyme
ECM29	Proteasome-associated protein ECM29 homolog
EGF	Epidermal growth factor
E-MAP	Epistatic miniarray profiles
EPS15	Epidermal growth factor receptor substrate 15
ESI	Electrospray ionization
FBS	Fetal bovine serum

FDR	False discovery rate
FN	False negatives
FP	False positives
FRET	Förster resonance energy transfer
GFP	Green fluorescence protein
GST	Glutathione-S-transferase
HA	Hemaglutinin
HCD	Higher-energy collisional activation dissociation
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
HRS	Hepatocyte growth factor-regulated tyrosine kinase substrate
iBAQ	Intensity-based absolute quantification
ICAT	Isotope-coded affinity tag
IgG	Immunoglobulin G
ILK	Integrin-linked protein kinase
iTRAQ	Isobaric tag for relative and absolute quantitation
LC-MS-MS	Liquid chromatography-tandem mass spectrometry
LDH	Lactate dehydrogenase
LOPIT	Localization of organelle proteins by isotope tagging
LTQ	Linear trapping quadrupole
m/z	Mass/charge
MALDI	Matrix assisted laser desorption/ionization
MAPK	Mitogen-activated protein kinase

miRNA	microRNA
MLST8	Target of rapamycin complex subunit LST8
MS/MS	Tandem mass spectrum
mTOR	Mammalian target of rapamycin
mTORC1	mTOR complex 1
mTORC2	mTOR complex 2
MWCO	Molecular weight cut off
p62	p62/sequestosome 1
PAC	Proteasome assembling chaperone
PARVA	Alpha-parvin
PCP	Protein correlation profiling
PDGF	Platelet-derived growth factor
PINCH	Particularly interesting new Cys-His protein 1
PLS	Partial least square
PMA	Phorbol 12-myristate 13-acetate
PTB	Phosphotyrosine binding
PTM	Post translational modification
PXN	Paxillin
Rac1	Ras related C3 bolulinum toxin substrate 1
RAPTOR	Regulatory-associated protein of mTOR
RF	Radio frequency
RICTOR	Rapamycin-insensitive companion of mTOR
ROC	Receiver-operator characteristics

RPN10	26S proteasome regulatory subunit RPN10
RPN13	Proteasome regulatory particle non-ATPase 13
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecyl sulfate poly acrylamide gel electrophoresis
SEC	Size exclusion chromatography
SH2	Src homology 2
SHIP2	SH2 domain-containing inositol 5'-phosphatase 2
SILAC	Stable isotope labeling by amino acids in cell culture
SRM	Selected reaction monitoring
STAM1	Signal transducing adapter molecule 1
STAM2	Signal transducing adapter molecule 2
STAT3	Signal transducer and activator of transcription 3
TAP	Tandem-affinity purification
TEV	Tobacco etch virus
TN	True negative
TP	True positive
TTI1	TELO2-interacting protein 1 homolog
UCHL5	Ubiquitin carboxyl-terminal hydrolase isozyme L5
UTR	3’ untranslated region
Vif	Viral infectivity factor
Y2H	Yeast-two-hybrid
YFP	Yellow fluorescence protein

Acknowledgements

I would like to thank my supervisor Dr. Leonard J. Foster for teaching me how to become a scientist. First, when I was in your laboratory seven years ago as an exchange student as part of my Danish master studies, for answering my multiple questions and for teaching me proteomics, mass spectrometry and cell culture. Second, during my PhD you have taught me to become an independent scientist and have always encouraged me to pursue my own ideas. It has been amazing to have a supervisor that could provide excellent supervision in very diverse areas such as bioinformatics, protein separation, cell biology and mass spectrometry.

Thanks also to the lab members who had to tolerate my sometimes extensive use of the mass spectrometer. A special thanks to Dr. Nickolay Stroynow and Dr. Joost Gouw for technical assistance with liquid chromatography and mass spectrometry, and Nat Brown for teaching me adobe illustrator and drawing the bathtubs throughout this thesis.

A particularly thank you to the members of my supervisor committee Dr. Joerg Gsponer and Dr. Juergen Kast for scientific input and relaxed European attitude. Dr. Joerg Gsponer in particular for the guidance with the bioinformatics, both the theory and the practical problems I encountered, and Dr. Juergen Kast for strongly supporting me when I doubted I would finish this PhD.

Most importantly, thanks to my family, Karina, Maja and Zoe, for being especially important to me. Thanks to Karina for having the courage to follow me to the Canada and being an excellent mother and Maja and Zoe for teaching me much about life.
Finally, thanks to the Danish tax-payers for supporting my education and to UBC and Center for blood research for scholarships.

To Maja and Zoe

Chapter 1: Introduction

1.1 Protein networks

Biological systems can be depicted as networks composed of nodes and edges, representing biomolecules and the interactions among them, respectively; since this thesis focuses on proteins, however, nodes and edges will here refer to proteins and protein-protein interactions, respectively [1]. Protein networks are not stable structures but are constantly changing due to internal programming or the continuous bombardment by external stimuli [2]. Protein networks can be regulated through changes in the expression of the proteins or through changes in the interactions between proteins, which results in different consequences. Removal of the protein will eliminate not only the function of the protein but also all interactions that the protein participates in, whereas removal of specific interactions will cause a more subtle effect on the network [3,4]. Therefore, I have divided this thesis up in two sections, where the first part handles the regulation of protein expression changes and the second part handles the regulation of protein interaction changes in response to external perturbation.

The regulation of protein expression is a relative slow and energetically expensive process, occurring as a consequence of changing the synthesis or the degradation rate of a protein, which I will discuss in section 1.3. The synthesis rate of a protein is regulated by numerous processes such as mRNA expression changes, microRNA (miRNA) and RNA binding proteins, whereas the degradation rate of proteins can be regulated by the ubiquitin proteasomal system or autophagy. The regulation of protein expression has mainly been investigated in a number of focused studies, allowing to decipher how a handful of proteins were regulated in response to perturbation of the cell: however, in recent years a number of

studies have emerged which study the relationship between mRNA and protein expression changes of thousands of proteins.

The regulation of protein interactions is a relative inexpensive process, which can happen in seconds in response to cellular perturbation, which I will discuss in chapter 1.4. Protein complexes, defined as a group of proteins in the interactome (all protein-protein interactions within a system) that is highly interconnected, are extremely important in the cell, since it is estimated that at least 90% of soluble proteins serve as components of protein complexes in both prokaryotes and eukaryotes, which implies the assembly of protein complexes is a fundamental principle of most, if not all, living organisms [5-7]. The complexity of these interactions covers a wide range, from simple pairs of proteins, or dimers, to large complexes such as the ribosome or the nuclear core complex. Protein complexes play central roles in all aspects of cellular functions including DNA replication, transcription, translation, metabolism, signaling transduction and structural organization.

The study of protein interactions has mainly been driven by technology; as new techniques emerged new features of networks were discovered. The founding technique for the study of protein interactions in systems-wide level was yeast-two-hybrid (Y2H) [8], which allowed identification of multiple interactions *in vivo* in a single experiment. Later, the interaction partners of a given protein could be identified using affinity purification followed by mass spectrometry [6,7,9-13]. Recently, a group of techniques that identify endogenous proteins belonging to the same macromolecular complex or organelle based on their co-elution through one or multiple separation techniques have been introduced [14-20].

Throughout this thesis the main method used to analyze the regulation of protein networks is mass spectrometry-based proteomics, which I will introduce in detail in chapter

1.2. Mass spectrometry is highly suitable for identifying proteins, since it does not require generation of fusion proteins or using antibodies, as other biochemical approaches do; instead, it can be used to directly detect peptides or proteins. Furthermore, by using stable isotopes of common elements found in proteins (e.g., C, N, O, H) in combination with mass spectrometry, it is possible to quantify how expression or interactions of proteins change in response to external perturbation, which can provide valuable functional information.

1.1.1 Protein network and disease

Phenotypic variations in an organism, especially those that result in a disease, can be seen as being caused by perturbations of the interaction network. This network view of diseases is often more appropriate than the linear causative model, which states that diseases happen due to failure in one gene. A growing body of evidence suggests that a number of major diseases, e.g. type 2 diabetes and coronary diseases, derive from small defects in a number of genes [21].

Mutations in different domains of a protein can affect different edges in a protein interaction network, a principle exemplified by mutations in two domains in tumor protein p63 that lead to two different developmental disorders [22]; mutations in the DNA binding domain cause ectrodactyly ectodermal dysplasia, whereas mutations in the SAM2 domain that is involved in protein interactions cause ankyloblepharon ectodermal dysplasia [22].

Network perturbations are also critical in infectious diseases. One archetypical example is *Salmonella*, a Gram-negative bacterium that secretes a number of virulence proteins through its type 3 secretion system into the host cell, where they target strategic parts of the host cell's protein interaction network. Among many mechanisms, SopE/E2 and

SptP from the bacteria act as guanine nucleotide exchange factors and GTPase activating proteins, respectively, for the human small G-proteins Cell division control protein 42 (CDC42) and Ras related C3 bolulinum toxin substrate 1 (Rac1); together with another bacterial effector, SopB, these proteins cause a massive actin rearrangement that leads to the bacteria being engulfed by the host cell [23-25]. Another example is the human immunodeficiency virus (HIV) that blocks part of the host cell's antiviral defenses by producing an accessory factor, Viral infectivity factor (Vif), which removes a key node of the host protein interaction network. This is accomplished through hijacking a host ubiquitin ligase complex that subsequently initiates the degradation of DNA dC->dU-editing enzyme APOBEC-3G (APOBEC3G), which normally play a critical role in viral defense by interfering with reverse transcription [26]. More recently, Vif was shown to have a wider impact on the host cell protein interaction network though its recruitment of host CBF-β to the ubiquitin ligase complex, disrupting CBF-β's normal interactions with RUNX DNA binding proteins [27,28]. This recruitment helps facilitate Vif folding and is required for binding and poly-ubiquitylating APOBEC3G.

The expression of a protein is extremely tightly regulated both by synthesis and degradation, since failure can result in diseases such as cancer [29] and neurodegenerative disease [30]. Gene amplification is often associated with oncogenic transformation through increasing the concentration of a protein in the cell [31]. In the case of the Epidermal growth factor (EGF) receptor, this will lead to constitutive kinase activation through increased concentration of the active dimers, leading to autophosphorylations of a number of tyrosine residues. Following activation of the receptor, a number of adaptor proteins will bind to the phospho-tyrosines of the receptor, mainly through Src homology 2 (SH2)- and

Phosphotyrosine binding (PTB)-domains, leading to activation of mainly the Ras/Mitogen-activated protein kinase (MAPK) pathway which results, among others, in cellular proliferation [32,33].

Deregulation of the expression of proteins can also take place at the level of protein degradation, which has been associated with neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease and Huntington's disease: these are all diseases where a particular protein or a group of proteins misfold and form aggregates [34]. The importance of lysosomal proteolysis in Alzheimer's disease was recently shown in mouse models of Alzheimer's disease, where enhancing the lysosomal proteolysis improved clearance of amyloid beta and restored the memory and learning defects normally associated with Alzheimer's disease [35]. Furthermore, it has been observed that pharmacologically inhibiting autophagy or the proteasome lead to accumulation of protein aggregates in cells, which suggest that both these pathways are important in clearing aggregate prone proteins [36].

1.2 Mass spectrometry based proteomics

The word 'proteome' was coined more than 15 years ago and it refers to all the proteins expressed in a given organism or cell population at a given time [37,38]. The main technical development that paved the way for modern proteomics was the ionization of proteins either by electrospray ionization (ESI) [39] or matrix assisted laser desorption/ionization (MALDI) [40]. Later, it was discovered that proteins could be identified with much greater sensitivity by mass spectrometry than by Edman degradation, which quickly transformed mass spectrometry into the preferred method for identifying

proteins [41]. In the last decade, a number of technical improvements in mass spectrometers and in how samples are prepared has made it possible to identify complete proteomes and to quantify changes of both protein and post translational modifications in response to cellular perturbation [42-44].

The workflow of a typical proteomics experiment is composed of three steps: sample preparation, mass spectrometry and data analysis. Sample preparation involves incorporating cells or an organism with stable isotope labeled amino acids, if stable isotope labeling by amino acids in cell culture (SILAC) is used as a quantitation method, before proteins are extracted and eventually separated by for example sodium dodecyl sulfate poly acrylamide gel electrophoresis (SDS-PAGE). The proteins in the sample are subsequently digested into peptides, often by the use of trypsin since it cleaves polypeptide chains exclusively at carboxy-terminal of lysine and arginine [45]. Proteins are digested to peptides because mass spectrometers are more sensitive for molecules with a low molecular weight, and the identification of a protein can be based on a single unique peptide [46]. However, the disadvantages of digesting proteins to peptides are that the sample gets more complex, isoform information is often lost and it is difficult to unravel combinatorial post translational modification (PTM) information [47]. To reduce sample complexity, peptide mixtures are often separated in one or multiple dimensions prior to mass spectrometry by, e.g., isoelectric focusing or strong cation exchange (SCX) before being separated by reverse phase chromatography coupled directly to the mass spectrometer. The ionization method used most frequently in modern proteomics is ESI, which typically generates multiply-charged ions whose mass/charge ratios (m/z) can be analyzed by a mass spectrometer. The mass spectrometer will be programmed to first generate a survey spectrum to determine the m/z of

the peptides, and this will be followed by a number of product ion spectra (MS/MS) of the most intense ions. In tandem mass spectrometry, one of the precursor ions observed in the initial survey scan will be fragmented by colliding the ion with an inert gas, which increases the vibrational energy of the peptide that is dissipated very quickly through the breakage of different amide bonds, resulting in the generation of several new, smaller ions, known as product ions.

Following MS analysis, the proteins and peptides will be identified and quantified using dedicated software packages, which generally first identify and then quantify the peaks in the MS and MS/MS spectra, followed by database searching and correcting for false discovery rates [48].

1.2.1 Quantitative proteomics

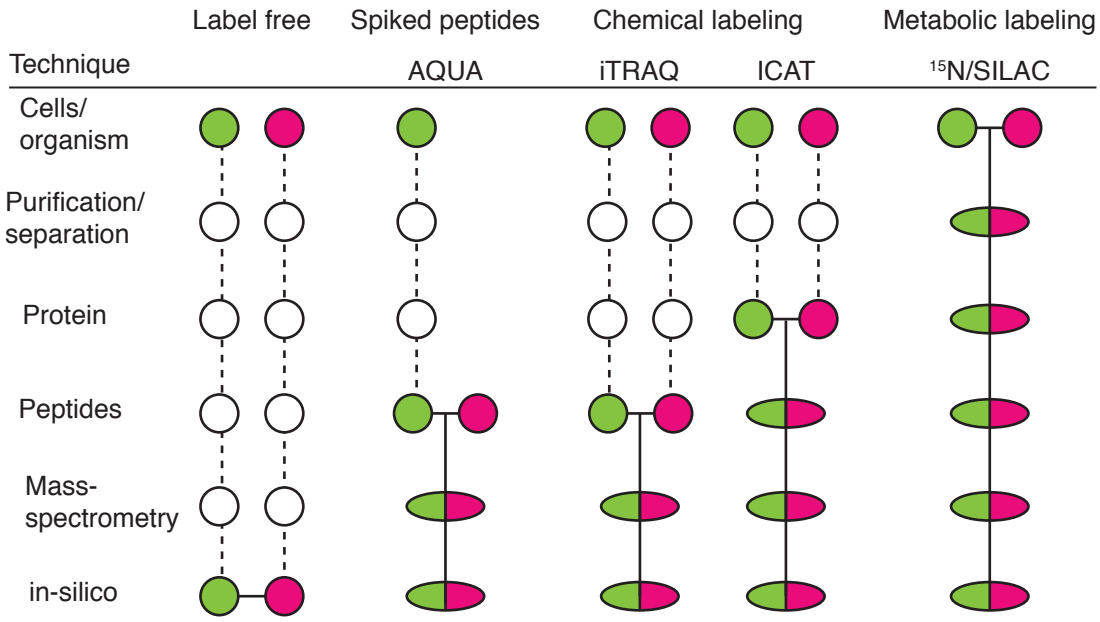
Quantification in proteomics can be performed by determining either the absolute amount of protein present in a sample or the relative amount between different samples. In absolute quantitation, the number of molecules present (i.e., copy number) or the concentration (e.g., mg ml^{-1} of a biomarker in serum) can be determined, whereas in relative quantitation the fold change of a protein in one condition versus another condition is determined [49]. The premise of performing quantitative proteomics with mass spectrometry is that the integration of the intensity of a peak in a mass spectrum, or in a series of mass spectra as the ion elutes from a chromatographic column, will be proportional to the amount, in the sample, of the particular peptide represented by that ion. This means that if there is double the amount of a particular peptide present the signal will be twice as big. On the contrary, the signal from two different peptides will not show this linear relationship mainly

because of differences in peptide ionization properties. The simplest approach to determine the absolute amount of protein, called absolute quantitation (AQUA), is to add a known amount of a synthetic peptide containing stable isotopes [50] as an internal standard. Since the stable isotope labeled peptide will have similar physico-chemical properties, other than molecular weight, to the endogenous peptide, the amount of endogenous peptide can be determined by comparing the signal measured for the endogenous versus synthetic peptides. However, AQUA suffers in throughput since it is only possible to identify the absolute amount of the proteins for which there is a synthesized peptide, and this has led to the development of techniques to determine the absolute amount of all the proteins in the proteome. One approach is intensity-based absolute quantification (iBAQ) that compares the total signal from all detected peptides from a protein divided with the theoretically number of observable peptides from that protein to a standard composed of 48 proteins covering a concentration range spanning six orders of magnitude; this technique allowed absolute quantitation of almost 2,000 proteins in a single proteomic experiment [51]. iBAQ was further used to estimate the protein amounts in a single cell line, revealing that just 122 protein contributed 50 % of the protein mass in the cell [42].

The other major approach in quantitative proteomics is to quantify the relative change of proteins between two different cellular states. The simplest approach is to analyze samples from the two states individually, comparing the signals for specific peptides; however, this approach suffers from lack of quantitation accuracy. To minimize the variation coming from separate but parallel sample preparation, a number of approaches that apply stable isotopes have been developed. The introduction of stable isotopes is accomplished by two approaches: chemical or metabolic labelling. One of the differences between these two approaches is how

early in the workflow the samples from the two cellular states can be combined, since each step that is required before the samples are combined potentially introduces viability and makes for less accurate quantitation (see Figure 1.1).

Figure 1.1 Stages of introducing stable isotopes



The green and red circles represent different samples being compared; the horizontal lines represent the stage where the samples are mixed and dashed lines represents stages where samples are processed in parallel, hereby potentially introducing variability. Inspired by [49].

In chemical labeling the proteins are digested into peptides and then chemically modified by adding isotopologues or isotopomers of a derivatizing agent, thereby making them distinguishable by mass spectrometry. Two forms of chemical tags exist: one where the different tags actually have the same mass (i.e., isobaric) but since they are isotopomers their fragments have different masses (e.g., isobaric tag for relative and absolute quantitation (iTRAQ [52]) and one where isotopologues are used so that the derivatized peptides have different masses and thus can be distinguished at the level of a precursor ion spectra (e.g.,

reductive dimethylation [53,54] or isotope-coded affinity tag (ICAT) [55]). Isobaric tags have the advantage that the precursor spectrum gets less complex and that it is possible to quantify up to 8 different biological states in a single mass spectrometry using iTRAQ. Isobaric tagging has a very significant downside, however: if there are interfering ions in the precursor spectrum the quantitation in the tandem mass spectrum is not precise. Also, if an ion is not selected for tandem mass spectra no quantitative information is acquired for that ion [56]. One of the most commonly used methods for chemical labeling involves reductive dimethylation of primary amines with isotopologues of formaldehyde, which has the advantage that it is very inexpensive [57].

In metabolic labeling the proteins are labelled by feeding the organism or cells with isotopologues of amino acids that then get incorporated into the proteins or with isotopes of specific elements that autotrophs can use to make their own labelled amino acids. The first approach of combining stable isotope labeling of proteins followed by quantitation by mass spectrometry was performed in *Saccharomyces cerevisiae* by letting one population of the yeast grow on media containing ^{14}N and one population grow on media containing ^{15}N and subsequently knocking down the expression of the G1/S-specific cyclin CLN2 (CLN2) protein in one of the populations [58]. By mixing the populations together, digesting the proteins to peptides and analyzing them by mass spectrometry, the impact of the CLN2 protein on the proteome could be determined by quantifying the ratios between ^{14}N and ^{15}N containing peptides. A $^{14}N/^{15}N$ ratio of a peptide pair different from 1.0 suggested that the CLN2 protein had an effect on the expression of the protein that that peptide was derived from. This metabolic labeling approach using ^{13}C or ^{15}N -labeling has been extended to whole organisms such as *Arabidopsis thaliana, Caenorhabditis elegans* and *Drosophila*

melanogaster and has the advantage that it is cheap, samples can be mixed together at an early stage and it does not require any genetic manipulation of the organisms [59-62]. The disadvantages of metabolic ^{13}C or ^{15}N -labeling include that the mass difference between the light and heavy labeled amino acids changes with the peptide sequence and that it can be hard to acquire high enough purity of the heavy isotopes; low isotopic enrichment tends to make the quantitation less accurate [57,60].

SILAC is a form of metabolic labeling where amino acid isotopologues are incorporated into the organism's proteome. Two populations of cells are cultured in media containing either heavy or light forms of essential amino acids and after at least five cell divisions the proteome will be completely incorporated with the supplied amino acid [63]. Originally, cells were incorporated using different versions of leucine [63], but now arginine and lysine are preferred if the proteins will be digested with trypsin, since then all tryptic peptides, except perhaps the carboxy-terminal peptide, will contain at least one labeled amino acid.

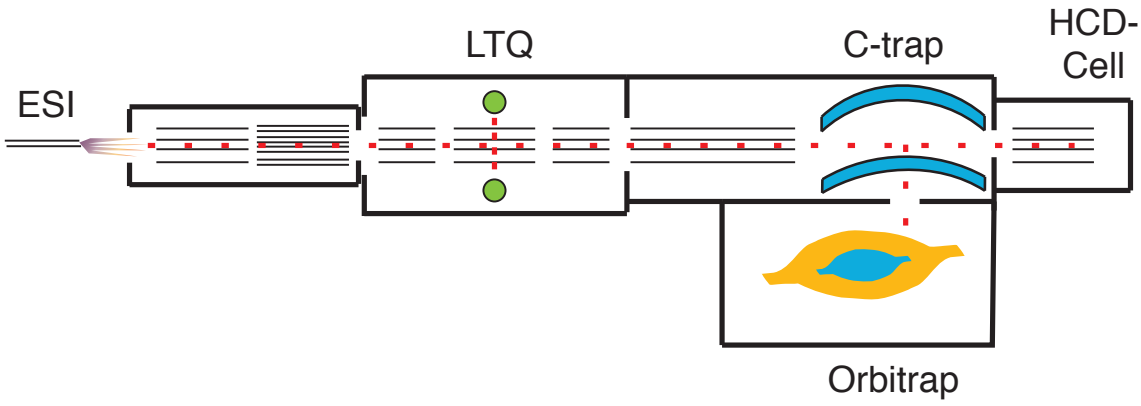
A clear advantage of SILAC and metabolic labeling in general is that the samples can be combined at a very early point during the handling, leading to less technical variability as earlier mentioned. Recently, SILAC has also been used in other organisms, such as mouse [64], *C. elegans* [65] and *Danio rerio* [66]. Since complete metabolic labeling is necessary for effective quantitation it is not feasible in humans, or indeed in any large organism, simply because of financial considerations; one way around this is an approach called Super-SILAC where mixtures of multiple cell lines deriving from the same tissue are incorporated with heavy isotopes [67]. A mixture of the proteins from these cell lines is therefore likely to contain virtually all proteins that might be expressed in the actual tissue and, therefore, the

mixture can be spiked into real tissue samples as an internal standard to quantify against. This approach has successfully been applied to distinguish, on the proteome level, two histologically similar subtypes of diffuse large B-cell lymphoma [68].

1.2.2 LTQ-Orbitrap mass spectrometer

Various models of a linear trapping quadrupole (LTQ)-Orbitrap mass spectrometer were used for the work in this thesis. This class of instruments is composed of two mass analyzers, the LTQ and the Orbitrap. The analyte is electrosprayed into the interface of the mass spectrometer, and then they are transported to the LTQ by a successive quadrupole and an octapole. In the LTQ, the ions can be counted using a feature known as automated gain control (AGC) [69], they can be fragmented and/or they can be sent along to the C-trap [70]. The C-trap is a bent quadrupole (in the shape of a 'C') from where the ions can be ejected to either the Orbitrap or to the higher-energy collisional activation dissociation (HCD) cell, where they can be fragmented as well [71] (Figure 1.2).

Figure 1.2 Schematic drawing of the LTQ-Orbitrap mass spectrometer



The LTQ-Orbitrap is composed of two mass spectrometers: the LTQ and the Orbitrap. The peptides are ionized by ESI before they are transported to the LTQ, where they can be fragmented and/or detected or transported further to the C-trap. From here, they can be transported to the HCD-cell, where ions can be fragmented, or transported to the Orbitrap, where they can be detected.

The LTQ, the first mass spectrometer the ions arrive at, has the advantage of being fast and sensitive but it has poor resolving power, low mass accuracy (~500 ppm) and a limited mass range compared to the Orbitrap. It consists of four hyperbolic rods cut in three axial sections with two of the center section rods, called the exit rods, having a slot through which the ions are ejected to the detector [72]. In each rod section; rods opposite each other are connected electrically, so it is reasonable to also state that the four rods in one section can be considered to be two pairs of two sections each. By keeping the potential in the front and back axial section higher than the center section and simultaneously changing the voltages of the rod pairs at a frequency of 1 MHz, it is possible to trap the ions in the center section. Here the peptides can be fragmented by collision-induced dissociation (CID), resulting in the semi-random breakage of peptide bonds that result in the formation of b- and y-ions. The fragment ions can subsequently be scanned out of the center rod slots to the detectors by

increasing the radio-frequency (RF)-voltage, which causes ions in resonance with the voltage to become unstable. Instead of fragmenting the ions in the LTQ, it is also possible to transport the ions through a transport quadrupole to the C-trap and from there to the Orbitrap mass analyzer [70].

The Orbitrap is characterized by its high mass accuracy (<1-2ppm using internal calibration), high resolution (60,000 at m/z 400 at a scan rate of 1 Hz), high sensitivity and high dynamic range; however, it is relatively slow when compared to the LTQ. It is composed of a spindle-like central electrode surrounded by a barrel-like outer electrode, which is split up in two halves [73]. The ions arrive from the C-trap as a small tightly clustered package but they will quickly distribute around the central electrode to form a narrow ring of ions that oscillates along the central electrode in the axial dimension. This oscillation frequency along the central electrode in the axial direction is given by the following equation: $\omega = \sqrt{m/z \cdot k}$, where k is proportional to the potential difference between the inner and outer electrode and m/z is the mass to charge ratio of the ion [73]. Hereby can be observed that by knowing the oscillation frequency along the central electrode, the m/z ratio of the ion can be identified. The axial oscillation frequency of the ions is detected by measuring the image current induced in the two outer electrodes, and since multiple ions are typically present, the signal can be Fourier transformed to obtain the individual frequencies.

1.2.3 Identifications and quantitation of peptides/proteins

The identification and quantitation of many proteins in one experiment (i.e., proteomics) is a much more complex process than, e.g., similar measurements of messenger

RNAs using micro-arrays [48]. It generally starts with extracting the features of both the precursor- and tandem mass spectrums, such as the m/z, charge state and intensities of the precursor ion and the derived fragment ions. More advanced programs will also recalibrate the precursor mass spectra for systematic errors by, e.g., using multiple consecutive precursor spectra to determine the mass error of individual peaks, and hereby the average mass accuracy obtained from Orbitrap data can be as low as 500 parts-per-billion (ppb). More simple programs will merely take the parent mass observed before the tandem mass spectrum in question was collected, resulting in mass accuracies that are typically at least 10-fold worse than what could be obtained.

The next step is to search the extracted data against theoretical data predicted from the known in silico digested proteins sequences of an organism; the most common approach to this, used by such popular search engines as Andromeda and Mascot, is to calculate the probability that the observed matches are occurring by chance [74]. The mass accuracy of the precursor ions is used to constrain the database matches since the theoretical number of matching peptides will be greatly reduced with increased mass accuracy [75,76]. A very common approach to determine criteria that give minimal false positive identifications is to also search the spectra against a target-decoy database, assuming that 'hits' against this decoy database are likely to be false and therefore allowing an estimate of a false discovery rate [77]. The identified peptides are afterwards assembled into proteins, a process that is non-trivial in shotgun proteomics since some peptides can belong to multiple proteins due to splice variants and sequence homology. Instead of identifying proteins, a number of algorithms instead identifies protein groups, which are groups of proteins that cannot be separated based on the peptides identified [78].

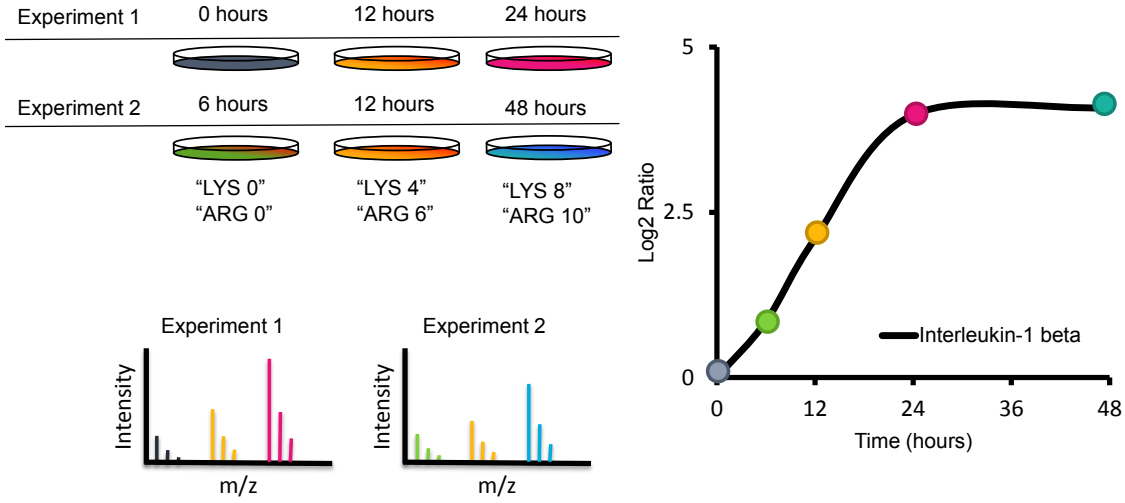
Quantitation of the peptides using SILAC is a relative simple process, where the ion intensities are calculated for each of the isotopic envelopes of the SILAC pairs, and the quantitation ratio of a protein group is the result of the median of all the peptide ratios.

1.2.4 Temporal proteomics

Proteomics holds an advantage over other methods of detecting proteins (e.g., immunoblotting) in that it can monitor thousands of proteins in a single experiment, thereby making it possible to pinpoint very accurately how a system responds to perturbations. By combining SILAC or other quantitative techniques with temporal stimulations it is possible to monitor how thousands of proteins and/or post translational modifications respond to the stimuli over time [79,80].

Typically, the approach taken to decipher how protein expression changes over multiple time points involves combining different SILAC experiments after they have been analyzed by mass spectrometry, which is achieved using a common time point as a reference in all of the individual experiments [44,81,82] (Figure 1.3).

Figure 1.3 Principle of combining two triple SILAC experiments to generate a five time point expression profile



Two cell populations are first triple SILAC encoded using three isotopologues of arginine (ARG) and three isotopologues of lysine (LYS). The cells are subsequently stimulated for 0, 12, 24 hours in experiment 1 and 6, 12, 48 hours in experiment 2. After MS-analysis, the two experiments can be combined by the common time point (12 hours) and protein expression profiles can be drawn.

One of the first studies characterizing temporal proteome changes used murine C2C12 cells that were differentiated from myoblast to myotubes for two or five days, respectively [63]. Here, undifferentiated cells were encoded with light isotope labeled amino acids and cells that had been differentiated for two or five days were encoded with heavy isotope labeled amino acids; proteins extracted from these cells were mixed together before being analyzed by mass spectrometry. Afterwards, the heavy-to-light ratio of the peptides were determined before the two SILAC experiments could be combined using the undifferentiated cells as a reference [63]. Later, the change in protein expression following differentiation of B-cells was investigated using five duplex SILAC experiments. Here a common time point at 0 h differentiation was used as a reference to combine the five SILAC experiments, thereby making it possible to follow 234 proteins at all five time points during differentiation out of some 1001 proteins identified overall. The protein expression profiles

were afterwards clustered using a self-organizing tree algorithm, revealing that functionally related proteins were displaying co-expression during beta cell differentiation [81]. Because of the semi-stochastic process of selecting peptide ions to be fragmented in a tandem mass spectrometer, in global proteomics the proteins that are actually identified will vary between experiments, and therefore the more experiments that are combined the less proteins are common between them, as illustrated above where less than one quarter of the proteins could be identified in all five experiments. This problem can be minimized by using three distinct types of amino acids for labeling in the SILAC experiments, which will enable five different conditions to be analyzed by combining two triple SILAC encoded experiments with a common reference point [44,82]. Taking this concept even further, five isotopically distinct forms of arginine were used to profile temporal proteins expression changes during adipocyte differentiation in a single experiment; this comes at a price, however, as introducing more labels means that spectra will be more complex [83].

The temporal information adds an extra dimension, and therefore extra value, to the data: Fuzzy-C-means clustering of profiles of phosphorylation dynamics in response to EGF stimulation revealed temporal profiles that seemed to correlate well with events in the signaling cascade [44]. Temporal information was also used to monitor the degradation of proteins in response to amino acid starvation, which revealed that different cytoplasmic compartments are degraded at different speeds by autophagy; this suggested that autophagy does not simply degrade proteins randomly [82].

1.2.5 Organelle proteomics

The most intuitive approach to identify a protein localized to an organelle is to purify the organelle of interest e.g. by density gradient centrifugation, and then to simply identify all the proteins in the final fraction by mass spectrometry [84]. This approach have been widely used to identify proteins of, e.g., autophagosomes [85], Golgi apparatus [86], and endosomes [87]; however, a major caveat of this approach is that just trace amount of contaminating proteins can lead to misclassification due to the high sensitivity of mass spectrometers. A more elegant approach to distinguish specific from unspecific organelle proteins, called protein correlation profiling (PCP), takes advantage of how proteins of an organelle will display the same distribution through density gradients whereas contaminating proteins will display a different distribution [16]. PCP is performed by first enriching a given organelle in some sort of gradient and then collecting several fractions spanning from the gradient (usually ≤10), followed by mass spectrometric analysis of the individual fractions. By identifying the relative amount of peptides in each fraction, using label free, intensity-based quantification, it is possible to construct a gradient distribution profile for every protein. Assignment of proteins to an organelle can be achieved by comparing the distribution profiles of the individual proteins to the elution profile of a marker protein of the organelle of interest. The PCP approach has been used to profile cellular compartments such as centrosome [16], peroxisomes [88], mitochondria [89] and membrane-enclosed cytoplasmic organelles [18].

Since the accuracy of label-free quantitation is sub-optimal (as mentioned above), a number of approaches have been developed that apply isotope based quantitation methods, such as ICAT, iTRAQ and SILAC, to improve the specificity of assignments made based on

the protein distribution profiles. ICAT have been used in an approach called localization of organelle proteins by isotope tagging (LOPIT) to map membrane proteins of *A. thaliana* [90]. Here, the different fractions from the density gradient were compared in pairs before principle component analysis was used to condense the dataset from six to two dimensions, whereby proteins located to the same cellular compartment showed similar distribution. The LOPIT approach was subsequently combined with iTRAQ to increase the resolution of the method by increasing the number of fractions and the number of quantifiable peptides; in this instance the assignment of proteins to organelles was accomplished using partial-least squares discriminatory analysis [91]. Finally, PCP was recently combined with SILAC, by an approach called PCP-SILAC, where two cell populations are encoded with light and heavy isotopic labeled amino acids, respectively, followed by gradient centrifugation of the two samples in parallel [19]. However, unlike PCP, after the gradient centrifugation all fractions of one sample are mixed, forming a homogeneous solution that is spiked into each of the other samples' fractions thereby functioning as an internal standard. By quantifying the SILAC ratios of the individual fractions, it was possible to draw profiles of the centrosome distribution in a density gradient, since the Malhalanobis distance between centrosome marker proteins and the other identified proteins was then used to classify proteins as being centrosomal or otherwise [19]. The PCP-SILAC approach has also been successfully applied to identify lipid raft and autophagosomal proteins as well [20,92].

1.2.6 Turnover proteomics

Most proteins in the cells are constantly being synthesized and degraded, leading to different turnover rates. Cell-wide protein synthesis rates have traditionally been studied by incorporating radioisotope-labeled amino acids such as ^{35}S-Met and then following the incorporation rate into the proteome by autoradiography [93]. A similar approach can be used to investigate the global degradation rates of proteins by a method referred to as pulse-chase, where proteins are incorporated with radioisotope-labeled amino acids followed by a change of media containing normal amino acids, and subsequently, the decrease of the radioisotope-labeled amino acids can be used as an estimate of the degradation rates of the proteins [94]. By fitting an exponential curve to the data points it is possible to estimate the absolute half-life of the proteins; however, to make this calculation accurate it is important to correct for the dilution- and growth-rate during cell division.

The degradation rate of individual proteins was recently investigated in tandem-affinity purification (TAP)-tagged yeast strains by blocking protein synthesis using cycloheximide and then performing quantitative western blotting [95]. Similarly, the turnover rates of individual proteins can be determined using bleach-chase, where cells expressing Yellow fluorescence protein (YFP)-tagged protein are bleached and then the recovery of YFP-signal are measured by time-lapse microscopy [96].

In the last decade, a number of studies that apply stable isotope-labeled amino acids and mass spectrometry to investigate the protein synthesis and/or degradation rate of the individual proteins have been reported. The principle here is the same as when using radiolabeled amino acids: the synthesis can be measured by the incorporation rate of stable isotope labelled amino acids and the degradation rate can be measured by the decrease of

incorporated stable isotope labelled amino acids. The main difference is the detection method. By using mass spectrometry it is possible to identify the individual proteins as well as to quantify the synthesis and degradation rates of that particular protein [97]. In the past, one downside of mass spectrometry was that it was less sensitive than autoradiography, but the recent introduction of click chemistry using azidohomoalanine makes it possible to study effects on synthesis rates down to 45 minutes or study newly synthesized secreted proteins [98,99].

A clear benefit of studying the synthesis and degradation rates using mass spectrometry is that it can be performed on endogenous, un-tagged proteins without inhibiting protein degradation or synthesis with drugs, thereby generating more physiological relevant data [97,100,101]. Another benefit is the possibility of investigating the synthesis and degradation rates in different cellular compartments because, interestingly, it has been observed that the same protein can have different turnover rates depending on where it is in the cell [97,101].

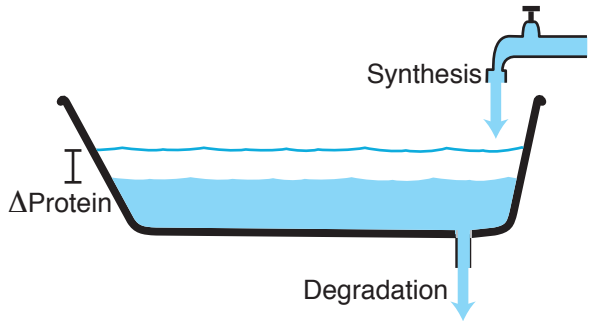
1.3 Regulation of protein expression

Throughout this thesis the term protein expression is used to describe changes in the relative amount of a protein, since this is the term generally used in proteomics literature. The regulation of protein expression changes can take place at the level of protein synthesis or protein degradation. An analogy of this is a bathtub with an open drain, where the amount of water corresponds to the amount of protein, the water inflow from the faucet resembles the synthesis rate, the draining of the water resembles the degradation rate and the relative water level change resembles the expression change of a protein (Figure 1.4). If the water inflow

and outflow rates are equal, the water level in the tub stays constant; however, the water still gets exchanged with a given velocity (i.e., the turnover rate for a protein). The bathtub representing each different protein will have different in- and outflow rates and, thus, each protein will have its own turnover rates. This means to change the water level in the bathtub, the inflow and/or outflow rate of the water will need to change, similarly the change of protein expression will also happen as a result of changing the synthesis and/or the degradation rate of the protein.

In the following two sections, I will briefly describe some of the cellular processes controlling the synthesis and degradation rates of a protein.

Figure 1.4 Bathtub model resembling the regulation of protein expression change



Water flow rate from faucet resembles protein synthesis rate, draining rate resembles protein degradation rate and water level change resembles protein expression change.

1.3.1 Processes regulating the translation rate

The processes controlling the translation rate of a protein can be divided into general or specific processes, which affect either a large or a small number of proteins, respectively. Proliferating cells will need to double their total mass of protein through each cell division, whereas senescent or differentiating cells will have much more modest requirements; clearly, the state of a cell is one of the major factors controlling the general synthesis rates of the

proteins in the cell. In line with this, different tissues in an organism will also display distinct synthesis rates for given proteins, and a fast proliferating tissue, such as gut epithelium, will have much faster synthesis rates than, e.g., heart tissue [64].

One of the main proteins regulating the general translation rates of proteins is mammalian target of rapamycin (mTOR), which receives input from mitogens, nutrition and energetic status of the cell through the AMP activated protein kinase [102,103]. When nutrition is plentiful, mTOR activates, through phosphorylation, a number of key regulators of translation such as eIF4E-binding protein 1, which is required for the initiation of cap-dependent translation, and S6 kinase 1, which activates a number of effectors leading to mRNA biogenesis as well as translation initiation and elongation [104]. The centrality of mTOR in the regulation of protein translation is highlighted by how the inhibition of mTOR, using Torin, suppresses translation of 99.8 % of the mRNA in the cell to some degree, with a mean reduction of translation of 61% [105].

A much more specific regulation of the translation rate of the individual proteins takes place through affecting the expression levels of precursor mRNA molecules, that code for the proteins in question. Intuitively, and according to the Central Dogma of Molecular Biology, the protein translation rate is dependent of the change in mRNA expression, suggesting that up regulation of mRNA leads to up regulation of the protein translation rate. However, in reality, only limited correlation has been observed between both the absolute abundance [55,106,107] and expression change [43,108-111] of mRNA versus protein. Often in these studies a number of reasons have been proposed for this discrepancy, such as that the mRNA and proteins were not collected from exactly the same cells [110], that post transcriptional regulation was occurring [108,109] and that substantial protein degradation

was taking place [108]; however, none of these studies provides actual evidence for these contributions. A few studies have used regression analysis to demonstrate the lack of correlation between mRNA and protein abundance. One study reported, by use of linear regression analysis, that the transcription and translation rates contributed the most to this discrepancy between the mRNA and protein abundance, whereas the degradation rate of the proteins contributed little [51]. This study, in addition, revealed that proteins on average are five times more stable than mRNA, since their median half-life was determined to be 46 and 9 hours, respectively. Another study revealed that mRNA abundance explain 25-30 % of the protein abundance whereas sequence features explain 30-40 % of the variance [112].

Another very important aspect of the correlation between mRNA and protein expression change is the temporal regulation of the two processes, which can be disclosed by performing transcriptomic and proteomic experiments in parallel in the same system and collecting samples at multiple time points after cellular perturbation. By applying this approach in yeast, it was revealed that protein expression changes after 6 h of rapamycin treatment correlated best with the mRNA expression change after 1 and 2 h stimulation, suggesting a 3-5 hours delay between when an mRNA molecule is synthesized and when protein starts to be expressed from it [109]. In another temporal study in yeast, mRNA and protein expression changes were followed in response to osmotic stress, and a transient 'burst' of mRNA expression was observed after just 30 min of stimulation, with the level of mRNA then gradually stabilizing at the final level [110].

Another specific process for regulating the protein synthesis rate of a protein is through miRNA. The first discovered miRNA was *lin-4* in *C. elegans*, which was observed to bind to the 3' untranslated region (UTR) of *lin-14* via antisense RNA-RNA interactions

and thereby inhibiting the synthesis of the LIN-14 protein [113,114]. By the use of quantitative proteomics, it has been observed that a single miRNA can regulate the synthesis rates of hundreds of proteins; however, the change of a protein in response to a miRNA was usually less than 4 fold [115].

Finally, a number of proteins are known to bind to mRNA and repress their translation. A noteworthy example of such a protein is CPE binding protein (CPEB), which binds to the mRNA of *Xenopus laevis* cyclin B1 in immature oocytes. This binding leads to recruitment of Maskin, which binds to initiation factor 4e, hereby blocking the binding site of initiation factor 4G and thus causing blockage of translation [116].

Another example of repression of mRNA happens at Processing-bodies (P-bodies), where mRNA and proteins form dynamic aggregates where mRNA can be translationally silenced. The P-bodies contain all the proteins required for degrading mRNA through the deadenylation-decapping -5'-3' decay pathway, as well, suggesting a role for P-bodies in mRNA degradation, but under some circumstances the mRNA can also exit and re-engage in translation [117].

1.3.2 Processes regulating the degradation rate of proteins

The two major cellular processes responsible for protein degradation in the cell are the proteasome-ubiquitin system and the lysosome [118,119]. The proteasome-ubiquitin system was originally described to be responsible for specific degradation of proteins, whereas the lysosome, through autophagy, was believed to be unspecific [119,120]. However, in recent years a number of specific degradation processes have been described to happen through autophagy; the assumption that degradation through the lysosome is

unspecific is therefore highly questionable [121,122]. A clear difference between these two processes is that through the ubiquitin-proteasomal system a single protein is tagged through poly-ubiqutinylation and subsequently degraded by the proteasome [119], whereas the lysosome, through autophagosomes, can degrade whole organelles and macromolecular complexes at once.

The proteasome-ubiquitin system involves conjugating a chain of ubiquitin molecules to a target protein, which will subsequently be degraded by the proteasome. The conjugation of ubiquitin to the target proteins happens via a three step mechanism, where initially an ubiquitin-activating enzyme hydrolyses ATP and forms a thioester bond between itself and ubiquitin [123]. Next, ubiquitin is transferred to an ubiquitin-carrier protein, which transport ubiquitin to an ubiquitin-protein ligase that catalyzes the covalent attachment of the ubiquitin to the target protein [123]. Each ubiquitin-protein ligase only recognizes a specific set of target proteins, which is reflected by the existence of 500-600 different ubiquitin-protein ligases [124] in the human genome. The activity of the attachment of ubiquitin to a target protein is counteracted by deubiquitination enzymes (DUBs) that remove the ubiquitin; approximately 95 DUBs are present in humans [125]. Since ubiquitin contains seven lysine residues (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 and Lys63) it can serve as a substrate for ubiquitination itself [126]. All seven residues can be ubiquitinated leading to chain formation; however, Lys48-linked poly-ubiquitin chains represent the signal that targets the protein for proteasomal degradation [127]. After poly-ubiquitinylation through Lys48, the target protein is recognized by the proteasome and subsequently degraded.

The proteasome is a macromolecular complex composed of a core particle attached to either one or two regulatory particles. The core particle is a cylindrical-barrel like structure

composed of heptameric rings of seven different alpha- and beta-subunits, arranged in a αββα stoichiometry. The assembly of the core subunit happens through assembly of the alpha rings with the help of the proteasome assembling chaperone (PAC)1-PAC4, whereas the proteasome maturation protein (UMP1) helps docking the beta subunits to the alpha ring [128]. The assembled core subunit can through three distinct active sites at the β1, β2, β5 proteins cleave the substrate through caspase, trypsin and chymotrypsin-like activity, respectively [129].

The regulatory particle is composed of at least 19 proteins, which are responsible for substrate recognition, de-ubiquitylation, unfolding, opening the gate of the core particle and translocation of the substrate into the core particle. In addition, a number of proteins are attached to the proteasome sub-stoichiometrically, such as Proteasome-associated protein ECM29 homolog (ECM29) [130]. The regulatory particle recognizes ubiquitylated substrates through the 26S proteasome regulatory subunit RPN10 (RPN10) and Proteasome regulatory particle non-ATPase 13 (RPN13); the RPN10 contains a ubiquitin interacting domain and RPN13 contains pleckstrin-like receptor of ubiquitin domain [131,132].

The other major place in the cell where degradation takes place is the lysosome, which is a membrane enclosed organelle characterized by having a pH of 4.6-5.0 and containing hydrolyzing enzymes [133]. A number of different membrane enclosed vesicles can fuse with the lysosome, such as autophagosomes, phagosomes and endosomes leading to degradation of their content [133].

Autophagy is an intracellular bulk degradation process in which a part of the cytoplasm is sequestered in a double membrane vesicle, known as an autophagosome, which through fusion with the lysosome will be degraded. It is believed to be the major cellular

pathway for degrading long lived proteins and organelles [134]. As mentioned earlier, it was originally believed to be a nonspecific process; however, in recent years a number of specific forms of autophagy been discovered such as mitophagy [135], ribophagy [136], pexophagy [137] which specifically degrade mitochondria, ribosomes and peroxisomes, respectively. Furthermore, amino acid starvation, which was earlier believed to induce nonspecific autophagy, was recently discovered, through quantitative proteomics, to cause different degradation rates for different cytoplasmic constituents, suggesting that autophagy is a specific process [82]. Through the use of PCP-SILAC and density centrifugation, it was recently shown how the autophagosome cargo changed according to cellular perturbation, which again suggest that autophagy is a specific degradation pathway [20]. At the molecular level of specific autophagy, it is believed that p62/sequestosome 1 (p62), which contains a LC3 interacting region and an ubiquitin associated (UBA) domain, plays a central role. One hypothesis is that the substrate for degradation gets ubiquitylated and then recognized by the p62 UBA domain, which will then sequester the substrate into large aggregates. The p62-substrate aggregate will afterwards, through binding of p62 to the autophagosomal membrane protein LC3, be recruited to the autophagosome, which will then be recruited to the lysosome leading, ultimately, to degradation of both p62 and the substrate [138]. This suggest that ubiquitin serves important roles of tagging proteins for degradation, both through the proteasome and through the autophagosome; presumably, this is because the ubiquitin conjugation system can then be reused by both processes [118].

1.4 Edges in a protein interaction network

The study of protein-protein interactions is not a new concept; it has been around for more than a century, even though in the beginning it was far from understood what was being observed. One of the first interactions discovered was trypsin bound to antitrypsin, from which it could dissociate [139]. Much later came the concept of protein interactions, including how these could serve as regulators of metabolism and how they were involved in critical elements of signal transduction [140]. In the last two decades it has become apparent that proteins form networks that can be modeled as graphs composed of nodes and edges, representing macromolecules and the interactions between them, respectively, where these networks seem to capture or represent phenotypic variation [141].

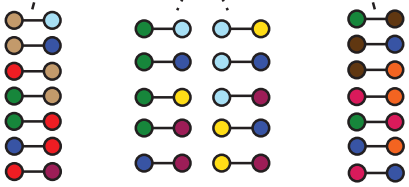
When the first protein interaction networks were mapped it was apparent that they were scale free, meaning that many proteins have few interaction partners, whereas few proteins interact with many proteins; proteins that interact with many partners became known as hubs [142]. By further investigating the characteristics of hub proteins based on the correlation of RNA expression among the proteins in complexes, it was realized that they could be divided into two classes: Party hubs (also known as intramolecular hubs) and date hubs (also known as intermolecular hubs) [143], although whether this is appropriate is still debated [144]. Party hubs are defined as stable complexes where proteins have higher than average correlating mRNA expression, and date hubs were defined as protein complexes with less than average correlating mRNA expression. In other words, party hubs are typically co-expressed with all their interacting partners, whereas date hubs are only co-expressed with a subset of their interacting partners in any given condition. It was further realized that date hubs interact with their partners at different times and/or conditions, whereas party hubs are

interacting with their partners at all times and conditions tested and hereby date hubs serve to connect functional modules that are composed of party hubs [4,145]. In addition, it has further been observed that date hubs are enriched in signaling domains such as tyrosine kinases, PDZ domains and Gα domains, and that mutations in these are more frequently associated with cancer [145].

1.4.1 Hierarchically levels of the protein-protein interactions

Defining a new terminology that is universally accepted is challenging, but we feel that it is relevant here in order to appreciate the differences between interactome methods. The idea behind this terminology is to capture the different features of protein interaction networks that are reported from the various techniques that are commonly used to measure them. The terminology is based on James G. Miller's ideas about general network theory: "The Universe contains a hierarchy of systems, each higher level of system being composed of systems of low levels" [146]. Similar to the terms ascribed to the various levels of protein structure, we believe protein interaction networks can also be depicted as having four levels of organization, primary, secondary, tertiary and quaternary (Figure 1.5), each of which is addressed by different methodological approaches.

Figure 1.5 Classification of the interactions in the protein interaction network

Interaction	Illustration	Description	Techniques
Quarternary		Union of tertiary interactions. Not necessarily present in the cell	Co-migration AP-MS Y2H PCA
Tertiary		Functional complexes. Different accessory proteins are added to secondary interactions	Co-migration Saturatated AP-MS
Secondary		Core complexes. Proteins that always share the same complex(es)	Co-migration Saturatated AP-MS
Primary		Pair-wise interactions. Also known as binary interactions.	PCA Y2H FRET CXMS

Pictographic representation of the protein interaction networks hierarchically levels (see text for detail).

The primary level, or the most basic type of information to come from interactome studies, is the knowledge of pairwise interactions among proteins. In principle, there is an even lower level describing interatomic interactions but such information is never captured in interactome studies. Pairwise interactions among proteins are commonly referred to as binary interactions and are typically the data output from *in vivo* techniques such as Y2H, split reporter protein assays and Förster resonance energy transfer (FRET), although *in vitro* techniques such as cross linking combined with mass spectrometry (CXMS) also report data in a similar way.

The secondary level of organization, sometimes referred to as core complexes [6,147], are proteins that always share the same complex(es) even though they might not interact directly. Expression patterns for such proteins would be highly correlated but by themselves would not necessarily form fully functional units. In addition, their biochemical role will be irreplaceable since deletion of any member of this group will typically lead to the loss of functional integrity of the whole complex [147], similar to party hubs. An illustrative example of a secondary interaction is the set of proteins mTOR, Target of rapamycin complex subunit LST8 (mLST8), DEP domain-containing mTOR-interacting protein (DEPTOR) and TELO2-interacting protein 1 homolog (TTI1), which are integral to both the mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2) [148] and can be identified as a single unit in interactome studies but they are not actually known to form a complex in isolation.

The tertiary level of organization within protein interaction networks, variably referred to as 'complex isoforms', is composed of functional complexes, where different accessory (or halo) proteins are added to the secondary interactions, or core complexes, leading to different protein assemblies in the cell [6,147]. The accessory proteins will therefore typically not share common deletion phenotypes, functional classification or cellular localization with the core proteins [147]. These peripherally attached proteins can, e.g., spatially or temporally affect the complex and its functions. This can again be illustrated with mTORC1 and mTORC2, where in mTORC1 the mTOR/mLST8/DEPTOR/TTI1 core is joined by the attachments; Regulatory-associated protein of mTOR (RAPTOR) and Proline-rich AKT1 substrate 1, while in mTORC2 the core is joined by the attachments Rapamycin-insensitive companion of mTOR (RICTOR), Target of rapamycin complex 2 subunit

MAPKAP1 and protein observed with Rictor-1/2. The implications of the different compositions of mTORC1 and mTORC2 are massive since mTORC1 gets activated by amino acids and ATP, inhibits autophagy and positively stimulates translation, whereas mTORC2 can activate several members of the AGC family of kinases including Akt [148].

Different versions of very similar tertiary assemblies can also exist, such as through alternative stoichiometries. This can be illustrated by the proteasome that can exist as a complex composed of one 20S core complex and two 19S regulatory subunits or a proteasome composed of one 20S core complex and one 19S regulatory subunit. In this case the actual proteins comprising the two different assemblies are the same, but there is clearly two discrete structures within the cell [15].

The quaternary level of interactions represents the union of the tertiary protein interactions and is therefore not necessarily present in any given cell but is nonetheless useful for gaining a bigger-picture view of the system. This view of the protein interaction network is generated computationally through clustering a matrix of primary interactions where 1's represent an interaction between two proteins and 0's represent no interaction. In this way proteins that interact with the same set of other proteins will cluster closely together [149,150]. The quaternary level can also be observed in non-saturated affinity purification (AP)-MS experiments, where not all proteins are tested as baits, leading to what are actually multiple tertiary complexes being identified as a single quaternary complex. To use the mTOR example again, affinity purification of any member of the mTOR/mLST8/DEPTOR/TTI1 core brings along with it all components of both the mTORC1 and mTORC2 complexes, and from such data it is not clear that there are actually two different structures present. Only by a saturating screen where every component of both

mTORC1 and mTORC2 are affinity purified and one sees, e.g., that RAPTOR and RICTOR are not in the same complex, can the composition of the two be resolved [151].

With this view of how interactions can be organized and described, we will now cover some of the different techniques used to target the different hierarchically organization levels of protein interaction network in the cell.

1.4.2 Primary protein interactions

Primary protein interactions are generally measured using *in vivo* assays, with the very large majority of studies having been conducted in yeast. Studying interactions *in vivo* is attractive since it can be more physiologically relevant and many of the sources of error involved *in vitro* approaches can potentially be avoided (e.g., interactions forced as a result of breaking the cell). The basic concept of *in vivo* techniques is that engineered fusions of the proteins of interest and reporter proteins are co-expressed and if an interaction occurs a detectable output signal will be generated. The most common applications of mass spectrometry typically generate quaternary-level interaction data (see below) but recent methodological and bioinformatic developments have led to great advances in the analysis of cross-linked proteins by mass spectrometry, which generates primary interaction data [152,153].

1.4.2.1 *In vivo* techniques

The Y2H assay, developed by Fields and Song more than 20 years ago, is the original *in vivo* method for studying protein interactions [8]. The core of the approach involves the fusion of the two hypothesized interacting proteins with two domains of the transcription factor GAL4, the DNA binding domain and the activation domain. An interaction between the two target proteins leads to reconstitution of the full transcription factor activity and synthesis of β-galactosidase proceeds; yeast colonies in which this has happened can then be identified colorimetrically since they will turn blue in the presence of X-gal. This approach has been used to test all of the yeast open reading frames (ORF) against each other, thereby generating a complete organism protein interaction network [154,155].

Since Y2H was first developed several other reporter systems have been developed. One of the most widely applied approaches is to split a reporter protein, e.g., ubiquitin [156], YFP [157] or dihydrofolate reductase [158], and then fuse each part to the proteins where an interaction is to be tested. If the two target proteins physically interact, the reporter fragments are brought close together and allowed to fold into their native structure, resulting in reconstitution of the reporter protein activity. For YFP this results in a fluorescent signal in the cells where the interaction takes place and for dihydrofolate reductase this results in a new selectable marker, where survival in the presence of methotrexate can be used to identify colonies in which an interaction is occurring [150]. A significant difference between Y2H and split reporter assays is that in the latter the interactions can presumably take place in the endogenous subcellular localization of the interacting proteins, whereas in Y2H the interaction must occur in the nucleus; additionally, split reporter assays can be used, in principle, in any organism that can be genetically modified, meaning that if they are

expressed off their endogenous promoter, the proteins should be present at natural levels and with the correct post-translational modifications.

Another imaging-based method that is widely used in focused[1] studies to examine protein-protein interactions is FRET, where two proteins are tagged with, e.g., cyan fluorescent proteins (CFP) and YFP, and co-expressed in the same cells; if the two proteins are within 10 nm of each other then excitation of CFP can result in resonance energy transfer to YFP, which then releases a photon at 535 nm, instead of the 480 nm expected for CFP [159]. This apparent shift in the emission wavelength of CFP indicates that the two proteins to which the fluorescence proteins are fused must be in such close proximity that the only reasonable explanation is that they are physically interacting. As mentioned, FRET is widely accepted as an exceptionally accurate method for confirming that two proteins interact but it typically requires significant optimization for any given bait and prey combination and has so far not been used for measuring whole protein interaction networks.

1.4.2.2 Cross linking combined with mass spectrometry

Crosslinking combined with mass spectrometry (CXMS) can be used to both discover new interactions and to characterize the structure of known interactions. In the latter use, it is more tolerant to impurities and low concentrations than are crystallography and NMR but of course it cannot compete in terms of resolution [160]. CXMS requires bifunctional chemical crosslinkers that can be applied either *in vivo* or *in vitro* to react with and covalently link neighboring proteins. The proteins are then digested to peptides and subjected to liquid

[1] We use the term 'focused' here to differentiate between 'omic'-scale studies that measure a more shallow but wider-reaching view of the system from studies that focus on one or a few proteins and try to understand their biology in greater detail.

chromatography-tandem mass spectrometry (LC-MS/MS) to detect the cross-linked peptides. In recent years, a number of linkers targeting different functional groups on proteins have been developed, with primary amines (mostly on lysine side-chains) being the preferred target as they are plentiful and mostly located at the surface of proteins. Since the abundance of cross-linked peptides are very low compared to linear peptides, the sensitivity of the approach can be boosted significantly be enriching crosslinked peptides prior to LC-MS/MS. A range of enrichment strategies have been proposed, most of which take advantage of the increased size and charge state of the cross-linked peptides using, e.g., SCX chromatography, size exclusion chromatography (SEC) or C4 reverse phase chromatography. The main challenge in CXMS, however, is interpreting the LC-MS/MS data to correctly identify the crosslinked peptides. Even in a clean system with only two proteins, the number of possible species that must be considered is the square of the number of linear peptides, resulting in a vastly enlarged search space. New algorithms that can deal with this issue have been developed [161,162] and these have partly contributed to a number of groundbreaking studies using CXMS to identify primary interactions among proteins in multi-protein complexes. One of the first large complexes analyzed was the RNA polymerase II complex, where Chen *et al.* identified 220 high quality interactions with approximately one third being primary interactions between proteins in the complex [152]. Hereby it was possible to observe the interaction between Pol II and transcription factor IIF with peptide resolution. Very recently the human protein phosphatase 2A complex was subjected to crosslinking followed by mass spectrometry, revealing an interaction with immunoglobulin binding protein 1 and the topology of Protein phosphatase 2A regulatory subunit binding to chaperonin [153].

1.4.3 Identifying secondary and tertiary protein interactions

Historically, protein complexes were identified one at a time by purifying them based on enzymatic activity and a complex was inferred when proteins co-migrated through multiple chromatographic methods. This approach provided information about the number of proteins in a complex and the size of the complex from SDS-PAGE and SEC, respectively [163]. However, this approach was really only useful when assays and/or regents existed for the complex in question and it was not practical until developments in mass spectrometry-based proteomics provided the tools for reconstructing migration profiles of many proteins in parallel.

1.4.3.1 Characterizing protein complexes using SEC

SEC was developed more than fifty years ago and has been widely used for studying protein complexes since it resolves analytes based on their Stokes radius, which is roughly proportional to molecular weight [164]. It relies on a column packed with a porous matrix that analytes can diffuse in and out of while an isocratic flow drives them through the column. The pore sizes define what range of Stokes radii can be separated, since smaller analytes effectively have a larger volume to diffuse into because they can enter the pores while larger analytes cannot. By comparing the elution times of standards of known molecular weight to measured values of unknowns, the molecular weight of an analyte can be estimated, since the elution volume decreases with the logarithm of the Stokes radius. This assumes, of course, that all analytes are the same shape, typically globular; while not precisely accurate this is a reasonable first assumption for most protein complexes. The big advantage of SEC over other chromatographic techniques for studying protein complexes is

that SEC provides very gentle conditions since no binding and release from the stationary phase is required and the proteins can be maintained in a physiologically relevant buffer the entire time.

The first unbiased large-scale mapping of stable protein complexes using chromatography was done on an *Escherichia coli* lysate by performing three different chromatographic separation steps, followed by iTRAQ to determine the relative migration of the single proteins, which identified 13 protein complexes based on co-migration of component proteins [165]. Around the same time, it was observed that one third of *S. cerevisiae* proteins were eluting at a much higher than expected weight from SEC [166]. Later, SEC was used to follow the co-migration of protein complexes in chloroplast stroma of A. *thaliana* by first fractionating chloroplast stroma by SEC, followed by mass spectrometry of the individual fractions where adjusted spectral counting was used to estimate levels of individual proteins in each fraction [167]. The spectral count chromatograms could subsequently be hierarchically clustered with the assumption that proteins clustering together should belong to the same complex; several complexes were characterized in this way, including the RuBisCO complex, ribosomes and the chaperone 60 complex. Finally, it was shown by combining SEC with an array of antibody beads that the CDK-cyclin complexes had different profiles in quiescent and proliferating cells [168].

1.4.3.2 Other co-migration approaches for characterizing protein complexes

Instead of just fractionating protein complexes in one dimension, Havugimana *et al.* took biochemical fractionation to the extreme, using weak anion exchange, isoeletric focusing and density gradients to generate 1163 different fractions from HEK293 and HeLa

S3 cells that were then analyzed by mass spectrometry [14]. Using the spectral counts observed for each protein in the 1163 fractions, the results were filtered using machine learning, where additional data from the literature such as evolutionary rates, messenger RNA co-expression, domain co-occurrence and physical associations reported earlier in worms and yeast were taken into account. This resulted in the identification of 13,993 co-complex interactions and 622 protein complexes, which represents the largest experimentally derived catalog of human protein complexes to date.

The PCP approach has also been successfully applied to membrane proteins by separating them on blue native PAGE, cutting the gel lane into several slices and analyzing each slice by LC-MS/MS; following this, label free quantitation was used to identify oxidative phosphorylation complexes I-V [169]. Recently, a comprehensive mitochondrial protein interaction network was identified using blue native and large-pore blue native gel electrophoresis combined with label free protein quantification [170]. Among the identified complexes was a novel complex, given the name ‘mitochondrial complex I assembly complex’, which highlights the advantage of studying protein complexes using non-hypothesis driven techniques.

1.4.4 Identifying quarternary interactions using AP-MS

The AP-MS approach is very widely used in focused studies to identify the interacting partners of a protein of interest. The method is scalable and partially automatable so with substantial resources one can use it to characterize a whole protein interaction network. In the simplest approach, the protein of interest (the bait) is enriched using an antibody and then the proteins bound to it are identified by mass spectrometry; after

accounting for potential non-specific interactions using an appropriate negative control the specific interacting partners can be identified. This approach has the advantage that no genetic manipulation is required and that it can be conducted in all organisms were an antibody is available. On the other hand, the success of this approach depends greatly on the specificity and affinity of the antibody towards the protein of interest, which is often sub-optimal. In addition, antibodies are not available for the majority of proteins, even in well-studied organisms, and those that are available are not of consistent quality. Broad panels of antibodies generated and validated under uniform conditions such as from the Human Protein Atlas project [171] will likely make this approach more popular. Recently, this technique was used to identify numerous protein interactions in human cells [172].

The most popular approach in AP-MS is to fuse the protein of interest to an affinity tag. This can relatively easy be done in yeast by homologous recombination and several whole genome collections are available for yeast. Homologous recombination is not possible in human systems though, so there the general approach has been to introduce the tagged version of the gene as a cDNA-based transgene, but these transgenes lack endogenous noncoding regulatory information, such as introns and 3'-untranslated regions, resulting in non-physiological expression levels in most cases. Recently, the bacterial artificial chromosome (BAC) TrangeneOmics approach was introduced that uses BACs to include most of the regulatory elements upstream of the gene of interest in order to achieve near-endogenous expression levels [173]. Beyond expression issues, there are many affinity tags to choose from, ranging from small peptides such as 6xHis, Hemaglutinin (HA) or FLAG to whole proteins such as green fluorescence protein (GFP) and glutathione-S-transferase (GST). Smaller tags seem less likely to disrupt protein complexes than a protein-sized tag,

but GFP has the obvious advantage that it can also be used to image the protein of interest to gain additional information such as localization, movement within the cell and co-localization [174,175]. The most common affinity system used in global protein interaction network studies is the TAP-tag, originally formulated as a fusion of a calmodulin binding peptide, a Tobacco etch virus (TEV) cleavage site and two Immunoglobulin G (IgG) units of protein A from *Staphylococcus aureus* [10], although several variants on this idea now exist [176]. The tagged protein and its interactors are first enriched using IgG-coated beads, then released from those beads by cleaving the tag with TEV protease and then further enriched on calmodulin beads [10]. TAP-MS has been widely used for global screens in yeast [6,7,9,12], as well as in more targeted analyses in human systems [11,177]. The great advantage of TAP is that the two rounds of enrichment help to minimize the non-specific proteins that would otherwise co-purify in a single step enrichment; on the flip side, the process still requires tagging and the longer and more complicated procedure to enrich interactors means that only very stable interactions survive to be detected.

Tagging proteins can be problematic: not only can the tags be challenging to introduce, they have also been shown to disrupt protein complexes and can potentially lead to changes in localization and function. For example, it was only possible to partly recover the chaperonin complex in a study using AP-MS because the ends of the proteins that were tagged were located in the interaction interface [6]. Moving the tag to the opposite terminus can often solve such issues but the majority of high-throughput AP-MS studies only tag one termini, meaning that many real interactions may be undetectable as a result of the experimental design. Tags can also affect the subcellular localization of proteins, exemplified by a recent report in *Caulobacter* where only one third of proteins localized properly

regardless of which end the protein was tagged on, suggesting that the tag disrupts terminal structures or interaction motifs of the protein [178]. Similar, it was observed in yeast that tagging and mild overexpression of a number of mitochondria proteins led to them failing to properly localize to the organelle and instead were delivered to the endoplasmatic reticulum, cytosol or nucleus [179]. Finally, in HeLa cells similar observations have been reported using the BAC transgenomics approach where e.g., Rab5c only localized correctly at the endosomes when tagged at the N-terminus and Aurora kinase B showed correct physiological localization dynamics through the cell cycle only when tagged at its C terminus [173].

Non-specific interactions can result in a protein being identified as an interactor when it is, in fact, a false positive. TAP minimizes (but does not completely eliminate) such problems through two rounds of enrichment but another approach is to perform the AP-MS in a quantitative fashion. This can be accomplished using stable isotope dilution or label free approaches. In SILAC [63,180] two populations of cells are first metabolically labeled and then used for the control and specific pull downs respectively, before the samples are combined, digested to peptides and analyzed by MS [181]. This approach have been used in numerous studies and has the great advantage that it is simple to perform and can easily distinguish specific from non-specific interactions [23,174,182]. For organisms/cell lines that are not amenable to SILAC, chemical labeling methods [182] or label free [175,183] approaches can be used to distinguish specific from non-specific interactions since the quantity of a specific interactor should be higher in the specific condition than in the negative control.

Finally as previously mentioned, one disadvantage to all affinity purification approaches is that they are blind to the arrangement of all the interactors identified. When

one identifies interactions in this way, there is no indication from the data whether the interactions are all occurring together (tertiary protein interactions) or whether the average composition of several different complexes is being observed (quarternary protein interactions). If exhaustive reciprocal affinity enrichments are performed with all the interactors then bioinformatic treatment of the entire network can sometimes tease apart such subtleties [6,7].

1.4.5 Temporal protein interaction network mapping

Mapping of protein interaction networks has taught us much about the overall design of cell systems but the vast majority of the data is static; that is, the networks are measured in a single experimental condition so our view is at best a snapshot in time. As with everything else in cells, protein interaction networks are dynamic systems that are constantly changing as cells are continuously bombarded by external stimuli and following internal programming. Changes in the networks can occur at both nodes and edges: e.g., nodes can change as a result of increased translation of that particular protein, whereas edges can change in response to stimuli such as when an adapter binds to a receptor in response to ligand binding. Measuring such changes is relatively straightforward in focused studies of one or a few proteins but the typical methods one would use in this case are not easily scalable, meaning that temporal protein interaction networks are rarely reported.

One approach temporal protein interaction network changes can be inferred is to use genetic interaction data, such as epistatic miniarray profiles (E-MAPs), which use a library of strongly down-regulated genes to screen both essentially and none-essential genes [184]. By conducting E-MAPs under two conditions, it was possible to map 873 differentially genetic

interactions during DNA damage response, from which the researchers could identify new roles for Slt2 kinase, Pph3 phosphatase and histone variant Htz1 and observe that protein complexes seem to be stable to perturbation [185].

Other proteomic methods have also been developed that directly measure changes in physical interactions. One of the earliest approaches combined AP-MS and SILAC, thereby making it possible to track dynamic changes among the affinity purified proteins [186]. The authors used HeLa cells labeled with triple SILAC and stimulated with EGF for different lengths of time, before antibodies against phospho-tyrosine were used to affinity purify tyrosine phosphorylated proteins and closely related binders. By quantifying the SILAC ratios for these proteins, it was possible to deduce how the stimulation affected a given protein's phosphorylation state. Later, this method was used to distinguish the effect of EGF and Platelet-derived growth factor (PDGF) on human mesenchymal stem cells, revealing that while more than 90 % of the signaling proteins were used by both pathways, phosphatidylinositol-3-kinase was exclusively activated by PDGF [187]. Finally, by conducting quantitative AP-MS of cyclins E1, A2 and B2 at multiple time points during cell cycle, another study found 295 specific interaction partners, which displayed significant enrichment of proteins with cell-division defects, suggesting a clear link between temporal protein interaction networks and functional roles [188]. These kinds of approaches provide an additional, valuable dimension to interaction data and often reveal novel connections that could not have been predicted otherwise; e.g., the YMER protein, which had similar activation profiles as other proteins involved in early EGF signaling [186] was later shown to function as an inhibitor for the down regulation of the EGF receptor [189].

On a similar scale, targeted proteomic methods can also be used to measure temporal interaction changes in protein interaction networks. Bisson *et al.* [190] used an approach called AP-selected reaction monitoring (SRM), where GRB2 was affinity purified and its interaction partners assayed by SRM. By integrating the peak areas of the peptides from each partner it was possible to measure the temporal protein dynamics at six different time points after EGF stimulation and additionally follow the changes in GRB2 protein interaction network caused by five other growth factors. SRM is much more sensitive than discovery-driven LC-MS/MS, it has large quantitative dynamic range, it does not require any *in vivo* or chemical labeling and it does not rely on stochastic sampling, leading to more reproducible data. The downsides, however, are that developing the assay is rather time consuming [190,191] and it is not easily scaled to many baits.

1.5 Research aims and hypothesis

The two different experimental parts of this thesis each have their own hypotheses:

Hypothesis one: by acquiring data for the cellular processes controlling protein expression, it is possible to estimate the contribution of the individual processes to protein expression and determine how these processes change in response to differentiation.

Hypothesis two: by applying SEC combined with PCP-SILAC we can simultaneously identify protein complexes and monitor how they change in response to EGF stimulation.

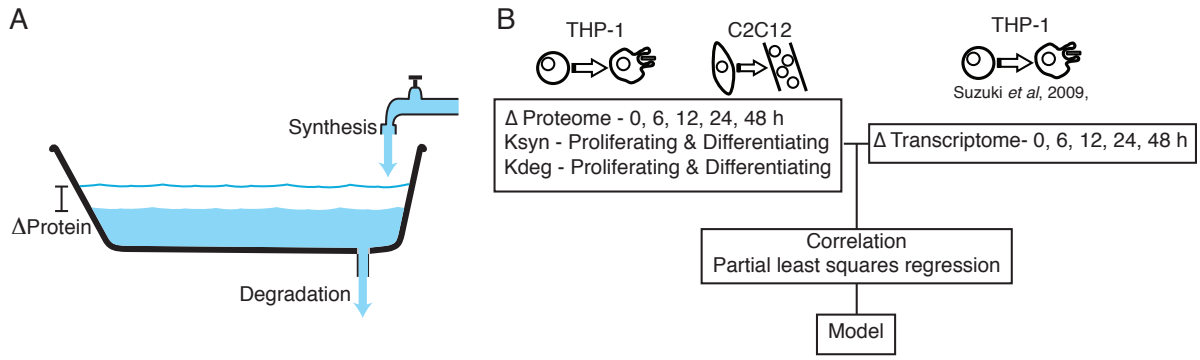
Chapter 2: Protein synthesis rate is the predominant regulator of protein expression during differentiation

2.1 Introduction

Proteins are not stable constituents in the cell; instead, they are continuously synthesized and degraded, leading to different turnover rates for individual proteins. An analogy of this, for any given protein, is a bathtub with an open drain, where the amount of water in the tub corresponds to the amount of protein, the water inflow from the faucet represents the synthesis rate, the water exiting through the drain represents the degradation rate and the change in the water level with time represents the change in expression of the protein (Figure 2.1A). If the inflow and outflow rate are equal, the level in the tub stays constant; however, the water is still exchanged (turned over) with a given velocity and similarly, proteins will also have different turnover rates.

In order to become more fit for a new state brought on by external perturbation, the cell needs to change expression level of many proteins through the regulation of a number of cellular processes, including transcription, protein synthesis and protein degradation. The protein synthesis rate has been shown to be regulated by miRNAs [115], mRNA change and different mRNA sequence features, whereas the protein degradation rate is predominantly regulated by the ubiquitin-proteasomal system [101,192]. Proteins with fast turnover rates are generally characterized by having low abundance [51,97], being intrinsically unstructured [193,194], aggregation prone [195,196] and involved in signal transduction and transcriptional activation [97,197,198].

Figure 2.1 Schematic drawing of regulation of protein expression change and data generated



A) Schematic drawing of the regulation of protein expression change. Water flow rate from faucet resembles protein synthesis rate, draining rate resembles protein degradation rate and water level change resembles protein expression change. B) Changes in overall protein expression, protein synthesis and protein degradation rates were measured during proliferation and differentiation of THP-1 and C2C12 cells. These data were analyzed on their own and also in combination with available transcriptomic data for THP-1 differentiation [199].

The last decade has seen the discovery of a number of characteristics defining the control of absolute expression level in bacteria, mouse and human under steady-state conditions, revealing how absolute expression level is mainly controlled by the protein synthesis rate, whereas the degradation rate has only a minimal contribution [51,112,200-202]. Much less information is available regarding the contribution of these processes to the regulation of protein expression change when the proteome needs to be rearranged, such as in response to external perturbation. Interestingly, studies that have examined the contribution of mRNA changes to changes in protein abundance have generally found a relatively poor correlation between them [43,109-111,201].

Protein metabolism has traditionally been investigated using isotope-labelled amino acids [93]; for radionuclides these were typically detected by autoradiography. Recently, it has become possible to measure the synthesis rate and/or the degradation rate of individual

proteins through the use of stable isotope labeled amino acids and mass spectrometry [51,97,115,203,204]. These experiments are often very time intensive, since they require incorporating the isotope labeled amino acids for varied length of time followed by quantifying the amount of incorporation. At the end, the synthesis rate can be calculated as a product of the incorporation of labeled amino acids over time. These types of experiments have provided numerous new biological insights, for instance that a single microRNA can repress the translation rate of hundreds of proteins [115], that the degradation rate of proteins is conserved between organisms [204] and that the degradation rate can vary for a protein according to the cellular compartment in which it is present [97,205]. Such efforts, however, have only focused on one aspect of the system, considering neither how synthesis and degradation may be related nor how these parameters are changed in response to stimuli.

Cellular differentiation is accomplished through a precisely orchestrated process where a proliferating cell stops dividing and acquires a new phenotype changing the proteins that it is expressing. The human THP-1 myelomonocytic leukemia cell line has been widely used as a model system to characterize the process of differentiation, since stimulation with phorbol 12-myristate 13-acetate (PMA) induces differentiation into a mature macrophage-like phenotype that no longer proliferates and is characterized by its ability to adhere to substrata, release O_2^-, and perform phagocytosis [206]. Similarly, the murine C2C12 myoblast cell line has also been widely used to characterize the differentiation process, since it can be differentiated by depriving confluent cells of serum, resulting in multinucleated myotubes.

Here our aim has been to characterize the various factors affecting regulation of protein expression change during cellular differentiation by measuring the synthesis and

degradation rates of proteins in proliferating and differentiating THP-1 and C2C12 cells using quantitative proteomics. By comparing these parameters to mRNA and protein expression change at multiple time points, we are able to construct a predictive model that highlights the contribution of different cellular processes to the regulation of the expression change of individual proteins during differentiation (Figure 2.1B).

2.2 Experimental procedures

2.2.1 Protein expression changes during differentiation of THP-1 and C2C12 cells

THP-1 cells were grown in RPMI media with 10 % dialyzed fetal bovine serum (FBS), 1 % glutamine, 1 % non essential amino acids, 1% penicillin/streptomycin and either (L-[U-$^{13}C_6$,$^{14}N_4$]arginine and L-[$^{2}H_4$]lysine or L-[U-$^{12}C_6$,$^{14}N_4$]arginine and [$^{1}H_4$]lysine or L-[U-$^{13}C_6$,$^{15}N_4$]arginine and L-[U-$^{13}C_6$,$^{15}N_2$]lysine (Cambridge Isotope Labs, Cambridge, MA). Cells were grown for at least five doublings to ensure ≥98% incorporation of labeled amino acids, before differentiated by applying 25 nM PMA for the given amount of time. Adherent and suspended cells were harvested by scraping and centrifugation at 600 relative centrifugal force (rcf) before being washed three times in PBS, lysed in 1 % deoxycholate, boiled for 5 min and, finally, equal amounts of protein from the three cell populations were mixed together.

C2C12 cells were grown in Dulbecco's Modified Eagle's medium (DMEM) with 20 % dialyzed FBS, 1 % glutamine, 1% penicillin/streptomycin and either (L-[U-$^{13}C_6$,$^{14}N_4$]arginine and L-[$^{2}H_4$]lysine or L-[U-$^{12}C_6$,$^{14}N_4$]arginine and [$^{1}H_4$]lysine or L-[U-$^{13}C_6$,$^{15}N_4$]arginine and L-[U-$^{13}C_6$,$^{15}N_2$]lysine (Cambridge Isotope Labs, Cambridge, MA). The cells were split from 50% confluence to 100 % confluence at time 0 h while being

transferred to similar media as described above except that 2 % FBS was added. Afterwards, the cells were washed three times in PBS, scraped off the plate in PBS, lysed in 1 % deoxycholate, boiled for 5 min and, finally, equal amounts of protein from the three cell populations were mixed together.

2.2.2 Measurement of relative synthesis and degradation rates

Two populations of cells were grown in the light and the medium SILAC media as described above for the respective cell lines. The "light" cells were harvested at time 0 h and frozen, while simultaneously the "medium" cells were washed three times with PBS and transferred to the "heavy" form of amino acids. After 48 h of proliferation or differentiation, (24 h for the proliferating C2C12 cells, which grow much faster) the cells were harvested, lysed and mixed with the "light" lysate before proceeding as below.

2.2.3 Lysate preparation

The lysates were treated with benzoase before being reduced, alkylated and digested as previously described [207] and then a total of 100 μg was separated by isoelectric focusing (Agilent technology) according to the manufactures instructions. The peptides were subsequently cleaned up [208] and analyzed by LC-MS/MS, as previously described [15].

2.2.4 Identifying the synthesis and degradation rates of proteins in macromolecular sub-complexes

Two populations of C2C12 cells were fully incorporated with "light" and "medium" amino acids before the "medium" population was switched to media containing "heavy"

amino acids for 24 h. The “light” and “medium/heavy” cells were subsequently mixed together in a 1:4 ratio (to correct for the growth of the light cells), before the cells were lysed with a Dounce homogenizer in SEC buffer (20 mM Tris, 50 mM sodium acetate, 50 mM KCl, pH 7.2) including Halt protease and phosphatase inhibitors cocktail (Thermo Scientific). The lysate was concentrated using ultrafiltration (100,000 molecular weight cutoff, Sartorius Stedim) and loaded onto on the preparative HPLC (Agilent technology) equipped with HPLC-SEC columns (300x7.8 mm Yarra-4000 (Phenomenex)) and fractionated into 48 fractions at 8°C, at a flow rate of 0.5 mL/min. The proteins of the individual fractions were digested to peptides and analyzed by mass spectrometry as described previously [15].

2.2.5 Data processing of the MS data

Tandem mass spectra were extracted, searched and quantified by MaxQuant (v2.3.0.5) [48]. The search was performed against Uniprot (proteomes) human (69906 sequences) 21/6/2011 or mouse (55269 sequences) 20/4/2012 with common serum contaminants and enzyme sequences added, 1% FPR on protein and peptide levels, trypsin/P cleavage rule with a maximum of 2 missed cleavages, 0.5 Da tolerance for MS/MS, Carbamidomethylation on cysteines as the sole fixed modification and oxidation of methionines and acetylation on protein N-terminal as variable modifications. Finally, we used a ratio count of 2 for quantification and the match-between-run-feature with a window of 2 min for quantification.

2.2.6 Identifying enriched signatures for differentially expressed proteins

Proteins whose abundance changed significantly during differentiation were identified by applying ANOVA between the five time points with the following settings (Permutation-based FDR, P=0.05, S0=1, 250 randomizations) using Perseus. Increasing and decreasing proteins were defined by clustering the data into two clusters using fuzzy C mean clustering by the following settings (c=2, m=2). Functional enrichment analysis was done using the Fisher exact test (P=0.05) the Benjamini-Hochberg correction with a minimum of 5 proteins per category using GPROX [209].

To investigate if proteins that were changing in expression were different in abundance from the non-regulated proteins, a Wilcoxon-Mann-Whitney test was applied to compare summed up eXtracted ion Current of the two groups at 0 h differentiation.

2D enrichment analysis ($P<0.05$) of the biological processes (Uniprot Keywords) that were enriched in any dimension of protein expression changes after 48 h in THP-1 and C2C12 cells were generated using Perseus according to [210], with a Benjamini-Hochberg FDR for truncation.

2.2.7 Data processing of synthesis and degradation data

The medium/light and heavy/light ratios were normalized by dividing by their median values before they were $\log_2$ ($-\log_2$ for medium/light) and Z-transformed. This approach to identifying relative synthesis and degradation rates resembling that used in cDNA microarray data analysis where a Z-transformation is also used to compare transcript changes between different experiments [211].

To identify proteins whose relative synthesis or degradation rates were significantly altered, we performed two-sided T-tests ($P<0.05$, $S0=1$) between the normalized data in proliferating and differentiating cells, where a Permutation-based FDR (250 randomizations) was used for truncation.

We investigated the characteristics for proteins with similar synthesis and degradation rates by measuring the Euclidian distances in synthesis/degradation space between all the proteins assigned to the same biological process (using Uniprot keywords). The P-value was calculated using Wilcoxon-Mann-Whitney test, to test if the distances in a biological process were significantly different from distances between random chosen proteins (50 iterations). A similar approach was used to investigate if there were any proteins in a macromolecular complex (defined by the CORUM database [212]) that displayed significantly different synthesis/degradation rates than the rest of the proteins in a macromolecular complex. Here, Euclidian distances were calculated among all synthesis/degradation rates of the proteins in a complex and Wilcoxon-Mann-Whitney test was applied to investigate if any proteins displayed significantly different distances than the rest of the proteins.

To investigate if proteins in macromolecular complexes (defined by the CORUM database [212]) display more similar synthesis and degradation rates than the rest of the proteins, we calculated the Euclidian distances between all proteins in a macromolecular complex and investigated if these distances to proteins not participating in macromolecular complexes were significantly smaller by Wilcoxon-Mann-Whitney test.

2.2.8 Identifying enriched signatures for proteins with fast/slow degradation rates

Degradation rates measured in proliferating human THP-1 and mouse C2C12 cells were paired between the two organisms based on their unique gene names taken from the HGNC homepage (http://www.genenames.org/) [213]. The proteins were considered to have fast degradation rates if the rate was among the fastest 20% in both organisms, and vice versa for slow degradation rates.

To investigate if proteins that had slow and fast degradation rates, respectively, displayed differences in abundance, we performed Wilcoxon-Mann-Whitney test between the two groups of the ion intensities from the light label.

To investigate if proteins that had slow and fast degradation rates, respectively, displayed differences in disordered structure, we performed Wilcoxon-Mann-Whitney test between the two groups using the % disorder per protein as calculated from the human structure using disorder2 software [214].

For KEN motif analysis, we calculated the number of KEN motifs in each protein sequence using an in house algorithm in Matlab (matworks.com) that searched for the canonical KENXXX[NDEQ] motif [215]. A Fisher's exact test between the two populations was used to calculate significance.

2.2.9 Data analysis for identification of synthesis and degradation rate of the proteins in macromolecular sub-complexes

Proteins were identified and quantified using MaxQuant with similar settings as above except label free quantitation was enabled. Chromatograms of the individual proteins were constructed using the label free quantitation value of the light label in each fraction, and

by smoothing these curves using a moving average of three consecutive fractions in Matlab. The relative synthesis and degradation rates of the proteins in the individual fractions were calculated from the normalized values of H/L and M/L which was $\log_2$ and $-\log_2$ transformed, respectively.

2.2.10 Identifying the correlation between mRNA and protein expression change

The mRNA dataset derived from [199] was filtered for a detection value larger than 0.95. The counts for 6, 12, 24 and 48 h differentiation were divided by the counts at 0 h before the ratios were $\log_2$ transformed. Changing mRNA was defined similarly to the proteins (see above) using Perseus. Gene-name identifiers were used to pair mRNA and protein expression values.

The Spearman correlation was calculated using Perseus between mean for the mRNA and proteins expression values for the gene having two out of three expression values in both datasets. Changing genes was defined as changing expression both on the mRNA and protein level using the above criteria.

The Spearman correlation between synthesis rate and mRNA and protein expression was calculated using the data from the three biological replicates of each of the parameters resulting in nine correlation values each, from which we could draw boxplots.

2D enrichment analysis ($P<0.05$) of the biological processes (Uniprot Keywords) that were enriched in any dimension of protein or mRNA expression change after 48 h THP-1 differentiation were generated using Perseus similar to [210] with a Benjemini-Hochberg FDR for truncation.

2.2.11 Partial least square regression analysis

The partial least square (PLS) regression analysis was performed in Matlab using the *plsregress* function. We calculated PLS regression for genes that had transcriptome expression data after 48 h differentiation, proteome expression data after 48 h differentiation and synthesis and degradation rate data in two out of three biological replicates using the mean of the values; the increasing and decreasing proteins were defined as in the expression experiment as above. The P-values in Table 2.2 was calculated by the pls-package in R.

2.2.12 Determining the effects of perturbation on degradation and synthesis rates

The correlations between synthesis and degradation rates in differentiating versus proliferating cells were calculated using the data from the three biological replicates of each of the parameters resulting in nine correlation values each, from which we could draw boxplots.

Biological processes (Uniprot Keywords) that were enriched ($P<0.05$) in any data dimension of synthesis or degradation rates between differentiating and proliferating cells were generated using Perseus and Benjemini-Hochberg FDR for truncation.

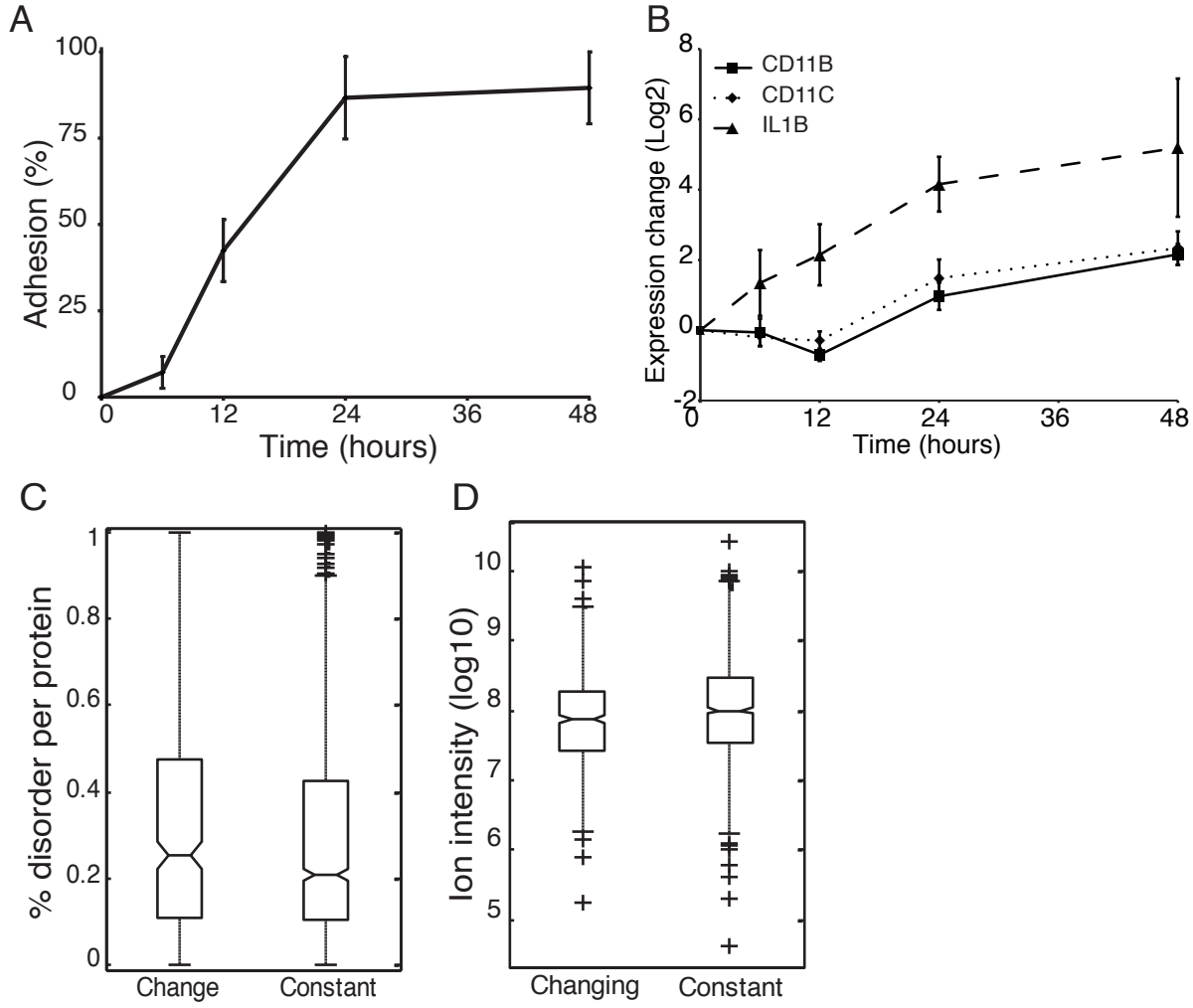
To produce the heatmap for the differentially regulated proteins and proteins with differentially regulated synthesis and degradation rates, we performed enrichment analysis of the biological processes (Uniprot keywords) using GPROX [209] against the increasing and decreasing protein's expression, synthesis and degradation rates, using Benjamini and Hochberg correction with a minimum of 5 proteins per category ($P<0.05$).

2.3 Results

2.3.1 Quantitative proteomics reveals mechanisms behind phenotypic changes

The primary system used here was the human THP-1 myelomonocytic leukemia cell line that can be differentiated from monocytes into macrophage-like cells by stimulation with 25 nM PMA. The measurements made here require metabolic labeling so we turned to SILAC [63], using two triplex SILAC experiments to cover 5 time points of differentiation in THP-1 cells. The cells were stimulated with PMA for 0, 6, 12, 24 and 48 h, which causes several morphological changes, including adherence (Figure 2.2A); biological triplicates were performed for all experiments. Whole cell lysates from the different conditions were combined as appropriate and the samples were split in two, with the two halves being pre-fractionated, either on the protein level by SDS-PAGE and on the peptide level by isoelectric focusing, before analysis on a LTQ-OrbitrapXL mass spectrometer. A total of 4977 proteins could be identified from the tandem mass spectra at a false discovery rate (FDR) below 1 %. Among the identified proteins were very well-known markers for monocyte to macrophage differentiation such as CD11B, CD11C and IL1B (Figure 2.2B), demonstrating that the cells were differentiating as expected and that the method is sensitive enough to detect relevant changes.

Figure 2.2 Proteome changes reflect phenotype



A) Fraction of THP-1 cells adhered versus time after differentiation was initiated by the addition of 25 nM PMA. B) Relative expression changes of differentiation markers during THP-1 differentiation, as measured by SILAC in the proteomic data. Error bars in A) and B) indicate standard deviation of three biological replicates. C) Proteins changing expression during differentiation are significantly (P=0.03, Wilcoxon-Mann-Whitney test) enriched in intrinsically disordered proteins. D) Proteins changing expression following perturbation are significantly ($P=2.8 \cdot 10^{-5}$, Wilcoxon-Mann-Whitney test) enriched in low abundant proteins. For the boxplots in C) and D), the line denotes the median, the box denotes the 1st and 3rd quartile, the notch denotes the 5 % significance level and the whiskers extend to the most extreme values, excluding outliers.

To identify proteins with significant differential expression, we applied a strict statistical test (ANOVA $p<0.05$, S0=1) among the five time points and found 457 proteins with significant changes in expression during differentiation, which could serve as potentially

new markers for the differentiation process. Functional enrichment analysis of these two groups of proteins revealed that the majority of the originally described phenotypic changes associated with differentiation, such as halting of the cell cycle, increased adhesion and increased lysosomal capacity, could be observed (Table 2.1). In addition, those proteins altered by differentiation tended to be more intrinsically disordered (P=0.03, Wilcoxon-Mann-Whitney test) (Figure 2.2C) and were generally lower in abundance when using ion intensities as a proxy for abundance ($P=2.8 \cdot 10^{-5}$, Wilcoxon-Mann-Whitney test) than the unregulated proteins (Figure 2.2D).

Table 2.1 Functional enrichment analysis of proteins increasing or decreasing in response to THP-1 differentiation

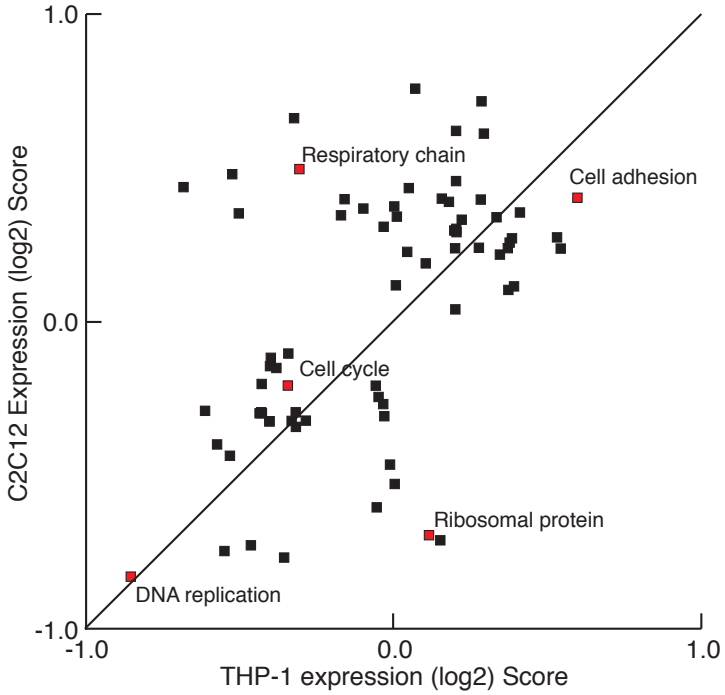
Increase	P-value	Decrease	P-value
Cell membrane	1.90E-07	DNA replication	6.00E-11
Cell adhesion	2.30E-06	Cell cycle	2.70E-06
Calcium	2.40E-05	Cell division	5.10E-03
Secreted	2.70E-05	Secreted	9.80E-04
Lysosome	4.80E-02	Zinc-finger	3.30E-02

Selected biological processes enriched among the increasing and decreasing proteins. The P-values have been Benjamini and Hochberg corrected.

To complement these data with a completely unrelated system of differentiation, we next performed a similar study using mouse C2C12 cells that were induced to differentiate from myoblasts, by deprivation of serum at confluence, to myotubes for 48 h. Even though the myoblasts-to-myotubes and monocytes-to-macrophages systems are quite distinct and from different organisms, many of the same mechanisms appear to be employed in order to achieve the required changes. Some similar classes of proteins were also involved in the two processes, such as a general depression in the level of cell cycle proteins. The starting and

end points in the two systems are quite disparate, however, so there were also a number of differences (Figure 2.3); e.g., many mitochondrial proteins, particularly those involved in the respiratory chain, increased during myotube differentiation, consistent with increased energy needs of these cells, whereas proteins involved in cell junctions and adhesion increased during macrophage differentiation. Overall the general patterns and specific markers observed to change in both systems were entirely consistent with what is already known, suggesting that the data faithfully reflects the changes in the proteome that underlie the morphological changes.

Figure 2.3 2D enrichment analysis of the protein expression in differentiating C2C12 and THP-1 cells



2D enrichment analysis was performed similarly to Cox et al. [210] ($P<0.05$) and detects proteins in the two cell lines that display consistent behavior in any of the data dimensions versus the rest of the proteins in the dataset. The cellular processes located on the diagonal suggest these are similarly regulated in THP-1 and C2C12 cells, whereas the processes located off-diagonal are regulated differently between THP-1 and C2C12 cells.

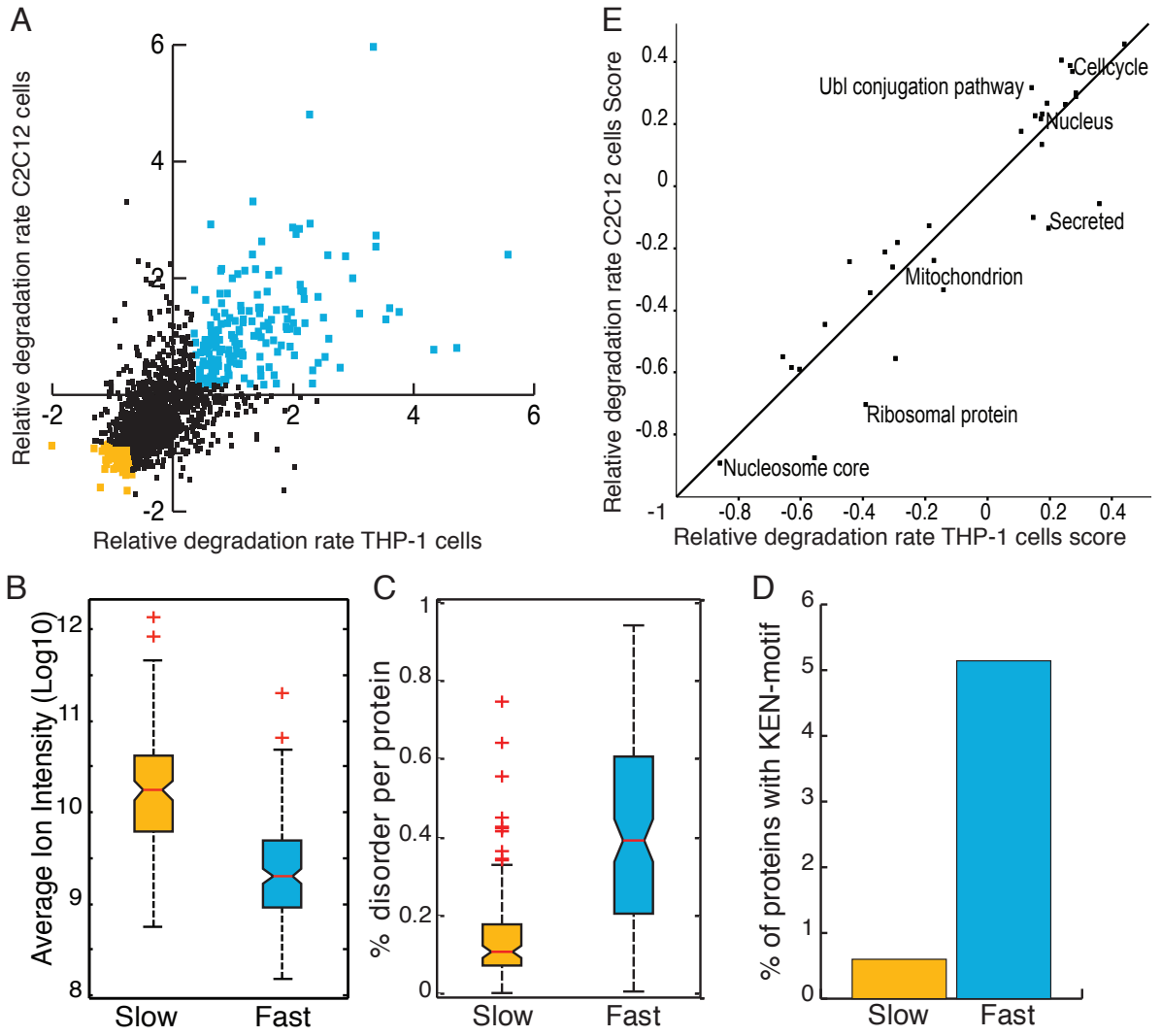
2.3.2 Measuring relative synthesis and degradation rates

To gain insight into the factors that regulate protein expression change, we decided to measure the synthesis and degradation rates of proteins in both proliferating and differentiating THP-1 and C2C12 cells. The rationale behind this was that it should allow us to detect how the synthesis and degradation rates change for individual proteins in different organisms and different differentiation pathways, hereby allowing us to deduce some of the general regulatory mechanisms at work here.

To increase the throughput of this study we decided to measure the relative synthesis and degradation rates of the individual proteins by slightly modifying a previously described approach [97]. We first completely incorporated two cell populations using light and medium amino acids. Then, by switching the medium population to heavy amino acids at the same time the induction of differentiation, all newly synthesized proteins are made with heavy forms of amino acids and all proteins being degraded can be monitored as a decrease in medium amino acids, with the light-labelled population acting as an internal control against which the medium and heavy can be compared. These measurements were made in proliferating myoblasts and monocytes, as well as during the differentiation of each cell type, thereby allowing us to deduce the changes in the synthesis and degradation rates of the individual proteins between proliferating and differentiating cells. Biological triplicates of all the above experiments resulted in the identification of 5721 proteins with a FDR below 1 %. We subsequently calculated the relative degradation and synthesis rates of the individual proteins by Z-transforming the medium/light and heavy/light ratios of the proteins, respectively.

With these data in hand, we first asked if the approach of identifying relative degradation and synthesis rates could identify some of the turnover characteristics previously described in studies that had measured absolute protein turnover. We therefore defined unstable proteins as the 20% of proteins with the fastest degradation rate and stable proteins as the 20% of proteins with the slowest degradation rate in both murine and human cells (Figure 2.4A). By using the average intensity from the mass spectrometry data as an approximation of the protein amount, we were able to confirm what has been reported previously, that stable proteins are more abundant than unstable proteins ($P=1.7 \cdot 10^{-30}$, Wilcoxon-Mann-Whitney test) (Figure 2.4B).

Figure 2.4 Relative protein degradation rates capture known protein turnover characteristics



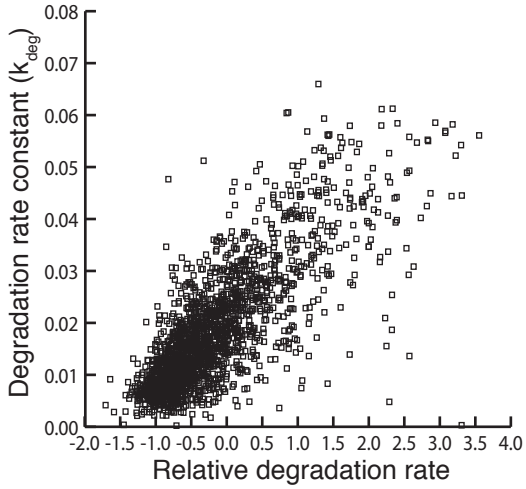
A) A scatterplot of degradation rates in proliferating THP-1 and C2C12 cells shows the values roughly cluster around the diagonal. We arbitrarily defined the fastest (blue) and slowest (yellow) 20% as unstable and stable proteins selected, respectively. B) The most stable proteins defined in A) are significantly ($P=1.7\cdot10^{-30}$, Wilcoxon-Mann-Whitney test) more abundant, as determined by the average intensities of their peptides, than unstable proteins. C) Conversely, unstable proteins have less intrinsic structure than stable proteins ($P=1.3\cdot10^{-27}$, Wilcoxon-Mann-Whitney test) based on the Disorder2 algorithm [214]. Related to this, D) unstable proteins also have more KEN-box motifs than stable proteins (P=0.02, Fisher's exact test). E) Degradation rates for functional classes of proteins are largely similar ($P<0.05$) between THP-1 and C2C12, based on 2D enrichment analysis [210].

We next measured the percentage of disorder of the individual proteins using the Disorder2 algorithm [214], since it has been reported that intrinsic disorder could facilitate faster turn-over of proteins [193,194]. Consistently, we find that unstable proteins are significantly enriched in intrinsic disorder ($P=1.3\cdot10^{-27}$, Wilcoxon-Mann-Whitney test) (Figure 2.4C). Similarly, we found that proteins recognizes by the anaphase promoting complex which are characterized of containing a KEN-box motif are significantly enriched (P=0.02, Fisher exact test) among proteins with a fast degradation rate, as has also been reported previously [215] (Figure 2.4D).

We then investigated how conserved the degradation rates for individual proteins and the general biological processes involved in differentiation were between the murine and human cell lines. For the individual proteins, we observed a high correlation of the degradation rates (0.65; Spearman correlation) between the murine and the human proliferating cells, suggesting that the turnover rates of individual proteins are strongly influenced by features encoded by the protein sequence [204]. Previous observations show how proteins involved in different biological processes display different stabilities so, to test this, we performed two-dimensional enrichment analysis between the proliferating C2C12 and THP-1 cells [210]. Briefly, this investigates if cellular processes are displaying consistent behavior in any of the data dimensions versus the rest of the proteins in the dataset. This again revealed a very consistent regulation between the two cell lines, since the majority of biological processes are located on the diagonal (Figure 2.4E). The relative degradation rates of the individual cellular processes also display similar regulation to what has been reported before: proteins involved in the ubiquitin pathway and the cell cycle have high degradation rates, whereas proteins of the nucleosome core (histones), ribosome and mitochondria

display slow degradation rates [97,204]. As a final verification of this approach, the relative degradation rates measured here for C2C12 correlate very well (0.77, Spearman correlation) with absolute rates measured in the same cell line, albeit for fully differentiated cells [204] (Figure 2.5).

Figure 2.5 Correlation of absolute and relative degradation rate



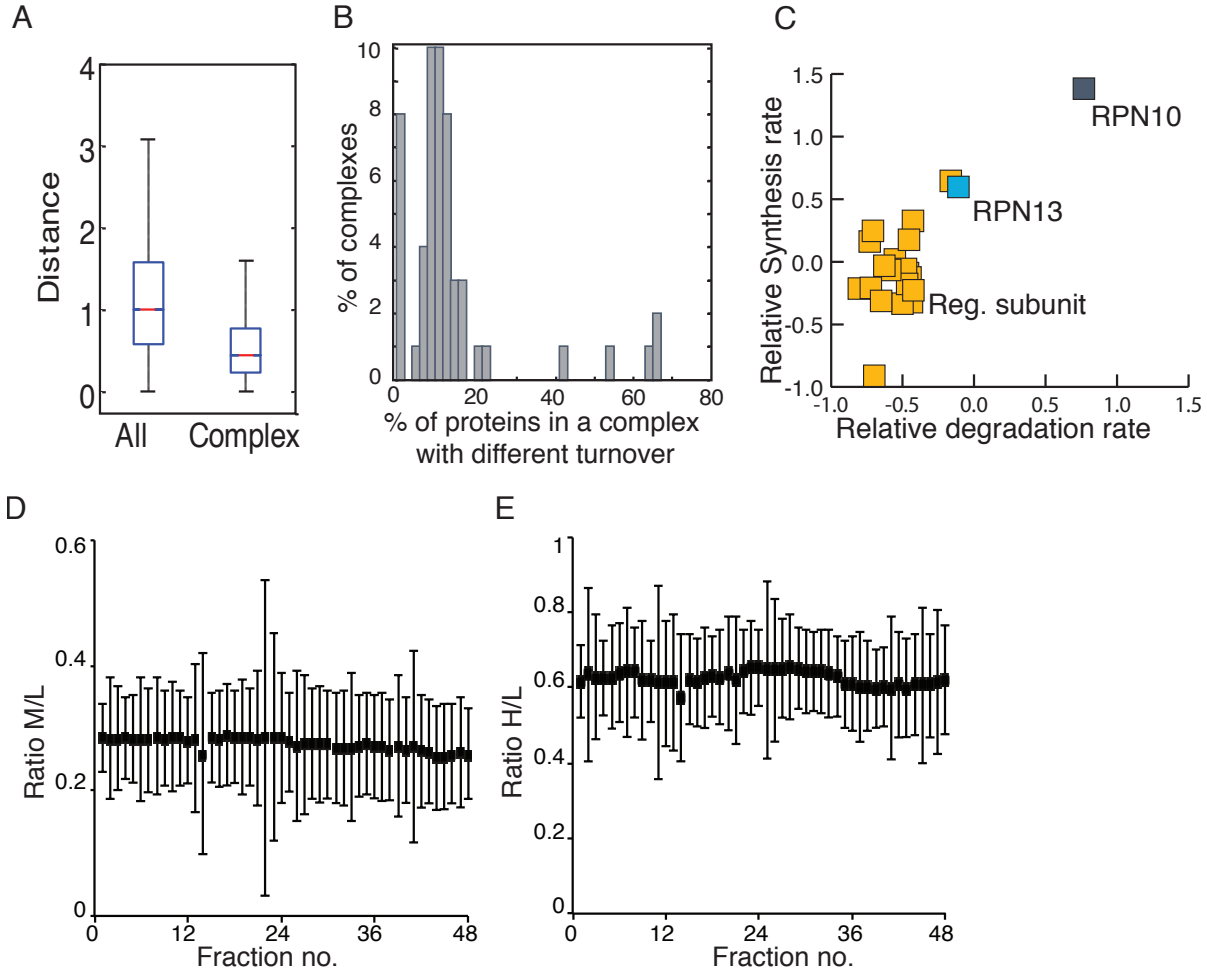
By comparing our identified relative degradation rate with the degradation rate constant (k_{deg}) for the individual proteins identified by Cambridge *et al*. [204] we observed a correlation of 0.77 (Spearman correlation), which strongly suggests that the relative degradation rate is a valid approach to investigate protein degradation. The study by Cambridge *et al*. was performed in C2C12 cells as well, albeit only in fully differentiated cells.

To this point all our observations of these data indicate that measuring relative synthesis and degradation rates of the individual proteins faithfully captures the characteristics of the system while increasing the throughput compared to conventional determinations of absolute rates.

2.3.3 Turnover rates of macromolecular sub-complexes

Having establish the veracity of the data from the parallel measurement of synthesis and degradation rates under both proliferating and differentiating conditions in two different cell lines, we next asked if there were any characteristics common to the proteins that display similar synthesis and degradation rates. As a measurement of similarity of the synthesis and degradation rates of the proteins, we calculated the Euclidian distances between proteins involved in the same biological process in synthesis-degradation space and determined if this distance was significantly shorter for particular groups of proteins than random chance would dictate. Two proteins with similar synthesis and degradation rates should therefore have short Euclidian distances, whereas protein with very different synthesis and degradation rates will be farther apart. By measuring the distances between proteins involved in the same biological process of both proliferating and differentiating THP-1 and C2C12 cells, it became especially clear that proteins participating in macromolecular complexes had very similar synthesis and degradation rate. A more specific way of showing this correlation is seen in Figure 2.6A, which depicts the Euclidian distances between the synthesis and degradation rates for all the proteins in complexes present in the CORUM database. Proteins within a complex are vastly ($P<10^{-99}$, Wilcoxon-Mann-Whitney test) closer in Euclidian space than the rest of the proteins (Figure 2.6A).

Figure 2.6 Turnover characteristics of macromolecular complexes



A) The Euclidean distance (in synthesis-degradation space, see Methods) between proteins involved in macromolecular complexes tends to be much shorter ($P<10^{-999}$, Wilcoxon-Mann-Whitney test) than the distance between random proteins. B) Distribution of proteins in macromolecular complexes having significantly ($P<0.01$) different synthesis and/or degradation rates than the rest of the proteins within the macromolecular complex. C) The relative synthesis and degradation rates of the components of the proteasome regulatory particle, which are generally tightly clustered in a scatterplot of synthesis vs. degradation rates. Spread of relative degradation (D) and synthesis rates (E) of all the proteins detected in the individual fractions eluting from the SEC column

To examine this co-regulation of interaction partners in more detail, we asked if this observation was constantly true for all the members of a macromolecular complex or if certain proteins in a complex had markedly different synthesis and/or degradation rates than the rest. This question was also answered by calculating the Euclidian distances among all

members of a complex and testing if any proteins displayed significantly different synthesis and/or degradation rates ($P<0.01$, Wilcoxon-Mann-Whitney test). Of the proteins known to be participating in macromolecular complexes, 30±1 % had significantly different synthesis and/or degradation rates than the other members of the same complex. However, this distribution was far from uniform (Figure 2.6B), since the fraction of proteins displaying synthesis and degradation rates distinct from those of their binding partners ranged from only one member (RPN10) of the proteasome regulatory particle (Figure 2.6C) to 50±15 % of the 60S ribosomal subunit. This observation was consistent under proliferation and differentiation conditions across both cell types (Data not shown).

Since macromolecular complexes often are comprised of different sub-complexes or have certain subunits that are not always present, we decided to take advantage of the throughput of the above described approach to directly measure synthesis and degradation within individual complexes using PCP-SILAC and SEC [15]. This was accomplished by first completely incorporating two C2C12 cell populations using light and medium amino acids and subsequently switching the medium population to heavy amino acids and allowing the cells to proliferate for 24 h. However, this time the cells were lysed without the use of detergents and the resulting lysate was separated by SEC into 48 fractions, before the proteins in each individual fraction were digested to peptides and analyzed by LC-MS/MS. Hereby, the individual sub-complexes will be separated out by SEC and elution profiles of the individual proteins can be constructed, using label free quantitation [48], from the ion intensities of the light form of each peptide, whereas the relative synthesis and degradation can be determined by the H/L and M/L ratios, respectively. By this approach we were able to identify 2423 proteins in 48 SEC fractions, from which we observed proteins from all major

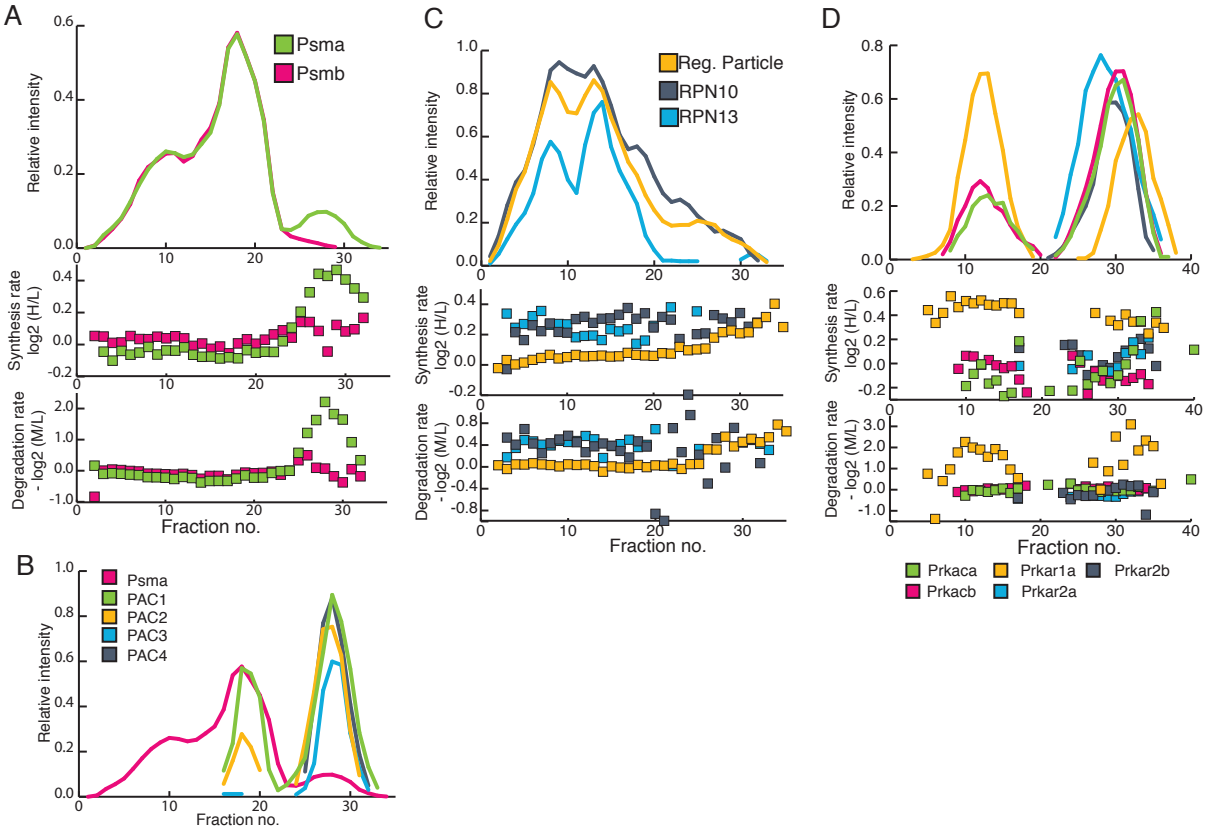
cytosolic protein complexes such as the ribosome, signalosome and proteasome; however, we also discovered protein complexes involved in signaling transduction such as the Protein kinase A (PKA) and the NFκB complex.

We first investigated if proteins participating in large macromolecular complexes had different synthesis and degradations rates than smaller macromolecular complexes or than the proteins in their free form, by taking the median of the synthesis and degradation rates of the proteins in each fraction. This did not reveal any correlation, suggesting that the synthesis and degradation rates of the macromolecular complexes are independent of the size of the complex (Figure 2.6D-E).

Since the protein composition of the proteasome is very well characterized on the sub-complex level, we decided to investigate the synthesis and degradation rates of each of the sub-complexes in detail. We had earlier discovered that the proteins of the proteasome core particle displayed very similar synthesis and degradation rates, so we investigated if that was still true at the sub-complex level. Plotting the elution profile and synthesis and degradation rates of the alpha and beta- subunits revealed similar synthesis and degradation rates when these co-elute with regulatory subunit, suggesting that the 26S proteasome is degraded as an intact complex (Figure 2.7A). Interestingly, we noticed that around fraction 29 the alpha subunit displayed an additional peak, which co-eluted with the peaks for the proteasome assembling chaperones (PAC1-PAC4) (Figure 2.7B**)** and is characterized by only containing newly synthesized proteins, suggesting that we are able to capture the dynamic process of assembly of the proteasomal alpha ring from newly synthesized proteins [128]. This means that even though the proteins of the proteasome core particle display similar

synthesis and degradation rates on the global level, the individual sub-complexes that make up the 26S can be subject to very different regulation.

Figure 2.7 Synthesis and degradation rates are consistent among members of protein complexes



A) Top, the median size exclusion chromatograms of the alpha- and beta-subunits of the proteasome core particle. Middle, the synthesis rate of the alpha- and beta-subunits of the proteasome core particle. Bottom, degradation rate of the alpha- and beta-subunits of the proteasome core particle. B) Psma and Pac1-4 co-elute in size exclusion chromatography. The alpha subunits of the proteasome core particle typically elute from size exclusion chromatography in at least three distinct peaks, with the peak around fraction 29 co-eluting with the proteasome assembling chaperones (PAC1-PAC4). C) Top, the median size exclusion chromatogram of the regulatory particle of the proteasome, RPN10 and RPN13. Middle, synthesis rate of regulatory particle of the proteasome, RPN10 and RPN13. Bottom, the relative degradation rates of the regulatory particle of the proteasome, RPN10 and RPN13. D) Top, size exclusion chromatogram of Prkaca, Prkacb, Prkar1a, Prkar2a, Prkar2b. Middle, the synthesis rates Prkaca, Prkacb, Prkar1a, Prkar2a, Prkar2b. Bottom, degradation rates of Prkaca, Prkacb, Prkar1a, Prkar2a, Prkar2b.

In our initial measurements of synthesis and degradation rates, we observed that the ubiquitin receptor RPN10 of the proteasome regulatory particle is synthesized and degraded faster than the rest of the proteins of the regulatory subunit (Figure 2.6C). This could be a result of RPN10 participating in additional complexes that are regulated differently than the proteasome and thus result in distinct average rates for RPN10 or because RPN10 is simply turned over differently than all the other components of the proteasome. We therefore compared the elution profile and turnover rates of RPN10 to the rest of the proteins within the regulatory particle, revealing that the elution profile of RPN10 is very similar to the rest of the regulatory particle, yet its turnover rate even in that region of the size exclusion chromatogram was still significantly faster than the rest of the proteins of the regulatory particles ($P = 2.8 \cdot 10^{-8}$, Wilcoxon-Mann-Whitney test) (Figure 2.7C). Intriguingly, we noticed that the other ubiquitin receptor of the proteasome RPN13 displayed similarly high turnover rates while bound to the regulatory particle ($P = 3.7 \cdot 10^{-8}$, Wilcoxon-Mann-Whitney test), suggesting that both the ubiquitin receptors RPN10 and RPN13 can exchanged with the free forms of RPN10 and RPN13, and that this approach allows provides a completely novel ability to probe such details.

Finally, we investigated the synthesis and degradation rates for the PKA complex that is composed of different combinations of homo- or heterodimers of the regulatory subunit associated with two catalytic subunits. We observed that the elution profile of the PKA complex showed at least two distinct peaks from the SEC column; one peak corresponding to the catalytic subunits (Prkaca and Prkacb) co-eluting only with Prkar1a, and one corresponding to the catalytic subunits co-eluting with Prkar1a, Prkar2a and Prkar2b (Figure 2.7D). Interestingly, the catalytic subunits displayed identical synthesis and degradation rates

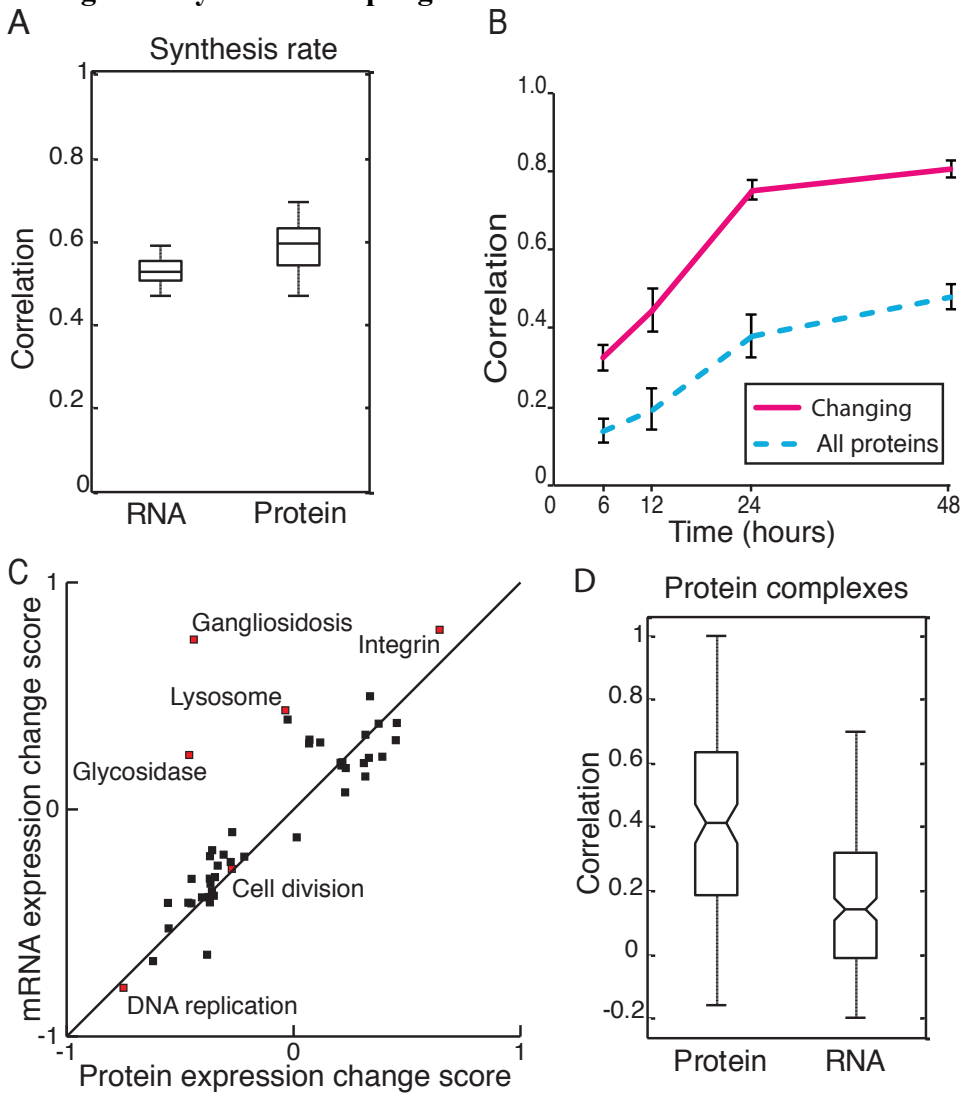
in both complexes, and these were significantly slower than the synthesis and degradation rates for the Prkar1a alone. On the other hand, the synthesis and degradation rates for Prkar2a and Prkar2b were similar to the catalytic subunit when they co-eluted, suggesting that the synthesis and degradation rates on the sub-complex level provides important modular information of the composition of the macromolecular complex.

These data extend the well-known concept of modularity of macromolecular complexes to highlight that production and recycling of such complexes can be regulated at the sub-complex level and that looking at average rates for proteins across the whole cell will obscure important functional details.

2.3.4 Temporal correlation between the transcriptome and the proteome

Many studies have found a poor correlation between changes in mRNA and protein levels in response to perturbation [43,109] but the Central Dogma still suggests that there must be a link. We therefore decided to correlate our datasets of protein expression change and the relative synthesis rate data with a recent published transcriptomic study, which also studied differentiation of THP-1 cells using similar stimuli and time points as our experiments [199]. First, we investigated how protein synthesis correlated with the mRNA and protein expression changes, which both revealed moderate correlation of 0.52±0.07 and 0.59±0.07, respectively (Figure 2.8A). This highlights how protein synthesis is regulated by a number of post-transcriptional processes, and that the levels of proteins within a cell are regulated by processes beyond just protein synthesis.

Figure 2.8 Correlation between mRNA and protein expression increases over time during monocyte to macrophage differentiation



A) Spread of the correlation between the protein synthesis rate and final changes in mRNA or protein expression changes after 48 hours THP-1 differentiation for three biological replicates of each parameter. The center horizontal line represents the median, the box spans the 1st through the 3rd quartiles and the whiskers span the most extreme values, excluding outliers. B) Plot of the level of correlation between mRNA and protein expression changes across 48 h of differentiation. The solid line represents genes that are changing on the mRNA and protein level while the dotted line represents all the genes. Error bars denote the standard deviation of three biological replicates of both transcriptomic and proteomic experiments. C) 2D enrichment ($P<0.05$) of functional classes according to the expression changes observed in the proteome vs. the transcriptome. D) Correlation of mRNA versus protein expression changes for members of CORUM complexes during differentiation of THP-1 cells reveals that protein expression change are more considerably better predictors for interactions than are mRNA change ($P = 7.9 \cdot 10^{-14}$, Wilcoxon-Mann-Whitney test).

Few studies have investigated how the correlation between expression change of mRNA and proteins changes over time in response to perturbation, and the correlation has only been investigated after relatively short-term perturbations [109,110]. When we examined the correlation of mRNA and proteins expression change at each of the five time points, we observed that the relationship nearly reached a steady state after 24h differentiation, suggesting that extensive post transcriptional regulation is taking place during early differentiation (Figure 2.8B). Interestingly, however, proteins and mRNAs that were being up- or down-regulated during differentiation were much more highly correlated, with less than 6 % of the genes showing opposite regulation at 48 hours differentiation, suggesting that post-transcriptional regulation is mainly just fine tuning the levels of these proteins. To investigate the regulation of the different cellular processes on the transcriptional and proteome level, we turned to 2D enrichment analysis. This revealed a generally consistent regulation of the cellular processes in both the transcriptome and the proteome of differentiating THP-1 cells; e.g., both the level of proteins involved in DNA replication and cell cycle and their mRNA decrease concomitantly upon differentiation (Figure 2.8C).

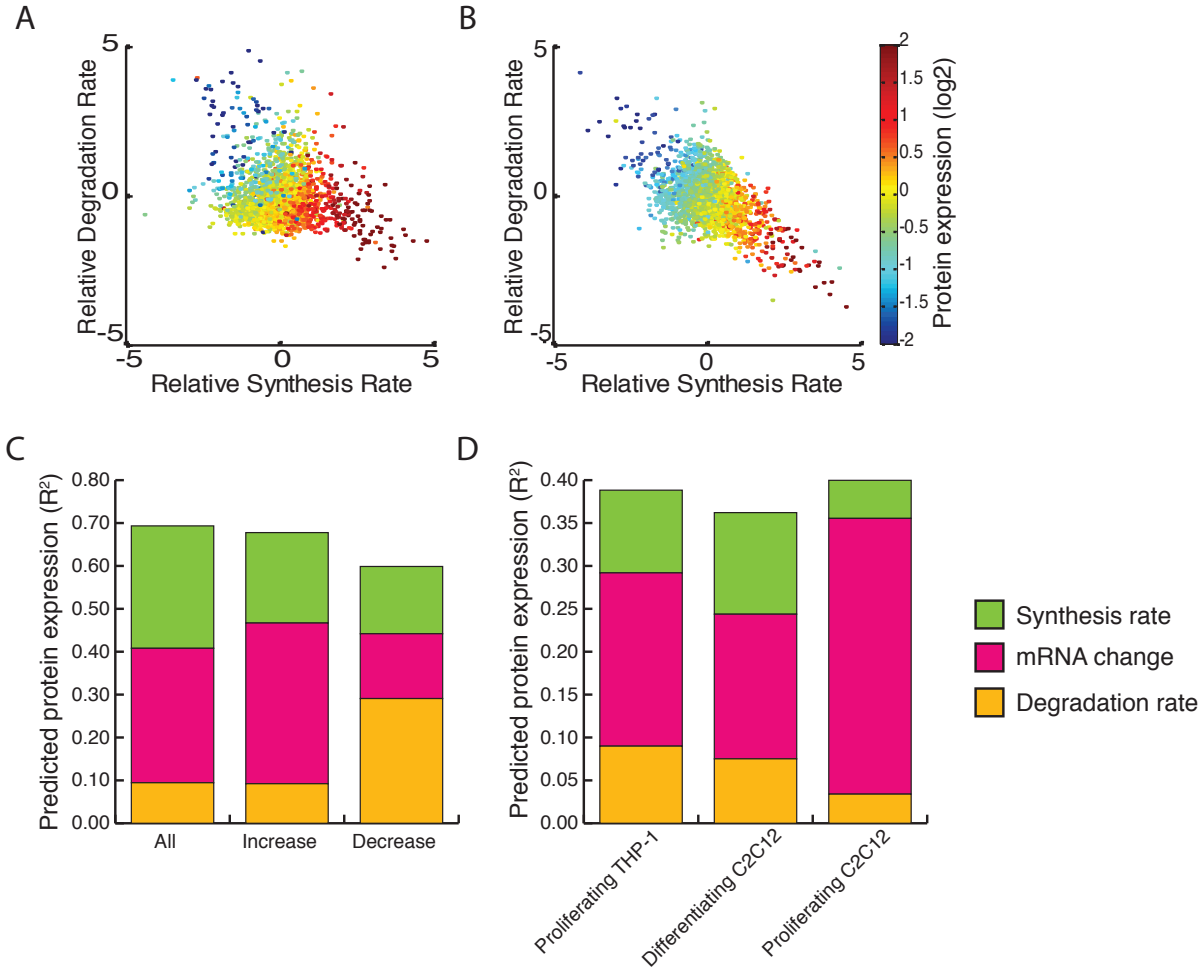
Messenger RNA expression changes have been widely used to predict protein-protein interactions since proteins with similar mRNA expression change are more likely to interact than a random selection of proteins [216]. Since we have mRNA- and protein-expression change recorded in the same system at similar time points, we explored what parameters are most predictive for interactivity. We focused only on those proteins annotated as components of complexes in the CORUM database, as this is the most widely accepted gold-standard interaction set. A comparison of how closely the mRNA versus protein expression change tracked for members of CORUM complexes reveals that protein expression change are vastly

more predictive for interactions than are mRNA change ($P = 7.9 \cdot 10^{-14}$, Wilcoxon-Mann-Whitney test) (Figure 2.8D). This suggests that if protein expression change can be measured, then they would lead to much more accurate predictions than what mRNA expression change would yield.

2.3.5 Modeling the control of protein expression

A protein's expression change is regulated by a number of processes, such as RNA transcription, protein synthesis and protein degradation but little information exists about the contribution of each or the combination of any of these processes in the control of protein expression change. If one looks first at the effect of synthesis and degradation rates on overall protein expression, it is obvious that if the synthesis and degradation rates of a protein are equal then the expression of the protein is stable, a.k.a. steady state (Figure 2.9A-B). It also shows, unsurprisingly, that if a protein has a high degradation rate it can still have a stable expression level if its synthesis rate is also high; however, if the degradation and synthesis rates are different, then the protein will change in expression.

Figure 2.9 Parameters controlling protein expression changes during cellular differentiation



A) Scatterplot of the synthesis and degradation rates in THP-1 after 48 h of differentiation from monocytes to macrophages, with the resulting change in overall expression encoded by color. B) Scatterplot of the synthesis and degradation rates in C2C12 after 48 h of differentiation from myoblasts to myotubes, with the resulting change in overall expression encoded by color. C) Protein expression change predictions of all proteins, significant increasing and significant decreasing proteins, respectively derived from partial least squares (PLS) regression of the various parameters controlling protein expression change during THP-1 differentiation, using the mean of at least two out of three biological replicates per parameter. D) Predicted protein expression change during THP-1 differentiation derived from PLS regression of the various parameters controlling expression change using the synthesis and degradation rates deriving from proliferating THP-1 cells, differentiating C2C12 cells and proliferating C2C12 cells, respectively. mRNA change derives from differentiating THP-1 cells similarly to C. For C-D) a value of 1 would indicate that the model perfectly predicts the expression change.

We then applied partial least square regression to our own data from differentiating THP-1 cells as well as transcriptional changes data from Suzuki *et al* [199] to examine the contributions of various factors across the board of all proteins to the ultimate changes in protein expression. This exercise revealed that after 48 h of THP-1 differentiation, the transcriptional change (49±4 %) was the single best predictor, followed by protein synthesis rate (46±7 %) and degradation rate (15±3 %), for the experimentally confirmed protein expression change. Combining all three parameters gives a very respectable predictive power of 69±7 % for any given replicate and this could be increased to 74 % by simply taking the mean of the parameters from the biological replicates. A breakdown of the contributions of the individual parameters in this model using one component is as follows: synthesis rate - 41 %, transcriptional change - 45 % and degradation rate - 14 % (Figure 2.9C and Table 2.2), which indicates that the synthesis rate contributes more than the degradation rate to the variance not explained by the transcription change. Similarly, we observed that during C2C12 differentiation the synthesis rate predicted the protein expression change better than the degradation rate of the proteins (Data not shown).

Table 2.2 The contributions of the different biological processes to protein expression change

	1 component	2 component	3 component	R^2-combined
Predicted protein expression R^2	0.693	0.038	0.005	0.736
P-value	<2e-16	<2e-16	2.15E-09	
Degradation rate	0.136	0.655	0.235	
mRNA change	0.452	0.267	0.328	
Synthesis rate	0.412	0.078	0.438	

Prediction of protein expression for all proteins during 48 h THP-1 differentiation, using mRNA change, synthesis and degradation rates from differentiating THP-1 cells. Notice using one component explains the majority of the variance of the model.

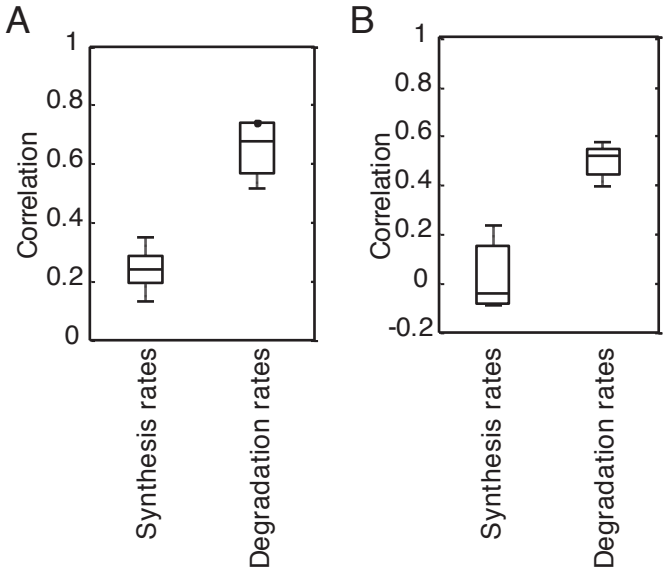
To investigate if proteins with increasing or decreasing expression changes were regulated similarly, we performed partial least square regression on the proteins significantly increasing or decreasing during differentiation. This revealed that the transcription change and synthesis rate makes the biggest contribution for proteins whose expression increases (55 % and 31 %, respectively), while the degradation rate accounted for 13 % of the variance explained by the model (Figure 2.9C). The opposite was observed for the proteins whose expression decreased, since here the transcription change and synthesis rate accounted for only 25 % and 26 %, whereas the degradation rate accounted for 49 % of the variance explained by the model (Figure 2.9C).

Next, we investigated if the addition of degradation and synthesis rates obtained under a different cellular state could improve the prediction of protein expression changes from transcriptional changes. We therefore used the degradation and synthesis rate obtained in proliferating THP-1 cells to predict the protein expression in differentiating THP-1 cells. Interestingly, this revealed that the both the synthesis and the degradation rates increased the prediction of change in protein expression (Figure 2.9D). Since we earlier observed a high conservation between degradation rates in the mouse and human cells, we next investigated if knowing the synthesis and degradation rate of differentiating or proliferating murine C2C12 cells would improve the predictive power versus RNA expression change alone. Indeed, both the synthesis and degradation rates improve the model considerably in differentiating murine C2C12 cells, although the effect was much more modest in proliferating murine C2C12 cells (Figure 2.9).

2.3.6 Protein synthesis rate is intensively regulated in differentiating cells

How the synthesis and degradation rates for individual proteins respond to external perturbation has long been an open question that our data now allow us to address directly. A comparison of the synthesis and degradation rates in differentiating versus proliferating cells reveals a significantly poorer correlation for synthesis rates than for relative degradation rates in both THP-1 cells and C2C12, (Figure 2.10A-B).

Figure 2.10 Correlation of protein synthesis and degradation rates for differentiating and proliferating cells

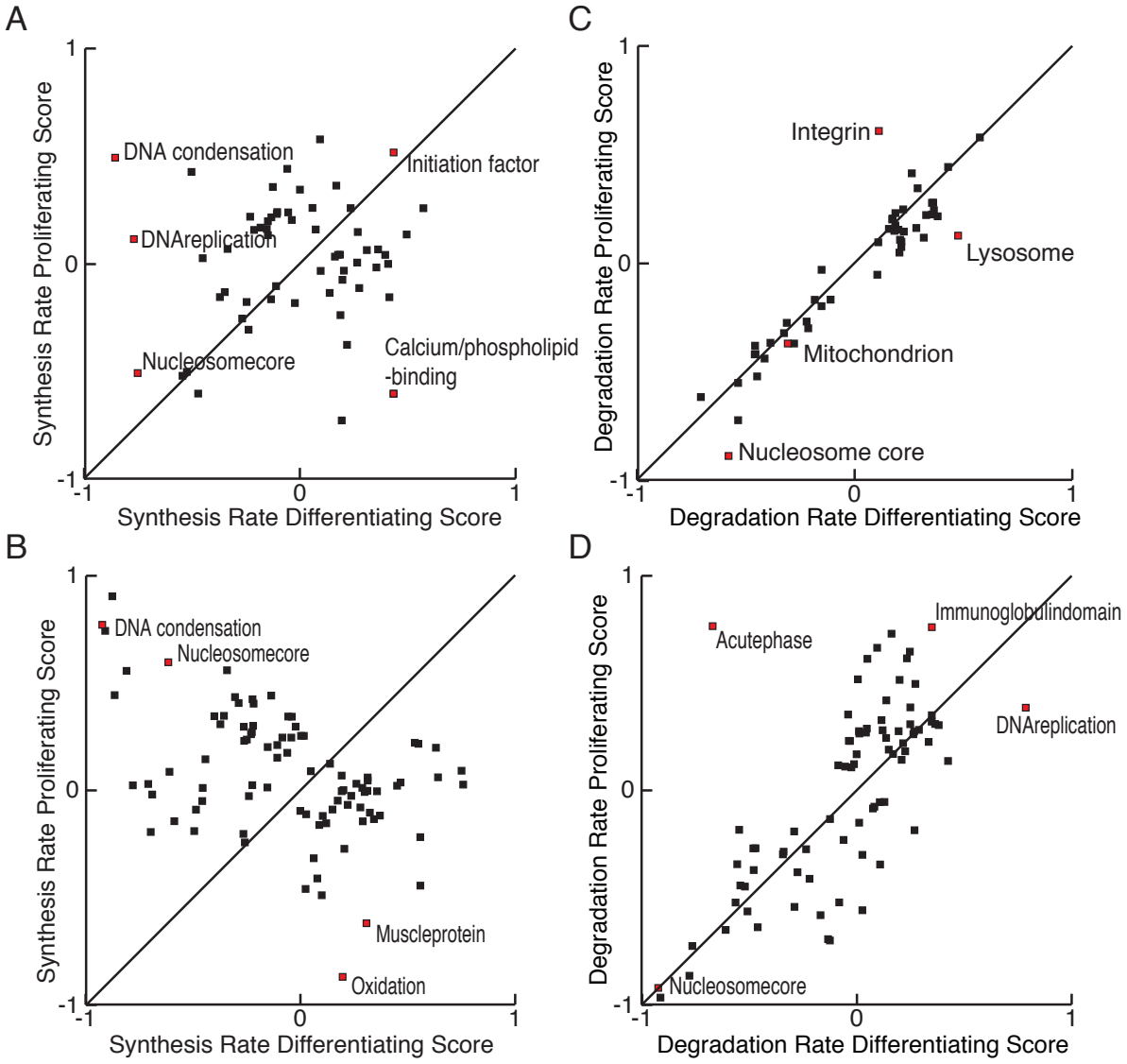


Boxplots of the spread of correlations (Spearman) from three biological replicates between the synthesis and degradation rates in differentiating and proliferating A) THP-1 and B) C2C12 cells.

To investigate this regulation more in detail, we performed 2D enrichment analysis ($P<0.05$) for the different biological processes between the relative synthesis rates of differentiating and proliferating cells (Figure 2.11A-B). This clearly revealed that many biological processes have different synthesis rates in proliferating and differentiating cells as can be seen by their off–diagonal location. For example, we observed a decrease in the

relative synthesis rates of proteins involved in DNA condensation and cell cycle in differentiating versus proliferating cells. On the contrary, we observed a high correlation for the relative degradation rates between proliferating and differentiating cells in both THP-1 and C2C12 cell lines, suggesting that only fairly small changes to the relative degradation rates of the proteins happens in response to differentiation. To investigate the lack of regulation of the relative degradation rate in more detail, we again performed 2D enrichment analysis ($P<0.05$) of the different biological processes between the relative degradation rates in differentiating and proliferating cells (Figure 2.11C-D). This revealed a very different picture than for the synthesis rates, since most enriched biological processes were located on the diagonal, suggesting the relative degradation rates of the different biological processes are not different in the differentiating and proliferating cells.

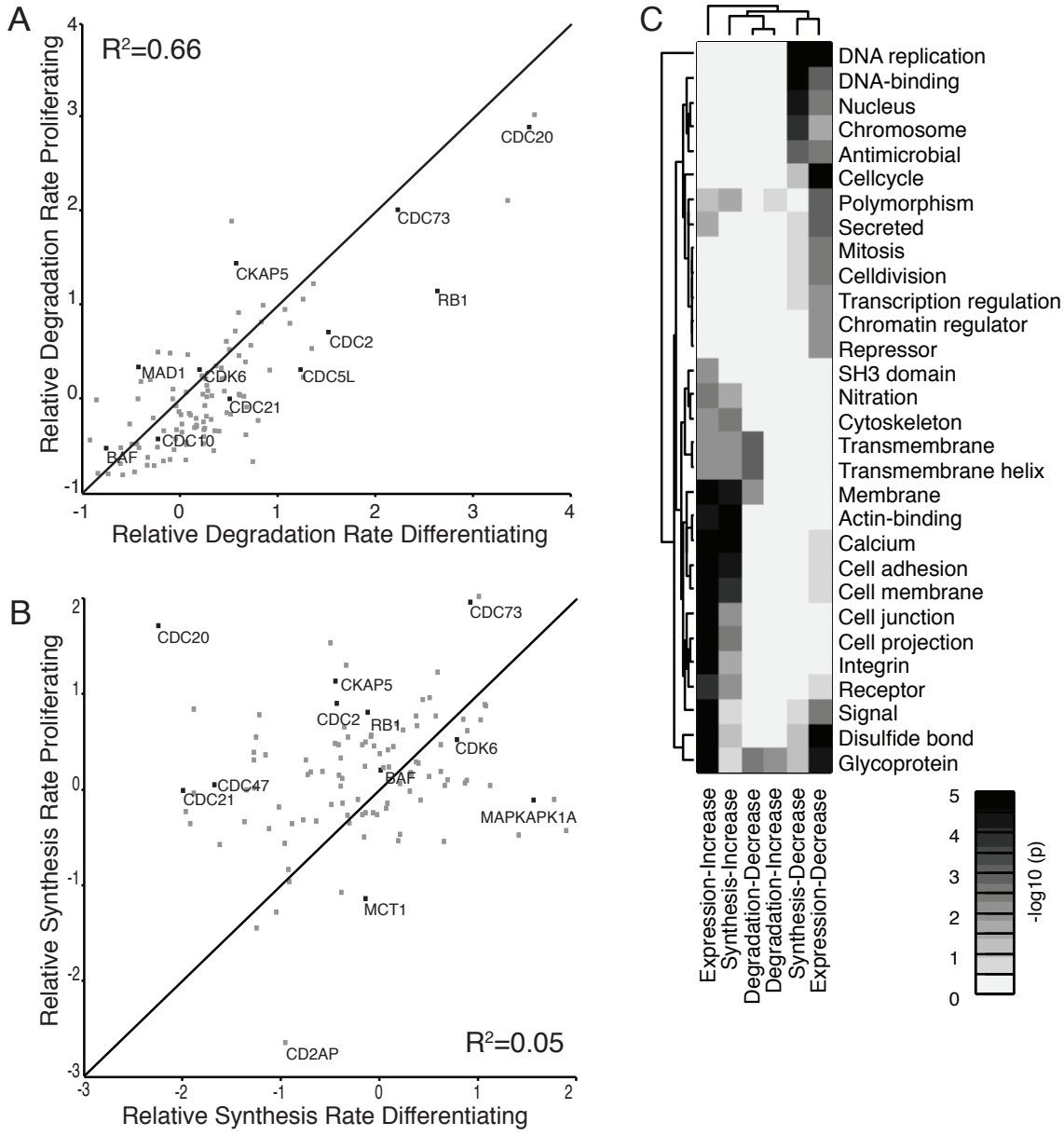
Figure 2.11 Regulation of protein synthesis and degradation rates in response to differentiation in THP-1 and C2C12 cells



2D analysis (P<0.05) of the functional classes of proteins enriched in different regions of proliferating vs. differentiating space of THP-1 and C2C12 cells. A) Synthesis rates in THP-1 cells. B) Synthesis rates in C2C12 cells. C) Degradation rates in THP-1 cells. D) Degradation rates in C2C12 cells.

As an example, the biological process 'cell cycle' was highly affected by differentiation in THP-1 cells and if one examines the proteins assigned to this class in detail, it is obvious that they have very different synthesis rates but very similar degradation rates (Figure 2.12A-B).

Figure 2.12 Regulation of protein synthesis and degradation rates for proteins involved in specific cellular processes



Scatterplots of A) degradation and B) synthesis rates for cell cycle-related proteins in proliferating vs. differentiating THP-1 cells. Degradations rates have a higher correlation coefficient suggesting that there is limited regulation of the degradation rates but dramatic regulation of the synthesis rates of individual proteins in response to differentiation. C) Clusterogram of the $-\log_{10}$(P-value) from enrichment analysis of the proteins whose expression, synthesis and degradation rates changed significantly ($P<0.05$) during THP-1 differentiation from monocytes to macrophages. The cellular processes enriched among those proteins with increased synthesis rates also tended to be enriched in those with increased protein expression change, whereas virtually no classes were enriched among those with altered degradation rates.

Finally, we decided to perform enrichment analysis of the biological processes of the proteins that displayed significant changes in synthesis or degradations rates and compare that to the biological processes that displayed significant change of protein expression during THP-1 differentiation. This revealed that the biological processes that displayed significant change of protein expression did also display significant regulated synthesis rate however not significant regulated degradation rate in respond to the differentiation (Figure 2.12C). Taken this together strongly suggests that the cell changes a protein's synthesis rate up or down either increase or decrease the amount of that protein.

2.4 Discussion

The regulation of protein expression change in response to external stimuli is fundamental to survival of all living organisms yet, to our knowledge, no quantitative assessment of the contributions that various cellular processes make to such a response has been made. Here we set out to model how protein expression change is regulated during differentiation using two un-related model systems so as to support more generalized conclusions. That unstructured, lower-abundance proteins were most dramatically affected by differentiation seems designed to allow very fast regulation of a large part of the signal transduction network. This is also consistent with observations that the differences among cell types are largely a result of lower-abundance proteins [217] and that intrinsically unstructured proteins have more interaction partners and are involved in cellular signaling [218,219].

The relative synthesis and degradation rates measured here compare very well with the turnover characteristics of the same proteins, previously discovered through studies

identifying the absolute rates [51,97,203,204] but the higher-throughput afforded by relative measurements allows us to assess how synthesis and degradation are affected by external perturbation and how sub-populations of proteins involved in specific protein complexes can be processed differently. In the example mentioned above, the alpha subunits of the proteasome core particle are being formed wholly from newly synthesized proteins rather than being re-assembled from recycled subunits, which implies that the proteasome core particle is also degraded as an intact complex. Furthermore, we used this data to show that the established modularity of macromolecular complexes [6] is also present during synthesis and degradation of the complexes, since the different modules of a complex will display similar synthesis and degradation rates, thereby adding an additional layer to the established assumption that proteins display different synthesis and degradation rates in different cellular compartments [97,205]. This then leads us to speculate that synthesis and degradation rates of a protein could be used as discriminatory parameters to improve the assignment of protein interactions, similar to mRNA co-expression [216].

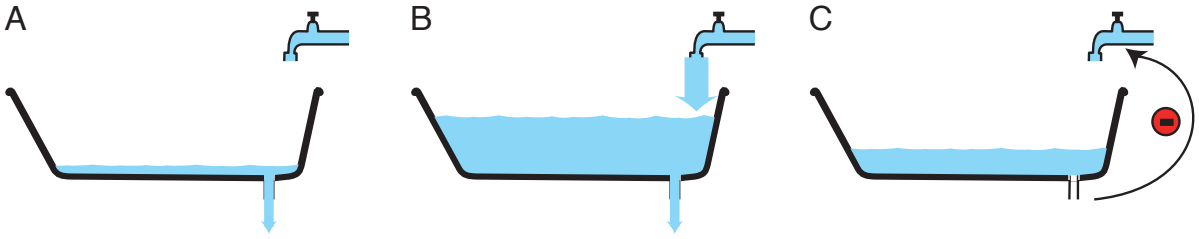
Towards our primary goal of modeling the contributions various cellular processes make to an eventual proteome, predictions of protein expression changes can be significantly improved by taking into account both the synthesis and the degradation rates of the proteins, contributing 41 % and 13 % respectively to the variance not explained by changes in transcription. That post-transcriptional regulation of synthesis and degradation rate are having such an impact after 48 h of differentiation, where the highest correlation between transcriptome and proteome expression changes was observed, suggests that the factors are even more important earlier in differentiation when RNA and protein levels are very poorly correlated. The importance of post-transcriptional regulation of protein expression changes

during THP-1 differentiation was illustrated in a recent paper that describes how overexpressing four microRNAs lead to partial differentiation [220]. Our results that synthesis rate contribute more than degradation rate to the prediction from transcript to protein expression changes is in agreement to what has been observed for these processes' contributions in predicting protein abundance, where it was observed that protein synthesis rate was a better predictor than protein degradation rate [51].

We observed clear differences in the extent to which synthesis and degradation rates contributed to the changes of protein expression during differentiation; perhaps unsurprisingly the synthesis rate was the best predictor for those proteins that increased in expression whereas the degradation rate contributed the most in predicting decreasing proteins. This does suggest though that failure to consider degradation rates seems like the most likely explanation for why a poor correlation is generally observed between transcriptome and proteome expression changes for proteins whose expression decreases [110]. That we observe a strong correlation between degradation rates in proliferating and differentiating cells suggests that one does not even have to measure the degradation rate in the same cellular state of the system of interest in order to take such a parameter into account when using transcriptome changes to predict protein expression changes; degradation rates are now being recorded in the newly developed Encyclopedia of Proteome Dynamics [101]. Perhaps the most striking finding here is that the changes in any given protein's expression during differentiation are largely due to changes in the synthesis rate, while the degradation rates remain constant. Using the bathtub analogy from the introduction, this means that the water level in the bathtub will be controlled mainly from the faucet either by increasing or decreasing the flow, whereas the draining will be at a nearly constant flow (Figure 2.13A-B).

This has at least two conceptual consequences: 1) The drain size for an individual protein is likely to vary little, i.e. the drain size is less adjustable than the faucet and; 2) the faucet plays a key role in preventing overflows. On the protein level, consequence 1 translates into a strong "default" component in the degradation of many proteins, which we propose is based on the protein's sequence and can either be dependent or independent of the ubiquitylation [221-223]. Consistent with this are previous observations that many proteins involved in signal transduction and transcriptional regulation have fast basal turnover rates and, consequently, they can be cleared quickly in response to perturbation; this would be energetically favorable to the cell since these proteins are typically expressed at low levels and therefore less energy would need to be consumed [194,198]. Interestingly, the regulation of changes in mRNA levels are also predominantly regulated at the level of transcription while mRNA degradation rate are generally constant [224]. Even though we state here that a change of synthesis rate is the main driver of expression change during differentiation, this does not exclude an important role for the ubiquitin-proteasomal system. The degradation rate in differentiating and proliferating cells does not correlate perfectly, meaning that the degradation rates of some proteins are regulated. Furthermore, differentiation is a relative slow process so perhaps changes in degradation rates could play a larger role in regulating protein expression changes in response to acute stimuli such as heat shock or inflammation [225,226].

Figure 2.13 Model of protein expression change



A) Proteins decreasing in expression are a result of decreased protein synthesis rate, whereas the protein degradation rate is constant. B) Proteins increasing in expression are a result of increased protein synthesis rate whereas protein degradation rate is constant. C) Inhibition of the protein degradation rate, will lead to feed back inhibition of the synthesis rate.

Consequence 2 means that a tight regulation of the synthesis of proteins is key to prevent dangerously high and potentially toxic levels of some proteins (Figure 2.13C). In agreement with this are previous observations showing that accumulation of malfolded proteins causes dephosphorylation of eIF2α, which leads to repression of translation [227,228]. Furthermore, it has been observed that mTOR gets sequestered in polyglutamine aggregates during Huntington's disease, resulting in decreased mTOR dependent TOP translation and introduction of autophagy, thereby serving a protective function [229].

These findings then raise the captivating question: why has evolution favored regulation of synthesis rather than of degradation? One possibility is that altering degradation rates might too easily induce protein aggregation, similar to what is seen when autophagy or the proteasome are inhibited [230-232]. One way the cell avoids aggregate formation is by keeping the lifetimes of aggregate-prone proteins short [195,196], which would be more difficult if the lifetime of a protein could be extended in response to perturbation.

Chapter 3: A high-throughput approach for measuring temporal changes in the interactome

3.1 Introduction

Growth factor signaling cascades have been studied intensely for the last thirty years due to their involvement in various cancers and metabolic disorders. Activation of growth factor receptors at the cell surface initiates signal propagation through various effector molecules that can lead to cell proliferation, differentiation and migration [31]. This process is strictly regulated via post-translational modifications, e.g., ubiqutination and phosphorylation, and involves the dynamic regulation of numerous protein-protein interactions [233].

The study of the interactome holds great promise for answering fundamental questions about biological pathways, not only with respect to their basic organization but also, more importantly, how they are rearranged following perturbation. Existing methods for studying interactomes require tagging or the creation of fusion proteins of all open reading frames of interest to provide a measurable readout or enable purification of the protein complex and subsequent identification of the interacting proteins by mass spectrometry [6,7,150,155]. This process is especially challenging in mammalian systems where it is difficult to systematically tag each protein and, at least for AP-MS approaches, the detection of all interacting proteins involves a huge amount of work. In addition, existing large-scale methods are not easily amenable to addressing how an interactome responds to stimulation, thereby missing the amazing cellular adaptations we know exist from a multitude of more focused studies.

PCP-SILAC has been used to profile organelle proteins across sucrose gradients and thereby assign localizations to those proteins; it does not rely on an organelle being purified to homogeneity, but instead relies on proteins localized to the same organelle displaying similar profiles across a density gradient [16,18,19]. The same concept could apply to protein complexes but protein complexes are poorly resolved on density gradients; size exclusion chromatography, however, is a universally accepted method for resolving protein complexes and assigning composition based on co-eluting enzymatic activity and/or immunoblot profiles. It is an optimal method for resolving complexes in their native state since it involves no binding and release from the stationary phase, the separation is performed using gentle buffer conditions and, importantly, there is no loss of sample. Traditionally, targeted detection assays have been used downstream of SEC [163], and its use in global monitoring of protein complexes have been limited [167].

3.2 Experimental procedures

3.2.1 Cell culture

Three populations of HeLa cells were SILAC labeled using Arg and Lys-free Dulbecco's Modified Eagle's medium (DMEM) with 1% glutamine, 1% penicillin/streptomycin and 10% dialyzed fetal bovine serum and either (L-[U-$^{13}C_6$,$^{14}N_4$]arginine and L-[$^{2}H_4$]lysine or L-[U-$^{12}C_6$,$^{14}N_4$]arginine and [$^{1}H_4$]lysine or L-[U-$^{13}C_6$,$^{15}N_4$]arginine and L-[U-$^{13}C_6$,$^{15}N_2$]lysine (Cambridge Isotope Labs, Cambridge, MA). The cells were grown for at least five doublings to ensure 100% incorporation of labeled amino acids and subsequently washed three times with phosphate-buffered saline (PBS) before being scraped into PBS.

3.2.2 SEC-PCP-SILAC

The cells were lysed in a Dounce homogenizer in size exclusion chromatography (SEC) mobile phase (50 mM KCl, 50 mM $NaCH_3COO$, pH 7.2) including protease inhibitors without EDTA (Roche) and phosphatase inhibitors (1 mM Na orthovanadate; 5mM sodium pyrovanadate; 0.5 mM pervanadate). Two millilitres of each lysate were clarified of very large material by a 15 min ultracentrifugation (100,000 rcf) to enrich soluble, cytosolic complexes, before being concentrated to 50 μL using ultrafiltration (100000 MWCO, Sartorius Stedim). The ultrafiltration served three purposes: 1) to reduce the volume, 2) to enrich for high molecular weight complexes and 3) to generate sharper peaks during SEC by minimizing the band loaded on-column. The medium/heavy fractions were recombined just prior to loading (100 μL) onto a 1200 Series semi-preparative HPLC (Agilent Technologies, Santa Clara, CA) equipped with a 600 x 7.8 mm BioSep4000 Column (Phenomenex) (resolving power 62257, 79522 and 109287 plates/meter, for the three iterations of the columns used in experiments 1, 2 and 3, respectively) controlled at 12°C and a flow rate of 0.5 mL/min. Fractions were collected at a rate of 2/min from 20 to 30 min and at 3/min from 30 to 40 min. The relatively low salt concentration was used since it has been reported that some protein complexes will dissociate even in physiological salt buffers [234]. The light SILAC population was similarly separated by SEC, after which all fractions were recombined and mixed thoroughly before being aliquoted equally into each of the medium/heavy fractions.

3.2.3 Protein digestion and mass spectrometry

To each of the combined fractions, sodium deoxycholate was added to a final concentration of 1% and then each sample was boiled for 5 min. Subsequently, the fractions were in-solution digested as described [207] and afterwards acidified by 1% TFA in 1% acetonitrile and the precipitated cholic acid was pelleted at 16,000 relative centrifugal force (r.c.f.) for 10 min. The individual fractions were cleaned up as described previously [235] and analyzed by LC-MS/MS. Peptides were separated by a 180 min gradient (5-35% acetonitrile in 0.5 % acetic acid) on an 1100 Series HPLC system (Agilent), using in-house packed C18 capillary column (75 μm ID, packed with 3 μm Reprosil-Pur (Dr Maisch)). Eluate was electrosprayed into an LTQ-Orbitrap XL that was operated with the following settings: One full scan (Resolution 60,000; m/z 300-1600) followed by five MS/MS scans using CID in the linear ion trap (Min. Signal Required 500; Isolation width 3, Normalized collision energy 35; Activation Q 0.25; Activation time 30 ms) using dynamic exclusion (Repeat count 1, Repeat duration 30 sec, Exclusion list size 200, Exclusion duration 80 sec).

3.2.4 Mass spectrometry data processing

Tandem mass spectra were extracted from the data files using the most recently available version of MaxQuant (v1.0.13.13- v2.2.2.5) [48] and searched against the human IPI database (v3.69, 74,854 sequences for experiment 1 and 2 or Uniprot 21/6/2011 69924 sequences for experiment 3) with common serum contaminants and enzyme sequences added. The results were then quantified and identified using MaxQuant with the following settings: 1% FPR on protein and peptide levels, trypsin/P cleavage rules with a maximum of 2 missed cleavages, 0.5 Da tolerance for MS/MS.

3.2.5 Data analysis

The three biological replicates were processed independently using the Curve Fitting Toolbox in Matlab (www.mathworks.com) to deconvolve chromatograms into component Gaussian curves. Prior to curve fitting though, chromatograms were filtered using two rules: 1) only datapoints in a group of at least five consecutive datapoints were retained, and 2) the remaining datapoints needed at least three consecutive points with a medium/light (M/L) ratio greater than 0.5 (signal:noise filter). Briefly, an interative .m script was written that fits from one to five Gaussian curves to each chromatogram, depending on the number of fractions with a M/L ratio above 0.5 (<6 fractions 1 gaussian; <9 fractions 2 gaussians; <12 fractions 3 gaussians; <15 fractions 4 gaussians; 15< fractions 5 gaussian) using the non-linear least squares method with the following lower and upper bounds for height, center and width: [1, 0.1, 0.3] and [(max ratio medium/light) 50, 8]. Then, for each successful fit a leave-one-out cross-validation was performed where one point from the chromatogram was dropped prior to re-fitting the data. The squares of the error (SSE) between the dropped datapoint and the re-fit curve was then summed across 500 such iterations and the number of Gaussians with the smallest SSE for a given chromatogram was considered the best fit.

3.2.6 Receiver-operator characteristics and recall-precision curves

We next calculated receiver-operator characteristics and recall-precision curves for the three biological replicates independently, since they were analyzed on three different SEC columns, using an in-house Matlab script for distances between center, height and width of the Gaussian curves and in addition for the Euclidian distances between chromatograms for two proteins. We used the CORUM database [212] as a validated set of true interactions,

generating all possible binary interactions within each contained complex to make it compatible with our data; this list contained 5571 interactions and represents all possible true positive (TP) interactions we could potentially find in our data. For true negative interactions, we first took all the proteins we identified here and that were also contained in CORUM and generated all possible interactions among them. From this we then subtracted all the true interactions contained in CORUM, leaving 139,689 interactions in the true negative (TN) set. False positives (FP) were defined as all interactions minus TP, false negatives (FN) were defined as the interactions in CORUM database not being found and finally the recall, precision, true positive rate (TPR) and false positive rate (FPR) was calculated exactly as described [236].

3.2.7 Assigning binary interactions and protein complexes

To assign binary protein-protein interactions we used two types of information: 1) First, we calculated the pairwise Euclidian distance (which is defined by the sum of ΔM/L-ratio at each fraction) to all other chromatograms, with the assumption that two proteins which always are together in the same complex would have similar chromatograms and thus would have small distances. Second, we used the Gaussian curves deriving from the deconvolved chromatograms, with the assumption that interactions among proteins that are not always in a complex together and proteins with incomplete chromatograms, should have similar Gaussian curves in part of the chromatogram. If two chromatograms were very similar (Euclidian distance resulting in a precision>0.8), the two proteins were assigned as having a binary interaction. However, if the distances between the curves were large (0.2<precision<0.8), additional criteria were used: very strict limits for center and width were

applied since they are not affected by differences in stoichiometry and wider limits for height in order to catch stoichiometry differences. Afterwards the binary interactions from the three independent biological replicates were combined and TPR, FPR, recall and precision were calculated, being 0.15, 0.0067, 0.15 and 0.53, respectively, for the combined dataset.

These interactions were subsequently converted to base-2 numbers where 1 indicates an interaction and 0 indicates no interaction; these data were then clustered using the dist package and hclust package in R, with a distance of 0.825 generating 291 complexes containing between 2 and 43 proteins.

All the scripts used for this analysis are available from our FTP site (ftp://foster.chibi.ubc.ca/Download/PCP-SILAC/) and lab website (http://www.chibi.ubc.ca/faculty/foster/software), together with some sample data and instructions on their use.

3.2.8 Validation of complex components by antibody-based SEC elution shift

Three SILAC populations of cells were grown as described above for dynamic PCP-SILAC. The cells were lysed and concentrated by ultrafiltration (100000 MWCO) before 10 μg of 14-3-3γ polyclonal antibody (C-16) (Santa Cruz Biotechnology) or HRS polyclonal antibody (C2C3) (GeneTex Inc.) was added to the heavy population and incubated for 30 min on ice. The medium and heavy SILAC populations lysates were fractionated independently by SEC prior to the fractions being combined and having the aliquots of the pooled light fractions spiked in. Since the medium and heavy populations are combined after SEC, the antibody has no opportunity to alter the elution times in the medium population. The fractions were analyzed as described above and proteins having medium/light (M/L)

ratios smaller than 2 and heavy/medium (H/M) ratios larger than 1.5 in fraction 34 and M/L ratios larger than 2 and H/M smaller than 1 in fraction 36 was assigned as interacting with 14-3-3γ.

3.2.9 Validating interactions using AP-MS of SEC fractions

Two SILAC populations of cells were grown, before the cells were lysed and concentrated by ultrafiltration (100000 MWCO) before being separated individually by SEC as described above. Fractions 19 through 24 from each population were combined and subjected to AP using 100 μl anti-rabbit Dynabeads and 20 μg of HRS polyclonal antibody (C2C3) (GeneTex Inc.) or 20 μg rabbit IgG individually. The pull down was washed three times in PBS, before being combined, eluted by LDL sample buffer (Invitrogen) and separated by SDS-PAGE. Finally the sample was "in-gel digested" as described [237] prior to the peptides being analyzed by MS and quantified and identified as described above by MaxQuant.

3.2.10 Determination of protein stoichiometries

The areas of the individual Gaussians derived from a protein's chromatogram are a direct representation of how much a protein participated in each of the individual subcomplexes. To calculate if any protein with a complex had significantly different stoichiometries, we first assigned the individual peaks to subcomplexes and calculated the relative stoichiometry of each protein, which is done by calculating the areas of the individual peaks. Next we performed principle component analysis of the different stoichiometries for the proteins and calculated the T2 value, from which using the F

cumulative distribution (n <50) the P-values could be identified. All calculations were performed in Matlab using the statistical toolbox.

3.2.11 Analysis of spatiotemporal changes following EGF stimulation

Proteins were assigned as changing their protein-protein interactions if the medium/heavy ratio changed 1.5 fold in three consecutive fractions and in addition had a medium/light ratio larger than 0.75. Two biological replicates were generated for the EGF stimulated cells and one for unstimulated cells as a control.

The dataset of proteins changing interaction after EGF treatment was compared to the following high throughput datasets:

1. Olsen *et al.* global phosphorylation dataset [44], where we assigned proteins as differentially regulated if a single phosphopeptide changed kinetics twofold at any of the timepoints.
2. Blagoev *et al.* phosphotyrosine proteome, where all proteins from supplementary table 1 were used as differentially regulated following EGF stimulation [186].
3. Argenzio *et al.* EGF ubiproteome, where proteins from the 'endogenous approach' were used since this experiment was carried out in Hela cells. We used the proteins assigned as differentially regulated and the steady state from the endogenous approach as not changing following EGF stimulation [238].

To investigate possible positive correlations with proteins known to bind to the EGFR we extracted the interactions from the IntAct database [239] with the following query: EGFR AND species:human, which resulted in 238 unique proteins from 46 publications.

3.2.12 Statistical tests

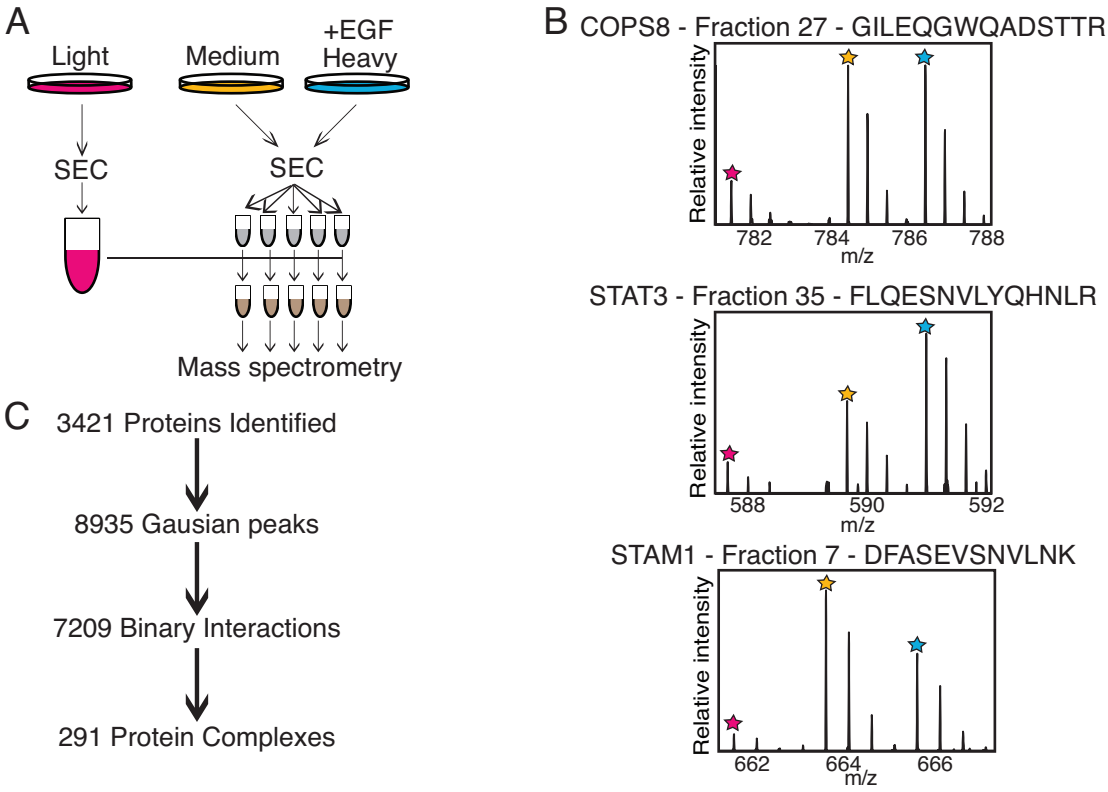
For comparison with the other large-scale experiments (IntAct, global phosphorylation, P-tyrosine phosphorylation and ubiquitination), the two-tailed Fisher's exact test was used. For investigating significantly different protein stoicheometries in a complex, we performed principal components analysis, calculated the Hotelling's T2-value, from which using the F cumulative distribution ($n < 50$) the P-values could be identified, which were all done in Matlab.

3.3 Results

3.3.1 Monitoring the interactome using PCP-SILAC

To overcome some of the limitations of present interactome-scale techniques, here we combine PCP-SILAC with high performance liquid chromatography using a SEC column with a theoretical plate count exceeding 100,000 plates/meter to determine the composition of the human interactome, as well as the global changes that occur in the interactome following EGF stimulation. Briefly, proteins in three populations of cells are mass-encoded using arginine and lysine isotopologs, allowing them to be resolved by mass spectrometry (Figure 3.1A) [63,186]. The heavy population is stimulated with EGF while the medium and light population are left untreated and then all three populations are mechanically lysed. An enrichment of the cytosolic protein complexes from each lysate was performed to minimize irrelevant interactions among proteins in different cellular compartments and then each lysate was separated on a SEC column with optimal resolution between 150 KDa and 2 MDa. The fractions from the light sample are then pooled together and aliquots are mixed into each of the medium/heavy fractions prior to tryptic digestion and mass spectrometric analysis. In this scheme, the light-labelled proteins act as internal standards and any interactome changes following EGF stimulation are monitored with the medium/heavy ratio (Figure 3.1B).

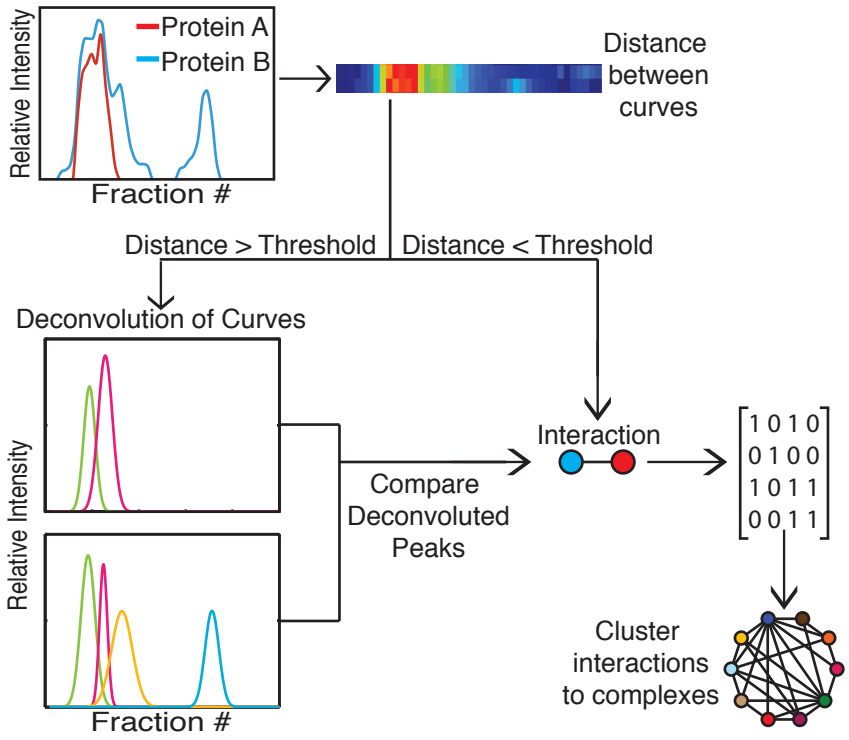
Figure 3.1 Identification of spatiotemporal changes in the interactome following EGF stimulation



A) Three populations of Hela cells are metabolically labeled with amino acid isotopologs, and the heavy population is stimulated with 150 ng/ml EGF for 20 min. The cells are lysed and the high MW fraction enriched by ultrafiltration prior to size exclusion chromatography (SEC). SEC fractions from the light cells are pooled and subsequently aliquoted into each fraction from medium/heavy fractions as an internal standard prior to LC-MS/MS. B) Mass spectra of three peptides that display different spatiotemporal interactions changes following EGF stimulation. The medium:light ratio (M/L) is used to generate chromatograms, whereas the heavy to medium (H/M) ratio represents the impact of EGF stimulation on that protein. The top panel is a mass spectra from SEC fraction 27 of the peptide `GILEQGWQADSTTR` from COPS8 in fraction 27. The monoisotopic peaks from the light, medium and heavy envelopes are marked by red, yellow and blue stars respectively. The M/L ratio would represent the intensity value of the chromatogram at fraction 27 and, in this example, the M and H peaks are essentially equal, indicating that EGF stimulation has no effect on it's interactions at this point in the chromatogram. In the middle panel, the H/M ratio for the peptide `FLQESNVLYQHNLR` from STAT3 in fraction 35 indicates that STAT3 gets recruited to a complex in fraction 35 following EGF stimulation. In the bottom panel, the peptide `DFASEVSNVLNK` from STAM1 in fraction 7 shows a H/M ratio much less than 1.0, indicating that it dissociates from a complex in fraction 7 following EGF stimulation. C) Summary of results from each analytical step.

Thirty-four hundred proteins were identified from three independent biological replicates using this approach and a chromatogram was reconstructed for each protein based on the light/medium ratios in the individual fractions using MaxQuant [48] (Figure 3.1C). The high resolution of SEC enabled us to easily determine that many of the chromatograms had multiple peaks, indicating that proteins very frequently participate in more than one complex or in similar complexes with different stoichiometries. We then used two approaches to systematically address the interactions represented in these chromatograms: First, for every chromatogram we calculated the Euclidian distance to all other chromatograms, with the assumption that two proteins that always occur together in the same complex(es) would have similar chromatograms and, thus, small distances. Second, we deconvolved each chromatogram into component Gaussian curves, with the assumption that for large complexes, which are made of independent, stable and observable subcomplexes, the constituent proteins might only show similarities in part of the chromatogram (overlapping Gaussian curves) because they are not all always in a complex together (Figure 3.2).

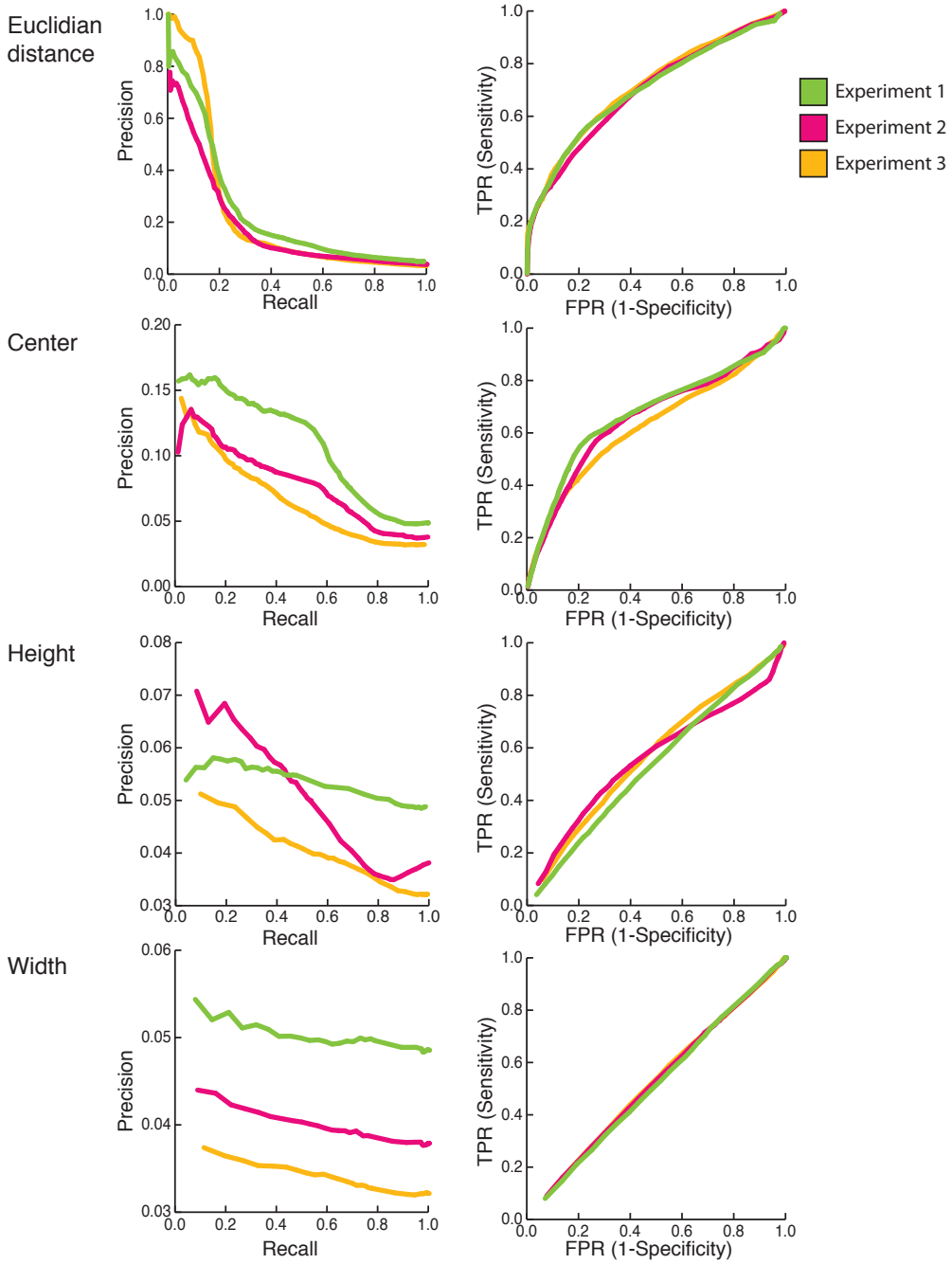
Figure 3.2 Data analysis approach



If the Euclidian distance between any two chromatogram resulted in a precision larger than 0.8 the two proteins represented by those chromatograms were considered to be in a binary interaction. For distances below this threshold, the chromatograms were deconvoluted into Gaussian components resulting in three coefficients per curve; center (C), height (H) and width (W). Then, if the Gaussian coefficients for any curves of two different proteins fell within a given window of one another, those two proteins were considered to be in a binary interaction. These interactions, represented in a binary matrix, were then clustered to determine which proteins were in a complex together.

This two-pronged approach resulted in four parameters - distances between chromatograms, as well as center, height and width for each deconvolved peak in a chromatogram - for which receiver-operator characteristics (ROC) and precision-recall curves could be generated (Figure 3.3).

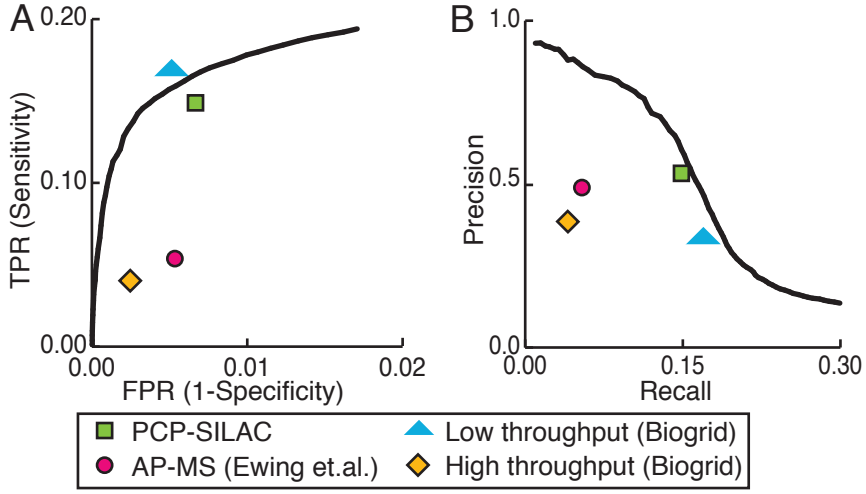
Figure 3.3 Precision-Recall and ROC-curves of the different parameters



Precision-Recall and ROC-curves were generated for the three independent biological replicates using the Corum database as true positives, and all interactions among identified proteins that were not true positive as true negatives. Euclidian distance and center are clearly the most discriminatory parameters (e.g., their ROC curves were most different from a diagonal line).

By selecting a combination of parameters that yielded a false positive rate of less than 0.7% and a precision of 53% (Figure 3.4A-B), 7209 binary protein interactions could be identified, which when hierarchically clustered resulted in the identification of 291 protein complexes with an average of 4.1 distinct proteins per complex. This is similar to the number of interactions identified in other high throughput techniques such as AP-MS but generates two orders of magnitude fewer samples for LC-MS analysis and avoids the tedious tagging and purification steps [6,7].

Figure 3.4 Evaluation of the PCP-SILAC approach to identify protein-protein interactions



A) The receiver operator characteristic (ROC) curve for only one of the four parameters, Euclidian distance, describing the tradeoff in identifying interactions misclassified as positive (FPR) versus the proteins correctly classified as positive (TPR). B) The precision-recall curve for the Euclidian distance describes the tradeoff in identifying interactions correctly called as positive (recall) vs. the fraction of called positives that are truly positives (precision). The points in A and B represent the following datasets: PCP-SILAC is the value for all 7209 binary interactions identified in this study after applying all limits (not just Euclidian distance), AP-MS (Ewing) is [177], Biogrid low and high throughput is acquired from the Biogrid database version 3.1.82 [240].

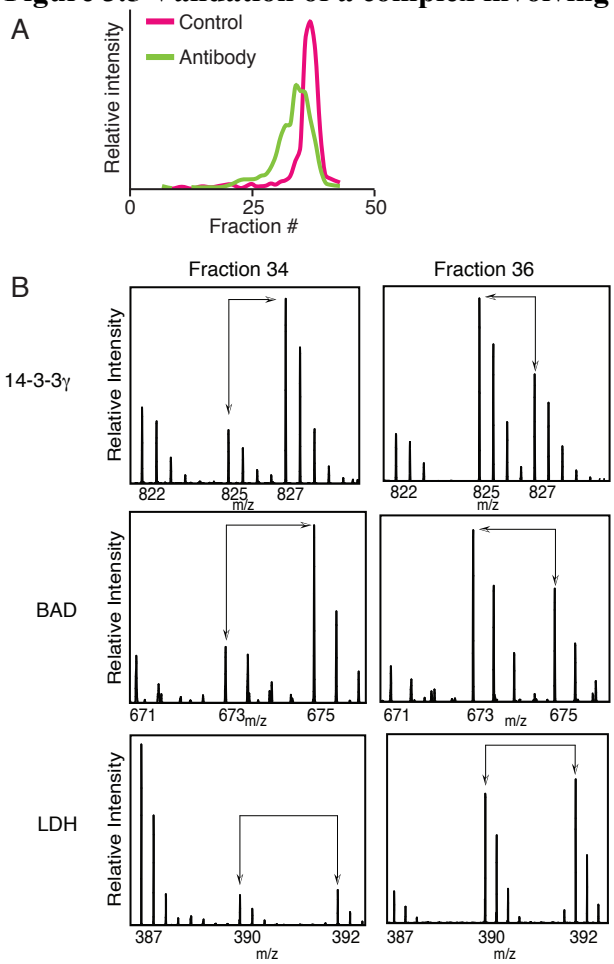
In addition, we were also able to detect relatively transient interactions that have been characterized previously, such as the binding of UCHL5 and RPN13 to the proteasome [241]. This is likely because during SEC the proteins in a complex move as a narrow, concentrated band, thereby allowing transient interactions to persist during the analysis time, similar to the process in analytical ultracentrifugation [242], whereas during affinity purification the equilibrium of an interaction is heavily shifted towards the free form of the proteins during the washing steps.

3.3.2 Antibodies can shift protein complex retention time

To validate one of the identified complexes we made further use of the high resolution of the SEC column to test the impact on retention time of adding an antibody against a specific protein found to be part of a complex; this approach should increase the Stokes radius of any complex containing the target of the antibody, resulting in earlier elution from the column. By adding an antibody against 14-3-3γ only to the heavy lysate and then resolving the heavy and medium cytosols before combining the equivalent fractions from the two separations, the heavy/medium ratio allows us to track the impact of the antibody on elution time. Indeed, the antibody increased the Stokes radius of the complex 14-3-3γ participated in to such an extent that it eluted at least two fractions earlier (Figure 3.5A) and, concomitant with this shift, the elution profiles of 14-3-3α/β, 14–3–3ε and BAD (Figure 3.5B) were also shifted by a similar amount. All three proteins have been previously reported to be binary interaction partners of 14-3-3γ [243,244] but as they migrate together and to a point in the chromatogram that is considerably larger than the sum of any two of their molecular weights, our data suggest they actually form a single larger complex or oligomers

of the individual pairs. Importantly, chromatograms of several other proteins that also peaked at or near fraction 36 in the initial experiment, including LDH (Figure 3.5B) were not affected by the addition of the antibody.

Figure 3.5 Validation of a complex involving 14-3-3γ

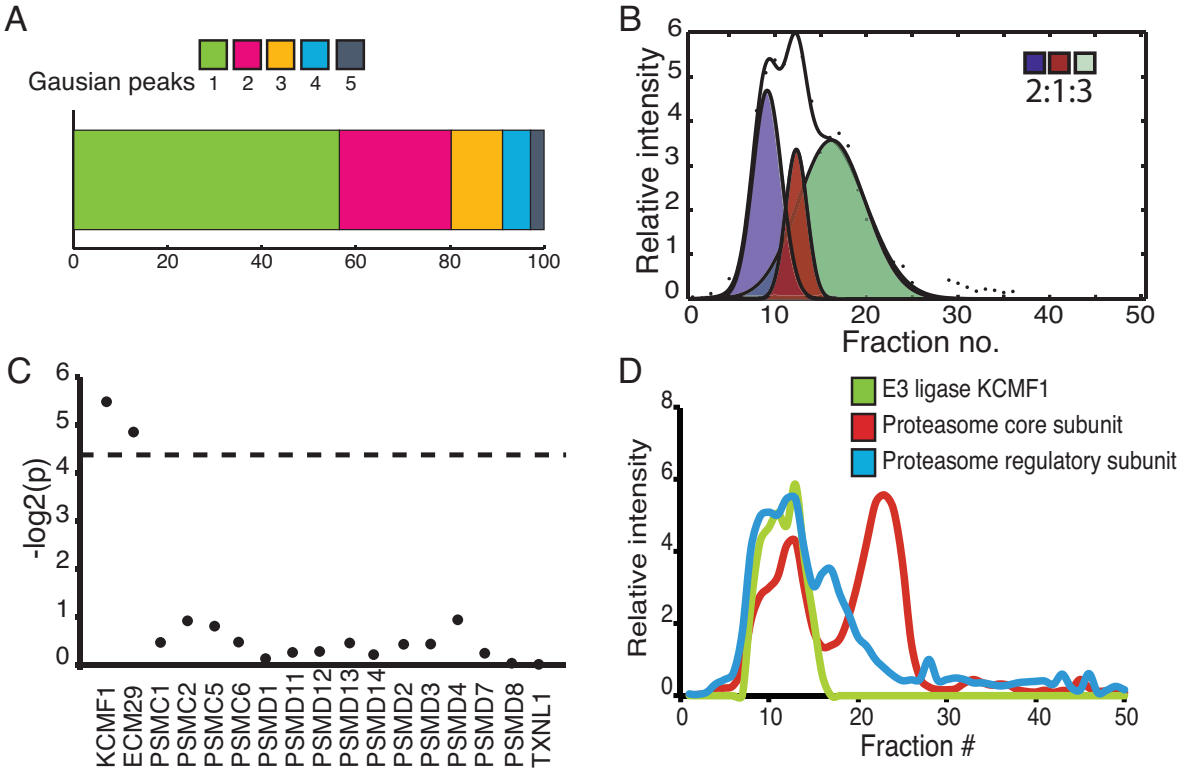


A) To experimentally validate one of the complexes discovered by PCP-SILAC, antibodies against 14-3-3γ (YWHAG) were added to heavy lysate prior to SEC. The chromatograms of 14-3-3γ in the absence (red line) and presence (green line) of exogenous IgG are shown. B) The shift in elution time caused by the addition of an antibody against 14-3-3γ can be seen in SILAC ratios at specific points in the chromatograms. The top panels are spectra of a peptide from 14-3-3γ from fractions 34 and 36 with the change in the medium and heavy isotopologues indicated by arrows. The middle panels are spectra of a peptide from BAD with similar medium:heavy ratios as observed for 14-3-3γ, while the bottom panels are spectra of a peptide from lactate dehydrogenase (LDH) showing essentially no shift in response to the addition of antibodies against 14-3-3γ.

3.3.3 Determination of stoichiometries between complexes

Combining PCP-SILAC and SEC allows the determination of an interactome but it also highlights the heterogeneity of complexes within the cell. Because the various complexes a protein might be involved in are not resolved, how a protein is concurrently distributed among complexes goes undetected in conventional interactome approaches yet this distribution can be important for biological outcomes; e.g., a scaffold protein was shown to either inhibit or activate a cascade depending on its concentration [245]. Forty-three percent of the protein chromatograms measured here deconvolved into more than one Gaussian peak (Figure 3.6A), suggesting that these proteins bind to multiple different proteins. Since the chromatograms are quantitative, the relative stoichiometry of a protein binding to its various partners should therefore be calculable from the areas of the individual Gaussian curves, similar to the information obtained from multisignal sedimentation velocity analytical ultracentrifugation [242]. An interesting example that arises from this analysis is the proteasome: Our data reveals that on average there are 1.7± 0.6 moles of 19S regulatory particle proteins in the doubly-capped proteasome for every mole in a singly-capped proteasome (Figure 3.6B). Since each doubly-capped proteasome molecule would have twice as many regulatory proteins as a singly-capped proteasome, this means that the singly- and doubly-capped complexes are approximately equally abundant, an observation that supports an estimated stoichiometry of one between the double- and single-capped proteasome in the cytosol reported by others [246].

Figure 3.6 Determination of stoichiometry using SEC-PCP-SILAC



A) Deconvolution of all the chromatograms into Gaussian curves reveals that 43% of proteins elute in multiple peaks, suggesting that these proteins bind to multiple different proteins or at least to multiple distinct assemblies of proteins. B) The areas under the fitted curves can be used to determine the relative stoichiometry of a protein in different complexes. The three curves fitted (R^2 = 0.99, SSE = 0.15) to the chromatogram of PSMD8 (26S proteasome non-ATPase regulatory subunit 8) had approximate relative areas of 2:1:3, representing the fraction of PSMD8's participation in the full, 26S, doubly-capped proteasome, in the singly-capped proteasome and in the isolated 19S regulatory subunit respectively. C) Proteins with significantly different stoichiometries in a complex could be determined by employing principle components analysis and Hotelling's T2-statistic. In this way, for a complex made up of multiple subcomplexes, if any component protein's distribution was different from the members of a subcomplex it would show as an outlier: KCMF1 and ECM29 had significantly different distributions among proteins involving the regulatory subunit of the proteasome than did the canonical components of the regulatory particle (dotted line represents P=0.05). D) Plotting the average chromatogram for the proteasome core subunit proteins, regulatory subunit proteins and the E3 ligase KCMF1 reveals that all three only overlap at fractions 9 and 12, representing the double (i.e., 26S) and single capped proteasome, but not at fractions 16 and 22 where the regulatory subunit and core subunits are by themselves, suggesting that KCMF1 only associates with capped forms of the proteasome.

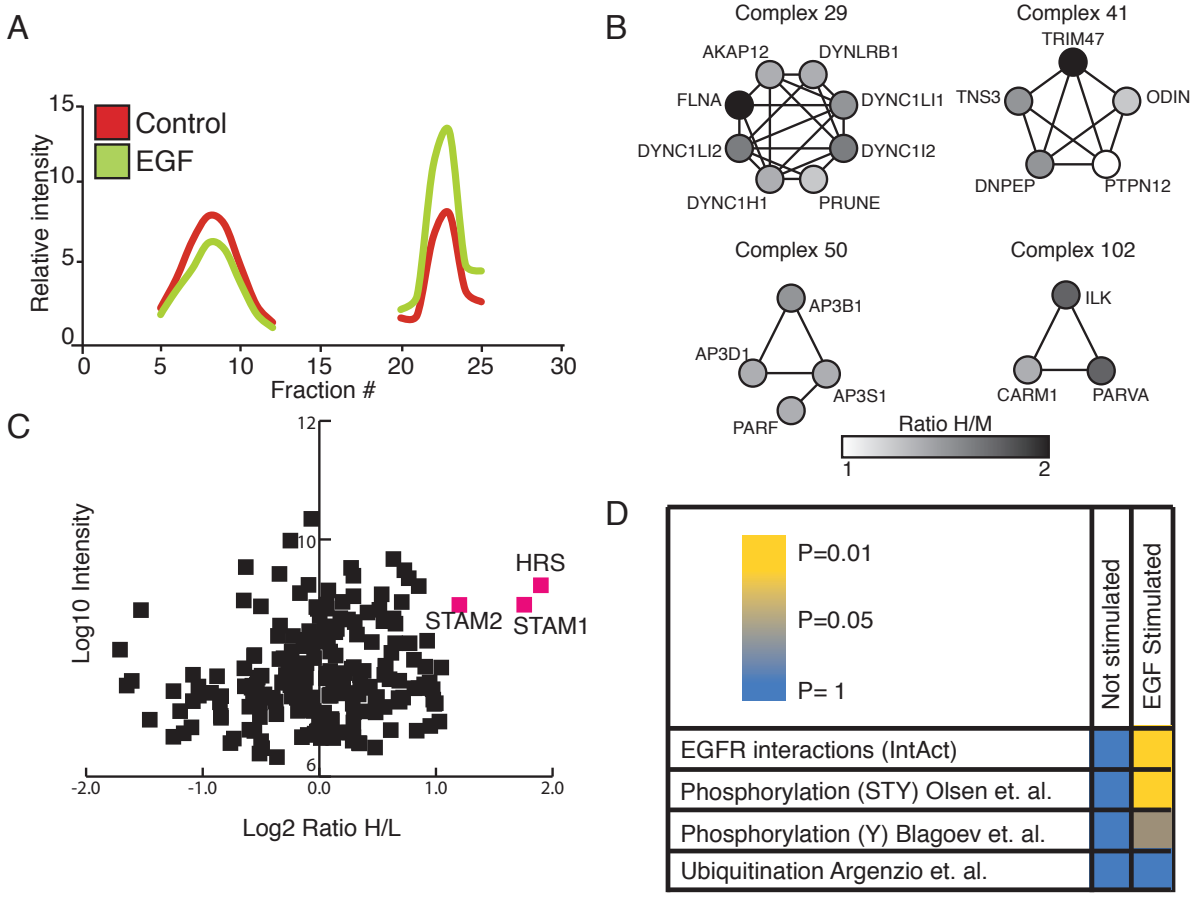
Furthermore, within the same protein complex we can distinguish proteins with different stoichiometries, by first resolving the stoichiometries for each protein by principle components analysis and then detecting outliers by Hotelling's T2 statistic (Figure 3.6C). This highlights that the E3 ubiquitin-protein ligase KCMF1, which recently has been shown to bind to the proteasome by AP-MS [246], has significantly different stoichiometry than the rest of the 19S regulatory particle proteins (P=0.02, Hotelling's T2 statistic) and a closer inspection of the chromatogram revealed that it is only bound to the 19S regulatory particle when the 19S is in turn bound to the 20S core particle (Figure 3.6D). This could imply that this E3 ligase is only delivering ubiquitinated proteins to the 19S regulatory particle when that is associated with the 20S core particle forming the 26S proteasome. As the above examples highlight, this PCP-SILAC approach to the interactome can provide additional valuable detail about the composition of complexes that can then inform subsequent experiments.

3.3.4 Unraveling the dynamic interactome

Interactome rearrangement must occur in order for a cell to respond to stimuli but large-scale detection of such rearrangement has not previously been reported. Incorporation of a third SILAC label into the PCP-SILAC scheme (Figure 3.1B) allows the heavy/medium ratio be used to quantify the temporal changes in the interactome following 20 min EGF stimulation. This approach reveals 351 proteins whose association with a complex is increased or decreased by the EGF challenge (Figure 3.7A). Among these changes were well documented proteins known to be a part of the EGF signaling cascade: EPS15, SHIP2, STAM1/STAM2 and HRS capable of binding to the EGF receptor (EGFR) itself, STAT3 is

an important transcription factor downstream of the receptor, and finally ILK, PXN, PARVA and PINCH are all involved in integrin-mediated signaling that synergizes with growth factor signaling, including EGF [247]. We also observed that all members of an AP-3 complex to be affected by EGF (Figure 3.7B). AP-3 associates with clathrin, colocalizes at endosomes and assists in delivering proteins to late endosomes and lysosomes [248] so its regulation by EGF suggests that it could be functioning similarly to the well-documented AP-2 complex in the endocytosis and down-regulation of the EGF receptor.

Figure 3.7 Determination of interactome response to EGF stimulation



A) Chromatogram deriving from STAM1 highlighting the EGF-induced spatiotemporal change. The red line represents the chromatogram from un-stimulated cells (i.e., the plot of M/L as in Fig. 1b), while the green line represents the chromatogram from the EGF-stimulated cells (i.e., the plot of H/L). EGF stimulation causes the complex with STAM1 at fraction 7 to dissociate and STAM1 to associate with a different complex eluding in fraction 22. B) Example of four protein complexes being regulated by EGF stimulation; the color key is the average H/M ratio in the three fractions where the complexes peaked. The AP-3 complex (complex 50) is known to bind clathrin and has earlier been described to be involved in endosome to lysosome trafficking. The three other complexes are an ILK complex (complex 102), a dynein complex (complex 29) and a novel complex (complex 41) where we observe regulation of a number of proteins in responds to EGF. C) Proteins identified as being specific interacters of HRS by AP-MS of fractions 19 through 24. By conducting quantitative affinity purification-mass spectrometry of fractions 19-24 where HRS is eluting, we identified three proteins with SignificanceB value [48]<0.01 and a log2 ratio>1. D) Comparison of proteins with spatiotemporal changes following EGF stimulation with other high throughput datasets. The IntAct dataset [239] are proteins known to bind to EGFR deriving from 46 different publications, the phosphorylation (STY) dataset from Olsen et al. containing global phosphorylation changes caused by EGF stimulation[44], the phosphorylation (Y) dataset are from Blagoev et al. containing tyrosine phosphorylation changes following EGF stimulation [186] and finally the ubiquitination dataset is from Argenzio et al. which studied ubiquitination changes following EGF stimulation [238]. All of the last three studies are carried out in Hela cells and used a similar stimulation time as this study.

Next we validate one of the proteins that observed by PCP-SILAC to change its interactions following EGF stimulation by using two approaches: 1) By adding an antibody against HRS and thereby accelerating it and its interaction partners through the SEC column similarly to the above example with 14-3-3γ, and 2) by performing AP-MS of HRS from only those SEC fractions it eluted in. In both cases we identified only STAM1 and STAM2 (Figure 3.7C), which is known to associate with HRS in response to EGF stimulation [249].

As an unbiased test of the specificity of the proteins detected to be changing with EGF, we asked whether the enrichment of specific proteins observed in these 351 proteins compared to other sets of proteins known to be involved in EGF-mediated events. There was significant (P=0.01, Fisher's exact test) co-enrichment between the proteins detected here and the subset of proteins in the IntAct database that are recorded to bind to EGFR [239],

implying that PCP-SILAC of soluble complexes can still detect relevant signaling complexes even when membrane-bound proteins are specifically depleted by the exclusion of detergents. From many focused studies we know that post- translational modifications control many protein interactions and the EGF system is one of the best-studied mammalian signaling systems. While we do not know if the 351 affected proteins were phosphorylated in our study, there was significant co-enrichment between this set and two published phosphoproteomic analyses of EGF signalling. Many of the same proteins we detected were also found to have altered phosphorylation in one study that looked at global phosphorylations changes in response to EGF, (P=0.01; Fisher's exact test) [44] and in a second study focused on phosphotyrosine signaling (P=0.03; Fisher's exact test) [186] (Figure 3.7D). Interestingly, however, we did not observe a significant co-enrichment between our data and proteins whose ubiquitylation state was altered following EGF stimulation (P=0.58; Fisher's Exact test) [238], implying that phosphorylation is a stronger regulator of interactions than is ubiquitylation. The fact that we are able to observe significant correlation among three different independent datasets suggests that PCP-SILAC is accurate in identifying temporal changes following cell perturbation.

3.4 Discussion

Despite the impressive progress in unraveling the interactome in recent years, the majority of such data is derived from studies of a single experimental condition and can their collection was immensely labor intensive. PCP-SILAC combined with high resolution SEC provides a valuable tool for future interactome monitoring, since it is a quick approach that is complementary to existing techniques, and it does not rely on over-expression or creating

fusion proteins, but instead detects completely endogenous protein-protein interactions. This requires much less labor than conventional approaches and generates more biologically relevant data, since a tag can easily disrupt protein complexes; as recently reported: in *Caulobacter* only one third of proteins localized properly regardless of which end the protein was tagged at, suggesting that the tag disrupts terminal structures or interaction motifs of the protein [178]. Our approach makes it possible to investigate one cellular compartment at a time, resulting in only interactions among proteins present in a particular compartment being detected and thereby avoiding irrelevant interactions that can arise when all potential proteins interactions are tested in the cell. Since PCP-SILAC combined with high resolution SEC does not rely on constructing fusion proteins and analyzing thousands of samples by MS, it can be tailored to all organisms or cell systems amenable to SILAC labeling, including plants, flies and mice [64,250,251], making it possible to routinely detect protein interactions in their native compartment, tissue or cell type under many different conditions. SEC can also be performed in the presence of detergents so this should even be extendable to membrane interactomes.

Through the use of a third SILAC label to encode how the interactome rearranges following EGF stimulation, we were able to identify 351 proteins whose interactions were altered by the stimulation. Generally, interactomes are only constructed in one standard condition, leading to a static network; recently, however, Bandyopadhyay *et al.* used an approach called differentially epistasis mapping, leading to the identification of 873 differential genetic interactions following DNA damage response [185]. The advantages we see of using PCP-SILAC are that real physical interactions are detected and control and stimulated cells are mixed together at a very early step in the procedure, thereby minimizing

technical variability and ensuring that data is recorded for both populations. The time and resources that go into a PCP-SILAC experiment are reasonable enough that one could envision using it to screen the impact of a whole range of stimuli or pharmacological agents on the interactome, something that would be completely inconceivable with conventional methods. This could be used to look for off-target effects of drugs or to finally fill in one of the huge remaining gaps in data used for systems biology modeling.

Chapter 4: Conclusion

The research presented in this thesis has focused on the regulation of protein networks using mass spectrometry-based proteomics. Chapters 2 and 3 describe how protein networks can be deciphered in response to cellular perturbation; this was achieved through measuring protein interaction and expression changes.

The genome is a relative stable constituent throughout the lifetime of an organism, but proteins and the interactions between them will constantly change, making the organism fit for new states. The technology for studying protein interactions has mainly focused on identifying more interactions with lower false positive rates, very similar to how the field of proteomics mainly focused on identifying more proteins in the first decade. However, the study of protein interactions is now starting to change and focus on how the interactions change in response to external perturbation. This has mainly been achieved through performing AP-MS or by studying genetic interactions under two conditions with the doubling of the workload as a result [185,186,190]. In addition, both AP-MS and the in-vivo techniques require adding tags or generating fusion proteins, which has been reported to cause changes in the localization and/or activity of a protein [173,178,179]. Furthermore, while the introduction of tags or generating fusion proteins is manageable in organisms such as yeast, it is a much harder job in mammalian cells and impossible in humans.

In Chapter 3 of this thesis I described the development of a novel approach to investigate temporal protein interaction network that combines SEC with PCP-SILAC, making it possible to generate chromatograms for thousands of proteins. By then comparing these chromatograms to each other, it is possible to identify protein interactions based on co-migration. Finally, by including a third SILAC label into the PCP-SILAC scheme, I showed

that it is possible to identify temporal protein interactions changes in response to perturbation of the cells.

I believe my newly developed approach, SEC-PCP-SILAC, has many things to offer the protein interaction field, since it addresses both afore mentioned bottlenecks. It allows for the study of protein interactions under two conditions without increasing the workload, making it possible to identify protein interaction network changes in response to cellular perturbation, and it does not require protein tags, the generation of fusion proteins or overexpression. In addition, it allows the determination of the stoichiometry of a protein participating in multiple complexes and it can identify protein interactions with similar precision and false positive rates as AP-MS. I think the most interesting biological results from the SEC-PCP-SILAC approach were the strong co-enrichment of the proteins we found to change interaction with the previous published changes in phosphorylation, in response to EGF stimulation. This has been known from numerous focused studies, but showing this strong connection on a global level strongly suggests that a number of protein phosphorylations are functioning as switches that can turn interactions between protein on or off.

I envision SEC-PCP-SILAC, in the coming years, to play a major role in providing valuable information concerning protein interaction network topology and how it responds to external perturbation. The SEC-PCP-SILAC approach can also be extended to reveal other characteristics about proteins in macromolecular complexes. This I showed by combining it with pulse SILAC labeling to be able to measure the relative synthesis and degradation rate of individual proteins in sub-complexes. However, it could also be a valuable tool in identifying potential protein-drug interactions in a similar way as was recently published

where co-migration of drug and proteins were detected through chromatographic co-elution [252].

I would argue that some technologies for measuring protein interactions are nearing their maximum throughput efficiency, meaning that to identify protein interactions by AP-MS or by Y2H will, most likely, require the same workload in 10 years as it does now. The reason for this is that AP-MS will still rely on proteins being affinity purified followed by MS-analysis and Y2H will still rely on mating different yeast strains, without any obvious increase in number of identified interactions. On the other hand, I believe the SEC-PCP-SILAC approach has much more to offer, in part because numerous proteins we identified could not be assigned to any protein complex since they did not obey the stringent filtering criteria, yet they were still migrating at much higher than their calculated molecular weight and so were presumably part of a complex. Two technical developments would greatly improve SEC-PCP-SILAC: 1) higher resolution SEC (during this project the resolution from the columns I have used has doubled), allowing better fractionation to be obtained, and 2) faster and more sensitive mass spectrometers that can identify more proteins in more fractions with higher precision and lower false positive rates.

In Chapter 2 of this thesis I investigated the contributions of transcription change, protein synthesis and protein degradation to the protein expression change following differentiation in two different cell lines deriving from two different organisms. One of the interesting observations from this was that the general lack of correlation between transcriptome and proteome expression changes derives from not considering the synthesis and degradation rates of the proteins. In addition, I could show that the synthesis rates of the individual proteins changed dramatically in response to differentiation, whereas the

degradation rates of the individual proteins changed little. Finally, I confirmed that the degradation rates are conserved between organisms, consistent with previous reports [204], I think these results are very interesting with respect to how we often think about protein expression changes. Before I started this project, my perception of protein expression was that a protein increased in expression as a result of increased synthesis rate and decreased in expression as a result of increased degradation rate. However, that changes in the synthesis rate are such a strong variable in both the up and down regulation of protein expression was a big surprise to me, although as I read more literature this made more sense. Intriguingly, the mRNA expression change is regulated at the level of synthesis as well, whereas the degradation rate of mRNA is relatively constant [224].

That the degradation rates change little in response to differentiation and are conserved between organisms suggests a strong "default" component in the degradation of many proteins, which I propose is based on the protein's primary sequence and can either be dependent or independent of ubiquitination [221-223]. This default component is presumably evolutionary optimized; low abundant proteins will have faster turnover rates since it degrading fewer protein molecules is energetically more affordable than degrading many. That the degradation rate of a protein is changing only minimally suggests that a protein with high turnover rate can be removed faster in response to perturbations, which can be central for proteins regulating signaling transduction or transcription.

It will be interesting in the future to observe if protein expression is regulated in a similar manner in response to other perturbations or if it is a differentiation specific phenomenon.

References

[1] Barabasi, A., Albert, R., Emergence of scaling in random networks. *Science* 1999, 286, 509–512.

[2] Ideker, T., Krogan, N.J., Differential network biology. *Molecular Systems Biology* 2012, 8, 1–9.

[3] Madhani, H.D., Styles, C.A., Fink, G.R., MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation. *Cell* 1997, 91, 673–684.

[4] Vidal, M., Cusick, M.E., Barabási, A.-L., Interactome Networks and Human Disease. *Cell* 2011, 144, 986–998.

[5] Kuhner, S., van Noort, V., Betts, M.J., Leo-Macias, A., et al., Proteome Organization in a Genome-Reduced Bacterium. *Science* 2009, 326, 1235–1240.

[6] Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., et al., Proteome survey reveals modularity of the yeast cell machinery. *Nature* 2006, 440, 631–636.

[7] Krogan, N.J., Cagney, G., Yu, H., Zhong, G., et al., Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* 2006, 440, 637–643.

[8] Fields, S., Song, O., A novel genetic system to detect protein-protein interactions. *Nature* 1989, 340, 245–246.

[9] Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., et al., Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002, 415, 141–147.

[10] Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., et al., A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnology* 1999, 17, 1030–1032.

[11] Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.-O., et al., A physical and functional map of the human TNF-α/NF-κB signal transduction pathway. *Nature* 2004, 6, 97–105.

[12] Babu, M., Vlasblom, J., Pu, S., Guo, X., et al., Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae. *Nature* 2012, 489, 585–589.

[13] Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., et al., Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* 2005, 433, 531–537.

[14] Havugimana, P.C., Hart, T.G., Nepusz, T., Yang, H., et al., A Census of Human Soluble Protein Complexes. *Cell* 2012, 150, 1068–1081.

[15] Kristensen, A.R., Gsponer, J., Foster, L.J., A high-throughput approach for measuring temporal changes in the interactome. *Nature Methods* 2012, 9, 907–909.

[16] Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., et al., Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* 2003, 426, 570–574.

[17] Andersen, J.S., Lam, Y.W., Leung, A.K.L., Ong, S.-E., et al., Nucleolar proteome dynamics. *Nature* 2005, 433, 77–83.

[18] Foster, L.J., de Hoog, C.L., Zhang, Y., Zhang, Y., et al., A Mammalian Organelle Map by Protein Correlation Profiling. *Cell* 2006, 125, 187–199.

[19] Jakobsen, L., Vanselow, K., Skogs, M., Toyoda, Y., et al., Novel asymmetrically localizing components of human centrosomes identified by complementary

proteomics methods. *The EMBO Journal* 2011, 30, 1520–1535.
[20] Dengjel, J., Hoyer-Hansen, M., Nielsen, M.O., Eisenberg, T., et al., Identification of Autophagosome-associated Proteins and Regulators by Quantitative Proteomic Analysis and Genetic Screens. *Molecular & Cellular Proteomics* 2012, 11, M111.014035–M111.014035.
[21] Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., et al., Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls 2007, 447, 1–23.
[22] Zhong, Q., Simonis, N., Li, Q.-R., Charloteaux, B., et al., Edgetic perturbation models of human inherited disorders. *Molecular Systems Biology* 2009, 5, 1–10.
[23] Rogers, L.D., Kristensen, A.R., Boyle, E.C., Robinson, D.P., et al., Identification of cognate host targets and specific ubiquitylation sites on the Salmonella SPI-1 effector SopB/SigD. *Journal of Proteomics* 2008, 71, 97–108.
[24] Patel, J.C., Galan, J.E., Differential activation and function of Rho GTPases during Salmonella-host cell interactions. *The Journal of Cell Biology* 2006, 175, 453–463.
[25] Zhou, D., Chen, L.M., Hernandez, L., Shears, S.B., Galan, J.E., A Salmonella inositol polyphosphatase acts in conjunction with other bacterial effectors to promote host cell actin cytoskeleton rearrangements and bacterial internalization. *Mol. Microbiol.* 2001, 39, 248–259.
[26] Yu, X., Yu, Y., Liu, B., Luo, K., et al., Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* 2003, 302, 1056–1060.
[27] Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., et al., Global landscape of HIV-human protein complexes. *Nature* 2012, 481, 365–370.
[28] Jäger, S., Kim, D.Y., Hultquist, J.F., Shindo, K., et al., Vif hijacks CBF-β to degrade APOBEC3G and promote HIV-1 infection. *Nature* 2012, 481, 371–375.
[29] Joensuu, H., Pylkkänen, L., Toikkanen, S., Bcl-2 protein expression and long-term survival in breast cancer. *Am. J. Pathol.* 1994, 145, 1191–1198.
[30] Harris, H., Rubinsztein, D.C., Control of autophagy as a therapy for neurodegenerative disease. *Nat Rev Neurol* 2012, 8, 108–117.
[31] Blume-Jensen, P., Hunter, T., Oncogenic kinase signalling. *Nature* 2001, 411, 355–365.
[32] Lemmon, M.A., Schlessinger, J., Cell Signaling by Receptor Tyrosine Kinases. *Cell* 2010, 141, 1117–1134.
[33] Pawson, T., Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 2004, 116, 191–203.
[34] Paulson, H.L., Protein fate in neurodegenerative proteinopathies: polyglutamine diseases join the (mis)fold. *Am. J. Hum. Genet.* 1999, 64, 339–345.
[35] Yang, D.-S., Stavrides, P., Mohan, P.S., Kaushik, S., et al., Reversal of autophagy dysfunction in the TgCRND8 mouse model of Alzheimer's disease ameliorates amyloid pathologies and memory deficits. *Brain* 2011, 134, 258–277.
[36] Ravikumar, B., Duden, R., Rubinsztein, D.C., Aggregate-prone proteins with polyglutamine and polyalanine expansions are degraded by autophagy. *Hum. Mol. Genet.* 2002, 11, 1107–1117.
[37] Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., et al., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis

and amino acid analysis. *Biotechnology (N.Y.)* 1996, 14, 61–65.
[38] Cox, J., Mann, M., Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annu. Rev. Biochem.* 2011, 80, 273–299.
[39] Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, 246, 64–71.
[40] Hillenkamp, F., Karas, M., Beavis, R.C., Chait, B.T., Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* 1991, 63, 1193A–1203A.
[41] Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., et al., Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 1996, 379, 466–469.
[42] Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., et al., Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology* 2011, 7, 1–8.
[43] de Godoy, L.M.F., Olsen, J.V., Cox, J., Nielsen, M.L., et al., Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 2008, 455, 1251–1254.
[44] Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., et al., Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* 2006, 127, 635–648.
[45] Olsen, J.V., Ong, S.-E., Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell Proteomics* 2004, 3, 608–614.
[46] Steen, H., Mann, M., The abc“s (and xyz”s) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004, 5, 699–711.
[47] Wu, C., Tran, J.C., Zamdborg, L., Durbin, K.R., et al., A protease for “middle-down” proteomics. *Nature Methods* 2012, 9, 822–824.
[48] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* 2008, 26, 1367–1372.
[49] Ong, S.-E., Mann, M., Mass spectrometry–based proteomics turns quantitative. *Nat Chem Biol* 2005, 1, 252–262.
[50] Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., Gygi, S.P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* 2003, 100, 6940–6945.
[51] Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., et al., Global quantification of mammalian gene expression control. *Nature* 2011, 473, 337–342.
[52] Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., et al., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics* 2004, 3, 1154–1169.
[53] Hsu, J.-L., Huang, S.-Y., Chow, N.-H., Chen, S.-H., Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* 2003, 75, 6843–6852.
[54] Boersema, P.J., Aye, T.T., van Veen, T.A.B., Heck, A.J.R., Mohammed, S., Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* 2008, 8, 4624–4632.
[55] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., et al., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*

1999, 17, 994–999.
[56] Ow, S.Y., Salim, M., Noirel, J., Evans, C., et al., iTRAQ Underestimation in Simple and Complex Mixtures: “The Good, the Bad and the Ugly.” *J. Proteome Res.* 2009, 8, 5347–5355.
[57] Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., Heck, A.J.R., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols* 2009, 4, 484–494.
[58] Oda, Y., Huang, K., Cross, F.R., Cowburn, D., Chait, B.T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* 1999, 96, 6591–6596.
[59] Simonsen, K.T., Møller-Jensen, J., Kristensen, A.R., Andersen, J.S., et al., Quantitative proteomics identifies ferritin in the innate immune response of C. elegans. *virulence* 2011, 2, 120–130.
[60] Gouw, J.W., Krijgsveld, J., Heck, A.J.R., Quantitative Proteomics by Metabolic Labeling of Model Organisms. *Molecular & Cellular Proteomics* 2010, 9, 11–24.
[61] Gruhler, A., Schulze, W.X., Matthiesen, R., Mann, M., Jensen, O.N., Stable isotope labeling of Arabidopsis thaliana cells and quantitative proteomics by mass spectrometry. *Mol. Cell Proteomics* 2005, 4, 1697–1709.
[62] Krijgsveld, J., Ketting, R.F., Mahmoudi, T., Johansen, J., et al., Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. *Nature Biotechnology* 2003, 21, 927–931.
[63] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., et al., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* 2002, 1, 376–386.
[64] Krüger, M., Moser, M., Ussar, S., Thievessen, I., et al., SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell* 2008, 134, 353–364.
[65] Larance, M., Bailly, A.P., Pourkarimi, E., Hay, R.T., et al., Stable-isotope labeling with amino acids in nematodes. *Nature Methods* 2011, 8, 849–851.
[66] Konzer, A., Ruhs, A., Braun, H., Jungblut, B., et al., Stable isotope labelling in zebrafish allows in vivo monitoring of cardiac morphogenesis. *Molecular & Cellular Proteomics* 2013.
[67] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R., Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods* 2010, 7, 383–385.
[68] Deeb, S.J., D'Souza, R.C.J., Cox, J., Schmidt-Supprian, M., Mann, M., Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles. *Molecular & Cellular Proteomics* 2012, 11, 77–89.
[69] Finnigan Corporation, Method and Apparatus of Increasing Dynamic Range and Sensitivity of a Mass ... U.S. Patent 5572022, 1995.
[70] Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., et al., Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* 2006, 78, 2113–2120.
[71] Olsen, J.V., Macek, B., Lange, O., Makarov, A., et al., Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* 2007, 4, 709–712.
[72] Schwartz, J.C., Senko, M.W., Syka, J.E.P., A two-dimensional quadrupole ion trap

mass spectrometer. *J. Am. Soc. Mass Spectrom.* 2002, 13, 659–669.
[73] Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* 2000, 72, 1156–1162.
[74] Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., et al., Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 2011, 10, 1794–1805.
[75] Zubarev, R.A., Håkansson, P., Sundqvist, B., Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements. *Anal. Chem.* 1996, 68, 4060–4063.
[76] Mann, M., Kelleher, N.L., Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 18132–18138.
[77] Elias, J.E., Gygi, S.P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* 2007, 4, 207–214.
[78] Nesvizhskii, A.I., Aebersold, R., Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell Proteomics* 2005, 4, 1419–1440.
[79] Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., et al., Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. *Science Signaling* 2010, 3, ra3–ra3.
[80] Rigbolt, K.T.G., Prokhorova, T.A., Akimov, V., Henningsen, J., et al., System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Science Signaling* 2011, 4, rs3.
[81] Romijn, E.P., Christis, C., Wieffer, M., Gouw, J.W., et al., Expression clustering reveals detailed co-expression patterns of functionally related proteins during B cell differentiation: a proteomic study using a combination of one-dimensional gel electrophoresis, LC-MS/MS, and stable isotope labeling by amino acids in cell culture (SILAC). *Mol. Cell Proteomics* 2005, 4, 1297–1310.
[82] Kristensen, A.R., Schandorff, S., Høyer-Hansen, M., Nielsen, M.O., et al., Ordered organelle degradation during starvation-induced autophagy. *Molecular & Cellular Proteomics* 2008, 7, 2419–2428.
[83] Molina, H., Yang, Y., Ruch, T., Kim, J.-W., et al., Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. *J. Proteome Res.* 2009, 8, 48–58.
[84] Au, C.E., Bell, A.W., Gilchrist, A., Hiding, J., et al., Organellar proteomics to create the cell map. *Current Opinion in Cell Biology* 2007, 19, 376–385.
[85] Øverbye, A., Fengsrud, M., Seglen, P.O., Proteomic analysis of membrane-associated proteins from rat liver autophagosomes. *autophagy* 2007, 3, 300–322.
[86] Bell, A.W., Ward, M.A., Blackstock, W.P., Freeman, H.N., et al., Proteomics characterization of abundant Golgi membrane proteins. *J. Biol. Chem.* 2001, 276, 5152–5165.
[87] Duclos, S., Clavarino, G., Rousserie, G., Goyette, G., et al., The endosomal proteome of macrophage and dendritic cells. *Proteomics* 2011, 11, 854–864.
[88] Wiese, S., Gronemeyer, T., Ofman, R., Kunze, M., et al., Proteomics characterization of mouse kidney peroxisomes by tandem mass spectrometry and protein correlation profiling. *Mol. Cell Proteomics* 2007, 6, 2045–2057.
[89] Forner, F., Foster, L.J., Campanaro, S., Valle, G., Mann, M., Quantitative

proteomic comparison of rat mitochondria from muscle, heart, and liver. *Mol. Cell Proteomics* 2006, 5, 608–619.

[90] Dunkley, T.P.J., Watson, R., Griffin, J.L., Dupree, P., Lilley, K.S., Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell Proteomics* 2004, 3, 1128–1134.

[91] Dunkley, T.P.J., Hester, S., Shadforth, I.P., Runions, J., et al., Mapping the Arabidopsis organelle proteome. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 6518–6523.

[92] Zheng, Y.Z., Berg, K.B., Foster, L.J., Mitochondria do not contain lipid rafts, and lipid rafts do not contain mitochondrial proteins. *The Journal of Lipid Research* 2008, 50, 988–998.

[93] Schoenheimer, R., Ratner, S., Rittenberg, D., Studies of Protein Metabolism. *J. Biol. Chem.* 1939, 333–344.

[94] Goldberg, A.L., Dice, J.F., Intracellular protein degradation in mammalian and bacterial cells. *Annu. Rev. Biochem.* 1974, 43, 835–869.

[95] Belle, A., Tanay, A., Bitincka, L., Shamir, R., O'Shea, E.K., Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 13004–13009.

[96] Eden, E., Geva-Zatorsky, N., Issaeva, I., Cohen, A., et al., Proteome Half-Life Dynamics in Living Human Cells. *Science* 2011, 331, 764–768.

[97] Boisvert, F.M., Ahmad, Y., Gierlinski, M., Charriere, F., et al., A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Molecular & Cellular Proteomics* 2012, 11, M111.011429–M111.011429.

[98] Somasekharan, S.P., Stoynov, N., Rotblat, B., Leprivier, G., et al., Identification and quantification of newly synthesized proteins translationally regulated by YB-1 using a novel Click–SILAC approach. *Journal of Proteomics* 2012, 77, e1–e10.

[99] Eichelbaum, K., Winter, M., Diaz, M.B., Herzig, S., Krijgsveld, J., Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nature Biotechnology* 2012, 1–9.

[100] Ahmad, Y., Boisvert, F.-M., Lundberg, E., Uhlen, M., Lamond, A.I., Systematic analysis of protein pools, isoforms, and modifications affecting turnover and subcellular localization. *Molecular & Cellular Proteomics* 2012, 11, M111.013680.

[101] Larance, M., Ahmad, Y., Kirkwood, K.J., Ly, T., Lamond, A.I., Global subcellular characterization of protein degradation using quantitative proteomics. *Molecular & Cellular Proteomics* 2013, 12, 638–650.

[102] Sancak, Y., Peterson, T.R., Shaul, Y.D., Lindquist, R.A., et al., The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* 2008, 320, 1496–1501.

[103] Inoki, K., Zhu, T., Guan, K.-L., TSC2 mediates cellular energy response to control cell growth and survival. *Cell* 2003, 115, 577–590.

[104] Ma, X.M., Blenis, J., Molecular mechanisms of mTOR-mediated translational control. *Nat Rev Mol Cell Biol* 2009, 10, 307–318.

[105] Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., et al., A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* 2012, 485, 109–113.

[106] Washburn, M.P., Koller, A., Oshiro, G., Ulaszek, R.R., et al., Protein pathway and

complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* 2003, 100, 3107–3112.
[107] Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M., Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* 2003, 4, 117.
[108] Griffin, T.J., Gygi, S.P., Ideker, T., Rist, B., et al., Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. *Mol. Cell Proteomics* 2002, 1, 323–333.
[109] Fournier, M.L., Paulson, A., Pavelka, N., Mosley, A.L., et al., Delayed Correlation of mRNA and Protein Expression in Rapamycin-treated Cells and a Role for Ggc1 in Cellular Sensitivity to Rapamycin. *Molecular & Cellular Proteomics* 2010, 9, 271–284.
[110] Lee, M.V., Topper, S.E., Hubler, S.L., Hose, J., et al., A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Molecular Systems Biology* 2011, 7, 1–12.
[111] Munoz, J., Low, T.Y., Kok, Y.J., Chin, A., et al., The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Molecular Systems Biology* 2011, 7, 1–13.
[112] Vogel, C., de Sousa Abreu, R., Ko, D., Le, S.-Y., et al., Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Molecular Systems Biology* 2010, 6, 1–9.
[113] Wightman, B., Ha, I., Ruvkun, G., Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 1993, 75, 855–862.
[114] Lee, R.C., Feinbaum, R.L., Ambros, V., The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993, 75, 843–854.
[115] Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., et al., Widespread changes in protein synthesis induced by microRNAs. *Nature* 2008, 455, 58–63.
[116] Mendez, R., Richter, J.D., Translational control by CPEB: a means to the end. *Nat Rev Mol Cell Biol* 2001, 2, 521–529.
[117] Kulkarni, M., Ozgur, S., Stoecklin, G., On track with P-bodies. *Biochem. Soc. Trans.* 2010, 38, 242–251.
[118] Clague, M.J., Urbé, S., Ubiquitin: same molecule, different degradation pathways. *Cell* 2010, 143, 682–685.
[119] Hershko, A., Ciechanover, A., The ubiquitin system. *Annu. Rev. Biochem.* 1998, 67, 425–479.
[120] Klionsky, D.J., Autophagy: from phenomenology to molecular understanding in less than a decade. *Nat Rev Mol Cell Biol* 2007, 8, 931–937.
[121] Dengjel, J., Kristensen, A.R., Andersen, J.S., Ordered bulk degradation via autophagy. *autophagy* 2008, 4, 1057–1059.
[122] Kirkin, V., McEwan, D.G., Novak, I., Dikic, I., A role for ubiquitin in selective autophagy. *Molecular Cell* 2009, 34, 259–269.
[123] Hershko, A., Heller, H., Elias, S., Ciechanover, A., Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J. Biol. Chem.* 1983, 258, 8206–8214.

[124] Nakayama, K.I., Nakayama, K., Ubiquitin ligases: cell-cycle control and cancer. *Nature Reviews Cancer* 2006, 6, 369–381.
[125] Nijman, S.M.B., Luna-Vargas, M.P.A., Velds, A., Brummelkamp, T.R., et al., A genomic and functional inventory of deubiquitinating enzymes. *Cell* 2005, 123, 773–786.
[126] Ikeda, F., Dikic, I., Atypical ubiquitin chains: new molecular signals. "Protein Modifications: Beyond the Usual Suspects" review series. *EMBO Rep* 2008, 9, 536–542.
[127] Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., et al., A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* 1989, 243, 1576–1583.
[128] Hirano, Y., Hendil, K.B., Yashiroda, H., Iemura, S.-I., et al., A heterodimeric complex that promotes the assembly of mammalian 20S proteasomes. *Nature* 2005, 437, 1381–1385.
[129] Arendt, C.S., Hochstrasser, M., Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. *Proc. Natl. Acad. Sci. U.S.A.* 1997, 94, 7156–7161.
[130] Leggett, D.S., Hanna, J., Borodovsky, A., Crosas, B., et al., Multiple associated proteins regulate proteasome structure and function. *Molecular Cell* 2002, 10, 495–507.
[131] Husnjak, K., Elsasser, S., Zhang, N., Chen, X., et al., Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* 2008, 453, 481–488.
[132] Deveraux, Q., Ustrell, V., Pickart, C., Rechsteiner, M., A 26 S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* 1994, 269, 7059–7061.
[133] De Duve, C., Wattiaux, R., Functions of lysosomes. *Annu. Rev. Physiol.* 1966, 28, 435–492.
[134] Levine, B., Klionsky, D.J., Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Dev. Cell* 2004, 6, 463–477.
[135] Okamoto, K., Kondo-Okamoto, N., Ohsumi, Y., Mitochondria-anchored receptor Atg32 mediates degradation of mitochondria via selective autophagy. *Dev. Cell* 2009, 17, 87–97.
[136] Kraft, C., Deplazes, A., Sohrmann, M., Peter, M., Mature ribosomes are selectively degraded upon starvation by an autophagy pathway requiring the Ubp3p/Bre5p ubiquitin protease. *Nat. Cell Biol.* 2008, 10, 602–610.
[137] Iwata, J.-I., Ezaki, J., Komatsu, M., Yokota, S., et al., Excess peroxisomes are degraded by autophagic machinery in mammals. *J. Biol. Chem.* 2006, 281, 4035–4041.
[138] Johansen, T., Lamark, T., Selective autophagy mediated by autophagic adapter proteins. *autophagy* 2011, 7, 279–296.
[139] Hedin, S.G., Trypsin and Antitrypsin. *Biochem. J.* 1906, 1, 474–483.
[140] Taylor, S.S., cAMP-dependent protein kinase. Model for an enzyme family. *J. Biol. Chem.* 1989, 264, 8443–8446.
[141] Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., et al., The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 8685–8690.
[142] Pastor-Satorras, R., Smith, E., Solé, R.V., Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology* 2003, 222, 199–210.

[143] Han, J.-D.J., Bertin, N., Hao, T., Goldberg, D.S., et al., Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 2004, 430, 88–93.
[144] Agarwal, S., Deane, C.M., Porter, M.A., Jones, N.S., Revisiting Date and Party Hubs: Novel Approaches to Role Assignment in Protein Interaction Networks. *PLoS Comput Biol* 2010, 6, e1000817.
[145] Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., et al., Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnology* 2009, 27, 199–204.
[146] Miller, J.G., Living systems: basic concepts. *Behav Sci* 1965, 10, 193–237.
[147] Dezso, Z., Oltvai, Z.N., Barabási, A.-L., Bioinformatics analysis of experimentally determined protein complexes in the yeast Saccharomyces cerevisiae. *Genome Research* 2003, 13, 2450–2454.
[148] Laplante, M., Sabatini, D.M., mTOR Signaling in Growth Control and Disease. *Cell* 2012, 149, 274–293.
[149] Rives, A.W., Galitski, T., Modular organization of cellular networks. *Proc. Natl. Acad. Sci. U.S.A.* 2003, 100, 1128–1133.
[150] Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., et al., An in Vivo Map of the Yeast Protein Interactome. *Science* 2008, 320, 1465–1470.
[151] Haar, E.V., Lee, S.-I., Bandhakavi, S., Griffin, T.J., Kim, D.-H., Insulin signalling to mTOR mediated by the Akt/PKB substrate PRAS40. *Nature* 2007, 9, 316–323.
[152] Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., et al., Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO Journal* 2010, 29, 717–726.
[153] Herzog, F., Kahraman, A., Boehringer, D., Mak, R., et al., Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. *Science* 2012, 337, 1348–1352.
[154] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., et al., A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* 2001, 98, 4569–4574.
[155] Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., et al., A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 2000, 403, 623–627.
[156] Johnsson, N., Varshavsky, A., Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 1994, 91, 10340–10344.
[157] Hu, C.-D., Chinenov, Y., Kerppola, T.K., Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Molecular Cell* 2002, 9, 789–798.
[158] Remy, I., Michnick, S.W., Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *Proc. Natl. Acad. Sci. U.S.A.* 1999, 96, 5394–5399.
[159] Lippincott-Schwartz, J., Snapp, E., Kenworthy, A., Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol* 2001, 2, 444–456.
[160] Singh, P., Panchaud, A., Goodlett, D.R., Chemical Cross-Linking and Mass Spectrometry As a Low-Resolution Protein Structure Determination Technique. *Anal. Chem.* 2010, 82, 2636–2642.

[161] Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., et al., False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature Methods* 2012, 9, 901–903.
[162] Yang, B., Wu, Y.-J., Zhu, M., Fan, S.-B., et al., Identification of cross-linked peptides from complex samples. *Nature Methods* 2012, 9, 904–906.
[163] Wilk, S., Orlowski, M., Cation-sensitive neutral endopeptidase: isolation and specificity of the bovine pituitary enzyme. *J Neurochem* 1980, 35, 1172–1182.
[164] Porath, J., Flodin, P., Gel filtration: a method for desalting and group separation. *Nature* 1959, 183, 1657–1659.
[165] Dong, M., Yang, L.L., Williams, K., Fisher, S.J., et al., A "Tagless" Strategy for Identification of Stable Protein Complexes Genome-wide by Multidimensional Orthogonal Chromatographic Separation and iTRAQ Reagent Tracking. *J. Proteome Res.* 2008, 7, 1836–1849.
[166] Liu, X., Yang, W.-C., Gao, Q., Regnier, F., Toward chromatographic analysis of interacting protein networks. *Journal of Chromatography A* 2008, 1178, 24–32.
[167] Olinares, P.D.B., Ponnala, L., van Wijk, K.J., Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Molecular & Cellular Proteomics* 2010, 9, 1594–1615.
[168] Wu, W., Slastad, H., la Rosa Carrillo, de, D., Frey, T., et al., Antibody Array Analysis with Label-based Detection and Resolution of Protein Size. *Molecular & Cellular Proteomics* 2008, 8, 245–257.
[169] Wessels, H.J.C.T., Vogel, R.O., van den Heuvel, L., Smeitink, J.A., et al., LC-MS/MS as an alternative for SDS-PAGE in blue native analysis of protein complexes. *Proteomics* 2009, 9, 4221–4228.
[170] Heide, H., Bleier, L., Steger, M., Ackermann, J., et al., Complexome Profiling Identifies TMEM126B as a Component of the Mitochondrial Complex I Assembly Complex. *Cell Metabolism* 2012, 16, 538–549.
[171] Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., et al., Towards a knowledge-based Human Protein Atlas. *Nature Biotechnology* 2010, 28, 1248–1250.
[172] Malovannaya, A., Lanz, R.B., Jung, S.Y., Bulynko, Y., et al., Analysis of the Human Endogenous Coregulator Complexome. *Cell* 2011, 145, 787–799.
[173] Poser, I., Sarov, M., Hutchins, J.R.A., Hériché, J.-K., et al., BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature Methods* 2008, 5, 409–415.
[174] Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., et al., Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *The Journal of Cell Biology* 2008, 183, 223–239.
[175] Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., et al., Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *The Journal of Cell Biology* 2010, 189, 739–754.
[176] Collins, M.O., Choudhary, J.S., Mapping multiprotein complexes by affinity purification and mass spectrometry. *Current Opinion in Biotechnology* 2008, 19, 324–330.
[177] Ewing, R.M., Chu, P., Elisma, F., Li, H., et al., Large-scale mapping of human protein-protein interactions by mass spectrometry. *Molecular Systems Biology*

2007, 3, 89.
[178] Werner, J.N., Chen, E.Y., Guberman, J.M., Zippilli, A.R., et al., Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106, 7858–7863.
[179] Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., et al., The proteome of Saccharomyces cerevisiae mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* 2003, 100, 13207–13212.
[180] Ong, S.-E., Mann, M., A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols* 2007, 1, 2650–2660.
[181] Blagoev, B., Kratchmarova, I., Ong, S.-E., Nielsen, M., et al., A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nature Biotechnology* 2003, 21, 315–318.
[182] Cherkasov, A., Hsing, M., Zoraghi, R., Foster, L.J., et al., Mapping the Protein Interaction Network in Methicillin-Resistant Staphylococcus aureus. *J. Proteome Res.* 2011, 10, 1139–1150.
[183] Sowa, M.E., Bennett, E.J., Gygi, S.P., Harper, J.W., Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell* 2009, 138, 389–403.
[184] Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., et al., Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile. *Cell* 2005, 123, 507–519.
[185] Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M.K., et al., Rewiring of Genetic Networks in Response to DNA Damage. *Science* 2010, 330, 1385–1389.
[186] Blagoev, B., Ong, S.-E., Kratchmarova, I., Mann, M., Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature Biotechnology* 2004, 22, 1139–1145.
[187] Kratchmarova, I., Blagoev, B., Haack-Sorensen, M., Kassem, M., Mann, M., Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* 2005, 308, 1472–1477.
[188] Pagliuca, F.W., Collins, M.O., Lichawska, A., Zegerman, P., et al., Quantitative Proteomics Reveals the Basis for the Biochemical Specificity of the Cell-Cycle Machinery. *Molecular Cell* 2011, 43, 406–417.
[189] Tashiro, K., Konishi, H., Sano, E., Nabeshi, H., et al., Suppression of the ligand-mediated down-regulation of epidermal growth factor receptor by Ymer, a novel tyrosine-phosphorylated and ubiquitinated protein. *J. Biol. Chem.* 2006, 281, 24612–24622.
[190] Bisson, N., James, D.A., Ivosev, G., Tate, S.A., et al., Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nature Biotechnology* 2011, 29, 653–658.
[191] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology* 2008, 4.
[192] King, R.W., Deshaies, R.J., Peters, J.M., Kirschner, M.W., How proteolysis drives the cell cycle. *Science* 1996, 274, 1652–1659.
[193] Prakash, S., Tian, L., Ratliff, K.S., Lehotzky, R.E., Matouschek, A., An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat. Struct. Mol. Biol.* 2004, 11, 830–837.
[194] Gsponer, J., Futschik, M.E., Teichmann, S.A., Babu, M.M., Tight regulation of

unstructured proteins: from transcript synthesis to protein degradation. *Science* 2008, 322, 1365–1368.
[195] Gsponer, J., Babu, M.M., Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Rep* 2012, 2, 1425–1437.
[196] De Baets, G., Reumers, J., Delgado Blanco, J., Dopazo, J., et al., An Evolutionary Trade-Off between Protein Turnover Rate and Protein Aggregation Favors a Higher Aggregation Propensity in Fast Degrading Proteins. *PLoS Comput Biol* 2011, 7, e1002090.
[197] Yen, H.-C.S., Xu, Q., Chou, D.M., Zhao, Z., Elledge, S.J., Global protein stability profiling in mammalian cells. *Science* 2008, 322, 918–923.
[198] Legewie, S., Herzel, H., Westerhoff, H.V., Blüthgen, N., Recurrent design patterns in the feedback regulation of the mammalian signalling network. *Molecular Systems Biology* 2008, 4, 190.
[199] Suzuki, H., Forrest, A.R.R., van Nimwegen, E., Daub, C.O., et al., The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.* 2009, 41, 553–562.
[200] Brockmann, R., Beyer, A., Heinisch, J.J., Wilhelm, T., Posttranscriptional Expression Regulation: What Determines Translation Rates? *PLoS Comput Biol* 2007, 3, e57.
[201] Maier, T., Schmidt, A., ell, M.G.U., hner, S.K.U., et al., Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Molecular Systems Biology* 2011, 7, 1–12.
[202] Lu, P., Vogel, C., Wang, R., Yao, X., Marcotte, E.M., Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology* 2006, 25, 117–124.
[203] Pratt, J.M., Petty, J., Riba-Garcia, I., Robertson, D.H.L., et al., Dynamics of protein turnover, a missing dimension in proteomics. *Mol. Cell Proteomics* 2002, 1, 579–591.
[204] Cambridge, S.B., Gnad, F., Nguyen, C., Bermejo, J.L., et al., Systems-wide Proteomic Analysis in Mammalian Cells Reveals Conserved, Functional Protein Turnover. *J. Proteome Res.* 2011, 10, 5275–5284.
[205] Lam, Y.W., Lamond, A.I., Mann, M., Andersen, J.S., Analysis of Nucleolar Protein Dynamics Reveals the Nuclear Degradation of Ribosomal Proteins. *Current Biology* 2007, 17, 749–760.
[206] Auwerx, J., The human leukemia cell line, THP-1: a multifacetted model for the study of monocyte-macrophage differentiation. *Experientia* 1991, 47, 22–31.
[207] Rogers, L.D., Foster, L.J., The dynamic phagosomal proteome and the contribution of the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 18520–18525.
[208] Rappsilber, J., Mann, M., Ishihama, Y., Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* 2007, 2, 1896–1906.
[209] Rigbolt, K.T.G., Vanselow, J.T., Blagoev, B., GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data. *Molecular & Cellular Proteomics* 2011, 10, O110.007450.
[210] Cox, J., Mann, M., 1D and 2D annotation enrichment: a statistical method

integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* 2012, 13, S12.

[211] Cheadle, C., Vawter, M.P., Freed, W.J., Becker, K.G., Analysis of Microarray Data Using Z Score Transformation. *The Journal of Molecular Diagnostics* 2010, 5, 73–81.

[212] Ruepp, A., Waegele, B., Lechner, M., Brauner, B., et al., CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Research* 2010, 38, D497–501.

[213] Gray, K.A., Daugherty, L.C., Gordon, S.M., Seal, R.L., et al., Genenames.org: the HGNC resources in 2013. *Nucleic Acids Research* 2012, 41, D545–D552.

[214] Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., Jones, D.T., Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *Journal of Molecular Biology* 2004, 337, 635–645.

[215] Pfleger, C.M., Kirschner, M.W., The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes & Development* 2000, 14, 655–665.

[216] Jansen, R., Greenbaum, D., Gerstein, M., Relating whole-genome expression data with protein-protein interactions. *Genome Research* 2002, 12, 37–46.

[217] Lundberg, E., Fagerberg, L., Klevebring, D., Matic, I., et al., Defining the transcriptome and proteome in three functionally different human cell lines. *Molecular Systems Biology* 2010, 6, 1–9.

[218] Uversky, V.N., Oldfield, C.J., Dunker, A.K., Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* 2005, 18, 343–384.

[219] Babu, M.M., Kriwacki, R.W., Pappu, R.V., Versatility from Protein Disorder. *Science* 2012, 337, 1460–1461.

[220] Forrest, A.R.R., Kanamori-Katayama, M., Tomaru, Y., Lassmann, T., et al., Induction of microRNAs, mir-155, mir-222, mir-424 and mir-503, promotes monocytic differentiation through combinatorial regulation. *Leukemia* 2009, 24, 460–466.

[221] Asher, G., Reuven, N., Shaul, Y., 20S proteasomes and protein degradation "by default". *Bioessays* 2006, 28, 844–849.

[222] Tasaki, T., Mulder, L.C.F., Iwamatsu, A., Lee, M.J., et al., A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol. Cell. Biol.* 2005, 25, 7120–7136.

[223] Tasaki, T., Kim, S.T., Zakrzewska, A., Lee, B.E., et al., UBR box N-recognin-4 (UBR4), an N-recognin of the N-end rule pathway, and its role in yolk sac vascular development and autophagy. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, 3800–3805.

[224] Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., et al., Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature Biotechnology* 2011, 29, 436–442.

[225] Fang, N.N., Ng, A.H.M., Measday, V., Mayor, T., Hul5 HECT ubiquitin ligase plays a major role in the ubiquitylation and turnover of cytosolic misfolded proteins. *Nat. Cell Biol.* 2011, 13, 1344–1352.

[226] Bhoj, V.G., Chen, Z.J., Ubiquitylation in innate and adaptive immunity. *Nature* 2009, 458, 430–437.

[227] Jiang, H.-Y., Wek, R.C., Phosphorylation of the alpha-subunit of the eukaryotic

initiation factor-2 (eIF2alpha) reduces protein synthesis and enhances apoptosis in response to proteasome inhibition. *J. Biol. Chem.* 2005, 280, 14189–14202.
[228] Moreno, J.A., Radford, H., Peretti, D., Steinert, J.R., et al., Sustained translational repression by eIF2α-P mediates prion neurodegeneration. *Nature* 2012, 485, 507–511.
[229] Ravikumar, B., Vacher, C., Berger, Z., Davies, J.E., et al., Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nat. Genet.* 2004, 36, 585–595.
[230] Komatsu, M., Waguri, S., Chiba, T., Murata, S., et al., Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature* 2006, 441, 880–884.
[231] Hara, T., Nakamura, K., Matsui, M., Yamamoto, A., et al., Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature* 2006, 441, 885–889.
[232] Bence, N.F., Sampat, R.M., Kopito, R.R., Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* 2001, 292, 1552–1555.
[233] Zhong, Z., Wen, Z., Darnell, J.E., Stat3: a STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. *Science* 1994, 264, 95–98.
[234] Sancak, Y., Thoreen, C.C., Peterson, T.R., Lindquist, R.A., et al., PRAS40 is an insulin-regulated inhibitor of the mTORC1 protein kinase. *Mol Cell* 2007, 25, 903–915.
[235] Rappsilber, J., Ishihama, Y., Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* 2003, 75, 663–670.
[236] Davis, J., Goadrich, M., The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning* 2006, 233–240.
[237] Shevchenko, A., Tomas, H., Havli sbreve, J., Olsen, J.V., Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* 2007, 1, 2856–2860.
[238] Argenzio, E., Bange, T., Oldrini, B., Bianchi, F., et al., Proteomic snapshot of the EGF-induced ubiquitin network. *Molecular Systems Biology* 2011, 7, 1–15.
[239] Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., et al., The IntAct molecular interaction database in 2010. *Nucleic Acids Research* 2009, 38, D525–D531.
[240] Stark, C., Breitkreutz, B.-J., Chatr-aryamontri, A., Boucher, L., et al., The BioGRID Interaction Database: 2011 update. *Nucleic Acids Research* 2011, 39, D698–704.
[241] Wang, X., Huang, L., Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Mol. Cell Proteomics* 2008, 7, 46–57.
[242] Balbo, A., Minor, K.H., Velikovsky, C.A., Mariuzza, R.A., et al., Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation. *Proc. Natl. Acad. Sci. U.S.A.* 2005, 102, 81–86.
[243] Jin, J., Smith, F.D., Stark, C., Wells, C.D., et al., Proteomic, Functional, and Domain-Based Analysis of In Vivo 14-3-3 Binding Proteins Involved in

Cytoskeletal Regulation and Cellular Organization. *Current Biology* 2004, 14, 1436–1450.

[244] Lizcano, J.M., Morrice, N., Cohen, P., Regulation of BAD by cAMP-dependent protein kinase is mediated via phosphorylation of a novel site, Ser155. *Biochem. J.* 2000, 349, 547–557.

[245] Levchenko, A., Bruck, J., Sternberg, P.W., Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl. Acad. Sci. U.S.A.* 2000, 97, 5818–5823.

[246] Tai, H.-C., Besche, H., Goldberg, A.L., Schuman, E.M., Characterization of the Brain 26S Proteasome and its Interacting Proteins. *Front Mol Neurosci* 2010, 3.

[247] Vuori, K., Ruoslahti, E., Association of insulin receptor substrate-1 with integrins. *Science* 1994, 266, 1576–1578.

[248] Dell'Angelica, E.C., Klumperman, J., Stoorvogel, W., Bonifacino, J.S., Association of the AP-3 adaptor complex with clathrin. *Science* 1998, 280, 431–434.

[249] Bache, K.G., Raiborg, C., Mehlum, A., Stenmark, H., STAM and Hrs are subunits of a multivalent ubiquitin-binding complex on early endosomes. *J. Biol. Chem.* 2003, 278, 12513–12521.

[250] Schütz, W., Hausmann, N., Krug, K., Hampp, R., Macek, B., Extending SILAC to proteomics of plant cell lines. *The Plant cell* 2011, 23, 1701–1705.

[251] Sury, M.D., Chen, J.-X., Selbach, M., The SILAC fly allows for accurate protein quantification in vivo. *Molecular & Cellular Proteomics* 2010, 9, 2173–2183.

[252] Chan, J.N.Y., Vuckovic, D., Sleno, L., Olsen, J.B., et al., Target Identification by Chromatographic Co-elution: Monitoring of Drug-Protein Interactions without Immobilization or Chemical Derivatization. *Molecular & Cellular Proteomics* 2012, 11, M111.016642–M111.016642.