CHARACTERIZATION OF SIGNAL TRANSDUCTION PATHWAYS BY MS-BASED QUANTITATIVE PROTEOMICS

PHOSPHOTYROSINE INTERACTOMICS AND INSULIN SIGNALING

ABSOLUTE QUANTITATION OF PROTEINS BY MASS SPECTROMETRY

Dissertation

Stefan Hanke

aus

Rosenheim

2008

Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

CHARACTERIZATION OF SIGNAL TRANSDUCTION PATHWAYS BY MS-BASED QUANTITATIVE PROTEOMICS

PHOSPHOTYROSINE INTERACTOMICS AND INSULIN SIGNALING

ABSOLUTE QUANTITATION OF PROTEINS BY MASS SPECTROMETRY

Stefan Hanke

aus

Rosenheim

2008

Table of contents

1   INTRODUCTION

1.1   PHOSPHORYLATION IN CELL SIGNALING

Cells need to react quickly to changes in their environment and to this end they developed mechanisms to translate extracellular and intracellular stimuli into a biochemical code that can be read by the protein machinery. As fundamental requirement this often needs to occur fast and thus cannot only rest upon synthesis of new proteins. A large part of this code is therefore based on posttranslational modifications (PTMs) that are attached to specific sites in proteins. In order to maintain flexibility, many types of PTMs can be removed again. Phosphorylation is one of the most prominent reversible PTMs on proteins and typically occurs on serine, threonine and tyrosine. In prokaryotes, histidine and aspartate are also common amino acids for the attachment of phospho-moieties. Apart from proteins, lipids and sugars can also be phosphorylated, with phosphatidylinositol phosphates and phosphoinositides constituting important docking sites and second messengers, respectively. In agreement with the central and manifold role of phosphorylation, the human genome harbors more than 500 protein kinases and more than 150 protein phosphatases[1, 2].

1.1.1 Phosphorylation as regulatory switch in proteins

Phosphorylation can change the functional state of a protein either by triggering a conformational change or by facilitating or preventing intermolecular interactions. Regulated protein-protein interactions form the basis of cellular signal transduction, and frequently posttranslational modifications such as phosphorylation constitute the molecular switch to enable the association or dissociation of proteins. A phosphorylated motif can serve as a docking site for a specific binding domain. Prominent domains that recognize phosphoserine (pSer) and phosphothreonine (pThr) are the 14-3-3, WW, and FHA domains[3, 4]. 14-3-3 domains prefer motifs with a proline in the +2 position (with respect to the pSer/pThr), and a serine or aromatic amino acid in the -2 position, preceded by an arginine in the -3 or -4 position. WW domains typically recognize short proline sequences like PPXY, PPLP or PPR. Only a subset of WW domains binds to pSer/pThr followed by a proline. FHA domains are dedicated to the recognition of pThr-motifs, and they exploit the +3 position for selectivity. Phosphotyrosine (pTyr) engages separate domain families for recruitment, which will be highlighted in detail below.

1.1.2 Tyrosine phosphorylation

Phosphorylation is a prominent instrument in the toolbox of signaling, and especially tyrosine phosphorylation has attracted much attention in the propagation of signals downstream of receptor tyrosine kinases (RTKs) and cytoplasmic protein tyrosine kinases. The human genome harbors 90 genes for protein tyrosine kinases and 107 genes for protein tyrosine phosphatases[5]. Phosphorylated tyrosine sites serve particularly often as docking interfaces. Intramolecular interactions with a pTyr-binding module such as an SH2 domain can activate the protein (as in the case of Fes) or inactivate it (as in the case of Src and Abl)[6]. In intermolecular interactions, the mere binding to a phosphorylated motif can also modulate the enzymatic activity of the protein in some cases. However, the primary goal of the interaction is usually the effect on the localization of the recruited protein. By increasing its local concentration and scaffolding it together with its upstream or downstream effectors, signaling pathways can be initiated, streamlined and terminated in a concerted manner.

1.1.3 Interaction modules in phosphotyrosine signaling

Domains as independent entities within proteins form one of the pivotal pillars on which the interplay of proteins in eukaryotic cells is built. Approximately 70% of all human proteins contain at least one domain, and each individual domain appears between 10 and 300 times in the proteome. Domains involved in protein-protein interaction typically recognize short, unstructured sequence motifs[7]. From a pharmaceutical point of view, this makes them somewhat better suited for intervention using small molecule drugs than protein-protein interactions involving extended parts of both molecules[8].

There are two domain families that are well-known for their ability to bind to pTyr-containing sequences: the Src-homology 2 (SH2) domain and the Phosphotyrosine-binding (PTB) domain. The genomes of humans and mice contain 120 different SH2 domains in 110 different proteins. Ten of those proteins contain two SH2 domains[9]. PTB domains are represented 56 times in the genome. In addition there is a growing body of evidence for the existence of other domains with pTyr-binding capability.

1.1.3.1 SH2 domains

SH2 domains are made up of approximately 100 amino acids and mediate protein interaction in a plethora of different signaling pathways. Consequently, many mutations in genes of SH2 domains are known that contribute to human disorders, including cancers, diabetes,

developmental disorders, and immunodeficiencies[9, 10]. A large portion of the free binding energy of a pTyr-motif to an SH2 domain is provided by the phospho-moiety itself, laying the groundwork for a strong on/off-switch for binding. Even free phosphotyrosine and phenylphosphate bind relatively well. The other portion, and most importantly also the binding specificity, is contributed by interactions with the residues C-terminal to the phosphotyrosine. As determined largely via degenerate peptide library screening, motifs for SH2 domains typically encompass residues +1 to +4 relative to the pTyr[11]. Some SH2 domains also exploit amino acids at the N-terminal side for binding[11, 12]. In few cases even more extended contacts from -6 to +6 are formed[13, 14]. The SH2 domain of SAP, for example, engages so many residues that already the non-phosphorylated form shows considerable binding, and phosphorylation only enhances the binding 5-fold[15]. The structure of SH2 domains consists of a central β-sheet, flanked by two α-helices. In almost all SH2 domains a conserved arginine is responsible for binding of the pTyr[11]. The general binding mode of SH2 domains directs the sequence of the partner protein perpendicular to the central β-sheet of the SH2 domain in an extended conformation. Therefore the interaction is largely independent of the structural context in the native protein. This allows to study SH2-binding using short synthetic peptides[11]. Typical structures of SH2 and PTB domains and their peptide binding manners are shown in figure 1.
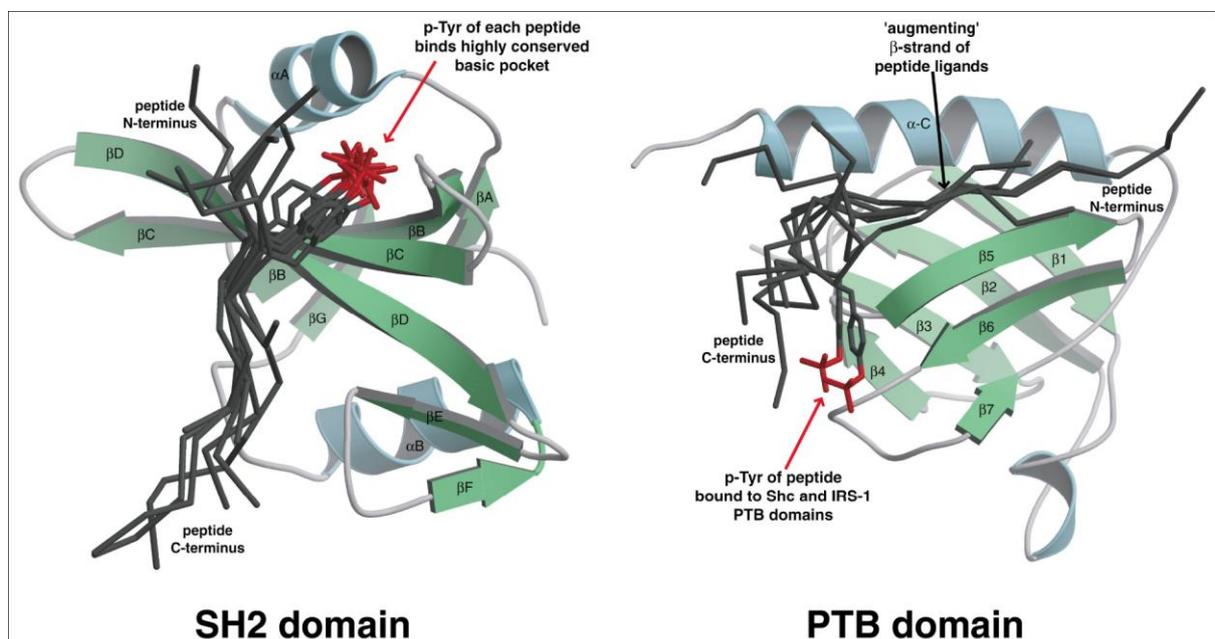
Figure 1: Different modes of peptide recognition by SH2 and PTB domains.
Ribbons representations are shown on the left for the Src SH2 domain [PDB code 1SPS] and on the right for the IRS-1 PTB domain [PDB code 1IRS]. Superimposed in each case are multiple different ligands, to show how they bind with respect to the SH2 or PTB domain fold. For SH2 domains, eight different SH2/phosphopeptide complex structures were overlaid, and for PTB domains six available PTB domain/peptide complex structures were taken. Reproduced from Schlessinger and Lemmon, Science, 2003.

The binding constants of SH2 domains lie in the range of several hundred nanomolar or less[16]. The half-life of SH2-mediated interactions is short, which fulfills the requirements of dynamic signaling[17]. Some SH2 domains may also bind nonpeptide ligands. The SH2 domains of phosphatidylinositol-3-kinase (PI3K) can bind phosphatidylinositol-3,4,5-triphosphate (PI(3,4,5)P_3), which interferes with recognition of pTyr-containing peptides and may thereby provide a feedback inhibition of SH2-phosphoprotein interactions[18]. A few proteins contain two SH2 domains in a tandem setup. This not only increases specificity, but also enhances the affinity up to 1000-fold compared to engaging just one domain[6]. The dendrogram in figure 2 illustrates the relationship between the 120 human SH2 domains based on sequence similarity.

While the major function of SH2 domains consists of their contribution to signal-dependent localization, they sometimes serve additional purposes. Binding of the PI3K regulatory subunit to pTyr-motifs, for example, not only relocates the constitutively associated catalytic domain to the membrane, but also relieves an autoinhibitory configuration. Similarly, the cytoplasmic protein tyrosine kinase Src is maintained in an autoinhibited configuration by intramolecular interactions mediated by its SH2 and SH3 domains. Besides dephosphorylation of the respective tyrosine and binding of the SH3 domain to a different protein, the interaction of the SH2 domain with another protein is a mechanism to activate Src[6].

SH2-containing proteins span a broad range in terms of the various downstream effects they initiate. Figure 3 displays those diverse functions, and at the same time it highlights the frequently observed multidomain architecture. Adaptor proteins like Grb2, Crk and Nck are typically composed of SH2 and SH3 domains, which they use for binding to downstream effectors with proline-rich motifs. While for example the N-terminal SH3-domain of Grb2 primarily binds to Sos, the C-terminal SH3-domain recruits Gab1. Gab1 then becomes phosphorylated, generating binding sites for PI3K. Thus, Grb2 can simultaneously activate the Ras/MAPK-pathway and the PI3K/Akt-pathway. The scaffold protein Shc contains a PTB domain in addition to the SH2 domain, which enables it to interact with two tyrosine-phosphorylated proteins simultaneously. pTyr-sites can also directly influence the phosphorylation status in their surrounding by recruiting kinases (e.g. Fyn, Src) or phosphatases (e.g. PTN11/SHP2). The N-terminal SH2 domain of the tyrosine phosphatase SHP2, for example, contains a loop that binds to the catalytic cleft of the phosphatase and prevents access of substrates. Binding of this SH2 domain to pTyr-motifs releases this autoinhibition. Signal regulation, for example via negative feedback, is carried out by proteins of the SOCS and Cbl families. They attenuate the signals generated at the membrane in response to cytokines, growth factors and hormones through competitive binding and recruitment of E3 ubiquitin ligases which mark proteins for proteasomal degradation. Also at
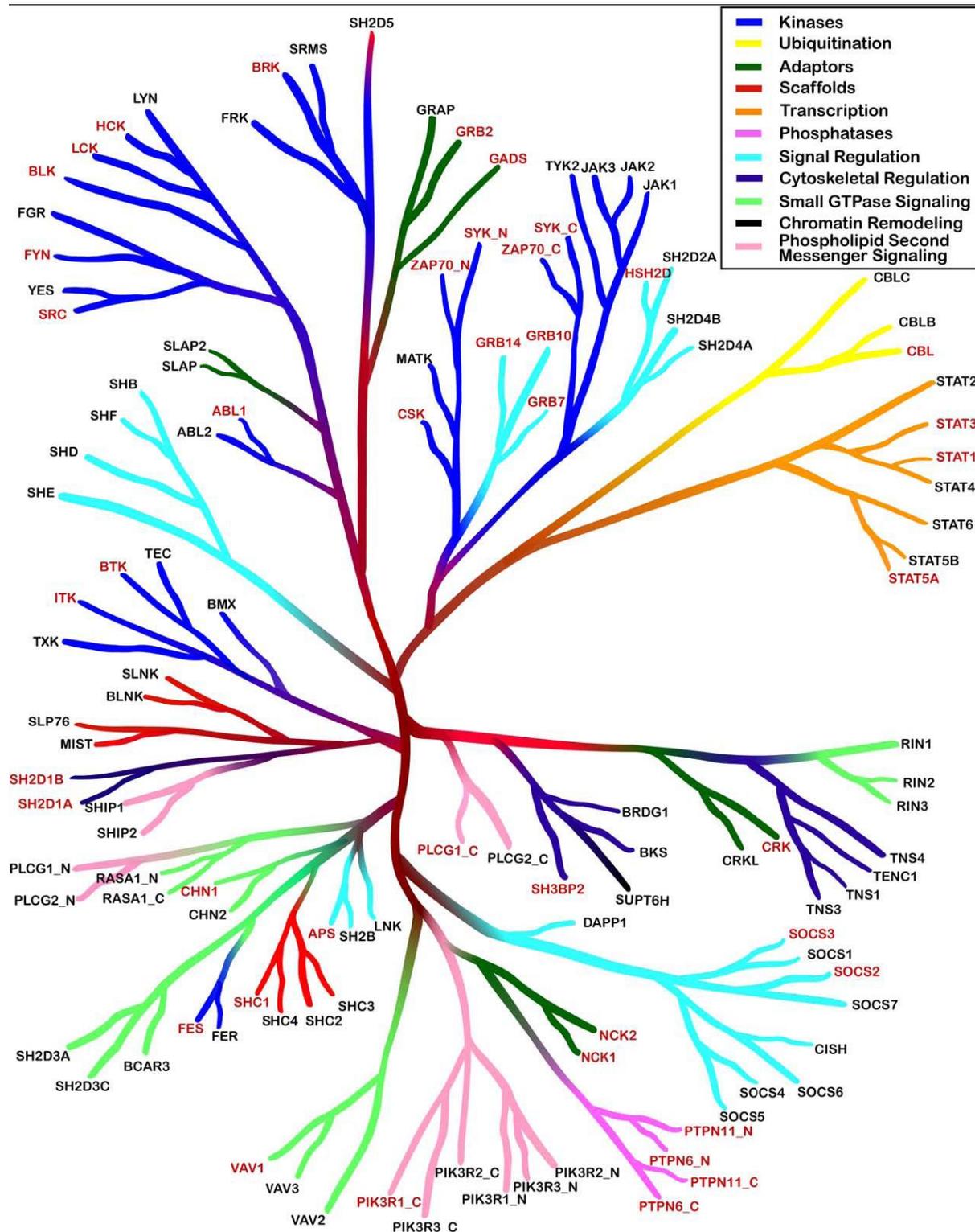
Figure 2: Dendrogram of the Human Complement of SH2 Domains.

The sequence similarity between the 120 known human SH2 domains is visually depicted. The initial branching pattern was built from a neighbor-joining tree derived from a ClustalW protein sequence alignment of the domains. Branches of the tree are colored according to the presumed function of the protein in which each SH2 domain is embedded according to the legend provided. The protein name is indicated in red if one or more structures exist as PDB files. Reproduced from Liu et al, Molecular Cell, 2006.
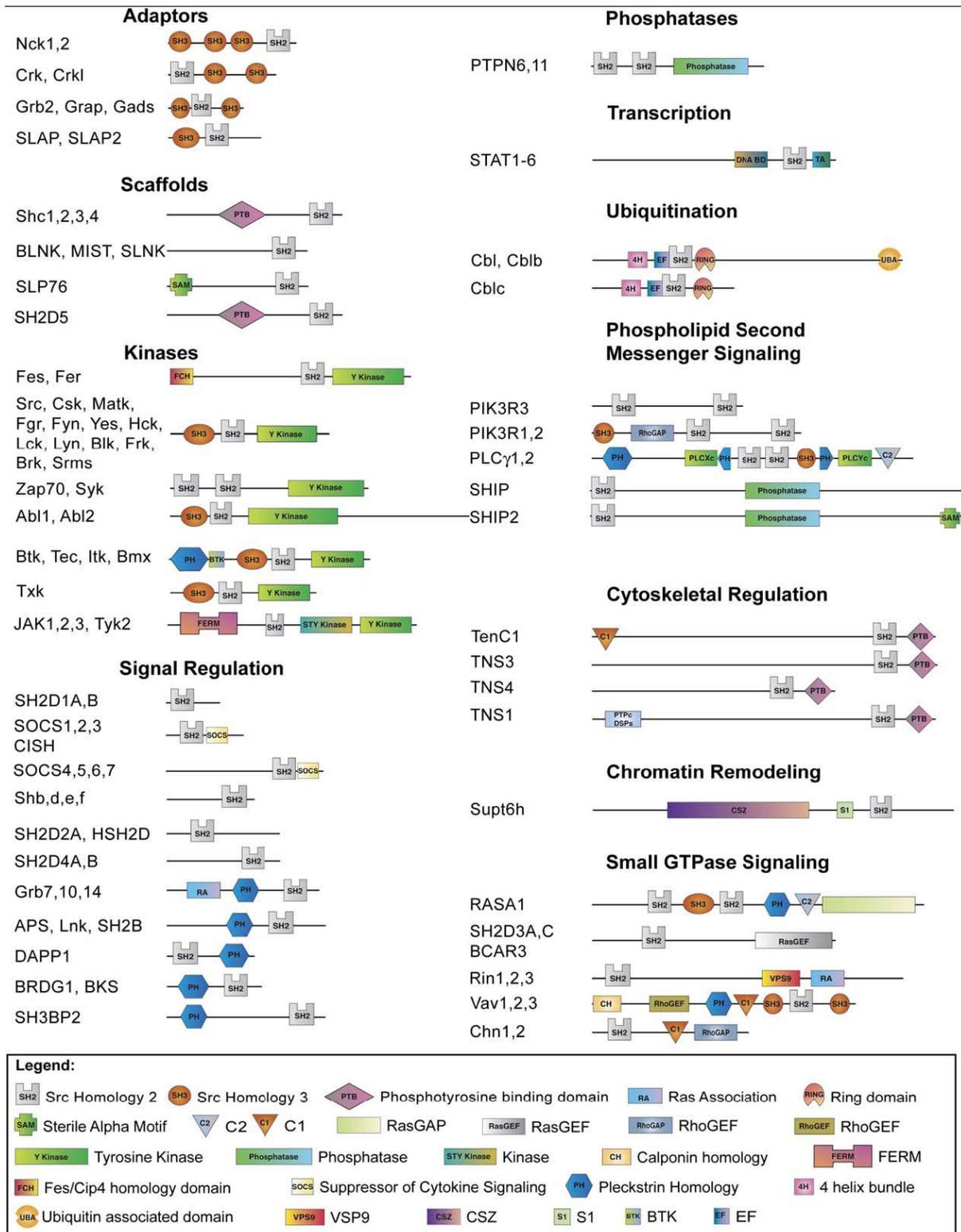
Figure 3: Modular Domain Organization of SH2-containing Proteins.
Classification and domain composition of the 110 nonredundant human SH2 domain-containing proteins identified in human and mouse. Reproduced from Liu et al, Molecular Cell, 2006.

the level of gene transcription SH2-containing proteins can play an immediate role. Certain cytokine receptors activate JAKs, which in turn phosphorylate STATs on tyrosine. Their SH2 domains then mediate mutual intermolecular binding. In this dimerized state they shuttle

into the nucleus and participate in the regulation of gene expression. Phospholipid signaling at the membrane can be modulated by respective kinases (PI3K) and phosphatases (SHIP2), as well as by lipases. Phospholipase Cγ (PLCγ) is recruited via its SH2 domains to pTyr- motifs in RTKs, and additionally via its PH domain to $PI(3,4,5)P_3$, so that those two anchor points attract it to the membrane. It then gets activated by phosphorylation and subsequently cleaves $PI(4,5)P_2$ to diacylglycerol and IP_3. Rearrangements in the cytoskeleton can be mediated by adaptors (Nck), kinases (Csk) and phosphatases (tensines) among SH2-proteins. An example from G-protein signaling, Ras-GAP, displays multidomain cooperation in a particularly profound way as shown in figure 4.
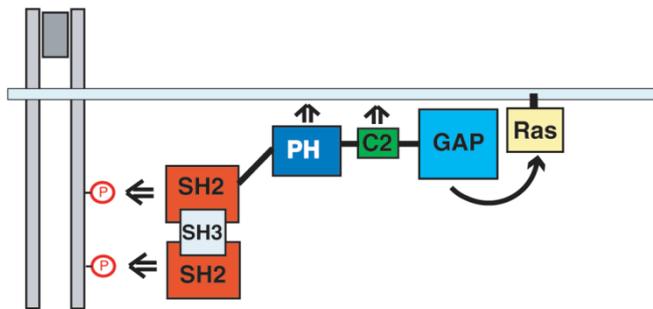
Figure 4: Multidomain cooperation in membrane targeting.
Efficient relocalization of Ras-GAP to the cell membrane is mediated by interactions between its two SH2 domains, PH domain, C2 domain, and perhaps also SH3 domain with tyrosine-phosphorylated receptors and other membrane components. The cooperation of multiple binding events at the cell membrane will position the GAP domain of Ras-GAP in proximity to Ras, which results in stimulation of GTPase activity and consequently silencing of Ras signaling. Reproduced from Schlessinger and Lemmon, 2003.

1.1.3.2 PTB domains

The type of binding of PTB domains can vary considerably and is less conserved than is the case for SH2 domains[19, 20]. The peptide motifs typically adopt a β-turn structure, which is often initiated by Asn and Pro in the -3 and -2 positions. The sequence then constitutes an additional β-strand, a mechanism termed "antiparallel β-sheet augmentation"[21]. The formation of this antiparallel β-strand is the most important characteristic for binding, even more important than the phosphotyrosine. Figure 1 shows the binding mode of PTB domains as compared to SH2 domains. PTB domains are divided into 3 classes: IRS1/DOK-like, Shc-like, and Dab-like. Despite the name PTB, only members of the first two classes do actually bind in a phosphorylation-dependent manner. Those account for just 25% of all PTB domains[6, 19]. Proteins from the DOK (downstream of kinase) and IRS (insulin receptor substrate) families utilize their PTB domains for association with RTKs. Shc-proteins act as scaffolds and participate in the assembly of protein complexes around RTKs.

1.1.3.3 Other phosphotyrosine-binding domains

Recently several reports have appeared that describe pTyr-dependent protein interactions which do not involve SH2 or PTB domains. While in one case this feature could be attributed to a C2 domain[22], the responsible modules are still elusive in other examples[23]. Even though it is still unclear whether those cases represent exceptions or rather more general principles, they certainly pave the way for a more open and flexible view of pTyr-binding domains.

1.2 INSULIN SIGNALING

1.2.1 Insulin signaling in health and disease

A particularly interesting and clinically relevant pathway that involves tyrosine phosphorylation is the insulin signaling network. The peptide hormone insulin is produced by β-cells in the pancreas and secreted into the blood stream to reach its target tissues. It controls large parts of metabolism, especially utilization and storage of molecules related to energy production. Malfunction of insulin signaling therefore has extensive negative effects, and can lead to type I diabetes if insulin is absent and to type II diabetes if insulin resistance cannot be compensated by pancreatic β-cells any longer. Moreover, a plethora of secondary diseases can be triggered once insulin signaling is perturbed in the body.

1.2.1.1 Systemic aspects

Functional specialization of individual tissues is reflected, among other factors, in their differential response to insulin. The main target tissues of insulin are skeletal muscle, liver and adipose tissue, with muscle accounting for more than 70% of glucose disposal[24]. In fact, it is one of the most prominent functions of insulin to promote glucose uptake from the circulation. For this reason, the translocation of the glucose transporter GLUT-4 from intracellular storage vesicles to the plasma membrane is among the most important metabolic effects of this hormone[25]. While the major consequences of insulin in muscle are glucose uptake and glycogen synthesis, the liver additionally reacts with protein and lipid synthesis, and inhibition of hepatic glucose release and VLDL secretion[26, 27].

Diabetes mellitus is a complex disorder that arises from various causes, including dysregulated glucose sensing or insulin secretion (maturity-onset diabetes of the young,

MODY), autoimmune-mediated β-cell destruction (type I diabetes), or insufficient compensation for peripheral insulin resistance (type II diabetes). Type II diabetes accounts for the vast majority of cases. It usually occurs in middle to old age, but is increasingly appearing in young people owing to the close association with obesity[28]. Initially, emerging insulin resistance can usually be counterbalanced by increased insulin secretion from β-cells. If this state is maintained for a long period, however, chronic hyperinsulinemia results in increased insulin resistance and in a decline of β-cell mass and function. Consequently insulin levels drop, which opens out into hyperglycemia and diabetes[29]. Diabetes is not only a problem for glucose homeostasis as such, but often entails a cohort of systemic disorders, like dyslipidemia, hypertension, cardiovascular disease, stroke, blindness, kidney disease, female infertility, and neurodegeneration[28].

While some monogenic forms of diabetes do exist such as maturity onset diabetes in the young (MODY), the common type II diabetes has metabolic causes rather than being primarily a genetic disease, apart from very few exceptions[30]. Along these lines, obesity is closely associated with the development of diabetes. The amount of evidence for a strong link between excess lipids in blood plasma, muscle and liver (so-called "lipotoxicity") on the one hand and reduced insulin sensitivity on the other hand is compelling[24]. The negative effect of obesity has two different reasons: First of all, excess lipids can unintentionally reach other tissues than adipose tissue, where their metabolites trigger insulin resistance. While lipids can serve as competitive substrates for oxidation and therefore reduce glucose utilization, their major negative effect is based on inactivating proteins involved in insulin signaling. Secondly, inflammatory cytokines like TNFα secreted by fat cells or macrophages among them disturb insulin signaling. In both cases the sphingolipid ceramide plays a key role in mediating the inhibitory effects at the molecular level. Ceramide is also involved in many accompanying diseases of diabetes, such as hypertension, arteriosclerosis or cardiac failure[31]. However, the key role of lipids in the development of diabetes is not limited to obese individuals, but also extends to adult onset type II diabetes as well as inherited type II diabetes. This correlation is thought to be based on the reduced numbers and activity of mitochondria in muscle tissue. As a consequence, lipid metabolites accumulate faster and can cause insulin resistance[32]. Furthermore, reactive oxygen species (ROS) can be generated by various stimuli as well as by metabolic imbalance (e.g. high ceramide levels), and subsequently activate certain signaling proteins which are known to downregulate insulin signaling[33].

1.2.1.2 The intracellular signaling network

Binding of insulin to the receptor triggers the spread of the signal inside of target cells. Figure 5 illustrates some of the proteins and pathways involved. IRS (insulin receptor substrate) proteins organize the distribution of the signal to a large extent. Among the most prominent pathways that are initiated is the mitogenic channel leading to activation of mitogen-activated protein kinase (MAPK). The adaptor protein Grb2 recruits Sos to the
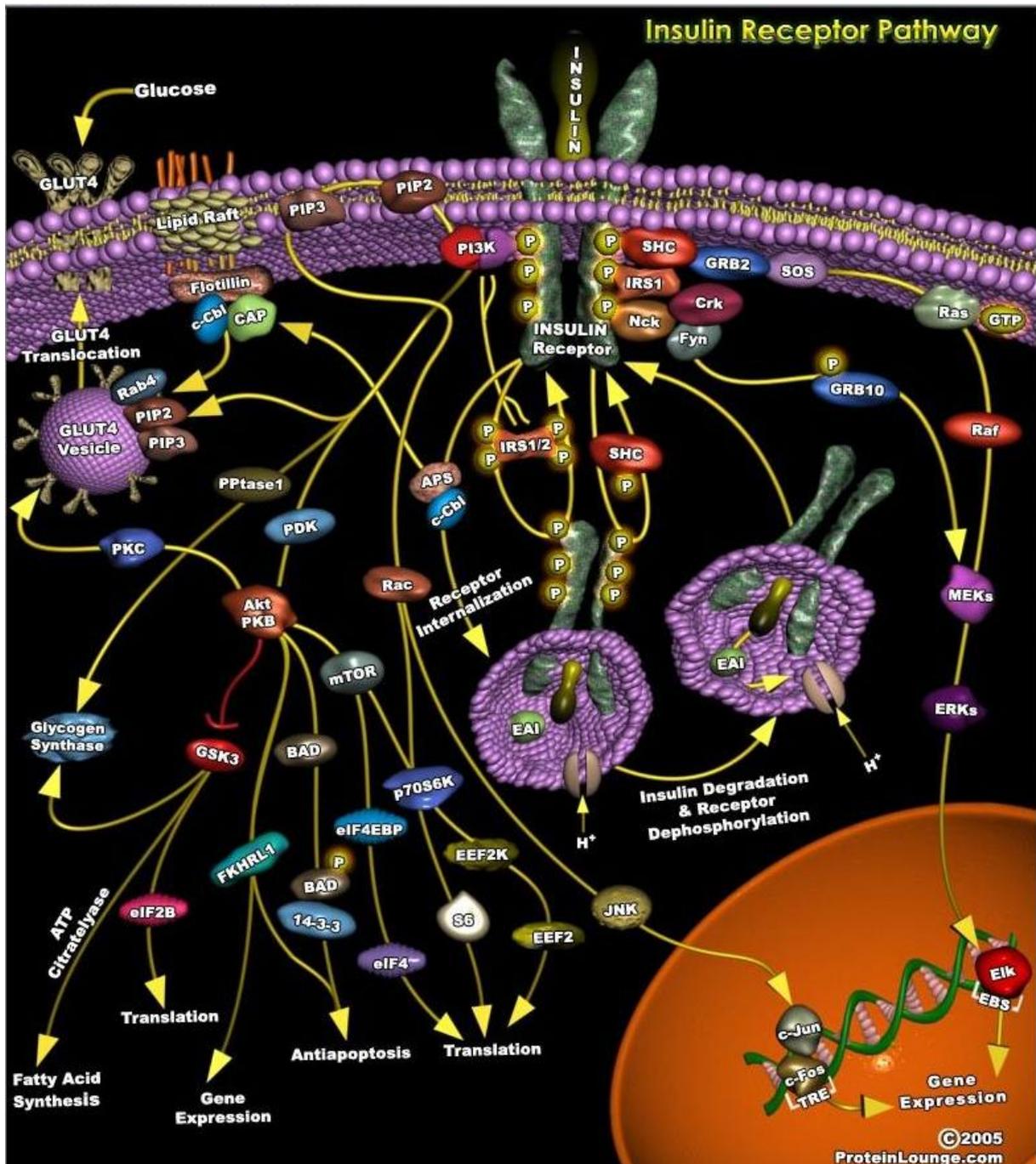
Figure 5: The insulin receptor pathway.

This schematic representation of the pathway contains many of the proteins involved in insulin signaling. Various components in the cytoplasm, the plasma membrane, vesicles and the nucleus are engaged to carry out the different effects of insulin. Obtained from proteinlounge.com

membrane, where it acts as guanine nucleotide exchange factor for the small G-protein Ras. Activated Ras in turn triggers the activation of the canonical cascade from Raf via MEK to ERK, which then initiates several mitogenic actions. In parallel, the insulin stimulus affects various metabolic routes, with glucose disposal being the most central effect. The rate-limiting step for glycogen synthesis is glucose uptake into cells. After entering the cell, glucose is phosphorylated by hexokinase and either stored as glycogen or oxidized to generate ATP[24]. While the transmembrane protein GLUT-1 carries out basal glucose import, GLUT-4 is responsible for insulin-induced glucose uptake. Insulin-stimulated activation of PI3K is closely linked to the translocation of GLUT-4 from intracellular vesicles to the plasma membrane. However, there is evidence for an alternative pathway besides the one triggered by insulin leading to the translocation of GLUT-4 to the cell surface. Muscle contraction by exercise has been shown to activate glucose transport and compensate for insulin resistance by triggering a PI3K-independent mechanism of GLUT-4 translocation. This has been proposed to function via 5′-AMP-activated kinase or increased cytoplasmic Ca^{2+}-concentrations. In addition to this, exercise has been demonstrated to result in increased expression of proteins involved in glucose metabolism as well as in enhanced insulin sensitivity[34]. Interestingly, different stimuli mobilize different pools of GLUT-4. This glucose transporter circulates in a complex fashion between trans-golgi-network (TGN), endosomes, GLUT-4 storage vesicles (GSVs), and the cell membrane. Under basal conditions, 30-40% of GLUT-4 resides in endosomes where it is not responsive to insulin. While insulin triggers the release from GSVs, physical exercise targets GLUT-4 in endosomes independently of PI3K. GSVs contain the v-snare VAMP2, and the cell membrane of muscle and fat cells harbors particularly high amounts of the t-snares syntaxin-4 and SNAP23. The complex of those three proteins mediates the exocytosis of GSVs[25].

Besides GLUT-4 translocation, PI3K promotes glycogen synthesis, antilipolysis, fatty acid synthesis, acetyl-CoA-carboxylase activity, synthesis of protein and DNA, growth and differentiation. This all relates to the PI3K-catalyzed generation of $PI(3,4,5)P_3$ in the membrane, which recruits proteins with PH domains and allows for their targeted activation. Those include AGC family protein kinases (like Akt/PKB), PKCs and SGK[35]. One of the best known examples is PDK1, which is responsible for the activation of Akt and atypical PKCs. Binding of Akt to $PI(3,4,5)P_3$ makes two sites accessible for phosphorylation through PDK1 and the mTOR/Rictor-complex, respectively. Phosphorylation of both sites is required for full activation of the enzyme. Akt then serves as a central signal distributor: It enhances the import of glucose and amino acids, it promotes the incorporation of incoming nutrients into glycogen, proteins and triacylglycerol, it inhibits apoptosis and stimulates translation of mRNA, and it regulates the transcription factors FOXO1 and FOXA2[31]. Phosphorylation of FOXO1 by Akt prevents its entry into the nucleus and thereby reduces the transcription of

genes involved in gluconeogenesis. FOXA2 is also inactivated by phosphorylation, which reduces the transcription of genes involved in fatty acid oxidation. Akt-mediated phosphorylation of the nutrient sensor mTOR promotes protein synthesis and growth[36]. While insulin stimulates protein synthesis in general, it especially promotes synthesis of growth-related proteins. This mechanism is carried out downstream of PI3K by phosphorylation of translation initiation factors like eIF4E and PHAS-I[37]. Specifically, mTOR/S6K mediates the enhancement of mRNA-translation of growth-regulated proteins, while PKCδ takes part in promoting protein synthesis in general[38]. However, several kinases including mTOR also exert negative feedback upstream of PI3K, which is important to restrain the cancerogenic potential of PI3K-signaling[39]. Negative regulation of insulin signaling is largely carried out through inhibitory phosphorylation on serine and threonine residues. However, these amino acids can also become modified by O-linked N-acetylglucosamine (O-GlcNAc). The donor molecule UDP-GlcNAc originates from hexosamine biosynthesis and correlates with the level of glucose. It therefore reflects the metabolic status, and the PTM O-GlcNAc is used as nutrient sensor. O-GlcNAc transferase binds via a phosphoinositide-binding domain to $PI(3,4,5)P_3$. Subsequently O-GlcNAc gets attached to the insulin receptor, IRS-1 and Akt. While this does not reduce the level of tyrosine phosphorylation on these proteins directly, it promotes phosphorylation on inhibitory serine sites in IRS-1 and thereby downregulates insulin signaling[40, 41].

Recent studies have shed light on the molecular details behind the negative effect of obesity-related factors on insulin sensitivity. Many of the triggers of insulin resistance like excess lipids[24, 31, 32], inflammatory cytokines[31] or reactive oxygen species[33], exert their undesired effect through activation of multiple kinases that phosphorylate insulin receptor substrate proteins (IRSs) on serine residues[24]. The fact that many different serine phosphorylation sites contribute to the inactivation of IRS proteins permits fine-tuning of the inhibition depending on the extent of the activation of the respective upstream kinases. Mitochondrial dysfunction is commonly observed along with the insulin resistant state, but it is still a matter of debate whether it is a cause or rather a consequence of insulin resistance. It is obvious, however, that defective mitochondria are highly detrimental for lipid and carbohydrate metabolism, and consequently enhance insulin resistance[42, 43]. Pancreatic β-cells have a special way to react to elevated lipid levels in the circulation. They express a receptor for free fatty acids (FFAs) called GPR40, which is responsible for both the acute stimulatory as well as the chronic inhibitory effect of FFAs on insulin secretion. The inhibitory effect is carried out via inhibition of transcription and translation of the insulin gene, expression changes of enzymes involved in glucose and fatty acid metabolism, opening of K^+-channels, uncoupling of the respiratory chain and thus decreasing ATP levels, and ultimately by promoting apoptosis[31].

1.2.2 The insulin receptor family

Insulin receptor (InsR), insulin-like growth factor 1 receptor (IGF1R) and insulin receptor-related receptor (IRR) are highly homologous cell surface receptors with distinct but partially overlapping functions. Although InsR and IGF1R share high similarity in sequence and structure, and as a result of this activate a comparable spectrum of signaling pathways, the IGF1R preferentially controls developmental and growth-related processes whereas the insulin receptor is a major player in maintenance of metabolic homeostasis[44]. Nevertheless, they contribute to each other's duties to some extent[45]. Experiments with chimeric receptors have shown that it is mainly the intracellular part that determines specificity, not the extracellular ligand binding domains[46-48]. Despite a similar degree of homology, the IRR is still an orphan receptor with obscure function, and its expression is limited to few tissues[49, 50]. It harbors the greatest differences compared to the other receptors in the C-terminal part, where it also lacks the last 50 amino acids.

The receptors are receptor tyrosine kinases (RTKs) and form dimers consisting of two α-chains and two β-chains. The extracellular α-chains bind their ligand, which triggers a conformational change in the intracellular part of the β-chains and activates the kinase domains. Full activation of the kinase is accomplished after mutual phosphorylation of the kinase domains on three tyrosines. Subsequently, additional tyrosines in the receptors become phosphorylated, and recruit various proteins that bind via their SH2 and PTB domains. Those can then become phosphorylated as well and propagate the signal within the cell. The possible formation of heterodimers generates hybrid receptors and can modulate the signal output. Alternative splicing of one exon in the α-chain of the InsR yields two different versions, InsR-α and InsR-β. InsR-α lacks exon 11 and is expressed mainly during development and later in a few specialized tissues, whereas InsR-β is the predominant version in the body. Figure 6 shows the ligand binding preferences of the different receptors. InsR-α binds insulin with moderate affinity, InsR-β with high affinity. Under normal circumstances insulin does not bind to the IGF1R, but during insulin resistance circulating insulin concentrations can rise high enough to activate it[51].
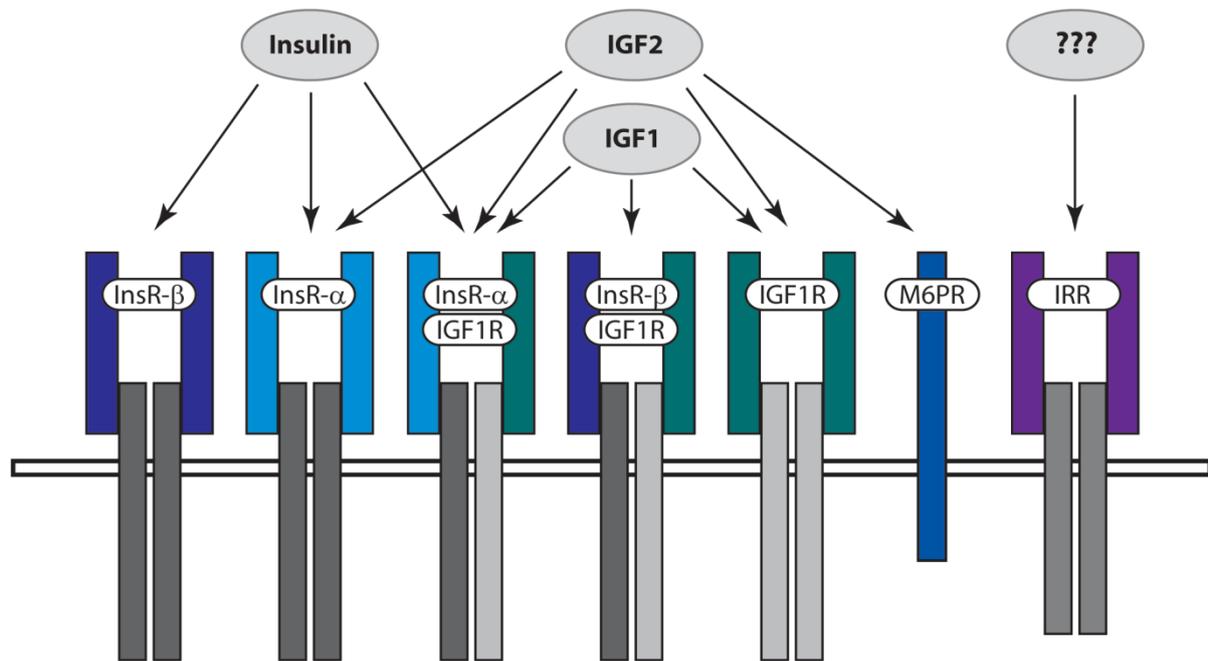
Figure 6: The insulin/IGF family and their receptors.

The insulin and IGF1 receptors can form hybrid receptors with individual binding properties. There are no reports so far on the ability of the IRR to integrate into hybrid receptors. The ligand(s) of the IRR are still not known. The IGF2 receptor is in fact the mannose-6-phosphate receptor (M6PR) responsible for transport of lysosomal enzymes from the Golgi complex and the cell surface to lysosomes. It can also bind IGF2, but it targets the hormone for degradation instead of signaling.

1.2.3 The insulin receptor substrate (IRS) family

A complex and still incompletely understood network of signaling proteins connects the insulin receptor family with its downstream effectors. In this context, IRSs play an indispensable role as interaction platforms that become phosphorylated on multiple tyrosines by the receptors and subsequently attract various signaling proteins to spread the stimulus within the cell. All of the 6 members of the IRS-family contain a PH domain and a PTB domain at the N-terminus, which they employ for binding to membranes and to NPXpY-motifs in receptors, respectively. However, they vary in the length of their C-terminal part. IRS-1 and IRS-2 are large proteins with many tyrosine phosphorylation sites and are expressed in many different tissues. IRS-3 is short and only present in rodents. IRS-4 has similar dimensions like IRS-1 and IRS-2, but is only expressed in very few and specialized tissues. IRS-5 and IRS-6 (also termed DOK-4 and DOK-5) possess a very short C-terminal part and consequently very few potential phosphotyrosine motifs[52, 53]. Figure 7 highlights the role of IRSs as interaction platforms and signal distributors.
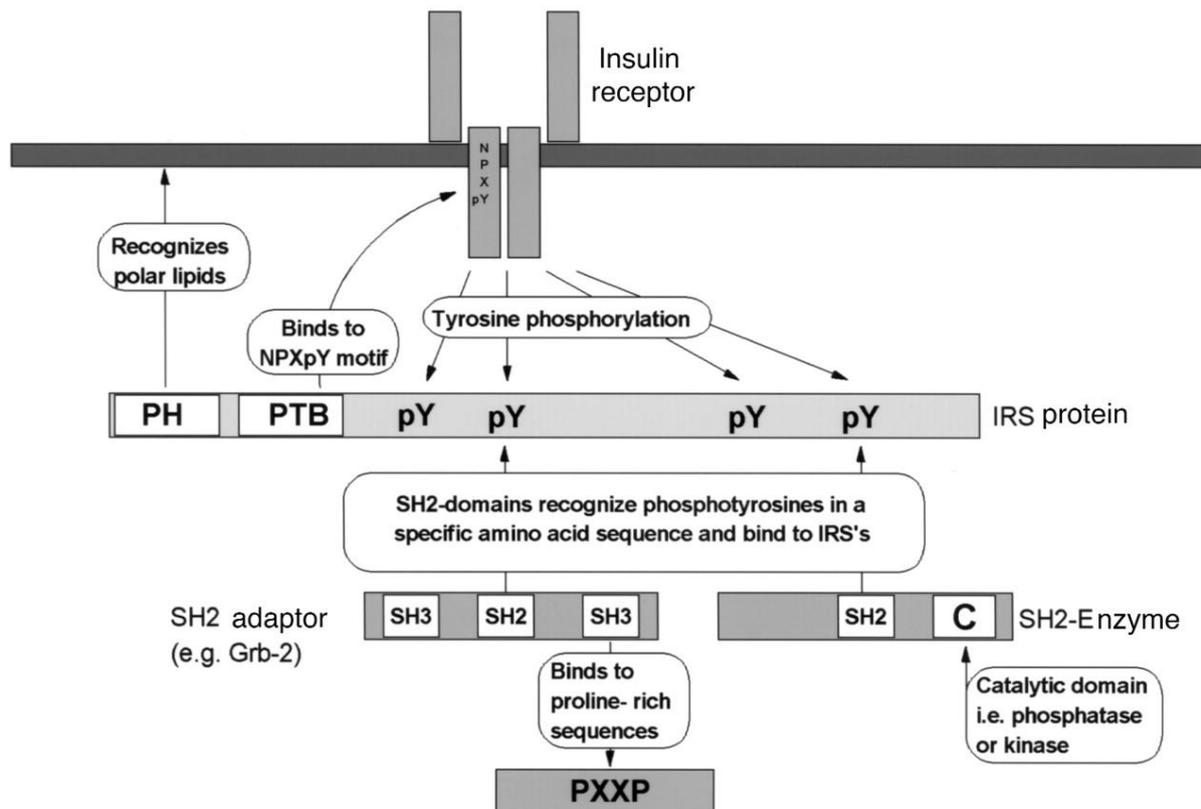
Figure 7: IRS proteins as interaction platforms connecting receptors to downstream signaling proteins. The PH and PTB domains anchor IRSs to their receptors. Subsequent phosphorylation on tyrosines generates docking sites for proteins with pTyr-recognizing domains. Adapted from Virkamäki et al, Journal of clinical investigation, 1999.

1.2.4 The pivotal role of IRS-1 and IRS-2 in insulin signaling

IRS-1 and IRS-2 gained their fame due to their unique role in mediating signals from members of the insulin receptor family, but also other receptors engage them for their purposes. Those include the prolactin[54], androgen[55], growth hormone[54, 56] and vascular endothelial growth factor (VEGF) receptors[57], as well as members of the integrin receptor family[58, 59] and several cytokine receptors[60]. These IRS-proteins are roughly 135 kDa in size, and represent major players in conducting the distribution of the insulin signal within the cell. They contain 34 and 36 tyrosines, respectively, and many of these can potentially become phosphorylated and subsequently serve as docking sites for a large variety of signaling proteins. Their ability to serve as scaffolds for the concerted assembly of many different kinases, phosphatases, adaptor proteins, and other kinds of signaling proteins is substantial. By being interaction platforms at the top of the signaling network, they can be seen as organizers at whose level the decision is taken which of the various possible downstream pathways will be activated to which extent. Apart from their role in metabolic signaling, they propagate proliferative and anti-apoptotic signals and are consequently overexpressed or activated more strongly in most cancers[61]. In their activated form they tend to associate with

cytoskeletal structures[62, 63], and IRS-1 has even been reported to enter the nucleus following stimulation via the IGF1R[64] or the androgen receptor[55]. They share 75% amino acid sequence identity in their N-terminal domains and 35% in their C-terminal part. Despite their high degree of homology and many similar tyrosine-phosphorylation motifs, studies in knockout mice and knockout cell lines indicate that these two IRS proteins serve complementary, rather than redundant, roles in insulin and IGF-1 signaling[65]. In general, IRS-1 plays a prominent role in growth, while IRS-2 is the main contributor for glucose homeostasis and is decisive for proper function of pancreatic β-cells[66]. Knockout of either of them results in insulin resistance, but only IRS-2-knockout leads to diabetes. The picture gets complicated by many tissue-specific differences between IRS-1 and IRS-2[67]. In muscle, IRS-1 is more closely associated with glucose uptake, whereas IRS-2 feeds the MAP-kinase pathway[68, 69]. In liver, both of them are involved in metabolic regulation, but IRS-2 has a more pronounced role in lipid metabolism and in confining hepatic glucose release into the circulation[70-72]. Adipose tissue engages IRS-1 mainly for differentiation while IRS-2 serves insulin-stimulated glucose uptake[27]. In metastatic breast cancer, IRS-1 is frequently inactivated, and compensatory overactivation of IRS-2 has been shown to enhance signaling via Akt and mTOR. This triggers increased translation of oncogenes, growth and survival signals, elevated cell motility and invasiveness, and angiogenesis[73].

In agreement with the unequal roles of IRS-1 and IRS-2, their phosphorylation kinetics are markedly different from each other. As investigated in skeletal muscle cells and adipocytes, IRS-2 is dephosphorylated more rapidly on tyrosines than IRS-1[63, 74].

Due to their central role, large parts of the regulation and fine-tuning of insulin signaling take place at the level of IRS proteins. Phosphorylation of IRS proteins on serine and threonine residues is the main „molecular handle' to exert negative feedback and set a limit to the activation of downstream pathways. In the basal state their phosphorylation level is already rather high, and insulin stimulation augments it even further. Those phosphorylation events can induce the dissociation of IRS proteins from the receptor, hinder tyrosine phosphorylation of IRS proteins, release IRS proteins from intracellular complexes that maintain them in close proximity of the receptor, or induce their degradation[24].

Kinases that carry out those inhibitory phosphorylations include mTOR (mammalian target of rapamycin), IKK-β (inhibitor kB kinase), JNK (c-Jun NH2-terminal kinase), ERK (extracellular signal-regulated kinase), S6K (ribosomal S6 kinase), PKCδ (protein kinase Cδ), PKCθ, PKCδ, GSK3 (glycogen synthase kinase 3), SIK2 (salt-inducible protein kinase 2) and casein kinase II[24, 39]. JNK even has a dedicated binding site in IRS-1 which is independent of tyrosine phosphorylation[75]. Nevertheless, some serine phosphorylations on IRS proteins can have a positive effect on insulin signaling, for example because they protect them from the

action of tyrosine phosphatases[76]. When talking about negative regulation, feedback inhibition through activation of those kinases following an insulin stimulus is just one aspect. Many of them can just as well be activated in a more constitutive way by common inducers of insulin resistance. In this respect, free fatty acids, diacylglycerol, acyl-CoA, glucose and TNFα contribute to their mobilization[52]. It appears that a large part of this effect is mediated through the generation of ceramide[77]. Similar to the activation of IRS proteins, their desensitization can also occur in a tissue- and IRS-specific way. In muscle, PKCθ exerts this function and primarily targets IRS-1, whereas in liver PKCε is a key effector and primarily phosphorylates IRS-2[32]. Figure 8 depicts the regulation of IRS activity via phosphorylation on serine and threonine residues.
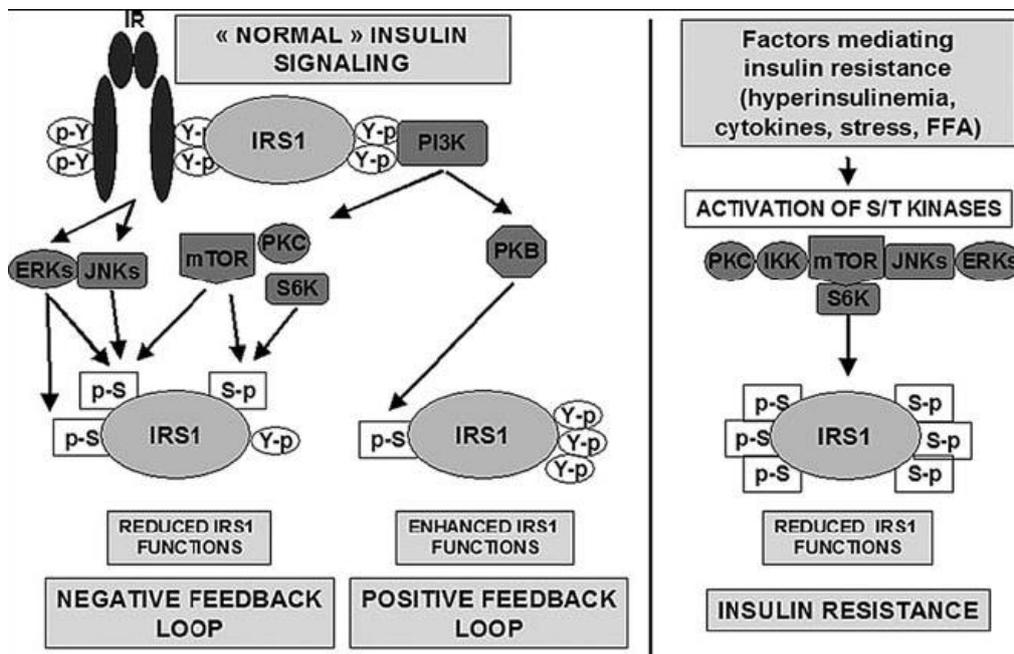
Figure 8: Phosphorylation of IRS proteins on serine and threonine residues regulates their functions. Phosphorylation of IRS proteins on serines and threonines usually occurs in the context of negative feedback inhibition. Several factors like free fatty acids, inflammatory cytokines and other stress factors exploit this mechanism to induce insulin resistance via activation of Ser/Thr-kinases. In a few cases serine phosphorylation can have a positive effect, for example through Akt/PKB. Reproduced from Gual et al, Biochimie, 2005.

1.3 MASS SPECTROMETRY-BASED QUANTITATIVE PROTEOMICS

1.3.1 Modern mass spectrometry applied to protein research

When studying the molecular components of life, a researcher can choose among a plethora of different techniques to pursue his or her endeavor. Typically, several methods are applied in combination to exploit synergies. Mass spectrometry (MS) is in a special position as compared to many other approaches due to several reasons. First, MS is an unbiased technique and therefore able to efficiently investigate mixtures of unknown compounds. In contrast to antibody-based approaches, for example, it is not constrained by hypotheses, which makes it particularly well-suited for novel discoveries[78]. Second, MS has a high throughput and can provide data for thousands of proteins per day. Accordingly, mass spectrometry has become the basis of proteomics as a new discipline that approaches biological questions with a global view. And third, proper application of MS results in excellent sensitivity, accuracy and reliability of data. It can measure proteins at very low levels that are otherwise only detectable by targeted immunoassays. The application of tandem mass spectrometry (MS/MS) yields sequence information on top of the accurate mass of the intact molecule, which supports unambiguous identification[79].

The typical workflow in MS-based proteomics is visualized in figure 9. Protein samples are digested with a protease before analysis to generate peptides that are amenable to straightforward mass spectrometric measurement. Information obtained at the peptide level can later be mapped back to their parent proteins during data analysis.
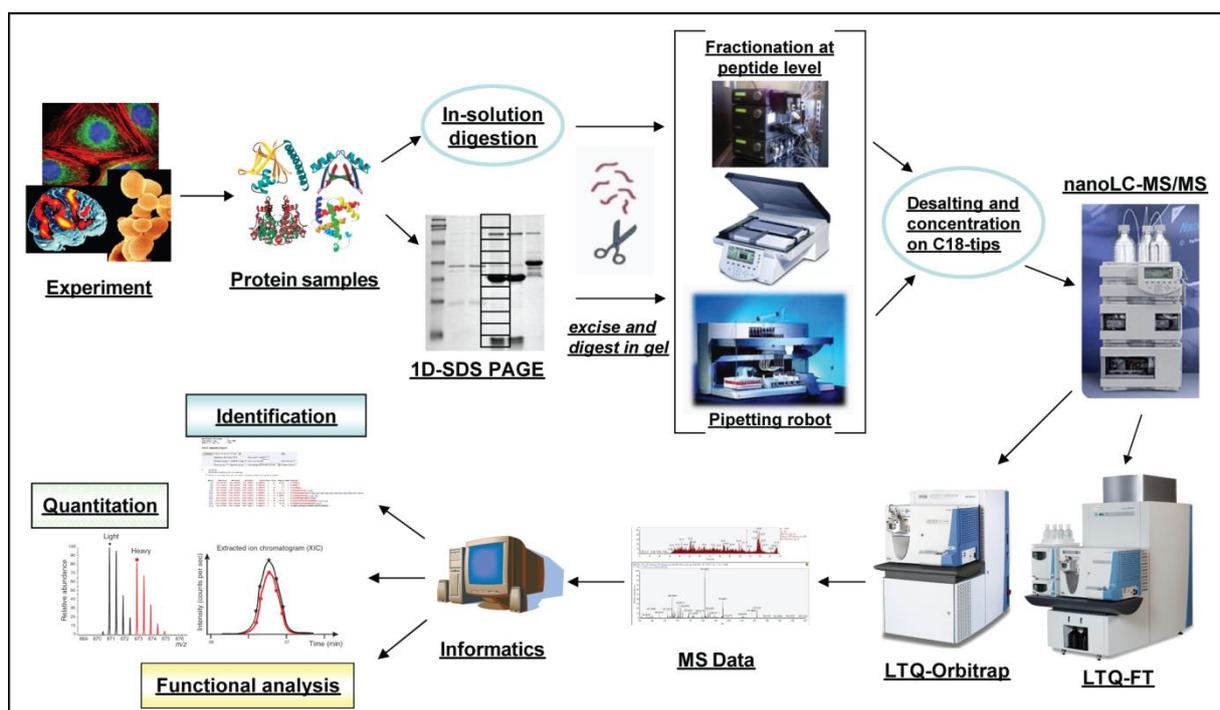
Figure 9: Workflow in MS-based proteomics.

Starting out with the initial biological experiment, proteins are extracted and fractionated as desired. Enzymatic digestion yields peptides that can be fractionated further and analyzed by MS. In most modern setups, samples are introduced into the mass spectrometer through online nanoscale reversed phase liquid chromatography.

1.3.2 Instrumentation

Mass spectrometers measure the mass-to-charge ratio of analytes. In order to equip each molecule with charges, samples are typically introduced into the machine in an acidic environment. An ion source is required to transfer the ions into the gas phase before they can enter the mass spectrometer. To this end, either MALDI (matrix-assisted laser desorption/ionization) or ESI (electrospray ionization) is employed as means to vaporize peptides in a mild way without destroying them. In MALDI, analytes are co-crystallized in a matrix, from where they are volatized by a laser beam[80]. ESI, on the contrary, generates gaseous ions through a strong difference in the electric potential between spray tip and orifice of the mass spectrometer. Tiny liquid droplets evaporate quickly after leaving the tip, increasing charge repulsion of the retained ions, which leads to further dispersion and ultimately to single ions free of solvent[81]. In the study described here, an HPLC-system was directly coupled to the mass spectrometer ("online" LC-MS), supplying peptide ions via reversed phase chromatography and nanoelectrospray. A picture of the ion source is given in figure 10.
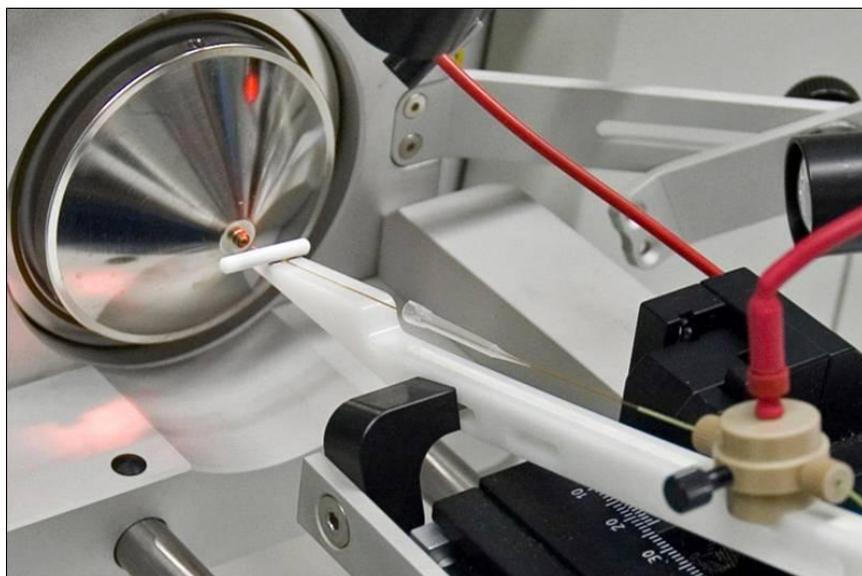
Figure 10: Nanoelectrospray ion source in an online LC-MS configuration.

The analytes are separated by reversed phase chromatography on a thin column (inner diameter 75 μm), and are directly transferred into the orifice of a mass spectrometer via nanoelectrospray throughout the whole LC-gradient.

The types of mass spectrometers used in this study are briefly described in the following section.

1.3.2.1 Linear ion trap

As its name already implies, an ion trap mass spectrometer is able to accumulate and confine ions. In addition, mass spectra can be acquired by linearly increasing the potential of electrodes, which leads to the ejection of ions from the potential well depending on their mass-to-charge ratio (m/z)[82]. Ion traps are known for their high sensitivity and speed. Linear ion traps are superior to 3D ion traps because they offer higher ion storage capacity and trapping efficiency. Drawbacks of ion traps are relatively low resolution and mass accuracy. These characteristics make them ideally suited as part of hybrid instruments. When fusing an ion trap together with a mass spectrometer that features high resolution and mass accuracy, the advantages of each machine can be intelligently exploited. In this study, hybrid instruments were employed that contain either an FT-ICR or an Orbitrap mass spectrometer following a linear ion trap.

The division of labor between the two consecutive mass spectrometers is as follows: The ion trap accumulates defined amounts of ions and passes them on to the FT-ICR or Orbitrap for high accuracy mass measurement. Based on the acquired mass spectrum, data dependent decisions are taken, typically about which ions shall be fragmented to gain sequence information. For fragment ion spectra, high mass accuracy is less critical. However, speed and sensitivity are very important to keep scan cycle times short. The ion trap fulfills these criteria, and parallel operation of ion trap MS/MS during full scan acquisition in the FT-ICR or Orbitrap further improves the performance with respect to the time invested. Fragment ion spectra are generated by collisionally induced dissociation (CID). The selected precursor ion is isolated and excited via an electric field, which increases its internal energy. Collision with inert gas molecules can then break the ion at one of the peptide bonds. Since the cleavage occurs semi-randomly, a fragment spectrum usually contains many different ions. Figure 11 shows the structure of a peptide with cleavage sites for CID. Calculation of the distance between two consecutive fragment ions in the mass spectrum yields the corresponding amino acid mass and therefore allows sequence assignments[83].
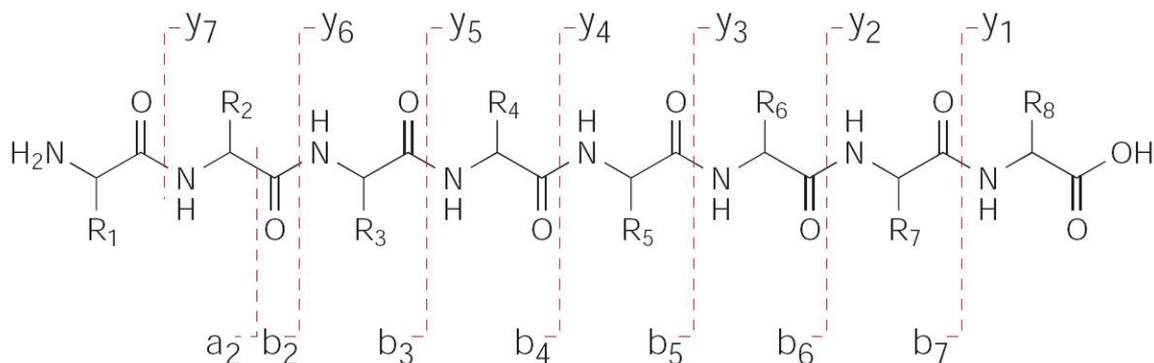
Figure 11: Peptide structure with sites for fragmentation by CID.
b-ions originate if the charge is retained by the C-terminal fragment, and y-ions if the charge is retained by the N-terminal fragment.

1.3.2.2 FT-ICR

The cyclotron detection mode of Fourier transform ion cyclotron (FT-ICR)[84, 85] mass spectrometers enables unmatched resolution and mass accuracy. Whereas in other mass spectrometers ions are filtered in a magnetic or electric field or selected by flight time, FT-ICR instruments detect them by their resonance frequency. ICR spectrometry is based on the principle of cyclotron motion in a uniform electric field. Ions are detected in a cyclotron cell, which is located inside a super-conducting magnet with a fixed field strength.

Figure 12 contains a scheme of a hybrid linear ion trap-FTICR mass spectrometer and of an ICR cell. Ions arriving inside the cyclotron cell are forced into an orbit by the uniform magnetic field. The front and end plates work as trapping plates to trap the ions inside the cyclotron cell. The two excitation plates are connected with a radiofrequency (RF) transmitter to excite the ions, and the detection plates register the induced current of the ions for detection. The transient signal from all ions is digitized and processed with a fast FT algorithm, resulting in a mass spectrum with ion abundance versus mass-to-charge ratio (m/z).
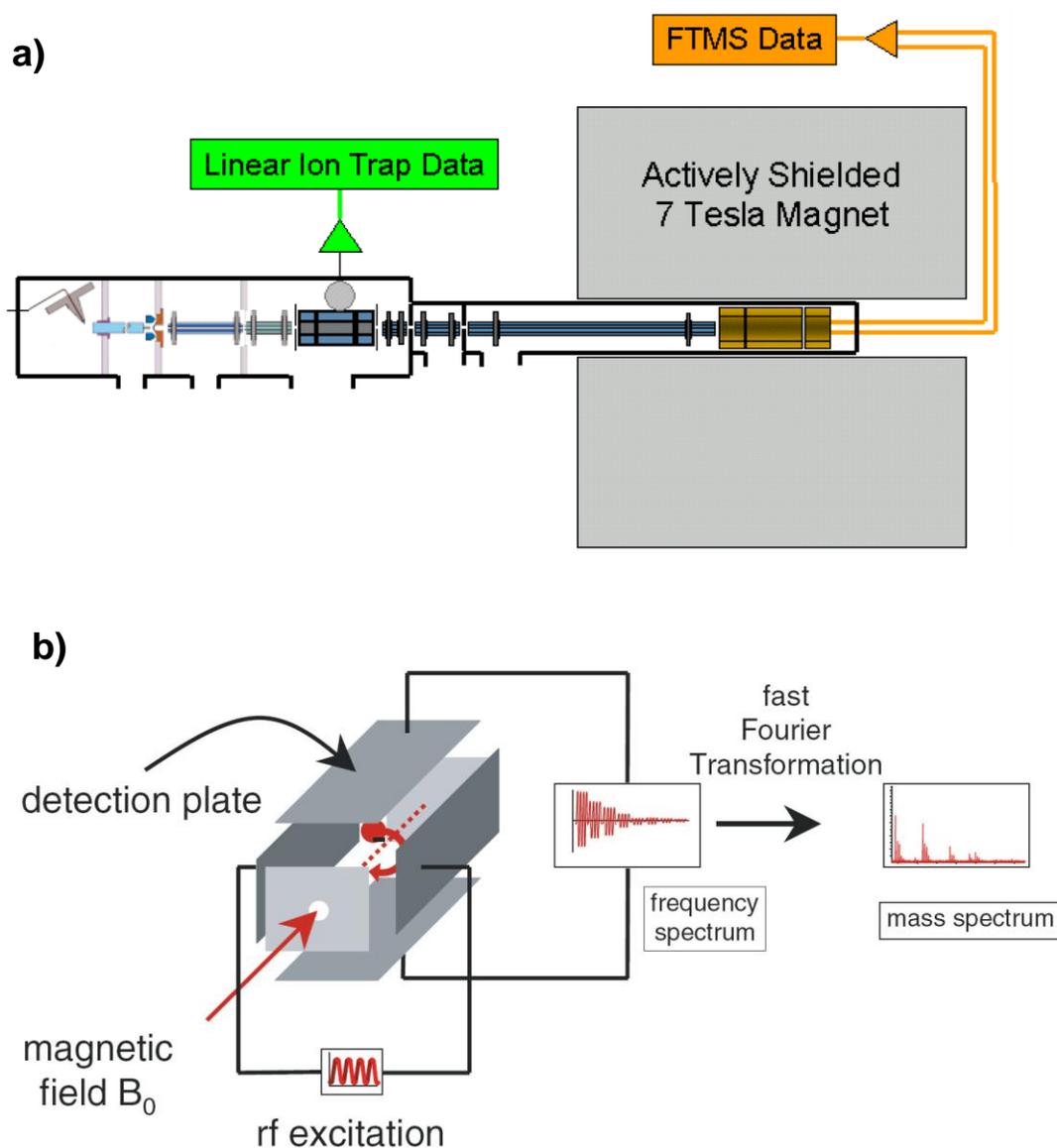
Figure 12: The principle of FT-ICR mass spectrometry.
a) A hybrid instrument comprised of a linear ion trap and an FT-ICR mass spectrometer is depicted. Their analytical strengths result in a synergy of speed, sensitivity, resolution and mass accuracy. **b)** The cyclotron motion of ions in the ICR cell generates a frequency spectrum that is transformed into a conventional mass spectrum by fast Fourier transformation. Adapted from Schrader and Klein, 2004.

1.3.2.3 Orbitrap

Similar to ICR mass spectrometers, also Orbitraps[86, 87] detect ions in a non-destructive fashion. Instead of employing a strong magnetic field, however, they use electrodes to guide the motion of ions in the cell. The electrostatic attraction towards the central electrode is compensated by a centrifugal force that arises from the initial tangential velocity of ions. Instead of the rotational frequency like in the case of ICR, an Orbitrap measures axial oscillations along the central electrode to obtain a mass spectrum. The outer electrodes register the axial oscillations, and Fourier transformation results in the determination of their

frequencies and subsequently the calculation of a mass spectrum. For a controlled injection of ions into the Orbitrap, they are first accumulated in a so-called C-trap. This is a curved (C-shaped) quadrupole that traps and cools ions, squeezes them into a small cloud, and finally injects them into the Orbitrap analyzer. The minimized distance between C-trap and Orbitrap reduces time-of-flight effects and enhances sensitivity. In addition, the C-trap allows for the co-injection of defined background ions for use as a so-called lock mass[88]. This internal mass calibration further improves mass accuracy, which serves higher identification confidence and lower false positive rates[89]. Figure 13 shows a scheme of a hybrid linear ion trap-Orbitrap mass spectrometer, and a close-up view of an Orbitrap cell.

a)

b)
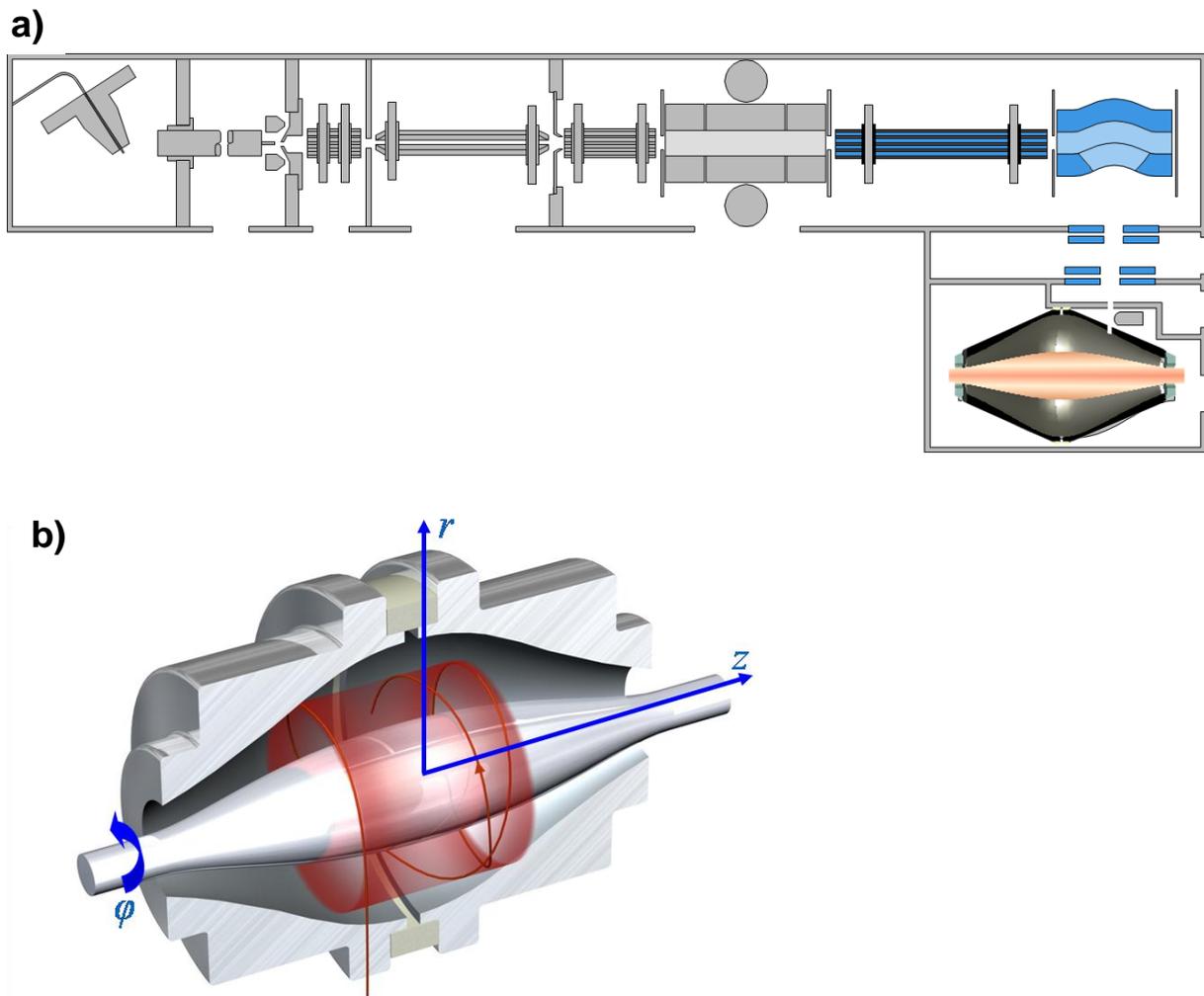
Figure 13: The principle of Orbitrap mass spectrometry.

a) A hybrid instrument comprised of a linear ion trap and an Orbitrap mass spectrometer is depicted. Particularly for complex mixtures, improved performance can be achieved as compared to linear ion trap FT-ICR.

b) Ion packages spin around the central electrode in the Orbitrap cell and generate a frequency spectrum that is transformed into a conventional mass spectrum by fast Fourier transformation. Adapted from Scigelova and Makarov, 2006.

1.3.3 Stable isotope labeling for quantitative readout

Mass spectrometry is largely applied as a qualitative method to identify components in a sample. In addition, it also possesses the ability to generate quantitative information. The signals of consecutive analyses can be compared, and differences in abundance of individual proteins between two samples can be extracted. The number of identified peptides[90], sequence coverage[91], spectral counting[92], identified peptides in relation to theoretically observable peptides[93, 94], and intensity or peak area of peptide signals[95, 96] have been used for this purpose. However, these methods often suffer from low accuracy and poor reliability. Truly accurate quantitative comparison between proteins of different samples can best be achieved if their peptides appear in the same spectra of one single analysis. This is possible if one of the two samples is labeled with heavy isotopes (usually ^{13}C, ^{15}N, ^{2}H), inducing a defined increase in mass. This way samples can be mixed prior to MS-analysis, since distinct signals are created in mass spectra for every proteolytic peptide present in labeled and non-labeled form. As depicted in figure 12, peptides appear as pairs, or even as triplets if an additional labeling state is applied.
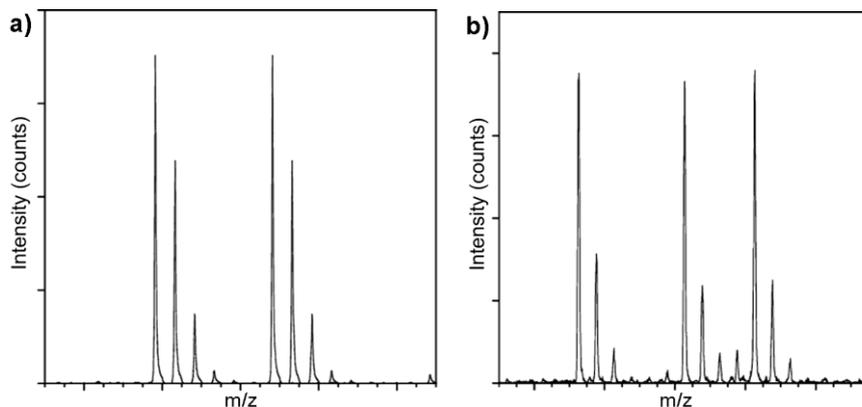
Figure 14: The principle of isotope labeling for quantitation between samples in mass spectrometry. Incorporation of heavy labels results in slightly increased mass. **a)** Peptides appear as pairs in mass spectra when labeled and unlabeled samples are mixed. **b)** If an additional labeling state was chosen, peptides are present as triplets.

By comparing peak areas of peptides originating from different samples, one can deduce the quantitative difference between the amounts of their parent proteins. Monitoring peptide peaks across their elution time results in extracted ion chromatograms (XIC) as illustrated in figure 15.
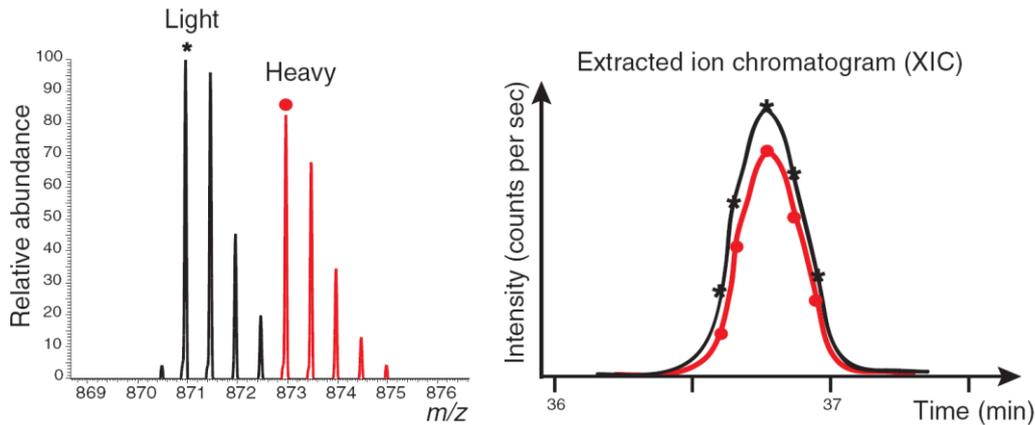
Figure 15: Relative quantitation of a peptide between two samples by stable isotope labeling.
As the peptides elute from the column, the signal is sampled several times, forming XIC-curves as depicted in the right panel. Modified from Ong and Mann, 2005.

1.3.3.1 Relative quantitation

Relative quantitation by isotope labeling between samples subjected to defined different experimental conditions is a widely used approach. Several methods have been developed for introducing heavy isotopes into a sample. Digestion in heavy water ($H_2^{18}O$)[97, 98], or chemical addition of labeled reagents like in ICAT[99], iTRAQ[100] or ICPL[101] integrate the label before, during or after protein digestion. Metabolic labeling techniques such as ^{15}N-labeling[102] and SILAC[103-105] even allow the incorporation of the label pre-harvest during culture. In cases where metabolic labeling is hard to accomplish, like with tissue samples, a modified SILAC-approach such as CDIT[106] can be applied.

Figure 16 summarizes different labeling types with respect to the stage at which the label is introduced in the experimental workflow.
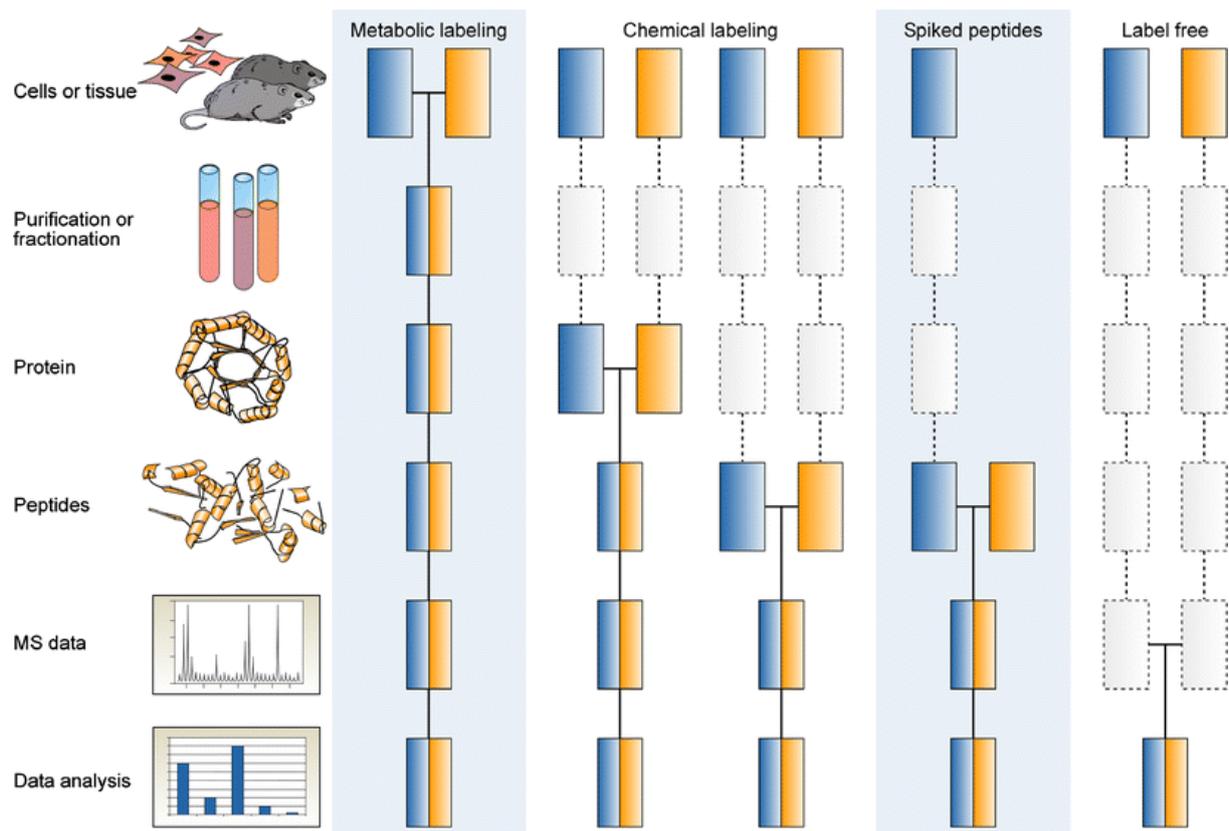
Figure 16: Different strategies for isotope labeling.
The various techniques differ in the timepoint of labeling in the experimental workflow. Metabolic labeling allows combination of samples directly after harvest, while other methods require parallel sample handling during the first steps. Reproduced from Bantscheff et al, 2007.

Variations in sample handling are difficult to control from a quantitative point of view. Thus, metabolic labeling is superior due to the very early combination of samples to be compared. Therefore SILAC was employed in the work presented here. SILAC stands for stable isotope labeling by amino acids in cell culture. Cells are cultured in medium that is deficient in one or more amino acids. These amino acids are supplemented in stable isotope labeled form, or in unlabeled form for the control culture to compare against. It is advisable to use labeled arginine and lysine, because trypsin cleaves C-terminal to these amino acids. This results in each proteolytic peptide containing a heavy label, and the presence at the C-terminus can aid in sequence assignment. Different labeling states of arginine that can be used for SILAC are shown in figure 17.
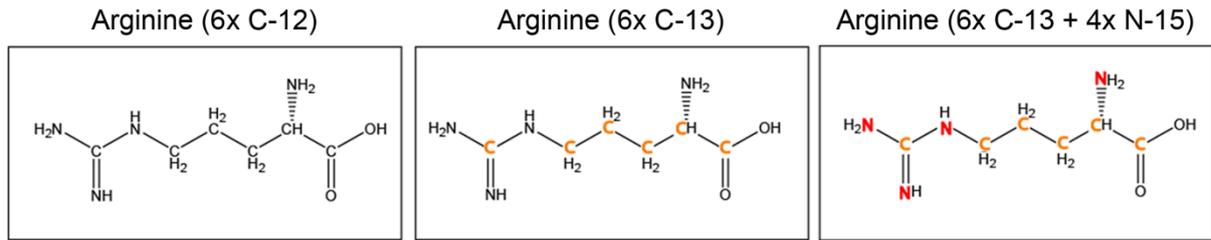
Figure 17: SILAC-labeling by different isotope-labeled versions of arginine.

If all ^{12}C-atoms are replaced by their heavy ^{13}C counterparts, the mass increment is 6 Da. If in addition all ^{14}N-atoms are replaced by ^{15}N, the mass increment is close to 10 Da.

During culture, cells incorporate the labeled amino acids into all newly synthesized proteins. After more than 5 cell doublings, labeling efficiency is close to 100%. Experiments with differential treatment can be conducted, such as comparison of hormone stimulation versus no stimulation. Subsequently, samples can be combined following cell lysis and protein extraction, because their different origin is encoded in the masses of their proteins. Any losses or variations in downstream sample processing have no influence on the stoichiometry between the mixed samples (figure 18).
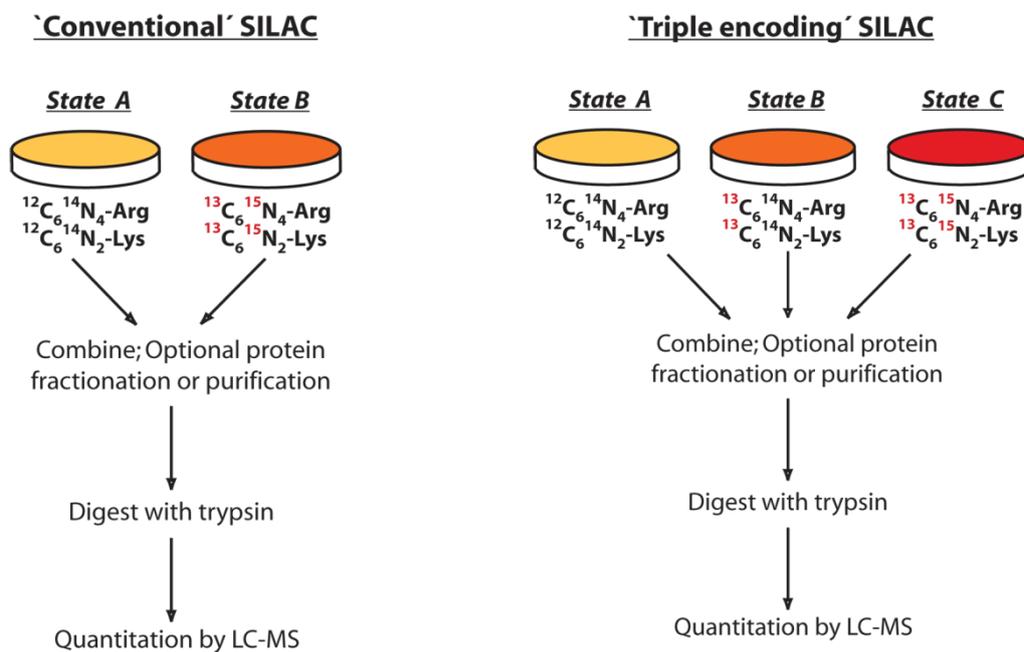
Figure 18: General workflow in SILAC-experiments.

The early combination of differentially labeled samples allows truly equal downstream sample processing.

1.3.3.2 Absolute quantitation

In addition to relative quantitation, several methods are available that combine isotope labeling and mass spectrometry for absolute quantitation of proteins. In each quantitation approach, a known amount of an isotope-labeled standard which resembles the protein to be quantified or a peptide thereof is spiked into the biological sample. Historically, digestion in heavy water was the first method used for absolute quantitation[107], followed by the use of synthetic peptide standards containing isotope-labeled amino acids for the quantitation of the parent protein[108], a concept expanded further and termed AQUA[109]. A protein-like standard in the form of concatenated peptides is employed in the QconCAT[110-112] or PCS[113] approach. Artificial genes comprising peptides from different proteins are expressed in medium containing $^{15}NH_4Cl$ or labeled amino acids. These techniques can supply quantitative data, however, their accuracy is not very high and suffers from several principal limitations. Difficulties of digestion in heavy water include the control of incorporation of one or two ^{18}O atoms as well as the necessity to add the standard at a very late stage in sample preparation. AQUA has the advantage that labeled synthetic peptides are easy to obtain, even though it should be noted that they are expensive reagents. However, when the absolute quantitation of a protein is based on only one or two peptides as is typically the case in AQUA and QconCAT, precision may suffer. Furthermore, not all peptides are amenable for use as standards because amino acids that could introduce side products need to be avoided. The chief limitation that these methods have in common is the relatively late addition of the quantitation standard to the sample. Any fractionation steps or other treatments that would result in differential effects on standard and endogenous protein need to be carried out before the standard can be added. This means that the absolute amount of a protein´s tryptic peptides that are present after sample digestion can be accurately determined. The actual purpose of the experiment, however, is to provide an accurate measure of the protein´s amount in the original sample. A truly accurate technique for absolute quantitation needs to be insensitive to protein losses and chemical modifications during sample processing, as well as incomplete digestion and recovery of proteolytic peptides.

One of the goals of this work was therefore the development of an MS-based absolute quantitation method which addresses these problems by spiking in the isotope labeled standard at the earliest possible stage. To achieve this, the standard needs to be highly similar if not identical to the endogenous protein to be quantified. The SILAC-approach that has proven to be very powerful for relative quantitation (see above) has therefore been adapted for its use in absolute quantitation of individual proteins.

An overview of several methods for absolute protein quantitation is given in table 1.

Table 1: Summary of approaches for absolute quantitation of protein abundance with mass spectrometry. Adapted from Kito and Ito, 2008.

	Standard Types	Measured Values	Spiking Time Point	Normalization	Accuracy	Coverage	Applicability to Post-Translational Modification	Noise Origins
Isotope labeling	Synthetic peptide	Ratio to standard	Before or after digestion	—	Low Medium	Low	Applicable	S/N of ion peak, Missed cleavage
	Intact protein	Ratio to standard	Just after protein extraction	—	High	High	Not	S/N of ion peak
	Peptide concatemer (QconCAT, PCS)	Ratio to standard	Prior to digestion	—	Medium	Low	Not	S/N of ion peak, Missed cleavage (QconCAT)
Label-free	—	Peak intensity	—	Average of most intense three peaks	Low	Medium	Not	Variation of ionization efficiency
	—	Spectral count	—	Observable peptides	Low	High	Not	Stochastically calculated index

1.4 GOALS OF THE STUDY

This work was conducted with two major goals in mind. One focus was on the characterization of pTyr-based protein interactions, mainly embarking on molecular interaction platforms in the insulin signaling pathway. The other objective was the development of a novel approach for protein quantitation, particularly for absolute quantitation.

The insulin signaling network contains critical nodes for signal distribution at the start of the intracellular pathway. The receptors are able to attract various proteins and integrate them into spreading the signal. Importantly, the receptors utilize IRS proteins as mediators for a large part of these interactions. A recently established protein interaction technology based on bait-fishing from whole cell lysates was therefore destined for the discovery of proteins that interact with the receptors (InsR, IGF1R and IRR) or IRS proteins (IRS-1 and IRS-2). Through a systematic and exhaustive profiling of all potential pTyr-dependent interaction sites in these key players of the pathway, a comprehensive and comparative picture of their interaction capabilities was envisioned. The design of the experimental setup was chosen in a way that allows obtaining site-specific information for known interaction partners as well as the discovery of novel interactions.

In a separate project, the same approach was to be employed to experimentally assess and validate pTyr-based interaction motifs originated from a new bioinformatic algorithm that clusters known phosphorylation sites according to their surrounding sequence.

Representative synthetic phosphopeptides were to be probed for their potential to mediate protein interactions. Particular emphasis was placed on previously undefined motifs that might constitute novel binding motifs for certain pTyr-binding domains.

The suboptimal accuracy of absolute quantitation of protein abundance in conventional MS-based techniques provided the motivation to develop a method to solve this problem. Based on its excellent properties in relative quantitation, the SILAC-technique was the fundament on which a novel technique was to be developed that offers high accuracy in absolute quantitation. In addition to the establishment of quantitation standard generation at the protein level, thorough characterization of the method and devising guidelines for optimal application formed an integral part of the project. This included the design of targeted data acquisition methods on the mass spectrometer that enhance identification rate, detection sensitivity, and quantitation quality.

The production of recombinant proteins in a SILAC-labeled fashion was primarily aimed at its use for absolute quantitation of proteins in regular bottom-up proteomics. With SILAC-labeled protein standards at hand, it was interesting to investigate their application in top-down proteomics as well. MS-analysis of intact protein pairs in labeled and unlabeled state was envisioned, aiming at a profound characterization of their behavior with respect to the effects of isotope labeling. Based on the results, conclusions for potential applications in quantitation, identification and assessment of posttranslational modifications were expected.

2 PROFILING OF PHOSPHOTYROSINE-INTERACTION PLATFORMS IN THE INSULIN SIGNALING PATHWAY

2.1 PUBLICATION: THE PHOSPHOTYROSINE INTERACTOME OF THE INSULIN RECEPTOR FAMILY AND ITS SUBSTRATES IRS-1 AND IRS-2

This manuscript presents the results from the pTyr-interactome project in the insulin signaling pathway. At the time of writing this thesis, it had been submitted for publication in the journal Molecular and Cellular Proteomics.

The following pages contain the submitted version of the manuscript.

The phosphotyrosine interactome of the insulin receptor family and its substrates IRS-1 and IRS-2

Stefan Hanke and Matthias Mann

Department of Proteomics and Signal Transduction, Am Klopferspitz 18, 82152 Martinsried, Munich, Germany

Corresponding author: Prof. Dr. Matthias Mann, Department of Proteomics and Signal Transduction, Am Klopferspitz 18, D-82152 Martinsried, Germany; tel: +49 89-8578-2557; fax: +49 89-8578-2219; e-mail: mmann@biochem.mpg.de

Running title: Phosphotyrosine interactome of insulin signaling

Keywords: stable isotope labeling, SILAC, proteomics, protein interaction, phosphorylation, SH2 domain, IRS, insulin receptor substrate, receptor tyrosine kinase, insulin receptor, insulin-like growth factor 1 receptor, IGF1R, insulin receptor related receptor, IRR, signal transduction

Abbreviations:

MS, mass spectrometry; LC, liquid chromatography; MS/MS, tandem mass spectrometry; LTQ, linear quadrupole ion trap; SILAC, stable isotope labeling by amino acids in cell culture; Arg, arginine; Lys, lysine; SH2, Src homology 2; PTB, phosphotyrosine binding; pTyr, phosphotyrosine, IRS, insulin receptor substrate, InsR, insulin receptor; IGF1R, insulin-like growth factor 1 receptor; IRR, insulin receptor related receptor

SUMMARY

The insulin signaling pathway is critical in regulating glucose levels and is associated with diabetes, obesity and longevity. A tyrosine phosphorylation cascade creates docking sites for protein interactions, initiating subsequent propagation of the signal throughout the cell. The phosphotyrosine interactome of this medically important pathway has not yet been studied comprehensively. We therefore applied quantitative interaction proteomics to exhaustively profile all potential phosphotyrosine (pTyr)-dependent interaction sites in its key players. We targeted and compared insulin receptor substrates 1 and 2 (IRS-1 and IRS-2) as central distributors of the insulin signal, the insulin receptor (InsR), the insulin-like growth factor 1 receptor (IGF1R), and the insulin receptor related receptor (IRR). Using the SILAC approach with phosphorylated vs. non-phosphorylated bait peptides, we found phosphorylation-specific interaction partners for 52 out of 109 investigated sites. In addition, doubly and triply phosphorylated motifs provided insight into the combinatorial effects of phosphorylation events in close proximity to each other. Our results retrieve known interactions and substantially broaden the spectrum of potential interaction partners of IRS-1 and IRS-2. A large number of common interactors rationalize their extensive functional redundancy. However, several proteins involved in signaling and metabolism interact differentially with IRS-1 and IRS-2 and thus provide leads into their different physiological roles. Differences at the receptor level are reflected in multisite recruitment of SHP2 by the IGF1R and limited but exclusive interactions with the IRR. In common with other recent reports, our data furthermore hint at non-SH2 or PTB domain mediated pTyr-binding.

INTRODUCTION

Regulated protein-protein interactions form the basis of cellular signal transduction, and frequently posttranslational modifications constitute the molecular switch to facilitate the association or dissociation of proteins. Phosphorylation is a prominent instrument in the toolbox of signaling, and tyrosine phosphorylation in particular is important in the upstream events following ligand binding to receptor tyrosine kinases (RTKs). The mere binding to a phosphorylated motif can modulate enzymatic activity in some cases. However, the primary effect of the interaction is usually to increase the local concentration of the recruited protein and scaffold it together with its upstream or downstream effectors. Common recognition modules for tyrosine phosphorylated sequences are the Src homology 2 (SH2) domain and the phosphotyrosine binding (PTB) domain (1, 2). A large part of the free binding energy to an SH2 domain is provided by the phospho-moiety itself. Another part, and most importantly the binding specificity, is contributed by interactions

with the residues C-terminal to the phosphotyrosine (pTyr). As determined by degenerate peptide library screening, motifs for SH2 domains typically encompass residues +1 to +4 relative to the pTyr (3). Some SH2-domains also exploit amino acids at the N-terminal side for binding (3, 4). In few cases even more extended contacts from -6 to +6 are formed (5, 6). The SH2-domain of SAP, for example, engages so many residues that already the non-phosphorylated form shows considerable binding, and phosphorylation only enhances this binding 5-fold (7). The general interaction mode of SH2 domains directs the sequence of the partner protein perpendicular to the central β-sheet of the SH2 domain in an extended conformation. Therefore the interaction is largely independent of the structural context in the native protein. This allows studying SH2-binding using short synthetic peptides (3). The genomes of humans and mice contain 120 different SH2 domains in 110 different proteins.

The other canonical interaction domain for pTyr, the PTB domain, occurs 56 times in the genome; its binding mode can vary considerably and is less conserved than is the case for SH2 domains (8, 9). PTB domains are divided into three classes: IRS1/DOK-like, Shc-like, and Dab-like. Only members of the first two classes, which account for just 25% of all PTB domains, actually bind in a phosphorylation-dependent manner (1, 8). Recently several reports have described pTyr-dependent protein interactions that do not involve SH2 or PTB domains. In one case binding was attributed to a C2 domain (10), but in other cases the responsible modules are still elusive (11). Even though it is still unclear whether those cases represent exceptions or instead more general principles, they demonstrate that pTyr mediated binding is not limited to the classical interaction domains.

A particularly important and clinically relevant pathway that involves tyrosine phosphorylation is the insulin signaling network. Malfunction of insulin signaling can lead to type II diabetes if insulin resistance cannot be compensated by pancreatic β-cells any longer. Initially, insulin resistance is counterbalanced by increased insulin secretion by β-cells, but chronic hyperinsulinemia results in increased insulin resistance and finally leads to a decline of β-cell mass and function, facilitating hyperglycemia and diabetes (12). The main target tissues of insulin are skeletal muscle, liver and adipose tissue, with muscle accounting for more than 70% of glucose disposal (13). One of the most important metabolic effects of insulin is the translocation of the glucose transporter GLUT-4 from intracellular storage vesicles to the plasma membrane (14).

A complex and still incompletely understood network of signaling proteins connects the insulin receptor (InsR) with its downstream effectors. Insulin receptor substrate proteins (IRSs) play a pivotal role as interaction platforms that become phosphorylated on multiple tyrosines by the InsR and subsequently attract various signaling proteins to spread the

stimulus within the cell. All six members of the IRS-family contain a PH domain and a PTB domain at the N-terminus, but they vary in the length of their C-terminal part. IRS-1 and IRS-2 are large proteins with many tyrosine phosphorylation sites and are ubiquitously expressed. IRS-3 is short and only present in rodents. IRS-4 has similar dimensions as IRS-1 and IRS-2, but is only expressed in very few and specialized tissues. IRS-5 and IRS-6 (also termed DOK-4 and DOK-5) possess very short C-terminal parts and consequently very few potential phosphotyrosine motifs (15, 16).

By far the most important players in insulin signaling are IRS-1 and IRS-2. Apart from their role in metabolic signaling, they propagate proliferative and anti-apoptotic signals and are consequently overexpressed or activated more strongly in most cancers (17). Furthermore many of the triggers of insulin resistance like excess lipids (13, 18, 19), inflammatory cytokines (18) or reactive oxygen species (20), exert their undesired effect through activation of multiple kinases that phosphorylate IRSs on serine residues (13). IRS-1 and IRS-2 share 75% amino acid sequence identity in their N-terminal domains and 35% in their C-terminal part. Despite their high degree of homology and many similar tyrosine-phosphorylation motifs, studies in knockout mice and knockout cell lines indicate that these two IRS proteins also serve complementary, rather than completely redundant, roles in insulin and IGF-1 signaling (21). In general, IRS-1 plays a prominent role in growth, while the main functions of IRS-2 are in glucose homeostasis and proper function of pancreatic β-cells (22). Knockout of either of them results in insulin resistance, but only IRS-2-knockout leads to diabetes. Tissue-specific differences between IRS-1 and IRS-2 add a further level of complexity (23). In muscle, IRS-1 is more closely associated with glucose uptake, whereas IRS-2 stimulates the MAP-kinase pathway (24, 25). In liver, both are involved in metabolic regulation, but IRS-2 has a more pronounced role in lipid metabolism (26, 27). Adipose tissue engages IRS-1 mainly for differentiation while IRS-2 serves insulin-stimulated glucose uptake (28).

Similarly, among the three different receptors – insulin receptor (InsR), insulin-like growth factor-1 receptor (IGF1R) and insulin receptor related receptor (IRR), relatively small differences in the C-terminal, cytosolic sequence lead to considerable diversity in the signaling output. The insulin receptor and its ligand maintain metabolic homeostasis, whereas IGF-1 and its receptor preferentially control developmental and growth processes (29). Despite a similar degree of homology, the IRR is still an orphan receptor with obscure function and is expressed in few tissues (30, 31). Its greatest differences compared to the other receptors are in the C-terminal part, where it lacks the last 50 amino acids compared to InsR and IGF1R.

Here we undertook a systematic and exhaustive profiling of all potential phosphotyrosine-dependent interaction sites in IRS-1, IRS-2, InsR, IGF1R and IRR. Since this kind of interaction is mediated by short, unstructured sequence motifs (2, 32, 33), and based on our previous results with ErbB-receptors (34), bacterial proteins and the histone code (35), we employed peptide bait fishing from cell lysates combined with state-of-the-art mass spectrometry as readout. Our peptide pulldown approach, combined with the SILAC technique (36), allows straightforward discrimination between specific interaction partners and background binders (37). In contrast to in vitro experiments using purified components (38), the interactions take place within the competitive environment of a whole cell lysate. In addition to the site-specific information obtained, the unbiased nature of the approach facilitates the discovery of unexpected interactions. This makes it possible, in principle, to uncover novel kinds of interactions mediated by modules other than the currently known SH2 and PTB domains in pTyr-dependent signaling.

Our large-scale experiment resulted in a global overview of specificity, redundancy and distribution of protein interaction sites in the insulin signaling platform. It furthermore allows comparing the principle signaling capabilities between IRS-1 and IRS-2 as well as between InsR, IGF1R and IRR. The pTyr-interactomes of the IRS-proteins were relatively similar, accounting for the large functional redundancy. However, we did observe specific differences in the binding of SH2/PTB domain containing proteins and intriguingly in some proteins with other functions such as fatty acid degrading enzymes.

A main difference between InsR and IGF1R turned out to be the recruitment of the tyrosine phosphatase SHP2 to more sites in the IGF1R. The cryptic function of the IRR could be attributed to lack of interactions in its C-terminal part, perhaps combined with the unique recruitment of a membrane-associated guanylate kinase discovered here. Finally, in some instances combinatorial tyrosine phosphorylation either abolished or enabled certain protein interactions.

EXPERIMENTAL PROCEDURES

SILAC cell culture and lysis – Murine C2C12 muscle cells were cultured in Dulbecco´s modified Eagle´s medium (DMEM) containing 4.5% glucose and deficient in arginine (Arg) and lysine (Lys), supplemented with 10% dialyzed fetal calf serum and antibiotics. One cell population was supplied with normal Arg and Lys, and the other one with the stable isotope labeled heavy analogues $^{13}C_6^{15}N_4$-Arg (or $^{13}C_6$-Arg) and $^{13}C_6^{15}N_2$-Lys from Sigma Isotec. For triple labeling experiments $^{13}C_6$-Arg and D_4-Lys as well as $^{13}C_6^{15}N_4$-Arg and $^{13}C_6^{15}N_2$-Lys were employed. Cells were expanded as myoblasts for at least five doublings and differentiation into myotubes was initiated by lowering serum content to 2% in

confluent dishes. After 8 days myotube cultures typically contained less than 15% mononucleated cells. Harvesting was carried out by washing dishes with PBS and adding ice-cold lysis buffer to the dishes for 15 minutes. Lysis buffer consisted of 1% Igepal (NP-40; v/v), 150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1 mM DTT, protease inhibitor cocktail (Roche complete tablets) and 1 mM sodium orthovanadate as tyrosine phosphatase inhibitor. Cells were scraped off the dishes and vortexed vigorously. Following centrifugation at 16,000 g for 15 minutes the supernatant was used for peptide affinity pulldown experiments.

Peptide synthesis – Peptides were synthesized as pairs in phosphorylated and non-phosphorylated form on a solid-phase peptide synthesizer using amide resin (Intavis, Germany). Peptides were designed as 15-mers with 8 amino acids N-terminal and six C-terminal of the (phospho)tyrosine. In addition, the dipeptide linker SG and a desthiobiotin were attached to the N-terminus. Identity and purity of the synthetic peptides were confirmed by mass spectrometric analysis.

Peptide pulldown procedure – Peptides were bound to streptavidin-coated magnetic beads (Dynal MyOne, Invitrogen), and cell lysate typically corresponding to 1.5 mg of protein (~5 mg/ml protein) was added to 75 µl of beads containing an estimated amount of 2 nmol peptide. Heavy SILAC-labeled lysate was incubated with the phosphorylated version of the peptide, while light SILAC-labeled lysate was added to the non-phosphorylated counterpart. In parallel a crossover experiment was conducted, where the incubation was inverted, that is heavy lysate was incubated with non-phosphorylated peptide and light lysate with phosphorylated version. After rotation at 4°C for six hours or overnight, the beads were washed for at least 3 times by vortexing with lysis buffer. Beads from each peptide pair were combined and bound proteins were eluted using 20 mM biotin. Eluted proteins were then precipitated by adding 5 volumes of ethanol together with sodium acetate and 20 µg glycoblue (Ambion, USA).

In-solution digestion of proteins – Proteins were resuspended in 20 µl 6M urea, 2M thiourea, 20 mM TrisHCl pH8,0 and reduced by adding 1 µg DTT for 30min, followed by alkylation of cysteines by incubating with 5 µg iodoacetamide for 20 min. Digestion was started by adding endoproteinase LysC (Wako). After three hours samples were diluted with four volumes of 50 mM NH_4HCO_3, and trypsin (Promega) was added for overnight incubation. Proteases were applied in a ratio of 1:50 to protein material and all steps were carried out at room temperature. Digestion was stopped by acidifying with TFA and the samples were loaded onto StageTips (39, 40) packed with RP-C_{18} Empore disks, 3M for desalting and concentration prior to LC-MS-analysis.

NanoLC-MS/MS – Digested peptide mixtures were separated by online reversed-phase (RP) nanoscale capillary liquid chromatography (nanoLC) and analyzed by electrospray tandem mass spectrometry (ES-MS/MS). Experiments were performed with an Agilent 1100 nanoflow system connected to an LTQ-Orbitrap or LTQ-FT mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). Binding and chromatographic separation of the peptides took place in a 15 cm fused silica emitter (75 μm inner diameter) in-house packed (41) with reversed-phase ReproSil-Pur C_{18}-AQ 3 μm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany).

Peptide mixtures were injected onto the column with a flow of 500 nL/min and subsequently eluted with a flow of 250 nL/min from 2% to 40% MeCN in 0.5% acetic acid, in a 100 min gradient. The mass spectrometer was operated in data dependent mode to automatically switch between MS and MS/MS (MS2) acquisition. Survey full scan MS spectra with m/z 300-1600 were acquired in the Orbitrap with a resolution of 60,000 at m/z 400 after accumulation to a target value of 1 million charges in the linear ion trap using the lock mass option for internal calibration of each spectrum (42). The five most intense ions were sequentially isolated for fragmentation in the linear ion trap using collisionally induced dissociation with normalized collision energy of 30% at a target value of 5000. The resulting fragment ions were recorded in the linear ion trap with unit resolution. Ions already selected for MS/MS were dynamically excluded for 60 seconds. For LTQ-FT measurement resolution was set to 25,000 for full scan, 5 million ions were accumulated and the top three ions were selected for sequencing. In FT-ICR, selected ion monitoring (SIM) scans were acquired before fragmentation for enhanced mass accuracy and signal-to-noise of the parent ion (resolution 50,000 at m/z 400, target value 50,000). MS^3 was performed on the most abundant fragment ion for increased certainty of identification.

Peptide identification and quantitation – Peak lists for database searching were generated from the raw data using in-house developed software called Raw2msm. From each fragment spectrum the six most intense peaks per 100 Th were extracted. Proteins were identified by automated database searching (Mascot version 2.1, Matrix Science) against an in-house curated version of the mouse IPI database (versions ranged from 3.00 to 3.37 and contained between 40,613 and 68,655 entries complemented with frequently observed contaminants like porcine trypsin and human keratins. Carbamidomethyl-cysteine was used as a fixed modification; variable modifications were oxidation of methionine, protein N-acetylation, deamidation of Asn and Gln, N-pyroglutamate and heavy versions of Arg and Lys. We required full tryptic specificity (cleavage at Arg-Pro and Lys-Pro as well as Asp-Pro was included), a maximum of two missed cleavages, and maximal mass deviation of 5 ppm for the parent ion and 0.5 Da for

fragment ions. As initial identification threshold for peptides we chose a false positive rate of max. 5% as judged by searching a concatenated database consisting of normal and reverse sequences. However, we required peptides with twice that score for final protein identification for phosphorylation-specific interaction partners (for example, MASCOT score 50 if score 25 corresponded to 5% false positive rate). Identification confidence of interaction partners was further enhanced by requirement of a SILAC ratio different from 1:1, and the identification of the binder in an independent, crossover experiment with inversed ratio. For the relative quantitation of SILAC peptide pairs our in-house developed software MSQuant was used (http://msquant.sourceforge.net). All peptides from proteins appearing with an elevated SILAC-ratio were verified and re-analyzed manually.

Determination of significant binding partners – A typical pulldown experiment yielded hundreds of identified proteins. The vast majority had a SILAC-ratio close to 1:1, indicating that they represent background binders of the beads irrespective of the phosphorylation state of the peptide. All protein ratios were normalized against the median. Proteins with a ratio more than 2-3 standard deviations above the median were considered as phosphorylation-specific binders, given that they had a correspondingly inverted ratio in the crossover experiment. It was not appropriate to decide on a fixed cut-off value, since the number of interactors with ratio different from 1:1 influences the standard deviation. In borderline cases (referring to the ratio or the amount of identified peptides), or whenever an unexpected binding partner was discovered (e.g. a protein without known pTyr-binding domain), the experiment was repeated at least once to verify the significance of binding.

RESULTS

Quantitative proteomics for unbiased identification of interactions – For the identification of site-specific, pTyr-dependent interaction partners, we applied our previously established peptide pulldown approach, which is based on quantitative proteomics (43). Briefly, we separately incubated cell extract with synthetic peptides in both the phosphorylated and non-phosphorylated version. These peptides were synthesized with an N-terminal desthiobiotin for coupling to streptavidin-coated beads and efficient elution via biotin. To enhance accessibility, a short flexible linker of one serine and one glycine preceded the actual sequence. In order to further account for steric limitations and to build on current knowledge of binding modes of SH2 and PTB domains, we chose the sequence stretch as a 15-mer with 8 residues N-terminal of the pTyr and 6 residues C-terminal.

Due to the central role of skeletal muscle in insulin signaling, we employed whole cell lysate from the murine muscle cell line C2C12, which we differentiated to myotubes before harvest. The SILAC technique was applied by incubating the phosphorylated peptide with lysate from cells grown in medium containing stable isotope labeled arginine and lysine, whereas unlabeled lysate was added to the control peptides. After washing steps and elution of bound proteins, LC-MS/MS of the combined samples yielded mass spectra in which unspecific background binders appeared with a signal intensity ratio of 1:1 between labeled and unlabeled versions of the peptides originating from protein digestion. Phosphorylation-specific binders were uncovered through their high ratios. The experimental design is depicted in Fig. 1. Every pulldown experiment was also carried out in a crossover fashion where cell lysates were swapped, resulting in inverted ratios for genuine, specific binders.
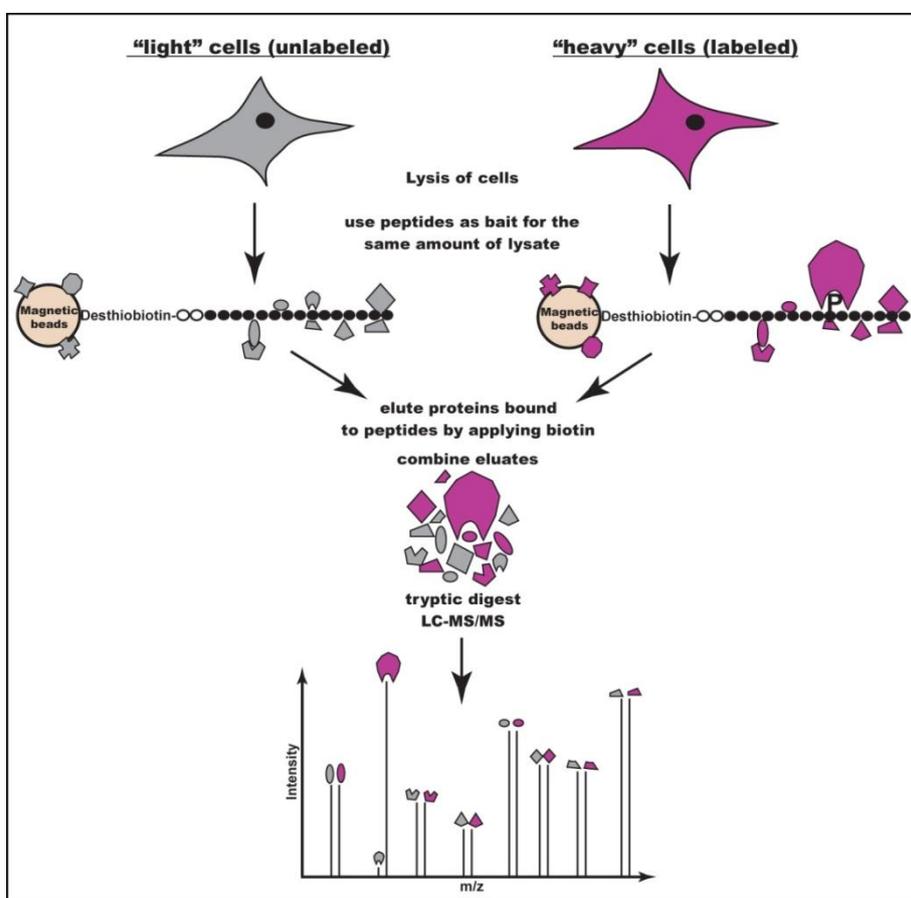
Fig. 1. **Proteomic screening for interaction partners of tyrosine phosphorylated sequences.** Peptides corresponding to potential tyrosine phosphorylation sites are synthesized in phosphorylated and non-phosphorylated form. Cell populations are metabolically labeled using the SILAC technique to allow discrimination based on different peptide masses. Cell lysate from the population labeled with heavy Arg and Lys is incubated with the phosphorylated version of the peptide, while the control cell lysate is incubated with the non-phosphorylated peptide. Eluted proteins from those parallel pulldown experiments are combined and digested with trypsin. Peptides from unspecific background binders appear as pairs with abundance ratios close to 1:1 in the mass spectra. Phosphorylation-specific binders are identified as such through their high abundance ratio between heavy and light labeling states. In a crossover experiment, the incubation scheme is swapped, resulting in inverted ratios.

Fig. 2 contrasts unbiased, quantitative proteomics with the classical candidate- and western blotting-based approach. When visualizing eluates from a pulldown experiment on a gel, it is impossible to distinguish a low abundant, phosphorylation-specific binder among abundant background proteins without targeted immunoassays. However, the unbiased mass spectrometric readout is ideally suited for discovering novel interactors since it does not depend on a priori hypotheses as is the case for western blotting (44). This is made possible by quantitative proteomics – in this case though SILAC –, which allows straightforward discrimination between phosphorylation-specific interactors and background binders.
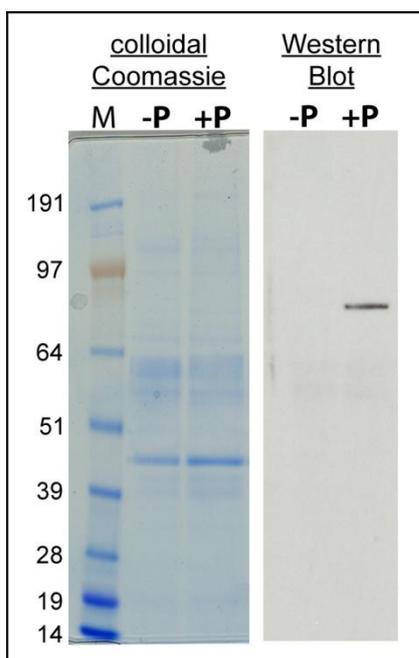
Fig. 2. Peptide pulldown experiment as visualized by SDS-gel and western blot. Instead of combining protein samples from phosphorylated and non-phosphorylated bait peptides followed by direct in-solution digestion for LC-MS/MS analysis, one pulldown sample was applied to an SDS-gel to illustrate sensitivity and importance of SILAC-based MS-readout. The sample contains roughly 1-2 µg of protein. An interaction partner of IRS-2 pY0649, PI3K p85α, was detected by western blotting after its identification by quantitative proteomics. The intense band at 45 kDa is actin.

Figure 3 depicts a typical result from a peptide pulldown experiment. The high abundance ratio between labeled and unlabeled state distinguishes phosphorylation-specific binders from hundreds of background binders. Note that specificity of binding is evident when a ratio is statistically different from the ratios of background binders. The value of the ratio itself, however, does not directly reflect the affinity or the stoichiometry of the interaction. In fact, repetitions of an experiment can yield fluctuations in the ratios, due to minor variations in the stringency of washing steps without, however, changing the significance of the interaction. Furthermore, in many cases the differences between binding to the bait and the control are so great that the unlabeled peptides (representing binding to the

control bait) become undetectable and hence unquantifiable (45). Even though this clearly indicates phosphorylation-specific interaction, it leads to imprecise ratios. For this reason, we recommend to consider the outlier significance (distance from mean) in addition to the absolute ratio.
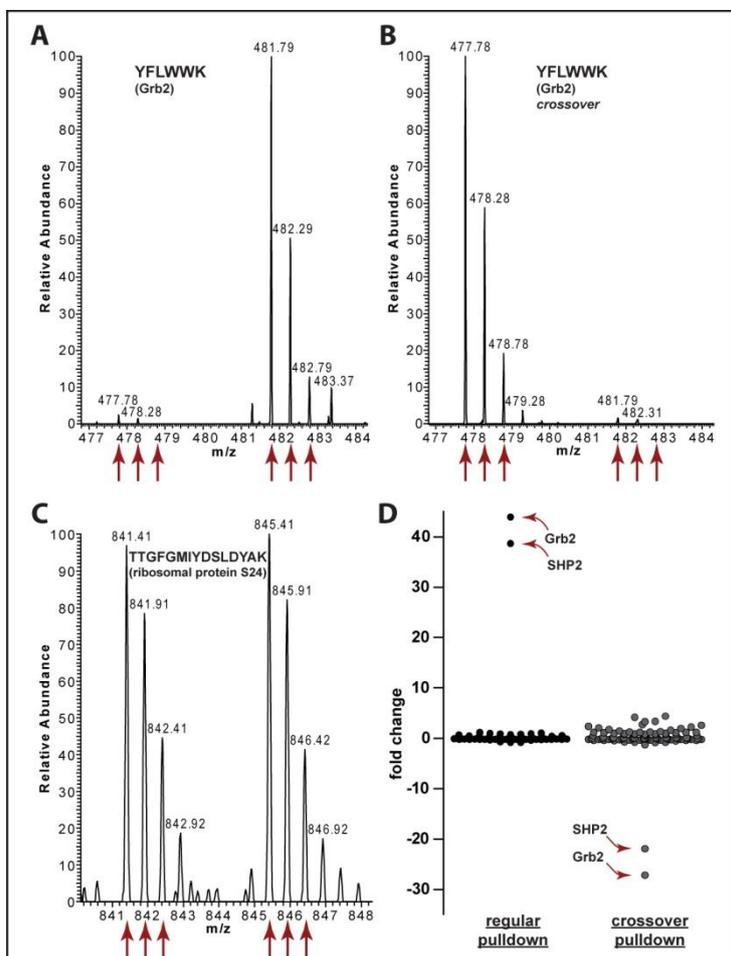
Fig. 3. Typical result of a SILAC peptide pulldown experiment, exemplified by the bait peptide IRS-1 Y0891. (A) A peptide derived from the interaction partner Grb2 is about 40 times more abundant in the heavy form than in the light form, indicating that Grb2 binds specifically to IRS1 phosphorylated on Y0891. **(B)** In a crossover experiment (swapped SILAC labels) the abundance ratio of the same peptide is inverted. **(C)** Most other proteins have a 1:1 ratio, indicating that they are unspecific binders to the peptides or the magnetic beads that are bound irrespective of phosphorylation. **(D)** A plot of the protein abundance ratios shows that Grb2 and SHP2 are significant outliers in both the Y0891 pulldown and the crossover experiment. Every dot represents one protein.

In total, we performed pulldowns for 109 individual sites in forward and crossover experiments. The majority of pulldowns were additionally repeated in separate experiments to validate novel interactors. Fourteen sites were furthermore targeted in combinatorial binding experiments (see below). The interaction data is provided in Tables 1-3 and summarized in Figures 4 and 5.

Table 1: pTyr-dependent interaction partners of IRS-1, IRS-2, InsR, IGF1R and IRR. Data from one representative experiment for each site is provided. Phosphorylation-dependent interactors of the bait sequences are listed along with their ratios, standard deviations distance from the median ratio, and numbers of identified and quantified unique peptides. Values of regular and crossover experiments are separated by a slash. Known phosphorylation sites for the mouse, rat and/or human sequence are indicated as stated on http://www.phosphosite.org and http://phospho.elm.eu.org. Values with an asterix (*) stem from experiments where one of the following reasons masked the true significance of the interaction: Either different degrees of clumping of the peptide-beads (due to a strong difference in solubility between phosphorylated and non-phosphorylated version), or partial protein precipitation in the cell lysate, led to a large standard deviation of background binder ratios. Note that four of the investigated tyrosines do not exist in the corresponding human sequence (IRS1 Y0489, IGF1R Y1282 and Y1283, IRR Y1291).

Protein	Site	Known P-site	Bait sequence	MS-instrument used	Experiments done	Phosphorylation-specific interaction partner	Protein ratio (regular/crossover experiment)	Std-Devs distance from median	Unique peptides identified	Peptides used for quantitation	Known pY-binding domain
IRS1	Y0018		SGFSDVRKVGYLRKPKS	LTQ-FT	1	x					
IRS1	Y0046	x	SGAGGPARLEYYENEKK	LTQ-FT		x					
IRS1	Y0047		SGGGPARLEYYENEKKW	LTQ-FT	1	Grb2, Q60631	4.5 +- 0.3 / 0.23 +- 0.03	2.7 / 2.3	8 / 14	5 / 7	SH2
IRS1	Y0087		SGKNKHLVALYTRDEHF	LTQ-FT	2	WIPI-3, Q9CR39	14 +- 14 / 0.076 +- 0.01	2.3 / 2.6 *	3 / 5	2 / 2	
IRS1	Y0107		SGSEAEQDSWVQALLQL	LTQ-Orbitrap	5	2.4-dienoyl-CoA reductase, Q9CQ62	28 +- 13 / 0.09 +- 0.07	4.4 / 4.9	20 / 24	13 / 16	
						Acyl-CoA dehydrogenase, very-long-chain specific, P50544	26 +- 0.09 / 0.07 +- 0.06	4.3 / 4.3	26 / 30	17 / 14	
						Acyl-CoA dehydrogenase, long-chain specific, P51174	11 +- 2 / 0.07 +- 0.02	3.2 / 4.3	3 / 3	3 / 3	
						Acyl-CoA dehydrogenase family member 9, Q8JZN5	12 +- 3 / 0.12 +- 0.03	3.2 / 3.5	14 / 21	8 / 13	
						Methylcrotonoyl-CoA carboxylase subunit alpha, Q99MR8	18 +- 7 / 0.08 +- 0.04	3.8 / 4.2	17 / 25	10 / 13	
						Methylcrotonoyl-CoA carboxylase subunit beta, Q3ULD5	23 +- 14 / 0.07 +- 0.06	4.1 / 4.3	16 / 23	10 / 8	
						Isobutyryl-CoA-dehydrogenase, Q9D7B6	16 +- 6 / 0.04 +- 0.02	3.6 / 5.3	16 / 15	10 / 6	
						Enoyl-CoA hydratase, Q8BH95	14 +- 1 / 0.10 +- 0.03	3.5 / 3.8	3 / 5	2 / 4	
						Glutaryl-CoA dehydrogenase, Q60759	9 +- 3 / 0.20	2.8 / 2.6	2 / 1	2 / 1	
						RNA polymerase I, II and III subunit RPABC1, Q80UW8	10 +- 1 / 0.20 +- 0.02	3.0 / 2.7	5 / 7	4 / 5	
						Unconventional prefoldin RPB5 interactor, Q3TLD5	10 +- 3 / 0.15 +- 0.02	3.0 / 3.1	10 / 12	8 / 8	
						Prefoldin subunit 2, O70591	9 +- 2 / 0.14 +- 0.03	2.9 / 3.1	3 / 3	2 / 2	
						Hydroxysteroid (17-beta) dehydrogenase 10, A2AFQ2	6.2 +- 1.2 / 0.28 +- 0.07	2.4 / 2.0	3 / 5	3 / 4	
						Septin 2, P42208	10 +- 3 / 0.16 +- 0.03	3.0 / 3.0	11 / 5	7 / 5	
						Septin 7, O55131	9 +- 2 / 0.14 +- 0.03	2.9 / 3.2	14 / 7	11 / 5	
						Septin 9, Q80UG5	8 +- 2.5 / 0.16 +- 0.03	2.7 / 2.9	13 / 6	8 / 4	
						Septin 11, Q8C1B7	12 +- 3 / 0.17 +- 0.04	3.3 / 2.9	9 / 13	6 / 9	
						WD-repeat containing protein 92, Q8BGF3	5.2 +- 0.6 / 0.06 +- 0.05	1.7 / 2.3 *	6 / 6	3 / 4	
IRS1	Y0147		SGGEAGEDLSYDTGPGP	LTQ-FT	1	x					
IRS1	Y0178		SGQTKNLIGIYRLCLTS	LTQ-Orbitrap	1	Hydroxysteroid (17-beta) dehydrogenase 10, A2AFQ2	12 +- 3 / 0.17 +- 0.07	1.5 / 1.4 *	3 / 5	3 / 4	
						WD-repeat containing protein 92, Q8BGF3	7 +- 1 / 0.32 +- 0.07	1.1 / 0.9 *	3 / 6	2 / 5	
IRS1	Y0426		SGGGFISSDEYGSSPCD	LTQ-FT	1	x					
IRS1	Y0460	x	SGRGEEELSNYICMGGK	LTQ-FT	1	PI3K reg. alpha subunit, P26450	40 +- 14 / 0.02 +- 0.01	2.4 / 2.7 *	10 / 7	6 / 3	2x SH2
						PI3K catal. alpha subunit, P42337	27 +- 9 / 0.03 +- 0.03	2.1 / 2.3 *	13 / 8	4 / 2	
						SHP2, P35235	18 +- 7 / 0.06 +- 0.04	1.8 / 1.8 *	6 / 3		
IRS1	Y0478		SGTLAAPNGHYILSRGG	LTQ-FT	1	x					
IRS1	Y0489		SGSRGGNGHRYIPGANL	LTQ-FT	1	x					
IRS1	Y0546		SGSSVASIEEYTEMMPA	LTQ-Orbitrap	1	RasGAP, Q91YX7	22 +- 9 / 0.05 +- 0.02	6.8 / 6.1	19 / 13	7 / 5	SH2
						PI3K reg. alpha subunit, P26450	2.6 +- 0.3 / 0.25 +- 0.07	2.2 / 2.9	10 / 11	9 / 6	2x SH2
						PI3K catal. alpha subunit, P42337	6 +- 1 / 0.15 +- 0.06	3.8 / 4.0	15 / 12	4 / 8	
						SHIP-2, Q6P549	7 +- 2 / 0.18	4.3 / 3.6	8 / 3	3 / 1	SH2
						See also pulldown with doubly phosphorylated sequence					
IRS1	Y0554		SGYTEMMPAAYPPGGGS	LTQ-FT	1	x					
IRS1	Y0567		SGGSGGRLPGYRHSAFV	LTQ-FT	1	x					
IRS1	Y0578		SGSAFVPTHSYPEEGLE	LTQ-FT	1						
IRS1	Y0608	x	SGSNLHTDGYMPMSPG	LTQ-FT	1	PI3K reg. alpha subunit, P42337	10 +- 8 / 0.13 +- 0.03	2.6 / 2.5 *	5 / 6	3 / 4	2x SH2
						PI3K catal. alpha subunit, P42337	7.5 +- 4.4 / ---	2.3 / --- *	3 / ---	2 / ---	
IRS1	Y0628	x	SGSNRKGNGDYMPMSPK	LTQ-FT	1	PI3K reg. alpha subunit, P26450	18 / 0.06 +- 0.03	6.8 / 6.1	6 / 3	1 / 2	2x SH2
IRS1	Y0658	x	SGPQRVDPNGYMMMSPS	LTQ-FT	1	PI3K reg. alpha subunit, P26450	20 +- 9 / 0.07 +- 0.01	5.0 / 3.9	7 / 5	4 / 3	2x SH2
						PI3K reg. beta subunit, O08908	17 / 0.09	4.7 / 3.5	1 / 1	1 / 1	2x SH2
IRS1	Y0690		SGSAAPSGSSYGKPWTN	LTQ-Orbitrap	2	RasGAP, Q91YX7	7.5 +- 0.2 / 0.06	3.5 / 4.6	4 / 2	3 / 1	SH2
						PI3K reg. alpha subunit, P26450	4.3 +- 0.23 / 0.26 +- 0.01	2.5 / 2.2	2 / 2	2 / 2	2x SH2
IRS1	Y0727	x	SGKLLPCTGDYMNMSPV	LTQ-Orbitrap	3	Grb2, Q60631	2.7 +- 0.3 / 0.6 +- 0.05	2.3 / 1.3	10 / 8	9 / 7	SH2
						PI3K reg. alpha subunit, P26450	1.50 +- 0.13 / 1.0 +- 0.2	0.9 / 0.0	4 / 4	2 / 4	2x SH2
						PI3K catal. alpha subunit, P42337	1.8 +- 0.1 / 1.2 +- 0.1	1.3 / 0.4	2 / 3	2 / 3	2x SH2
						PI3K reg. beta subunit, O08908	1.4 +- 0.0 / ---	0.7 / ---	2 / ---	2 / ---	2x SH2

Protein	Site	Known P-site	Bait sequence	MS-instrument used	Experiments done	Phosphorylation-specific interaction partner	Protein ratio (regular/crossover experiment)	Std-Devs distance from median	Unique peptides identified	Peptides used for quantitation	Known pY-binding domain
IRS1	Y0745		SGNTSSPSECYYGPEDP	LTQ-FT	1	x					
IRS1	Y0746		SGTSSPSECYYGPEDPQ	LTQ-FT	2	x					
IRS1	Y0759		SGPQHKPVLSYYSLPRS	LTQ-FT	1	SHP2, P35235	21 +- 6 / 0.06 +- 0.03	3.4 / 3.2	5/7	2/2	2x SH2
IRS1	Y0760		SGQHKPVLSYYSLPRSF	LTQ-Orbitrap	1	SHP2, P35235	11 +- 4 / 0.08 +- 0.03	3.7 / 3.5	26/20	18/16	2x SH2
IRS1						Crk-like protein, P47941	14 +- 5 / 0.06 +- 0.02	4.1 / 4.0	7/6	7/5	SH2
IRS1						Crk-like protein, P47941	10 +- 1.7 / 0.10 +- 0.04	3.6 / 3.3	10/6	7/6	SH2
						See also pulldown with doubly phosphorylated sequence					
IRS1	Y0795		SGSSSSGRLRYTATAED	LTQ-FT	1	x					
IRS1	Y0815		SGSSDSLGGGYCGARPE			SHP-2, Q6P549	16 +- 10 / 0.01 +- 0.01	5.5 / 7.7	6/8	4/4	SH2
IRS1	Y0891	x	SGPEPKSPGEVVNIEFG	LTQ-FT	2	Grb2, Q60631	45 +- 13 / 0.035 +- 0.02	5.9 / 4.4	13/12	13/10	SH2
						SHP2, P35235	40 +- 9 / 0.04 +- 0.02	5.7 / 4.2	20/12	12/5	2x SH2
IRS1	Y0903		SGEFGSGQPGYLAGPAT	LTQ-FT	1	SHP-2, Q6P549	58 +- 38 / 0.02 +- 0.01	6.9 / 6.9	13/19	8/8	SH2
IRS1	Y0935	x	SGREETGSEEYMNMDLG	LTQ-FT	2	Grb2, Q60631	9.1 +- 4.4 / 0.10 +- 0.01	2.8 / 2.7	3/3	2/2	SH2
						PI3K reg. alpha subunit, P26450	30 +- 21 / 0.03 +- 0.02	4.3 / 4.3	9/10	6/5	2x SH2
						PI3K reg. beta subunit, O08908	5.5 / 0.07	2.2 / 3.2	1/1	1/1	2x SH2
						PI3K catal. alpha subunit, P42337	21 +- 10 / 0.05 +- 0.03	3.9 / 3.6	13/11	5/5	
IRS1	Y0983	x	SGSVPNSRGDYMTMQIG	LTQ-Orbitrap	1	SHP-2, Q6P549	8.7 +- 4.7 / 0.13 +- 0.11	3.8 / 3.1	17/19	8/7	SH2
						PI3K reg. alpha subunit, P26450	4.7 +- 0.7 / 0.49 +- 0.06	2.7 / 1.1	3/5	3/5	2x SH2
						PI3K catal. alpha subunit, P42337	--- / 0.81 +- 0.74	--- / 0.3	--- / 6	--- / 4	
IRS1	Y0995		SGQIGCPRQSVVDTSPV	LTQ-Orbitrap	1	x					
IRS1	Y1006		SGTSPVAPVSYADMRTG	LTQ-FT	2	SHP2, P35235	31 +- 13 / 0.03 +- 0.02	6.4 / 4.9	10/9	3/5	2x SH2
IRS1	Y1173	x	SGGGLEKSLNYIDLDLA	LTQ-FT	2	SHP2, P35235	40 +- 26 / 0.05 +- 0.04	3.9 / 3.9	30/30	20/19	2x SH2
IRS1	Y1220	x	SGRSSEDLSNYASISFQ	LTQ-FT	2	SHP2, P35235	20 +- 8.5 / 0.05 +- 0.03	5.0 / 4.1	15/10	7/3	2x SH2
IRS2	Y0036		SGNHSVRKCGYLRKQKH	LTQ-FT	1	x					
IRS2	Y0075	x	SGPPQPPRLEYYESEKK	LTQ-Orbitrap	2	x					
IRS2	Y0076		SGPQPPRLEYYESEKKW	LTQ-FT	1	x					
IRS2	Y0111		SGKRADAKHKYLIALYT	LTQ-FT	2	x					
IRS2	Y0116		SGKHKYLIALYTKDEYF	LTQ-FT	1	x					
IRS2	Y0121		SGIALYTKDEYFAVAAE	LTQ-FT	1	x					
IRS2	Y0136		SGNEQEQEGWYRALTDL	LTQ-Orbitrap	4	2,4-dienoyl-CoA reductase, Q9CQ62	9 +- 4 / 0.06 +- 0.02	2.4 / 2.8 *	20/17	14/13	
						RNA polymerase I, II and III subunit RPABC1, Q80UW8	6 +- 4 / 0.07 +- 0.02	1.9 / 2.7 *	12/14	11/14	
						Unconventional prefoldin RPB5 interactor, Q3TLD5	9 +- 4 / 0.06 +- 0.04	2.4 / 2.9 *	11/10	6/9	
						Prefoldin subunit 2, O70591	8 +- 6 / 0.06 +- 0.04	2.3 / 2.8 *	7/4	4/3	
						WD-repeat containing protein 92, Q8BGF3	6.4 +- 1.9 / 0.03 +- 0.01	2.0 / 3.4 *	14/12	11/11	
						Septin 2, P42208	3.2 +- 1.2 / 0.13 +- 0.01	1.3 / 2.0 *	3/3	2/3	
						Septin 7, O55131	3.6 +- 0.1 / 0.12 +- 0.01	1.4 / 2.2 *	2/4	2/4	
						Septin 9, Q80UG5	5 +- 2.5 / ---	1.7 / --- *	2/ ---	2/ ---	
						Septin 11, Q8C1B7	4 +- 2 / 0.16 +- 0.05	1.7 / 1.8 *	8/6	6/6	
						Hydroxysteroid (17-beta) dehydrogenase 10, A2AFQ2	--- / 0.16	--- / 1.9 *	--- / 2	--- / 1	
IRS2	Y0181		SGGAAGCDDNYGLVTPA	LTQ-FT	2	x					
IRS2	Y0191		SGLVTPATAVVYREVWQV	LTQ-FT	1	SHP-2, Q6P549	11 +- 4 / 0.08 +- 0.05	4.1 / 4.6	3/6	2/4	SH2
						2,4-dienoyl-CoA reductase, Q9CQ62	4.2 +- 3.3 / 0.08 +- 0.06	2.5 / 4.5	2/3	2/2	
IRS2	Y0214		SGQSKNLTGVVRLCLSA	LTQ-Orbitrap	3	Glycylpeptide N-myristoyltransferase 1, O70310	10 +- 5 / 0.06 +- 0.01	1.9 / 2.6 *	16/17	12/15	
						Prefoldin subunit 2, O70591	5.6 +- 2.8 / 0.09	1.4 / 2.2 *	2/1	2/1	
						RNA polymerase I, II and III subunit RPABC1, Q80UW8	3.5 +- 0.7 / 0.14 +- 0.01	1.0 / 1.8 *	10/11	9/11	
						Unconventional prefoldin RPB5 interactor, Q3TLD5	--- / 0.07 +- 0.02	--- / 2.4 *	--- / 2	--- / 2	
						Hydroxysteroid (17-beta) dehydrogenase 10, A2AFQ2	3.0 +- 0.9 / 0.38 +- 0.25	0.9 / 0.9 *	5/4	5/4	
						WD-repeat containing protein 92, Q8BGF3	7 +- 2 / 0.05 +- 0.02	1.5 / 2.7 *	8/10	6/10	
IRS2	Y0456		SGSSGHGSGSYPLPPGS	LTQ-FT	1	x					
IRS2	Y0500		SGPGFMSLDEYGSSPGD	LTQ-FT	2	x					
IRS2	Y0536		SGRDGSGGELYGYMSMD	LTQ-FT	2	SHP-2, Q6P549	9 +- 2 / NQ	5.6 / ---	2/1	2 / NQ	2x SH2
						See also pulldown with doubly phosphorylated sequence					
IRS2	Y0538		SGSSGGGELYGYMSMDRP	LTQ-FT	1	PI3K reg. alpha subunit, P26450	17 +- 6 / 0.05 +- 0.02	5.2 / 4.2	5/4	3/4	2x SH2
						See also pulldown with doubly phosphorylated sequence					
IRS2	Y0552		SGPLSHCGRPYRRVSGD	LTQ-FT	1	x					
IRS2	Y0572		SGDRGLRKRTYYSLTTPA	LTQ-FT	1	x					
IRS2	Y0594	x	SGPSSASLDEYTLMRAT	LTQ-FT	1	PI3K reg. alpha subunit, P26450	15 +- 5 / 0.05	4.5 / 6.0	4/5	2/1	2x SH2
IRS2	Y0621		SGPASSPKVAYNPYPED	LTQ-FT	1	x					

| Protein	Site	Known P-site	Bait sequence	MS-instrument used	Experiments done	Phosphorylation-specific interaction partner	Protein ratio (regular/crossover experiment)	Std-Devs distance from median	Unique peptides identified	Peptides used for quantitation	Known pY-binding domain
IRS2	Y0624	x	SGSPKVAYNPYPEDYGD	LTQ-FT	1	x					
						RasGAP, Q91YX7	34 +- 28 / 0.05 +- 0.04	4.4 / 7.2	23 / 23	16 / 17	SH2
IRS2	Y0628	x	SGAYNPYPEDYGDIEIG	LTQ-Orbitrap	3	SHP2, P35235	18 +- 5 / 0.08 +- 0.04	3.6 / 6.1	13 / 15	12 / 12	2x SH2
						PI3K reg. alpha subunit, P26450	16 +- 3 / 0.06 +- 0.03	3.4 / 6.6	7 / 6	5 / 5	2x SH2
						PI3K catal. alpha subunit, P42337	21 +- 11 / 0.06 +- 0.04	3.8 / 6.6	3 / 7	2 / 4	
						PI3K reg. beta subunit, O08908	8.8 / ---	2.7 / ---	1 / ---	1 / ---	2x SH2
						PI3K catal. subunit beta, Q8BTI9	22 / ---	3.8 / ---	2 / ---	1 / ---	
						Dedicator of cytokinesis protein 6, Q8VDR9	11 +- 2 / 0.11 +- 0.04	3.0 / 5.2	13 / 13	7 / 10	
						Dedicator of cytokinesis protein 7, Q8R1A4	7.4 +- 1.6 / 0.19 +- 0.06	2.5 / 3.9	16 / 25	12 / 22	
						2,4-dienoyl-CoA reductase, Q9CQ62	3.8 +- 0.7 / 0.40 +- 0.04	1.7 / 2.1	17 / 20	17 / 18	
						LRCH3, Q8BVU0	8.8 +- 2.3 / 0.15 +- 0.00	2.7 / 4.4	2 / 4	1 / 3	
						See also pulldown with triply phosphorylated sequence					
IRS2	Y0649	x	SGSNLGADDGYMPMTPG	LTQ-FT	1	PI3K reg. alpha subunit, P26450	27 +- 10 / 0.06 +- 0.03	4.6 / 2.5	11 / 4	10 / 4	2x SH2
						PI3K catal. alpha subunit, P42337	43 +- 23 / 0.05 +- 0.04	5.3 / 2.7	5 / 3	3 / 2	
IRS2	Y0671	x	SGPNSCKSDDYMPMSPT	LTQ-FT	1	PI3K reg. alpha subunit, P26450	21 +- 12 / 0.07 +- 0.03	2.7 / 2.4*	11 / 10	7 / 5	2x SH2
						PI3K catal. alpha subunit, P42337	20 +- 7 / 0.03 +- 0.02	2.9 / 3.0*	8 / 11	3 / 2	
IRS2	Y0719		SGTFPVNGGGYKASSPA	LTQ-FT	1	x					
IRS2	Y0734	x	SGESSPEDSGYMRMWCG	LTQ-Orbitrap	4	PI3K reg. alpha subunit, P26450	22 +- 8 / 0.06 +- 0.06	4.5 / 4.8	23 / 25	10 / 15	2x SH2
						PI3K catal. alpha subunit, P42337	28 +- 15 / 0.05 +- 0.05	4.9 / 5.1	31 / 30	14 / 12	
						PI3K reg. beta subunit, O08908	16 +- 8 / 0.05 +- 0.01	4.1 / 5.0	5 / 9	4 / 6	2x SH2
						PI3K catal. subunit beta, Q8BTI9	26 +- 8 / 0.05 +- 0.02	4.8 / 4.9	18 / 16	6 / 9	
						SHP2, P35235	29 +- 14 / 0.04 +- 0.03	4.9 / 5.5	32 / 33	14 / 13	2x SH2
						RasGAP, Q91YX7	15 +- 6 / 0.06 +- 0.02	4.0 / 4.7	11 / 14	7 / 9	SH2
						Phospholipase C gamma, Q62077	28 +- 23 / 0.05 +- 0.04	4.9 / 5.1	46 / 51	20 / 14	2x SH2
						Shc1, P98083	20 +- 8 / 0.03 +- 0.03	4.4 / 6.1	23 / 23	9 / 8	SH2, PTB
						SHIP-2, Q6P549	26 +- 11 / 0.05 +- 0.04	4.7 / 5.1	52 / 61	21 / 25	
						2,4-dienoyl-CoA reductase, Q9CQ62	12 +- 3 / 0.10 +- 0.02	3.6 / 3.7	14 / 20	10 / 12	SH2
						Crk, Q64010	21 +- 7 / 0.03 +- 0.01	4.5 / 5.6	7 / 8	4 / 4	SH2
						Crk-like protein, P47941	15 +- 7 / 0.07 +- 0.02	4.0 / 4.4	6 / 9	4 / 4	SH2
						Grb2, Q60631	7 +- 2 / 0.21 +- 0.03	2.8 / 2.5	11 / 8	7 / 7	SH2
						Cullin-5, Q9D5V5	8 +- 1 / 0.18 +- 0.06	3.0 / 2.7	16 / 14	9 / 9	
IRS2	Y0758	x	SGPKLLPNGDYLNKSPS	LTQ-FT	1	PI3K reg. alpha subunit, P26450	34 / 0.04 +- 0.00	4.1 / 4.5	4 / 8	1 / 2	2x SH2
						PI3K catal. alpha subunit, P42337	41 +- 10 / 0.18 +- 0.15	4.4 / 2.4	3 / 3	2 / 2	
						Grb2, Q60631	9 +- 1 / 0.05 +- 0.01	2.6 / 4.1	2 / 5	2 / 3	SH2
						Phospholipase C gamma, Q62077	29 +- 21 / 0.05 +- 0.01	3.9 / 4.1	3 / 5	2 / 2	2x SH2
IRS2	Y0794		SGLKGIPGHCYSSLPRS	LTQ-FT	1	x					
IRS2	Y0801		SGCYSSLPRSYKAPCSC	LTQ-FT	1	x					
IRS2	Y0814	x	SGSCSGDNDQYVLMSSP	LTQ-FT	2	x					
						PI3K reg. alpha subunit, P26450	5.4 +- 0.9 / 0.15 +- 0.03	4.3 / 3.9	17 / 18	12 / 10	2x SH2
						PI3K catal. alpha subunit, P42337	7.7 +- 1.7 / 0.07 +- 0.02	5.2 / 5.4	19 / 21	10 / 13	
						PI3K reg. beta subunit, O08908	6.6 +- 2.3 / 0.15 +- 0.04	4.8 / 4.0	3 / 5	3 / 2	2x SH2
						PI3K catal. subunit beta, Q8BTI9	NQ / 0.11 +- 0	--- / 4.4	2 / 3	NQ / 2	
IRS2	Y0899		SGQTLPSMQEYPLPTEP	LTQ-FT	1	x					
IRS2	Y0911	x	SGTEPKSPGEYINIDPG	LTQ-FT	5	x					
						Phospholipase C gamma, Q62077	27 +- 13 / 0.09 +- 0.09	5.7 / 4.4	27 / 22	10 / 9	2x SH2
						Grb2, Q60631	40 +- 17 / 0.02 +- 0.02	6.4 / 6.8	14 / 14	8 / 6	SH2
						SHP2, P35235	13 +- 4 / 0.08 +- 0.03	4.4 / 4.6	5 / 3	4 / 2	2x SH2
						Grb2, Q60631	10 +- 0 / 0.05 +- 0.04	4.0 / 5.5	2 / 2	2 / 2	2x SH2
IRS2	Y0970	x	SGRQRSPLSDYMNLDPS	LTQ-FT	1	PI3K reg. alpha subunit, P26450	23 +- 12 / 0.05 +- 0.01	5.5 / 4.9	5 / 4	5 / 4	SH2
IRS2	Y1006		SGLSPEASSPYPPLPPR	LTQ-FT	1	x					
IRS2	Y1032		SGLPPAPGDLYRLPPAS	LTQ-FT	1	x					
IRS2	Y1061	x	SGSEPGDNGDYSEMAFG	LTQ-FT	1	PI3K reg. alpha subunit, P26450	14 +- 1 / 0.03	7.2 / 6.5	4 / 5	2 / 1	2x SH2
						PI3K catal. subunit beta, Q8BTI9	--- / 0.05	--- / 5.7	--- / 3	--- / 1	SH2
IRS2	Y1242		SGVGFQNGLNYIAIDVR	LTQ-FT	2	SHP2, P35235	10 +- 3 / 0.08 +- 0.08	3.9 / 4.1	9 / 5	3 / 2	2x SH2
						Acyl-CoA dehydrogenase, long-chain specific, P51174	8 +- 6 / 0.15 +- 0.05	3.4 / 3.1	5 / 3	2 / 3	
IRS2	Y1303	x	SGGALPSASTYASIDFL	LTQ-FT	2	SHP2, P35235	31 +- 17 / 0.04 +- 0.05	4.0 / 5.4	26 / 21	5 / 5	2x SH2
INSR	Y0982	x	SGPDGPMGPLYASSNPE	LTQ-FT	2	x, but see also pulldown with longer sequence in table 3					
INSR	Y0989	x	SGLYASSNPFEYLSASDV	LTQ-FT	1	x					
INSR	Y1001	x	SGSDVFPSSVVVPDEWE	LTQ-Orbitrap	2	2,4-dienoyl-CoA reductase, Q9CQ62	1.8 +- 0.4 / 0.5 +- 0.1	1.0 / 1.2	24 / 24	23 / 21	2x SH2
						RasGAP, Q91YX7	22 +- 10 / 0.07 +- 0.05	5.2 / 4.7	33 / 36	21 / 25	SH2

Protein	Site	Known P-site	Bait sequence	MS-instrument used	Experiments done	Phosphorylation-specific interaction partner	Protein ratio (regular/crossover experiment)	Std-Devs distance from median	Unique peptides identified	Peptides used for quantitation	Known pY-binding domain
						Phospholipase C gamma, Q62077	11 +- 7 / 0.07 +- 0.06	4.0 / 4.6	24 / 34	18 / 22	2x SH2
						SHP2, P35235	10 +- 1 / 0.09 +- 0.01	3.9 / 4.1	2 / 3	2 / 2	2x SH2
						Dedicator of cytokinesis protein 6, Q8VDR9	7 +- 4 / 0.16 +- 0.04	3.3 / 3.2	6 / 10	6 / 8	
						Dedicator of cytokinesis protein 7, Q8R1A4	3.5 +- 1.1 / 0.33 +- 0.10	2.1 / 1.9	16 / 17	14 / 16	
INSR	Y1028		SGGQGSFGMVYEGNAKD	LTQ-FT	2	CBL-B, Q3TTA7	4.3 +- 0.5 / 0.26 +- 0.05	2.5 / 2.4	3 / 6	3 / 6	SH2
INSR	Y1139		SGAEIADGMAYLNAKKF	LTQ-Orbitrap	2	CSK, P41241	5.6 +- 1.5 / 0.20 +- 0.12	2.8 / 2.8	2 / 3	2 / 3	SH2
						Grb2, Q60631	33 +- 13 / 0.03 +- 0.01	8.4 / 6.9	17 / 20	16 / 17	SH2
						Phospholipase C gamma, Q62077	17 +- 5 / 0.07 +- 0.03	6.8 / 5.4	6 / 14	4 / 7	2x SH2
INSR	Y1175	x	SGDFGMTRDIYETDYYR	LTQ-FT	2	x					
INSR	Y1179	x	SGTRDIYETDYYRKGGK	LTQ-FT	1	x					
INSR	Y1180	x	SGRDIYETDYYRKGGKG	LTQ-FT	2	x					
INSR	Y1227		SGITSLAEQPYQGLSNE	LTQ-FT	1	x					
INSR	Y1244		SGLKFVMDGGYLDPPDN	LTQ-FT	1	x					
INSR	Y1295	x	SGSFPEVSFFYSEENKA	LTQ-FT	2	x					
INSR	Y1345	x	SGSSLSIKRTYDEHIPY	LTQ-FT	1	x					
INSR	Y1351	x	SGRTYDEHIPYTHMNGG	LTQ-Orbitrap	1	PI3K reg. alpha subunit, P26450	29 +- 9 / 0.04 +- 0.01	7.8 / 6.8	8 / 6	6 / 3	2x SH2
						See also pulldown with doubly phosphorylated sequence					
IGF1R	Y0974	x	SGSRLGNGVLYASVNPE	LTQ-FT	1	SHP2, P35235	18 +- 10 / 0.09 +- 0.11	6.4 / 4.7	8 / 14	6 / 5	2x SH2
IGF1R	Y0981	x	SGLYASVNPEYFSAADV	LTQ-FT	1	x, but see also pulldown with longer sequence in table 3					
IGF1R	Y0988	x	SGPEYFSAADVYVPDEWE	LTQ-Orbitrap	3	2,4-dienoyl-CoA reductase, Q9CQ62	1.0 +- 0.2 / 1.2 +- 0.3	-0.1 / -0.4	27 / 28	26 / 27	SH2
						RasGAP, Q91YX7	10 +- 2 / 0.09 +- 0.02	3.3 / 5.4	4 / 5	4 / 5	2x SH2
						Phospholipase C gamma, Q62077	6.5 +- 1.3 / 0.06	2.7 / 6.6	2 / 3	2 / 1	2x SH2
						SHP2, P35235	9 / 0.11	3.1 / 5.1	2 / 2	1 / 1	2x SH2
						Dedicator of cytokinesis protein 6, Q8VDR9	2.8 +- 0.1 / 0.20 +- 0.03	1.5 / 3.6	3 / 11	2 / 8	
						Dedicator of cytokinesis protein 7, Q8R1A4	2.3 +- 0.4 / 0.44 +- 0.08	1.2 / 1.9	15 / 16	12 / 14	
IGF1R	Y1015		SGGQGSFGMVYEGVAKG	LTQ-FT	1	CBL-B, Q3TTA7	1.7 +- 0.3 / 0.72 +- 0.03	0.7 / 0.8	2 / 3	2 / 3	SH2
IGF1R	Y1091		SGMTRGDLKSYLRSLRP	LTQ-FT	2	x					
IGF1R	Y1127		SGGEIADGMAYLNANKF	LTQ-Orbitrap	2	Grb2, Q60631	9.4 / 0.08 +- 0.04	4.8 / 5.5	2 / 5	1 / 4	SH2
						SHP2, P35235	10 +- 2 / 0.11 +- 0.02	5.0 / 4.9	4 / 3	3 / 3	2x SH2
						Phospholipase C gamma, Q62077	8 / 0.07	4.6 / 5.8	2 / 3	1 / 1	2x SH2
IGF1R	Y1163	x	SGDFGMTRDIYETDYYR		1	x					
IGF1R	Y1167	x	SGTRDIYETDYYRKGGK		1	x					
IGF1R	Y1168	x	SGRDIYETDYYRKGGKG		1	x					
IGF1R	Y1215		SGIATLAEQPYQGLSNE		1	x					
IGF1R	Y1253		SGELMRMCWQYNPKMRP	LTQ-Orbitrap	2	SHP2, P35235	8 +- 2 / 0.17	5.6 / 4.5	6 / 1	4 / 1	2x SH2
IGF1R	Y1282	x	SGPSFQEVSFFYSEENK	LTQ-FT	1	x					
IGF1R	Y1283	x	SGSFQEVSFYYSEENKP	LTQ-Orbitrap	1	Acyl-CoA-dehydrogenase, short chain specific, Q91W85	32 +- 14 / 0.02 +- 0.01	9.8 / 10.7	18 / 15	8 / 12	2x SH2
						RasGAP, Q91YX7	8 / 0.22	6.0 / 4.1	1 / 1	1 / 1	SH2
IGF1R	Y1352	x	SGASFDERQPYAHMNGG	LTQ-Orbitrap	1	PI3K reg. alpha subunit, P26450	15 +- 8 / 0.02 +- 0.01	5.0 / 6.6	14 / 15	6 / 5	2x SH2
						PI3K catal. alpha subunit, P42337	14 +- 10 / 0.26	4.8 / 2.4	13 / 2	6 / 1	2x SH2
						PI3K reg. beta subunit, O08908	9 +- 0 / ---	3.9 / ---	3 / ---	2 / ---	
						PI3K catal. subunit beta, Q8BTI9	10 / ---	4.2 / ---	3 / ---	1 / ---	
IRR	Y0953		SGSRKRNSTLYTSVNPE	LTQ-FT		SHP2, P35235	16 +- 4 / 0.05 +- 0.01	4.5 / 4.6	9 / 8	5 / 6	2x SH2
IRR	Y0960	x	SGLYTSVNPEYFSASHM	LTQ-FT	1	x, but see also pulldown with longer sequence in table 3					
IRR	Y0967	x	SGEYFSASHMYVPDEWE	LTQ-Orbitrap	2	2,4-dienoyl-CoA reductase, Q9CQ62	5.9 +- 0.6 / 0.33 +- 0.12	4.2 / 4.4	7 / 6	4 / 5	
						RasGAP, Q91YX7	15 / ---	6.4 / ---	1 / ---	1 / ---	SH2
IRR	Y0994		SGGQGSFGMVYEGLARG		1	x					
IRR	Y1105	x	SGGEIADGMAYLAAKKF	LTQ-FT	2	x					
IRR	Y1141		SGDFGMTRDVYETDYYR		1	x					
IRR	Y1145		SGTRDVYETDYYRKGGK		1	x					
IRR	Y1146		SGRDVYETDYYRKGGKG		1	x					
IRR	Y1193		SGIVTLAEQPYQGLSNE		1	x					
IRR	Y1260	x	SGPSFRLCSFYSPECQ	LTQ-FT	2	x					
IRR	Y1261		SGSFRLCSFYSPECQR	LTQ-FT	1	x					
IRR	Y1291		SGPTLNGASDYSAPNGG	LTQ-FT	1	x					

The pTyr-interactome of IRS-1 and IRS-2 – IRS-1 and IRS-2 have been recognized as interaction platforms soon after their discovery, and literature reviews report approximately 10 known pTyr-dependent interaction partners, albeit sometimes with little supporting evidence (see DISCUSSION). Our study, for the first time, provides a systematic study of the pTyr interactome of these proteins. However, the interaction partners depicted in Fig. 4 – while specific to the phosphorylated baits – are not validated as biological interactions of the full length endogenous proteins in the cell, nor do we know for all pTyr sites if they are indeed phosphorylated upon insulin stimulation. Therefore we refer to them as potential interaction partners.

We first checked if our large-scale screen yielded known interaction partners. We indeed found these proteins, including the PI3K subunits, Grb2, SHP2, Nck, Csk and Crk. Additionally, the majority of interaction partners in Table 1 have SH2 or PTB domains, suggesting that these interactions are direct and specific. Thus we conclude that the screen performed well both in retrieving interaction partners known from literature as well as in expanding the interactome.

IRS proteins contain multiple interaction motifs for phosphatidylinositol-3-kinase (PI3K). For each of these motifs, we found the alpha isoform of the SH2-domain containing regulatory subunit, suggesting that each of the nine (IRS-1) or 11 (IRS-2) is indeed capable of binding to PI3K. This is not self-evident because some of these motifs are imperfect, i.e. they lack methionine in the +3 position. The beta isoform of the regulatory PI3K was found less often, perhaps due to its low expression levels. Likewise, the catalytic subunits of PI3K were frequently identified due to their tight association with the regulatory subunits.

The protein tyrosine phosphatase SHP2 and the adaptor protein Grb2 are also well-established IRS interactors. Here we mapped the specific sites on IRS-1 and IRS-2 that these effector proteins can dock to (7 and 4 sites for SHP2, respectively and 2 and 3 for Grb2; Fig. 4 and Table 1). In addition, the inositol-5-phosphatase SHIP-2 and the Ras GTPase activating protein RasGAP can bind to various sites on IRS-1 and IRS-2. In contrast, the adaptor molecules Crk and Crk-like protein were pulled down by only one specific tyrosine in each IRS-protein. Interestingly, this tyrosine occurs at a similar position (Y0760 in IRS-1 and Y0734 in IRS-2 in the mouse sequence) but the binding sequences are very different. Apart from Crk and Crk-like protein, the site in IRS-2, Y0734, attracts many different proteins simultaneously. Even though our washing steps were optimized to detect direct binders to the phosphorylated bait peptide, it is possible that some of the detected interactions occur via indirect binding such as noted above for the catalytic subunits of PI3K.
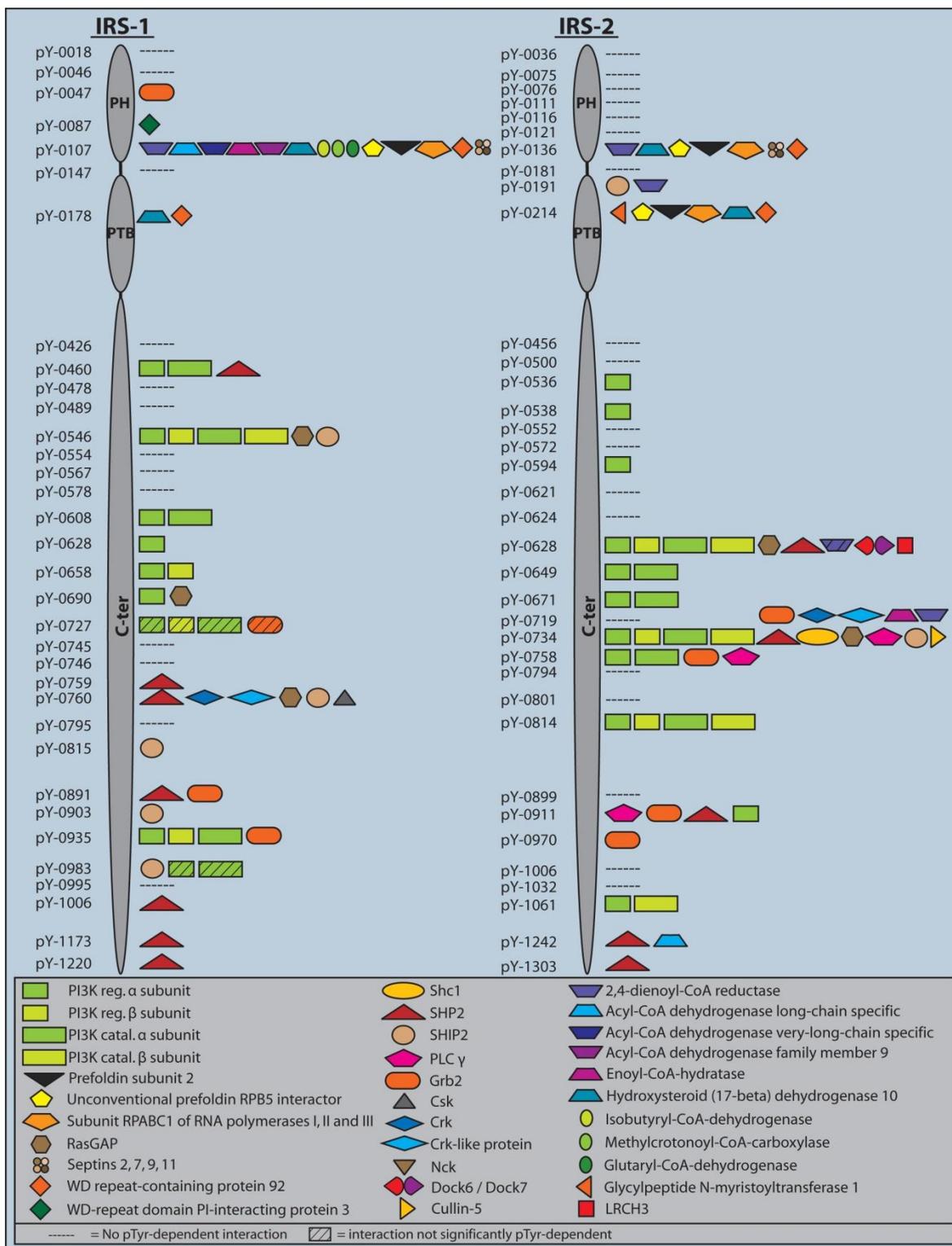
Fig. 4. The phosphotyrosine interactome of IRS-1 and IRS-2. pTyr-specific interaction partners obtained in peptide pulldown experiments are depicted as symbols along the primary structure of IRS-1 and IRS-2. Detailed data for every pulldown experiment is provided in Table 1.

Most interactors were recruited by both IRS-proteins. However, cases of exclusive interactions also occurred. Phospholipase C gamma (PLCγ), for example, docks to several sites in IRS-2 but none in IRS-1 in our experiment. Similarly, Shc, Cullin-5, LRCH3, DOCK-6 and -7 as well as glycylpeptide N-myristoyltransferase 1 were

48

exclusively recruited by IRS-2 derived peptides. IRS-1-specific binders include Csk, WD-repeat domain phosphoinositide-interacting protein 3, and several enzymes involved in fatty and amino acid degradation. Detailed results of all pulldown experiments are listed in Table 1.

The ubichinol-cytochrome c reductase complex chaperone CBP3 homolog (Swiss-Prot Q9CWU6) was exclusively and repeatedly found in pulldowns with the two homologous sites IRS-1 Y0107 and IRS-2 Y0136 with a slightly elevated ratio, but in most cases it was not significantly phosphorylation-dependent (and is hence not listed in Table 1). It could therefore be an indirect binder recruited via one of the fatty acid metabolizing enzymes. The protein tyrosine kinase Fyn has been reported as an IRS-2 interactor previously (46), and we similarly found consistent but relatively weak ratios when binding to Y0734 of IRS-2.

The phosphotyrosine interactome of InsR, IGF1R and IRR – As expected, the receptors yielded fewer interactors than the IRS proteins (Fig. 5). InsR and IGF1R share a very similar pTyr-interaction profile. However, the IGF1R contains multiple sites for recruitment of SHP2, and a site that binds RasGAP as well as short chain specific acyl-CoA dehydrogenase. The physiological importance of the latter site is questionable, however, because it is not conserved in the human IGF1R sequence. Strikingly, the IRR lacks interactions in its C-terminal part. InsR Y1001 and IGF1R Y0988 can recruit many different proteins, and apparently minor differences in the sequence N-terminal of the pTyr in the corresponding site IRR Y0967 impede some of these binding events. Interestingly, 2,4-dienoyl-CoA reductase shows phosphorylation-dependent binding only for the latter site but is strongly bound by the other sites irrespective of their phosphorylation state. This behavior was observed for a few sites in IRS proteins as well, and is indicated by skew stripes in the protein symbols in Figs. 4 and 5. Binding of IRS-1 and Shc to the NPEY-sites in the receptors was only detectable when the bait sequence was enlarged by six amino acids towards the N-terminal side. This observation can be rationalized on structural grounds: crystal structures of the IRS-1 PTB domain bound to the InsR reveal that a relatively long sequence at the N-terminal side of the pTyr is involved. Residues -8 to -3 form a beta-strand that establishes hydrogen bonds with a beta-strand of the PTB domain. The leucine at position -8 is particularly important and hydrophobic residues from -6 to -8 are favored by Shc and IRS-1 (47, 48). These findings suggest the desirability of providing longer and freely accessible N-terminal regions for sequences that are involved in PTB domain recruitment. However, IRS-1 was still only present with few peptides and IRS-2 was not detected. The PTB domains of Shc and IRS-1 are electrostatically polarized similar to PH domains, and can therefore associate with phosphoinositides in the membrane through their positively charged surface (1). Since these additional and
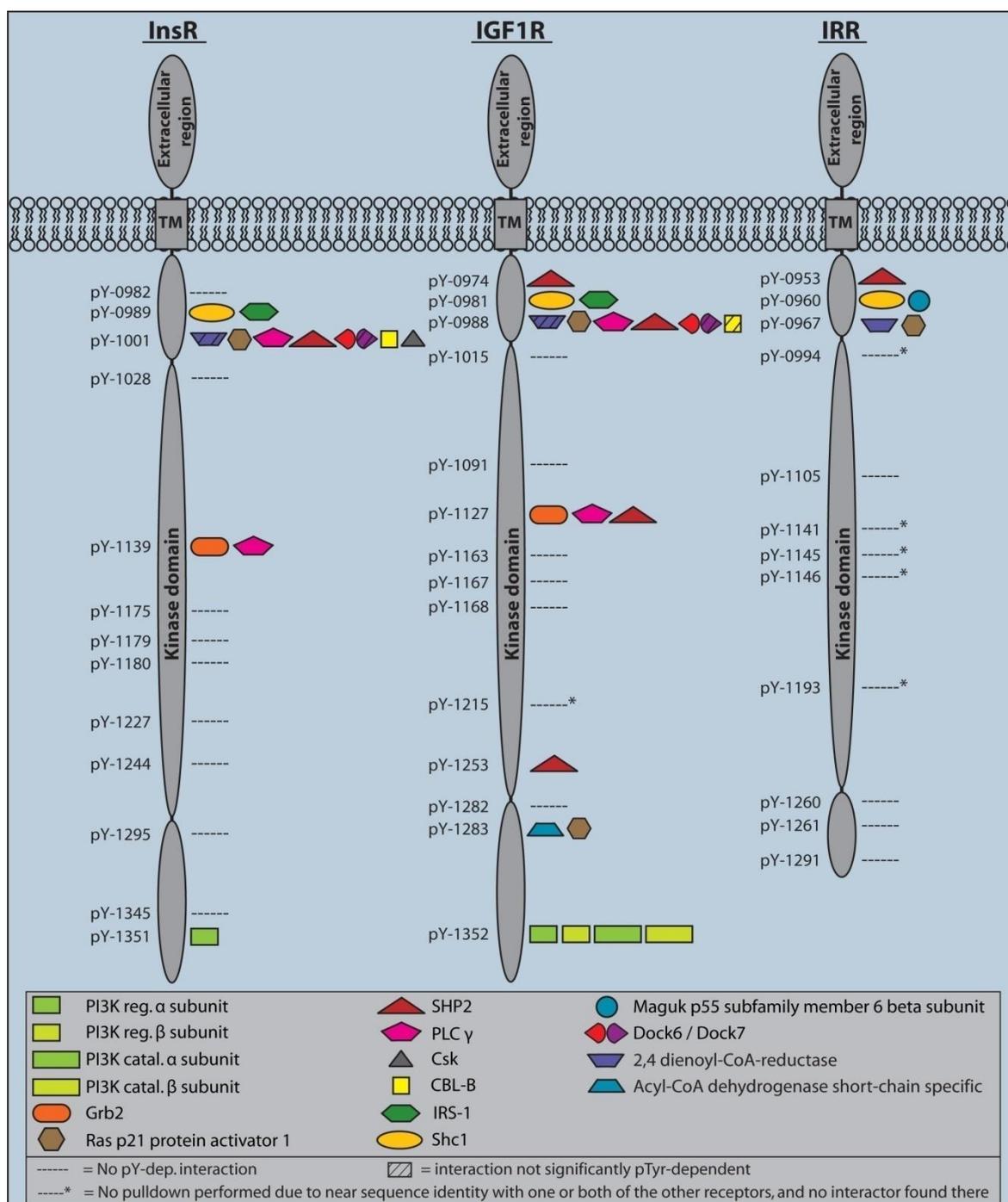
Fig. 5. The phosphotyrosine interactome of InsR, IGF1R and IRR. pTyr-specific interaction partners obtained in peptide pulldown experiments are depicted as symbols along the primary structure of the intracellular regions of the receptors. The cytoplasmic part starts at position 968 for the InsR, 961 for the IGF1R, and 944 for the IRR, according to the Swiss-Prot database. Detailed data for every pulldown experiment is provided in Table 1.

stabilizing interactions are not present in our experiment, reduced affinity (but not reduced specificity) is to be expected. We conclude that PTB domain interactions are more challenging to detect in our assay and that they might need further optimization. In contrast to its binding to InsR and IGF1R, IRS-1 was not detected in the experiment with the bait sequence of the IRR, likely due to the challenges of detection. However, the

membrane-associated guanylate kinase MAGUK p55 subfamily member 6 was identified as an interactor with IRR. We did not find interactors for a fourth receptor, IGF2R also called mannose-6-phosphate receptor that has two cytosolic tyrosines.

One of the advantages of conducting a large number of peptide pulldown experiments in the same system is facilitated discrimination of biochemical noise. We were initially puzzled by the identification of certain unexpected proteins with elevated ratios that passed the significance threshold in many pulldown experiments with completely unrelated bait sequences. We soon realized that the combination of frequent observation across multiple experiments and irreproducibility in repetitions with the same bait sequence classifies this kind of candidates as unlikely to be biologically relevant. Indeed, 7 out of those 8 „uninvited guests' are RNA-binding proteins, providing a ready explanation for their specific binding to phosphorylated peptides: the negative charges on the phosphorylated bait peptides introduce an ion exchange effect with the positively charged RNA-binding proteins, somewhat mimicking their binding to RNA. We list those proteins here, and advise to take care in categorizing such candidates in similar studies: 60S ribosomal protein L11, RNA-binding protein SiahBP homolog, Activated RNA polymerase II transcriptional coactivator p15 precursor, Eukaryotic translation initiation factor 5, Splicing factor U2AF 35 kDa and 65 kDa subunit, Peptidylprolyl isomerase B, ATP-dependent RNA helicase DDX3X.

Doubly and triply phosphorylated motifs – In order to assess the effect of phosphorylation events in close proximity to each other, we performed experiments with doubly or triply phosphorylated bait peptides for some of the sequence stretches that contain neighboring tyrosines (Table 2). We reasoned that combinatorial phosphorylation might either prevent binding events compared to the monophosphorylated counterpart, or even generate new binding interfaces. We indeed encountered some sites whose interaction capability was modified by additional phosphorylations. For example PI3K was not displaced by phosphorylations N-terminal of its binding sequences around IRS-2 Y0538 and InsR Y1351. However, it failed to interact with the IRS-2 Y0628 peptide sequence, which matches to a weaker consensus motif, if positions Y0621 and Y0624 were phosphorylated at the same time. Most strikingly, the triply phosphorylated peptide resembling the kinase activation loop of the InsR recruited several interactors (dedicators of cytokinesis 6 and 7, RasGAP, LRCH3), whereas a doubly and the singly phosphorylated versions did not recruit any proteins (see DISCUSSION).

For precise mapping of differences between mono- and doubly phosphorylated versions, we performed triple SILAC labeling experiments for selected cases (Table 3). In this type of pulldown experiments, non-phosphorylated peptide was incubated with unlabeled cell lysate (Arg0 + Lys0), monophosphorylated peptide was incubated with a medium labeled

Table 2: Interaction partners of doubly and triply phosphorylated pTyr-motifs. The data is presented as in Table 1.

| Protein	Site	Bait sequence	MS-instrument used	Experiments done	Phosphorylation-specific interaction partner	Protein ratio (regular/crossover experiment)	Std-Devs distance from median	Unique peptides identified	Peptides used for quantitation	Known pTyr-binding domain
IRS2	Y0536/Y0538	SGRDGSGGELYGYMSMDRP	LTQ-FT	1	PI3K reg. alpha subunit, P26450	16 +- 6 / 0.14 +- 0.03	3.7 / 2.8	14 / 10	10 / 7	2x SH2
					PI3K reg. beta subunit, O08908	11 / 0.12	3.2 / 3.0	1 / 1	1 / 1	2x SH2
					PI3K catal. alpha subunit, P42337	31 +- 12 / 0.04	4.6 / 4.7	6 / 1	5 / 1	
IRS2	Y0111/Y0116/Y0121	SGKRADAKHKYLIALYTKDEYFAVAAE	LTQ-FT	1	x					
IRS2	Y0621/Y0624/Y0628	SGSSPKVAYNPYPEDYGDIEIG	LTQ-Orbitrap	1	RasGAP, Q91YX7	11 / 0.04 +- 0.02	7.1 / 7.7	11 / 8	1 / 4	SH2
					SHP2, P35235	11 +- 2 / 0.26 +- 0.02	7.1 / 3.2	3 / 3	2 / 3	2x SH2
					Dedicator of cytokinesis protein 7, Q8R1A4	--- / 0.32 +- 0.14	--- / 3.2	--- / 2	--- / 2	
					2,4-dienoyl-CoA reductase, Q9CQ62	1.3 +- 0.2 / 0.9 +- 0.1	0.7 / 0.3	3 / 5		
INSR	Y1345/Y1351	SGSSLSIKRTYDEHIPYTHMNGG	LTQ-FT	1	PI3K reg. alpha subunit, P26450	5.8 +- 0.6 / 0.18 +- 0.03	3.9 / 3.1	6 / 2	5 / 2	2x SH2
					PI3K reg. beta subunit, O08908	6.3 / ---	4.1 / ---	1 / ---	1 / ---	2x SH2
INSR	Y1179/Y1180	SGFGMTRDIYETDYYRKGGKG	LTQ-Orbitrap	1	x					
INSR	Y1175/Y1179/Y1180	SGFGMTRDIYETDYYRKGGKG	LTQ-Orbitrap	2	Dedicator of cytokinesis protein 6, Q8VDR9	16 +- 4 / 0.08 +- 0.02	5.7 / 4.7	11 / 16	5 / 14	
					Dedicator of cytokinesis protein 7, Q8R1A4	9 +- 2 / 0.15 +- 0.04	4.5 / 3.6	20 / 30	16 / 24	
					RasGAP, Q91YX7	--- / 0.18	--- / 3.2	--- / 1	--- / 1	SH2
					LRCH3, Q8BVU0	9 +- 1 / 0.12 +- 0.02	4.5 / 3.9	2 / 6	2 / 5	

Table 3: Triple-labeling pulldown experiments to quantitatively assess combinatorial effects of two tyrosine phosphorylation events in close proximity. The data is presented as in Table 1. The phosphorylation site of the mono-phosphorylated peptide incubated with the medium label state is underlined. Values of mono-P to non-P, double-P to non-P, and double-P to mono-P are separated by slashes. The corresponding crossover values are preceded by a double slash and are listed in the order mono-P to double-P, non-P to double-P, and non-P to mono-P. The table also contains data from one pulldown with normal double-labeling, where an N-terminally prolonged bait sequence was employed. For ease of comprehension, the binding characteristics are summarized in the last column.

Protein	Site	Bait sequence	MS-instrument used	Experiments done	Phosphorylation-specific interaction partner	Protein ratio (regular/crossover experiment)	Std-Devs distance from median	Unique peptides identified	Peptides used for quantitation	Known pTyr-binding domain	Interaction with Mono-P	Interaction with Double-P
IRS1	Y0546/Y0554	SGSSVASIEEYTEMMPAAYPPGGGS	LTQ-Orbitrap	1	PI3K reg. alpha subunit, P26450	22 +- 13 / 15 +- 9 / 0.69 // 1.5 +- 0.26 / 0.05 +- 0.03 / 0.03	5.9 / 5.3 / 0.9 // 1.0 / 6.2 / 5.2	20 / 14	10 / 7	2x SH2	yes	yes
					PI3K reg. beta subunit, O08908	9 +- 2 / 6 +- 0.5 / 0.64 // 1.4 +- 0.5 / 0.04 +- 0.03 / 0.03	4.3 / 3.5 / 1.1 // 0.9 / 6.5 / 5.3	3 / 2	3 / 2	2x SH2	yes	yes
					PI3K catal. alpha subunit, P42337	15 +- 10 / 13 +- 8 / 0.88 // 1.3 +- 0.3 / 0.05 +- 0.03 / 0.04	5.2 / 5.0 / 0.3 // 0.6 / 6.0 / 4.9	24 / 24	16 / 16		yes	yes
					PI3K catal. subunit beta, Q8BTI9	4.3 +- 0.4 / 2.7 +- 1.2 / 0.63 // 1.6 / 0.12 / 0.08	2.8 / 2.0 / 1.2 // 0.4 / 4.3 / 3.8	3 / 1	1 / 1		yes	yes
					RasGAP, Q91YX7	11 +- 7 / 8 +- 5 / 0.70 // 1.4 +- 0.2 / 0.09 +- 0.04 / 0.06	4.5 / 3.9 / 0.9 // 0.9 / 4.9 / 4.2	4 / 7	4 / 5	SH2	yes	yes
					SHIP-2, Q6P549	9 +- 3 / 11 +- 3 / 1.1 // 0.9 +- 0.2 / 0.10 +- 0.07 / 0.10	4.3 / 4.6 / 0.3 // 0.1 / 4.6 / 3.4	12 / 12	8 / 10	SH2	yes	yes
IRS1	Y0546/Y0554	SGSSVASIEEYTEMMPAAYPPGGGS	LTQ-Orbitrap	1	PI3K reg. alpha subunit, P26450	0.75 +- 0.40 / 14 +- 7 / 19 // 0.08 + 0.05 / 0.06 +- 0.04 / 0.74	0.7 / 5.6 / 5.3 // 4.9 / 6.4 / 0.7	17 / 19	9 / 11	2x SH2	no	yes
					PI3K reg. beta subunit, O08908	0.70 / 10 / 15 // 0.06 / 0.11 / 1.85	0.8 / 4.9 / 4.8 // 5.4 / 4.9 / 1.4	4 / 2	1 / 1	2x SH2	no	yes
					PI3K catal. alpha subunit, P42337	0.63 +- 0.25 / 21 +- 15 / 33 // 0.05 +- 0.04 / 0.04 +- 0.04 / 0.85	0.9 / 6.3 / 6.2 // 5.6 / 7.0 / 0.0	20 / 25	7 / 7		no	yes
					PI3K catal. subunit beta, Q8BTI9	1.08 / 2.8 / 2.6 // NQ	0.2 / 2.2 / 1.7 // --	2 / 1	1 / NQ		no	yes
					RasGAP, Q91YX7	1.5 +- 0.6 / 11 +- 6 / 7 // 0.07 +- 0.03 / 0.07 +- 0.04 / 1.0	0.9 / 5.0 / 3.5 // 5.0 / 5.9 / 0.0	4 / 6	3 / 3	SH2	no	yes
					SHIP-2, Q6P549	1.3 +- 0.2 / 11 +- 4 / 8.5 // 0.09 +- 0.02 / 0.10 +- 0.06 / 1.04	0.6 / 7.6 / 3.7 // 4.3 / 5.2 / 0.1	14 / 16	9 / 11	SH2	no	yes
IRS1	Y0759/Y0760	SGPQHKPVLSYYSLPRSF	LTQ-Orbitrap	1	SHP2, P35235	3.4 +- 2.2 / 17 +- 13 / 5 // 0.12 +- 0.03 / 0.06 +- 0.04 / 0.46	1.9 / 4.0 / 2.4 // 2.7 / 4.2 / 1.1	19 / 22	12 / 9	2x SH2	little	yes
					RasGAP, Q91YX7	0.78 +- 0.31 / 20 +- 12 / 26 // 0.02 +- 0.02 / 0.03 +- 0.03 / 1.07	0.4 / 4.2 / 4.7 // 4.8 / 5.3 / 0.1	26 / 28	14 / 6	SH2	no	yes
					Crk, Q64010	1.41 +- 0.26 / 27 +- 7 / 19 // 0.05 +- 0.01 / 0.03 +- 0.03 / 0.55	0.5 / 4.6 / 4.3 // 3.8 / 5.1 / 0.8	12 / 12	3 / 3	SH2	no	yes
					Crk-like protein, P47941	1.05 +- 0.45 / 14 +- 5 / 13 // 0.08 +- 0.01 / 0.05 +- 0.03 / 0.69	0.1 / 3.6 / 3.7 // 3.3 / 4.2 / 0.5	11 / 11	8 / 6	SH2	no	yes
					SHIP-2, Q6P549	0.96 +- 0.37 / 14 +- 6 / 14 // 0.05 +- 0.03 / 0.06 +- 0.07 / 1.39	0.1 / 3.6 / 3.9 // 4.0 / 3.9 / 0.5	22 / 22	10 / 8	SH2	no	yes
					Grb2, Q60631	1.12 +- 0.63 / 10 +- 6 / 9 // 0.08 +- 0.02 / 0.16 +- 0.06 / 1.96	0.2 / 3.2 / 3.2 // 3.2 / 2.6 / 0.9	5 / 5	4 / 4	SH2	no	yes
					Phospholipase C gamma, Q62077	1.1 +- 0.55 / 6.4 +- 3.6 / 5.9 // 0.11 +- 0.03 / 0.11 +- 0.11 / 1.0	0.1 / 2.6 / 2.6 // 2.9 / 3.2 / 0.1	8 / 10	6 / 6	2x SH2	no	yes
					CSK, P41241	0.79 / 6.6 / 8.3 // 0.10 / 0.15 / 1.58	0.4 / 2.6 / 3.1 // 3.0 / 2.7 / 0.6	2 / 2	1 / 1	SH2	no	yes
					Nck2, O55033	0.35 +- 0.09 / 4.2 +- 0.6 / 11 // 0.09 / 0.04 / 0.40	1.6 / 2.0 / 3.6 // 3.0 / 4.7 / 1.3	3 / 3	2 / 1	SH2	no	yes
					PI3K reg. alpha subunit, P26450	1.68 / 8.6 / 5.1 // 0.09 / 0.22 / 2.28	0.8 / 3.0 / 2.4 // 3.0 / 2.2 / 1.2	3 / 1	1 / 1	2x SH2	no	yes
					CBL-B, Q3TTA7	--- / --- // 0.20 +- 0 / 0.27 +- 0.09 / 1.35	--- / --- / --- // 2.1 / 1.9 / 0.4	--- / 2	--- / 2	SH2	no	little

Protein	Site	Bait sequence	MS-instrument used	Experiments done	Phosphorylation-specific interaction partner	Protein ratio (regular/crossover experiment)	Std-Devs distance from median	Unique peptides identified	Peptides used for quantitation	Known pTyr-binding domain	Interaction with Mono-P	Interaction with Double-P
IRS1	Y0759/Y0760	SGPQHKPVLSYYSLPRSF	LTQ-Orbitrap	1	SHP2, P35235	9 +- 5 / 16 +- 8 / 1.8 // 0.50 +- 0.05 / 0.07 +- 0.03 / 0.14	2.8 / 3.4 / 1.0 // 1.3 / 3.3 / 1.4	22 / 27	11 / 12	2x SH2	yes	yes
					RasGAP, Q91YX7	7 +- 6 / 26 +- 20 / 3.7 // 0.24 +- 0.06 / 0.09 +- 0.10 / 0.39	2.5 / 4.1 / 2.2 // 2.6 / 3.0 / 1.1	28 / 34	11 / 17	SH2	little	yes
					Crk, Q64010	16 +- 3 / 24 +- 4 / 1.54 // 0.68 +- 0.05 / 0.05 +- 0.03 / 0.07	3.5 / 3.9 / 0.8 // 0.7 / 3.7 / 3.2	12 / 14	5 / 5	SH2	yes	yes
					Crk-like protein, P47941	17 +- 6 / 24 +- 11 / 1.47 // 0.82 +- 0.02 / 0.06 +- 0.02 / 0.07	3.6 / 4.0 / 0.7 // 0.3 / 3.5 / 3.4	10 / 14	5 / 4	SH2	yes	yes
					SHIP-2, Q6P549	9 +- 5 / 30 +- 15 / 3.28 // 0.28 +- 0.03 / 0.08 +- 0.05 / 0.28	2.8 / 4.2 / 2.1 // 2.3 / 3.2 / 1.6	18 / 29	11 / 8	SH2	little	yes
					Grb2, Q60631	1.13 +- 0.05 / 8 +- 1 / 7 // 0.14 +- 0.04 / 0.15 +- 0.03 / 1.05	0.1 / 2.6 / 3.5 // 3.6 / 2.4 / 0.1	3 / 8	3 / 6	SH2	no	yes
					Phospholipase C gamma, Q62077	0.63 +- 0.23 / 10 +- 3 / 16 // 0.04 +- 0.02 / 0.06 +- 0.07 / 1.7	0.6 / 2.8 / 4.8 // 3.4 / 3.5 / 0.7	6 / 8	3 / 4	2x SH2	no	yes
					CSK, P41241	6 / 7 / 1.18 // 0.98 +- 0.29 / 0.15 +- 0.05 / 0.15	2.3 / 2.5 / 0.3 // 0.0 / 2.4 / 2.4	1 / 2	1 / 2	SH2	yes	yes
					Nck2, O55033	0.64 +- 0.41 / 7 +- 3 / 11 // 0.12 / 0.04 / 0.30	0.6 / 2.5 / 4.2 // 3.9 / 4.2 / 1.5	2 / 4	2 / 1	SH2	no	yes
					PI3K reg. alpha subunit, P26450	1 / 11 / 11 // 0.14 / 0.22 / 1.6	0.0 / 3.0 / 4.2 // 3.7 / 1.9 / 0.6	1 / 1	1 / 1	2x SH2	no	yes
					CBL-B, Q3TTA7	0.67 / 3 / 4.5 // 0.25 / 0.46 / 1.85	0.5 / 1.4 / 2.6 // 2.6 / 1.0 / 0.8	1 / 1	1 / 1	SH2	no	little
INSR	Y0982/Y0989	SGDGPMGPLYASSNPEYLSASDV	LTQ-Orbitrap	1	Shc1, P98083	11 +- 2 / 1.2 / 0.11 // 6 +- 0.5 / 1.3 +- 1.4 / 0.23	3.4 / 0.5 / 3.5 // 2.8 / 0.6 / 2.9	4 / 4	2 / 2	SH2, PTB	yes	no
					IRS-1, P35569	4 / 4 / 1 // --- / --- / ---	1.9 / 3.3 / 0.1 // -- / --- / ---	1 / ---	1 / ---	PTB	yes	yes
					SHIP-2, Q6P549	1.7 +- 0.7 / 4.3 +- 2.2 / 2.5 // 0.33 / 0.09 / 0.28	0.8 / 3.6 / 1.5 // 1.8 / 1.3 / 1.1	4 / 1	3 / 1	SH2	little	yes
IGF1R	Y0974/Y0981	SGRLGNGVLYASVNPEYFSAADV	LTQ-Orbitrap	1	Shc1, P98083	12 +- 12 / 10 +- 11 / 0.83 // 1.1 +- 0.1 / 0.07 +- 0.04 / 0.07	5.4 / 5.6 / 0.4 // 0.1 / 5.4 / 4.9	6 / 6	6 / 4	SH2, PTB	yes	yes
					IRS-1, P35569	7 / 7 / 1 // 0.6 / 0.08 / 0.13	4.2 / 4.8 / 0.1 // 1.2 / 5.3 / 3.6	1 / 1	1 / 1	PTB	yes	yes
					SHP2, P35235	1.01 +- 0.33 / 22 +- 13 / 22 // 0.06 +- 0.02 / 0.08 +- 0.05 / 1.5	0.0 / 7.5 / 6.7 // 6.2 / 5.1 / 0.7	12 / 16	4 / 4	2x SH2	no	yes
IRR	Y0953/Y0960	SGRKRNSTLYTSVNPEYFSASHM	LTQ-Orbitrap	1	Shc1, P98083	--- / --- / --- // 0.72 / 0.03 / 0.04	--- / --- / -- // 0.7 / 5.0 / 3.9	--- / 1	--- / 1	SH2, PTB	yes	yes
					MAGUK p55 subfamily member 6, Q9JLB0	5.7 +- 0.03 / 1.3 / 0.23 // 3.6 +- 0.8 / 1.0 +- 0.9 / 0.27	2.5 / 0.4 / 3.0 // 2.7 / 0.0 / 1.6	4 / 3	2 / 2		yes	no
					SHP2, P35235	1.59 +- 1.18 / 36 +- 17 / 22 // 0.04 +- 0.02 / 0.07 +- 0.11 / 1.6	0.7 / 5.2 / 6.4 // 6.7 / 3.9 / 0.6	10 / 8	4 / 3	2x SH2	no	yes
IRR	Y0960_longer	SGRKRNSTLYTSVNPEYFSASHM	LTQ-Orbitrap	1	Shc1, P98083	22 +- 7 / 0.06 +- 0.05	6.4 / 5.4	7 / 14	5 / 11	SH2, PTB	yes	yes
					MAGUK p55 subfamily member 6, Q9JLB0	11 +- 3 / 0.12 +- 0.03	5.0 / 3.9	4 / 6	3 / 5			yes

state (Arg6 + Lys4) and doubly phosphorylated peptide was incubated with heavily labeled lysate (Arg10 + Lys8). Peptides appear as triplets in mass spectra, enabling accurate comparison between the three conditions. Using this assay we encountered differential effects of a phosphorylation located seven residues upstream of the NPEY motif in the receptors. The recruitment of Shc was not affected by this additional phosphorylation within the binding motif of the PTB domain in case of IGF1R and IRR. However, binding of Shc to the respective sequence in the InsR proved to be sensitive to this change, and the interaction was abolished (Fig. 6). In case of IRS-1 Y0546/Y0554, the second phosphorylation neither increased nor diminished the binding of interactors. On the other hand, for IRS-1 Y0759/Y0760 the double phosphorylation induced novel interactions. While SHP2, Csk, Crk and Crk-like protein were equally attracted by mono- and doubly phosphorylated version, the binding of RasGAP and SHIP-2 was enhanced by the additional phosphorylation. Some proteins were even exclusively recruited by the doubly phosphorylated sequence, namely Nck2, PLCγ, Grb2, PI3K and Cbl-B.
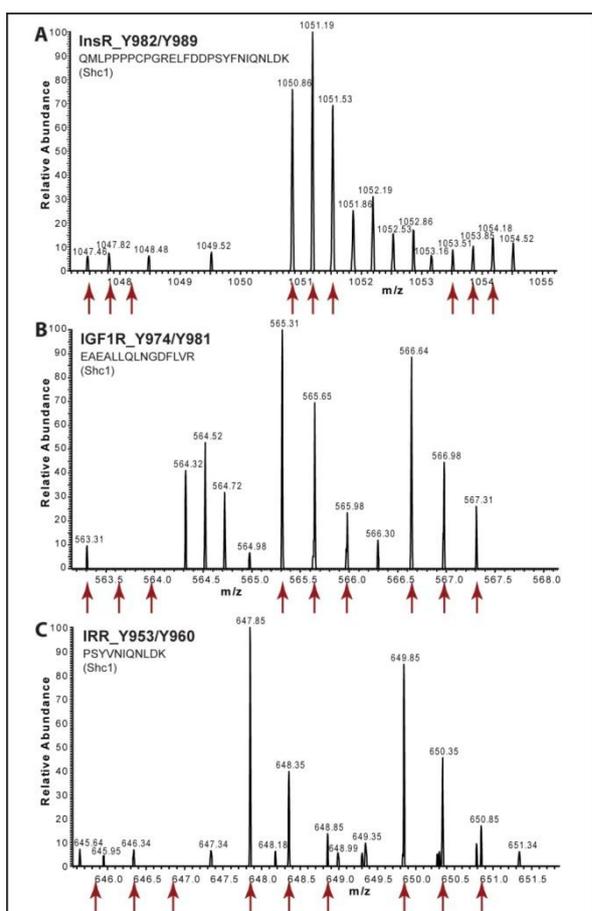
Fig. 6. Triple labeling pulldown reveals distinct combinatorial effects of double phosphorylation within the NPEY-motifs in the insulin receptor family. (A) When phosphorylated at the tyrosine within the NPEY-motif, the InsR recruits Shc, as demonstrated by the high ratio between the medium-label state and the unlabeled state. If the tyrosine located seven amino acids further N-terminal also carries a phosphorylation, the interaction with Shc is abolished. (B) and (C) When IGF1R or IRR exhibit this phosphorylation pattern, the interaction between Shc and the receptor is not influenced by the second phosphorylation event (Table 3.)

DISCUSSION

Capabilities and limitations of quantitative proteomics combined with peptide bait fishing – In recent years, MS-based quantitative proteomics has become a powerful and versatile tool for comparing complex proteomes and their modifications (49-51), and in particular, metabolic labeling by the SILAC technique has proven to be a very accurate method for this purpose (52). In addition, interaction proteomics using quantitative readouts has developed substantially during recent years (37). Our strategy aimed to uncover pTyr-dependent binding events on proteins that serve as interaction platforms in the insulin signaling pathway. To achieve this, we employed synthetic phosphorylated peptides as baits in pulldown experiments from whole cell lysates. An advantage of this approach is the application of sequence defined baits yielding site-specific interaction information. In comparison to purely in vitro experiments such as chip-based assays, the unbiased nature of the mass spectrometric approach enables discovery of novel interactors, unconstrained by prior hypotheses. Furthermore, our experiment involves the original expression level of the interacting proteins in the competitive environment of the whole proteome of the cell type studied. This mimics the true situation within a cell much more closely than in vitro assays do. Importantly, using quantitative proteomics enabled straightforward discrimination between specific phosphorylation-dependent interactors and background binders. However, even though all significant interactors found in our experiments are per definition specific for the phosphorylated form of the peptide, they can only take place in vivo if the site is accessible, becomes phosphorylated by a kinase, and if the binding partner is available for the interaction. Subcellular localization and tissue-specific differences will often play a decisive role in this respect. Affinity, relative expression levels, effects of the neighboring sequence and the involvement of additional interaction domains co-determine the actual occurrence of the pTyr-mediated interaction in vivo. For most interacting proteins there is additional information, such as known subcellular localization and the presence of known interaction domains that can help in determining the likelihood that the interaction is biologically relevant. Taken together our results represent a catalogue of interactions that are proven to be biophysically possible and – for interactions that fulfill the above criteria – likely to occur in vivo. In this regard, the present study is complemented by our recent, systematic investigation of the tyrosine phosphoproteome induced by insulin and IGF-1 stimulation (53). Below we first discuss the common and distinct interactomes of the IRS proteins and the three receptors that involve classical SH2 or PTB domain mediated interactions. We then discuss interactions with proteins not containing these binding modules.

Common interactors of IRS-1 and IRS-2 – Most of the interaction partners identified in this work contain an SH2 domain that mediates binding to the pTyr-sequence.

Recognition motifs for many SH2 domains have been determined by peptide library screening (3, 54). Despite their large number, SH2 domains do not bind to arbitrary sequences surrounding the pTyr but instead bind only to a subset of consensus motifs. Conversely, the same peptide sequence can bind to different SH2 domains (3), which is reflected here by the recruitment of a range of binding partners in many of our peptide pulldowns.

Among the proteins recruited to both IRS proteins, PI3K bound to the largest number of different sites. PI3K is a critical node in insulin signaling (28), which activates the Akt kinase through generation of phosphatidylinositol-3,4,5-triphosphate at the plasma membrane. Akt in turn has several downstream pathway branches important for growth and metabolism. Our prevailing observation of the alpha isoform of PI3K agrees with its predominant role in insulin signaling (55). The adaptor protein Grb2, which interacts with multiple phosphorylated peptides from IRS-1 and IRS-2, triggers signaling via the activation of the Ras to MAP-kinase axis. Intriguingly, the suppressor of Ras signaling RasGAP binds to multiple sites in IRS proteins. This novel finding suggests that the activity of RasGAP is regulated by recruitment to the IRS platform.

The tyrosine phosphatase SHP2 plays a dynamic role in the activity of the pathway by both providing negative feedback for PI3K activation via dephosphorylating pTyr-sites on IRS proteins that bind PI3K (56) and stimulating mitogenic signaling via insulin (57-59). Here we systematically map binding sites of SHP2 on the IRS signaling platforms.

The lipid phosphatase SHIP-2 dephosphorylates phosphatidylinositol phosphates at the 5´-position. Even though SHIP-2 action does not abolish the ability of PIPs to activate Akt, it clearly antagonizes insulin signaling (60). Here we mapped five interaction sites of SHIP-2 to IRS-1 and two to IRS-2.

Crk family adaptors mediate protein complex formation in various signaling pathways, and have been shown to influence mitogenesis, cytoskeletal rearrangements and insulin-stimulated glucose uptake (61, 62). While binding to IRS-1 and IRS-2 has been observed before (46), our data provides the specific docking sites.

Differential interactions between IRS-1 and IRS-2 – The physiological effects of insulin can vary greatly in different target tissues due to different modes of signal transmission and modulation inside cells. For example in liver synthesis of glycogen, proteins and lipids is triggered along with inhibition of hepatic glucose production and very low density lipoprotein (VLDL) secretion. In muscle, the response mainly involves glucose uptake and glycogen synthesis (28, 63). Previously many of these differences were attributed to stronger activation of IRS-1 or IRS-2. However, differences in binding partners between IRS-1 and IRS-2 are not only determined at the level of their primary sequence motifs, but

also by differing time-courses of phosphorylation, dose-response curves towards hormones, and intracellular compartmentalization (24). The expression level of IRSs and their interactors also play a role, and those factors can even result in opposite functions of an IRS protein depending on the cell type investigated (64). Illustrating this complexity, a recent study found that IRS-1 and IRS-2 trigger the same downstream signals, but IRS-1 was more active in the postprandial state, whereas IRS-2 was employed during fasting (65).

Given the extensive and often contradictory investigations into the differential roles of IRS-1 and IRS-2, our experiment at the least provides a large-scale data set delineating potential common and differential pTyr mediated interactors. In general, our results reinforce the notion of a large overlap in the signaling capabilities between IRS-1 and IRS-2. Most pTyr-dependent interaction partners involved in growth and metabolic signaling were recruited by both proteins. However, we also detected clear differences in the interactomes. For example, as noted above SHIP-2 binds to substantially more sites in IRS-1. Exclusive binding to IRS-1 was observed for Csk (C-terminal Src-kinase), which phosphorylates members of the Src family of kinases at C-terminal tyrosines, inhibiting their activity. Csk has been shown to bind to IRS-1 via its SH2-domain (66), whereas interaction or lack thereof with IRS-2 has not previously been noted. The Csk-mediated inactivation of Src-family kinases leads to an insulin-dependent decrease of pTyr on FAK and Paxillin, which reduces actin stress fibers and allows insulin to influence the reorganization of the cytoskeleton. Our observation that Csk is recruited only to IRS-1 derived peptides and not to IRS-2 derived peptides is in line with the observation that IRS-1, but not IRS-2, mediates actin remodeling in myotubes (24).

Shc and PLCγ were only found as interactors of IRS-2. Shc is an adaptor protein best known for its ability to bind Grb2 and accordingly promote Ras activation. PLCγ is the only PLC that contains SH2 domains. Binding via its N-terminal SH2 domain to a tyrosine kinase leads to phosphorylation of PLCγ on several tyrosines, thereby activating it and ultimately stimulating PKC activation (67). PLCγ has previously been shown to bind to IRS proteins in an insulin-dependent manner (46). Our finding that PLCγ is recruited to three sites in IRS-2 and none in IRS-1 suggests one mechanism for specific functions of the two signaling platforms.

The pTyr-interactome of InsR, IGF1R and IRR – The insulin signaling pathway is activated by the insulin receptor (InsR) as well as the insulin-like growth factor 1 receptor (IGF1R). Signaling through InsR has metabolic functions whereas IGF1R, as its name implies, has mainly growth and mitogenic effects. However, there is also a significant level of crosstalk between these functions. Experiments with chimeric receptors have shown that it is mainly the intracellular part that determines specificity, not the extracellular ligand

binding domains (68). The most apparent difference between the interactomes of IGF1R and InsR as determined here is the larger number of SHP2 binding sites to IGF1R (four vs. one). This observation helps to explain the stimulatory role of this phosphatase in growth signaling. Both are able to recruit Cbl, even though the interaction was only weakly pTyr-dependent in the case of the IGF1R. Cbl is an E3 ligase and allosterically activates an E2-enzyme for ubiquitinylation, which leads to the internalization and subsequent degradation of the receptor (69).

Insulin receptor related receptor (IRR) is thought to have a less prominent role in the insulin/IGF pathway and few functions have been described. We found that IRR is less prone to participate in pTyr-dependent binding events and detected just three interactors with SH2 domains. Only the membrane proximal sites recruit interaction partners. Our failure to detect IRS-1 binding to IRR might be due to very weak binding of the IRS PTB domain under the conditions of this assay (see RESULTS).

The NPEY-motif containing phosphopeptide derived from IRR did bind to MAGUK p55 (membrane-associated guanylate kinase MAGUK p55 subfamily member 6 – also termed PALS2), one of the non SH2 domain containing proteins otherwise discussed below. Little functional information is known about MAGUK p55 but it appears to be involved in proper targeting of receptors in polarized cells, as well as in stabilization of receptors and acting as signaling scaffold in non-polarized cells (70). This potential interaction would be unique to the IRR as compared to its other family members.

Doubly and triply phosphorylated motifs – Another interesting difference between the receptors became apparent when studying the combinatorial effects of an additional phosphorylation at the tyrosine located seven residues N-terminal of the NPEY-motif. In the doubly phosphorylated peptide, this site bound to SHIP-2 and SHP2 (to InsR and IGFR + IRR, respectively) just as in the case of the monophosphorylated peptide (Table 3). Likewise the doubly phosphorylated peptide still bound IRS-1. However, binding of Shc to the NPEY-motif was abrogated specifically for the InsR-sequence derived peptide. Since Shc activates MAP kinase signaling when associated with tyrosine kinases this is a possible mechanism of differential control – towards metabolic signaling – between the receptors.

In most other cases of combinatorial phosphorylation, we did not observe any changes. The kinase activation loop, interestingly, yielded several binding partners in its triply phosphorylated state. This may ensure that the interactions can only take place after full activation of the kinase. Selective binding to the triply phosphorylated sequence of the activation loop has previously been reported for APS and IRS-2 (71, 72). The sequence in

this region is identical between InsR and IGF1R and only has one Ile/Val substitution in IRR, therefore this result likely applies to all of the receptors.

IRS-1 has two tyrosines adjacent to each other (Y0759/Y0760). In pulldowns with singly phosphorylated peptides pY0760 interacts with SHP2 as well as five other proteins containing SH2 domains, however, pY0759 only interacts with SHP2. The doubly phosphorylated peptide still binds to all interactors and strikingly to several additional ones. These proteins - Grb2, PLCγ, Nck2 and Cbl – all contain SH2 domains and for two of them binding to doubly phosphorylated peptides has previously been described. PLCγ, for example, usually binds hydrophobic sequences with its C-terminal SH2-domain, but following a conformational change it creates a second pTyr-binding pocket and can then also bind doubly phosphorylated motifs, such as pYESPpYAD in the activated Syk tyrosine kinase (73). For Grb2, which binds selectively to the pYpY-sequence here, a similar observation has been made for a pYpY-motif in Shc (74). This doubly phosphorylated peptide is the only one binding Nck2 in our study. Nck has previously been observed to bind to IRS-1 in an insulin-dependent manner, and is engaged in cytoskeletal rearrangements and mitogenic signaling (75, 76), but the mode of its binding has not been described. The doubly phosphorylated version clearly bound the Src-kinase Fyn, albeit not in a significantly pTyr-dependent manner. Fyn has been reported to bind to IRS-1 and IRS-2 (46) and to tyrosine phosphorylated Cbl after insulin stimulation (77).

pTyr-binding independent of SH2 and PTB domains – The interactions described above are all readily explained by pTyr-SH2 or -PTB domain mediated binding. Most of them recapitulate and extend known interactions or can be readily understood in terms of the biology of this well studied pathway. A further goal of our experiments was to possibly detect novel interaction partners and interaction modes, not mediated by known interaction domains or interactions with known members of the insulin signaling pathway. Our results indeed contain a number of such interaction partners. Since indirect binding is possible in our experimental setup, they might have been recruited as secondary interactors. For some of the investigated sites, however, we exclusively found non-SH2 domain containing proteins. Furthermore, even for those sites where SH2-containing proteins might indirectly recruit other proteins, this may be unlikely since it should happen at all sites with which they interact. We therefore consider it unlikely that indirect binding is the only explanation for these binders.

As demonstrated by the relatively late discovery of the Cbl SH2 domain, genome annotation algorithms can sometimes miss an SH2 domain if it has an atypical sequence (78). More importantly, alternative pTyr-binding domains have recently been described. The C2 domain of PKCδ has been shown to bind pTyr (79). A very recent report by the Cantley group (also using the SILAC technology) showed that a yet uncharacterized

region in pyruvate kinase M2 binds specifically to pTyr containing peptides (11). This was interpreted to provide a direct link to mitochondria-based metabolic functions. In total we found 21 „non-classical' interaction partners and based on co-occurrence estimate that at least ten of these bind directly to pTyr containing bait peptides.

Note that some of the interactions might not necessarily occur in insulin or IGF1 signaling, because IRS-1 and IRS-2 can also be engaged by other receptors. Those include the prolactin (80), androgen (81), growth hormone (80, 82) and vascular endothelial growth factor (VEGF) receptors (83), as well as members of the integrin receptor family (84, 85) and several cytokine receptors (86).

Cytosolic interactors of the IRS proteins included two WD repeat proteins. The function of WDR92 (WD repeat-containing protein 92) is still elusive, but it has been suggested to act as a modulator of apoptosis (87). The function of WD-repeat domain phosphoinositide-interacting protein 3, which bound to a phosphopeptide derived from IRS-1, is also not yet known. It potentially binds to membrane phosphoinositides in addition to the interaction with Y0087 in the PH domain of IRS-1, which is membrane-associated, and thus could enhance membrane anchoring of IRS-1 after stimulation. Several septins were found as interactors of both IRS proteins. Septins behave like filaments or scaffold proteins and play a role in cytokinesis and in cytoskeleton and membrane organization (88).

Among the proteins exclusively interacting with peptides derived from IRS-2, Cullin-5 is a scaffold protein and part of an E3 ubiquitin ligase complex. As such it might be involved in proteasomal degradation of IRS proteins. Another cullin, cullin-7, triggers proteasomal degradation of IRSs (89). LRCH3 (leucine-rich repeat and calponin homology domain-containing protein 3) is a protein with unknown function, and DOCK-6 and -7 (dedicator of cytokinesis) act as GEFs (guanine nucleotide exchange factor) for small G-proteins of the Rho-family according to their UniProt annotation. This connects them to regulation of cytoskeletal changes, one of the known effects of insulin and IGF1 signaling. Glycylpeptide N-myristoyltransferase 1 attaches myristoyl-groups to proteins with a glycine at their N-terminus. Myristoylation of proteins equips them with a membrane anchor, targeting them to cellular membranes.

Intriguingly, we detected several effectors of metabolic regulation in insulin signaling, which bound to homologous sites in IRS-1 and IRS-2. We encountered a number of enzymes associated with fatty acid catabolism in mitochondria. Acyl-CoA dehydrogenases catalyze the first step in beta-oxidation, and enoyl-CoA hydratase is responsible for the second step. 2,4-dienoyl-CoA reductase participates in beta-oxidation by feeding unsaturated fatty acids into the pathway. Hydroxysteroid (17-beta) dehydrogenase 10 is crucial in the degradation of branched chain fatty acids and isoleucine as well as in the

metabolism of steroid hormones (90). The activity of these enzymes is known to exert an effect back on IRS signaling. Long-chain acyl-CoA can be metabolized to diacylglycerol, which activates PKCθ (91). PKCθ directly and indirectly (via JNK and IKK) leads to phosphorylation of IRS-1 on inhibitory serine-residues (92). A recent bioinformatic analysis even reports that two related acyl-CoA dehydrogenases (family members 10 and 11) contain kinase domains, pointing to a direct link between fatty acid catabolism and cell signaling (93).

Furthermore, several enzymes involved in amino acid catabolism were found to be able to interact with IRS-1 derived phosphopeptides. Methylcrotonoyl-CoA-carboxylase degrades leucine, isobutyryl-CoA-dehydrogenase degrades valine, and glutaryl-CoA-dehydrogenase degrades lysine and tryptophane. Branched chain amino acids play an important role in the regulation of translation via mTOR-signaling and in insulin signaling (94, 95).

Finally, a complex consisting of prefoldin, RPABC1/RPB5 and „unconventional prefoldin RPB5 interactor' binds to homologous sites in the IRS-1 and IRS-2 PH and PTB domains. This complex shuttles between cytosol and nucleus, where it is believed to mediate transcriptional effects of nutrient signaling via mTOR (96).

Figure 7 summarizes the various pathways that are potentially directly linked to IRS-1 and IRS-2 via their pTyr-dependent interaction partners as measured in this study.
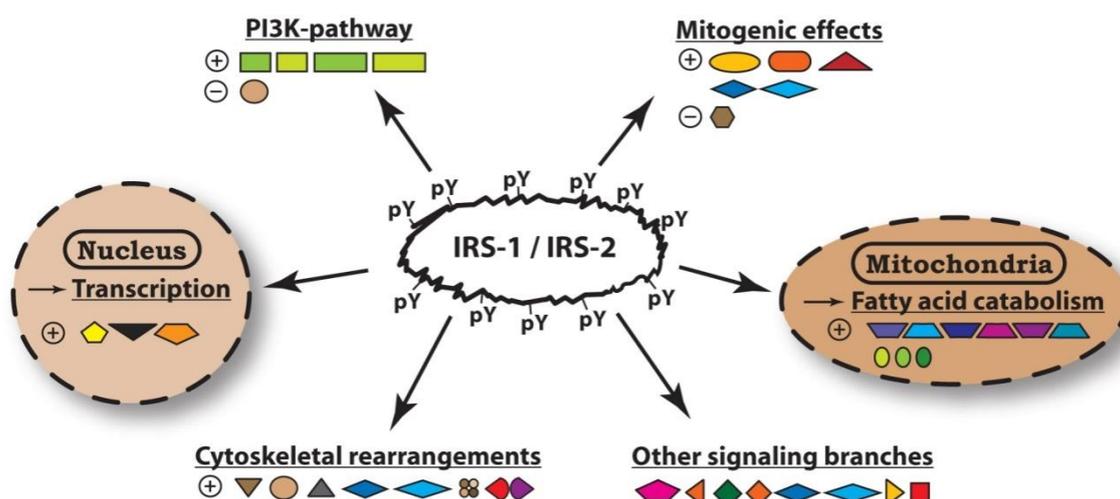
Fig. 7. Molecular functions potentially influenced by IRS-1 and IRS-2 through direct binding. The pTyr-dependent interaction partners of IRS-1 and IRS-2 identified in this study are grouped according to their function and subcellular location. Stimulatory and inhibitory effects on the pathways are indicated by ⊕ and Θ, respectively.

CONCLUSIONS

Here we have used quantitative interaction proteomics based on the SILAC technology in a systematic study of phosphotyrosine binding in the insulin signaling pathway. With this work we hope to have contributed new knowledge to the insulin signaling pathway, whose malfunction underlies diabetes, a disease that will soon afflict 300 million patients world-wide and that threatens the very functioning of national health systems (97). High accuracy mass spectrometry and relative quantitation between phosphopeptide pulldowns and control pulldowns from cell extracts ensured that our data represents specific phosphopeptide – protein interactions. Further studies are needed to validate these interactions in the context of endogenous, full-length proteins. However, for most of our interactors, previous knowledge of binding modes and involvement in the pathway make this extremely likely. For the potential interaction partners with non-traditional binding modes, our data raises interesting hypotheses that can be followed up in a directed way by researchers in the field. Recent work from a number of laboratories encourage us to believe that at least some of these interactions may point to novel and as yet unstudied mechanisms in insulin signaling.

REFERENCES

1.   Schlessinger, J., and Lemmon, M. A. (2003) SH2 and PTB domains in tyrosine kinase signaling. Sci STKE 2003, RE12.

2.   Pawson, T. (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. Cell 116, 191-203.

3.   Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., Pawson, T., Gao, Y., and Li, S. S. (2008) Defining the specificity space of the human SRC homology 2 domain. Mol Cell Proteomics 7, 768-784.

4.   Miller, M. L., Hanke, S., Hinsby, A. M., Friis, C., Brunak, S., Mann, M., and Blom, N. (2008) Motif decomposition of the phosphotyrosine proteome reveals a new N-terminal binding motif for SHIP2. Mol Cell Proteomics 7, 181-192.

5.   Hu, J., and Hubbard, S. R. (2005) Structural characterization of a novel Cbl phosphotyrosine recognition motif in the APS family of adapter proteins. The Journal of biological chemistry 280, 18943-18949.

6.   Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (1994) Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-gamma 1 complexed with a high affinity binding peptide. Cell 77, 461-472.

7.   Li, S. C., Gish, G., Yang, D., Coffey, A. J., Forman-Kay, J. D., Ernberg, I., Kay, L. E., and Pawson, T. (1999) Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A. Curr Biol 9, 1355-1362.

8.   Uhlik, M. T., Temple, B., Bencharit, S., Kimple, A. J., Siderovski, D. P., and Johnson, G. L. (2005) Structural and evolutionary division of phosphotyrosine binding (PTB) domains. Journal of molecular biology 345, 1-20.

9.   Smith, M. J., Hardy, W. R., Murphy, J. M., Jones, N., and Pawson, T. (2006) Screening for PTB domain binding partners and ligand specificity using proteome-derived NPXY peptide arrays. Molecular and cellular biology 26, 8461-8474.

10.   Sondermann, H., and Kuriyan, J. (2005) C2 can do it, too. Cell 121, 158-160.

11.   Christofk, H. R., Vander Heiden, M. G., Wu, N., Asara, J. M., and Cantley, L. C. (2008) Pyruvate kinase M2 is a phosphotyrosine-binding protein. Nature 452, 181-186.

12.   Saltiel, A. R. (2001) New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. Cell 104, 517-529.

13.   Sesti, G. (2006) Pathophysiology of insulin resistance. Best practice & research 20, 665-679.

14.   Bryant, N. J., Govers, R., and James, D. E. (2002) Regulated transport of the glucose transporter GLUT4. Nature reviews 3, 267-277.

15.   White, M. F. (2002) IRS proteins and the common path to diabetes. American journal of physiology 283, E413-422.

16.   Cai, D., Dhe-Paganon, S., Melendez, P. A., Lee, J., and Shoelson, S. E. (2003) Two new substrates in insulin signaling, IRS5/DOK4 and IRS6/DOK5. The Journal of biological chemistry 278, 25323-25330.

17.   Dearth, R. K., Cui, X., Kim, H. J., Hadsell, D. L., and Lee, A. V. (2007) Oncogenic transformation by the signaling adaptor proteins insulin receptor substrate (IRS)-1 and IRS-2. Cell cycle (Georgetown, Tex 6, 705-713.

18.   Summers, S. A. (2006) Ceramides in insulin resistance and lipotoxicity. Progress in lipid research 45, 42-72.

19.   Petersen, K. F., and Shulman, G. I. (2006) Etiology of insulin resistance. The American journal of medicine 119, S10-16.

20.   Houstis, N., Rosen, E. D., and Lander, E. S. (2006) Reactive oxygen species have a causal role in multiple forms of insulin resistance. Nature 440, 944-948.

21.   Sesti, G., Federici, M., Hribal, M. L., Lauro, D., Sbraccia, P., and Lauro, R. (2001) Defects of the insulin receptor substrate (IRS) system in human metabolic disorders. Faseb J 15, 2099-2111.

22.   Withers, D. J., Burks, D. J., Towery, H. H., Altamuro, S. L., Flint, C. L., and White, M. F. (1999) Irs-2 coordinates Igf-1 receptor-mediated beta-cell development and peripheral insulin signalling. Nature genetics 23, 32-40.

23.   Kido, Y., Burks, D. J., Withers, D., Bruning, J. C., Kahn, C. R., White, M. F., and Accili, D. (2000) Tissue-specific insulin resistance in mice with mutations in the insulin receptor, IRS-1, and IRS-2. The Journal of clinical investigation 105, 199-205.

24.   Huang, C., Thirone, A. C., Huang, X., and Klip, A. (2005) Differential contribution of insulin receptor substrates 1 versus 2 to insulin signaling and glucose uptake in l6 myotubes. The Journal of biological chemistry 280, 19426-19435.

25.   Araki, E., Lipes, M. A., Patti, M. E., Bruning, J. C., Haag, B., 3rd, Johnson, R. S., and Kahn, C. R. (1994) Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. Nature 372, 186-190.

26.   Taniguchi, C. M., Ueki, K., and Kahn, R. (2005) Complementary roles of IRS-1 and IRS-2 in the hepatic regulation of metabolism. The Journal of clinical investigation 115, 718-727.

27.   Withers, D. J., Gutierrez, J. S., Towery, H., Burks, D. J., Ren, J. M., Previs, S., Zhang, Y., Bernal, D., Pons, S., Shulman, G. I., Bonner-Weir, S., and White, M. F. (1998) Disruption of IRS-2 causes type 2 diabetes in mice. Nature 391, 900-904.

28.   Taniguchi, C. M., Emanuelli, B., and Kahn, C. R. (2006) Critical nodes in signalling pathways: insights into insulin action. Nature reviews 7, 85-96.

29.   Nakae, J., Kido, Y., and Accili, D. (2001) Distinct and overlapping functions of insulin and IGF-I receptors. Endocrine reviews 22, 818-835.

30.   Hirayama, I., Tamemoto, H., Yokota, H., Kubo, S. K., Wang, J., Kuwano, H., Nagamachi, Y., Takeuchi, T., and Izumi, T. (1999) Insulin receptor-related receptor is expressed in pancreatic beta-cells and stimulates tyrosine phosphorylation of insulin receptor substrate-1 and -2. Diabetes 48, 1237-1244.

31.   Klammt, J., Garten, A., Barnikol-Oettler, A., Beck-Sickinger, A. G., and Kiess, W. (2005) Comparative analysis of the signaling capabilities of the insulin receptor-related receptor. Biochemical and biophysical research communications 327, 557-564.

32.   Neduva, V., and Russell, R. B. (2006) Peptides mediating interaction networks: new leads at last. Current opinion in biotechnology 17, 465-471.

33.   Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N. P., Trave, G., and Gibson, T. J. (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. Front Biosci 13, 6580-6603.

34.   Schulze, W. X., Deng, L., and Mann, M. (2005) Phosphotyrosine interactome of the ErbB-receptor kinase family. Molecular systems biology 1, 2005 0008.

35.   Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., Baltissen, M. P., Stunnenberg, H. G., Mann, M., and Timmers, H. T. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell 131, 58-69.

36.   Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1, 376-386.

37.   Vermeulen, M., Hubner, N. C., and Mann, M. (2008) High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. Current opinion in biotechnology.

38.   Jones, R. B., Gordus, A., Krall, J. A., and MacBeath, G. (2006) A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature 439, 168-174.

39.   Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Analytical chemistry 75, 663-670.

40.   Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature protocols 2, 1896-1906.

41.   Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002) Microcolumns with self-assembled particle frits for proteomics. Journal of chromatography 979, 233-239.

42.   Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 4, 2010-2021.

43.   Schulze, W. X., and Mann, M. (2004) A novel proteomic screen for peptide-protein interactions. The Journal of biological chemistry 279, 10756-10764.

44.   Mann, M. (2008) Can proteomics retire the Western blot? Journal of proteome research 7, 3065.

45.   Hanke, S., Besir, H., Oesterhelt, D., and Mann, M. (2008) Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. Journal of proteome research 7, 1118-1130.

46.   Sun, X. J., Pons, S., Wang, L. M., Zhang, Y., Yenush, L., Burks, D., Myers, M. G., Jr., Glasheen, E., Copeland, N. G., Jenkins, N. A., Pierce, J. H., and White, M. F. (1997) The IRS-2 gene on murine chromosome 8 encodes a unique signaling adapter for insulin and cytokine action. Molecular endocrinology (Baltimore, Md 11, 251-262.

47.   Eck, M. J., Dhe-Paganon, S., Trub, T., Nolte, R. T., and Shoelson, S. E. (1996) Structure of the IRS-1 PTB domain bound to the juxtamembrane region of the insulin receptor. Cell 85, 695-705.

48.   Wolf, G., Trub, T., Ottinger, E., Groninga, L., Lynch, A., White, M. F., Miyazaki, M., Lee, J., and Shoelson, S. E. (1995) PTB domains of IRS-1 and Shc have distinct but overlapping binding specificities. The Journal of biological chemistry 270, 27407-27410.

49.   Ong, S. E., and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. Nature chemical biology 1, 252-262.

50.   Witze, E. S., Old, W. M., Resing, K. A., and Ahn, N. G. (2007) Mapping protein post-translational modifications with mass spectrometry. Nature methods 4, 798-806.

51.   Jensen, O. N. (2006) Interpreting the protein language using proteomics. Nature reviews 7, 391-403.

52.   Mann, M. (2006) Functional and quantitative proteomics using SILAC. Nature reviews 7, 952-958.

53.   Kruger, M., Kratchmarova, I., Blagoev, B., Tseng, Y. H., Kahn, C. R., and Mann, M. (2008) Dissection of the insulin signaling pathway via quantitative phosphoproteomics. Proceedings of the National Academy of Sciences of the United States of America 105, 2451-2456.

54.   Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., and et al. (1993) SH2 domains recognize specific phosphopeptide sequences. Cell 72, 767-778.

55.   Knight, Z. A., Gonzalez, B., Feldman, M. E., Zunder, E. R., Goldenberg, D. D., Williams, O., Loewith, R., Stokoe, D., Balla, A., Toth, B., Balla, T., Weiss, W. A., Williams, R. L., and Shokat, K. M. (2006) A pharmacological map of the PI3-K family defines a role for p110alpha in insulin signaling. Cell 125, 733-747.

56.   Myers, M. G., Jr., Mendez, R., Shi, P., Pierce, J. H., Rhoads, R., and White, M. F. (1998) The COOH-terminal tyrosine phosphorylation sites on IRS-1 bind SHP-2 and

negatively regulate insulin signaling. The Journal of biological chemistry 273, 26908-26914.

57. Li, W., Nishimura, R., Kashishian, A., Batzer, A. G., Kim, W. J., Cooper, J. A., and Schlessinger, J. (1994) A new function for a phosphotyrosine phosphatase: linking GRB2-Sos to a receptor tyrosine kinase. Molecular and cellular biology 14, 509-517.

58. Milarski, K. L., and Saltiel, A. R. (1994) Expression of catalytically inactive Syp phosphatase in 3T3 cells blocks stimulation of mitogen-activated protein kinase by insulin. The Journal of biological chemistry 269, 21239-21243.

59. Noguchi, T., Matozaki, T., Horita, K., Fujioka, Y., and Kasuga, M. (1994) Role of SH-PTP2, a protein-tyrosine phosphatase with Src homology 2 domains, in insulin-stimulated Ras activation. Molecular and cellular biology 14, 6674-6682.

60. Dyson, J. M., Kong, A. M., Wiradjaja, F., Astle, M. V., Gurung, R., and Mitchell, C. A. (2005) The SH2 domain containing inositol polyphosphate 5-phosphatase-2: SHIP2. The international journal of biochemistry & cell biology 37, 2260-2265.

61. Feller, S. M. (2001) Crk family adaptors-signalling complex formation and biological roles. Oncogene 20, 6348-6371.

62. Chang, L., Chiang, S. H., and Saltiel, A. R. (2004) Insulin signaling and the regulation of glucose transport. Molecular medicine (Cambridge, Mass 10, 65-71.

63. Saltiel, A. R., and Kahn, C. R. (2001) Insulin signalling and the regulation of glucose and lipid metabolism. Nature 414, 799-806.

64. Sadagurski, M., Weingarten, G., Rhodes, C. J., White, M. F., and Wertheimer, E. (2005) Insulin receptor substrate 2 plays diverse cell-specific roles in the regulation of glucose transport. The Journal of biological chemistry 280, 14536-14544.

65. Haeusler, R. A., and Accili, D. (2008) The double life of Irs. Cell metabolism 8, 7-9.

66. Tobe, K., Sabe, H., Yamamoto, T., Yamauchi, T., Asai, S., Kaburagi, Y., Tamemoto, H., Ueki, K., Kimura, H., Akanuma, Y., Yazaki, Y., Hanafusa, H., and Kadowaki, T. (1996) Csk enhances insulin-stimulated dephosphorylation of focal adhesion proteins. Molecular and cellular biology 16, 4765-4772.

67. Berridge, M. J., and Irvine, R. F. (1989) Inositol phosphates and cell signalling. Nature 341, 197-205.

68. Kim, J. J., and Accili, D. (2002) Signalling through IGF-I and insulin receptors: where is the specificity? Growth Horm IGF Res 12, 84-90.

69. Yaffe, M. B. (2002) Phosphotyrosine-binding domains in signal transduction. Nature reviews 3, 177-186.

70. Kamberov, E., Makarova, O., Roh, M., Liu, A., Karnak, D., Straight, S., and Margolis, B. (2000) Molecular cloning and characterization of Pals, proteins associated with mLin-7. The Journal of biological chemistry 275, 11425-11431.

71. Hu, J., Liu, J., Ghirlando, R., Saltiel, A. R., and Hubbard, S. R. (2003) Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. Molecular cell 12, 1379-1389.

72. Sawka-Verhelle, D., Tartare-Deckert, S., White, M. F., and Van Obberghen, E. (1996) Insulin receptor substrate-2 binds to the insulin receptor through its phosphotyrosine-binding domain and through a newly identified domain comprising amino acids 591-786. The Journal of biological chemistry 271, 5980-5983.

73. Groesch, T. D., Zhou, F., Mattila, S., Geahlen, R. L., and Post, C. B. (2006) Structural basis for the requirement of two phosphotyrosine residues in signaling mediated by Syk tyrosine kinase. Journal of molecular biology 356, 1222-1236.

74.   Velazquez, L., Gish, G. D., van Der Geer, P., Taylor, L., Shulman, J., and Pawson, T. (2000) The shc adaptor protein forms interdependent phosphotyrosine-mediated protein complexes in mast cells stimulated with interleukin 3. Blood 96, 132-138.

75.   Lee, C. H., Li, W., Nishimura, R., Zhou, M., Batzer, A. G., Myers, M. G., Jr., White, M. F., Schlessinger, J., and Skolnik, E. Y. (1993) Nck associates with the SH2 domain-docking protein IRS-1 in insulin-stimulated cells. Proceedings of the National Academy of Sciences of the United States of America 90, 11713-11717.

76.   Buday, L., Wunderlich, L., and Tamas, P. (2002) The Nck family of adapter proteins: regulators of actin cytoskeleton. Cellular signalling 14, 723-731.

77.   Virkamaki, A., Ueki, K., and Kahn, C. R. (1999) Protein-protein interaction in insulin signaling and the molecular mechanisms of insulin resistance. The Journal of clinical investigation 103, 931-943.

78.   Meng, W., Sawasdikosol, S., Burakoff, S. J., and Eck, M. J. (1999) Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase. Nature 398, 84-90.

79.   Benes, C. H., Wu, N., Elia, A. E., Dharia, T., Cantley, L. C., and Soltoff, S. P. (2005) The C2 domain of PKCdelta is a phosphotyrosine binding domain. Cell 121, 271-280.

80.   Yamauchi, T., Kaburagi, Y., Ueki, K., Tsuji, Y., Stark, G. R., Kerr, I. M., Tsushima, T., Akanuma, Y., Komuro, I., Tobe, K., Yazaki, Y., and Kadowaki, T. (1998) Growth hormone and prolactin stimulate tyrosine phosphorylation of insulin receptor substrate-1, -2, and -3, their association with p85 phosphatidylinositol 3-kinase (PI3-kinase), and concomitantly PI3-kinase activation via JAK2 kinase. The Journal of biological chemistry 273, 15719-15726.

81.   Lanzino, M., Garofalo, C., Morelli, C., Le Pera, M., Casaburi, I., McPhaul, M. J., Surmacz, E., Ando, S., and Sisci, D. (2008) Insulin receptor substrate 1 modulates the transcriptional activity and the stability of androgen receptor in breast cancer cells. Breast cancer research and treatment.

82.   Liang, L., Zhou, T., Jiang, J., Pierce, J. H., Gustafson, T. A., and Frank, S. J. (1999) Insulin receptor substrate-1 enhances growth hormone-induced proliferation. Endocrinology 140, 1972-1983.

83.   Senthil, D., Ghosh Choudhury, G., Bhandari, B. K., and Kasinath, B. S. (2002) The type 2 vascular endothelial growth factor receptor recruits insulin receptor substrate-1 in its signalling pathway. The Biochemical journal 368, 49-56.

84.   Vuori, K., and Ruoslahti, E. (1994) Association of insulin receptor substrate-1 with integrins. Science (New York, N.Y 266, 1576-1578.

85.   Shaw, L. M. (2001) Identification of insulin receptor substrate 1 (IRS-1) and IRS-2 as signaling intermediates in the alpha6beta4 integrin-dependent activation of phosphoinositide 3-OH kinase and promotion of invasion. Molecular and cellular biology 21, 5082-5093.

86.   Yenush, L., and White, M. F. (1997) The IRS-signalling system during insulin and cytokine action. Bioessays 19, 491-500.

87.   Saeki, M., Irie, Y., Ni, L., Yoshida, M., Itsuki, Y., and Kamisaki, Y. (2006) Monad, a WD40 repeat protein, promotes apoptosis induced by TNF-alpha. Biochemical and biophysical research communications 342, 568-572.

88.   Weirich, C. S., Erzberger, J. P., and Barral, Y. (2008) The septin family of GTPases: architecture and dynamics. Nature reviews 9, 478-489.

89.   Xu, X., Sarikas, A., Dias-Santagata, D. C., Dolios, G., Lafontant, P. J., Tsai, S. C., Zhu, W., Nakajima, H., Nakajima, H. O., Field, L. J., Wang, R., and Pan, Z. Q. (2008) The

CUL7 E3 ubiquitin ligase targets insulin receptor substrate 1 for ubiquitin-dependent degradation. Molecular cell 30, 403-414.

90.   He, X. Y., and Yang, S. Y. (2006) Roles of type 10 17beta-hydroxysteroid dehydrogenase in intracrinology and metabolism of isoleucine and fatty acids. Endocrine, metabolic & immune disorders drug targets 6, 95-102.

91.   Yu, C., Chen, Y., Cline, G. W., Zhang, D., Zong, H., Wang, Y., Bergeron, R., Kim, J. K., Cushman, S. W., Cooney, G. J., Atcheson, B., White, M. F., Kraegen, E. W., and Shulman, G. I. (2002) Mechanism by which fatty acids inhibit insulin activation of insulin receptor substrate-1 (IRS-1)-associated phosphatidylinositol 3-kinase activity in muscle. The Journal of biological chemistry 277, 50230-50236.

92.   Gual, P., Le Marchand-Brustel, Y., and Tanti, J. F. (2005) Positive and negative regulation of insulin signaling through IRS-1 phosphorylation. Biochimie 87, 99-109.

93.   Briedis, K. M., Starr, A., and Bourne, P. E. (2008) Analysis of the human kinome using methods including fold recognition reveals two novel kinases. PLoS ONE 3, e1597.

94.   Proud, C. G. (2002) Regulation of mammalian translation factors by nutrients. European journal of biochemistry / FEBS 269, 5338-5349.

95.   Patti, M. E., Brambilla, E., Luzi, L., Landaker, E. J., and Kahn, C. R. (1998) Bidirectional modulation of insulin action by amino acids. The Journal of clinical investigation 101, 1519-1529.

96.   Gstaiger, M., Luke, B., Hess, D., Oakeley, E. J., Wirbelauer, C., Blondel, M., Vigneron, M., Peter, M., and Krek, W. (2003) Control of nutrient-sensitive transcription programs by the unconventional prefoldin URI. Science (New York, N.Y 302, 1208-1212.

97.   Adeghate, E., Schattner, P., and Dunn, E. (2006) An update on the etiology and epidemiology of diabetes mellitus. Annals of the New York Academy of Sciences 1084, 1-29.

3 DEVELOPMENT OF A NOVEL METHOD FOR VERY HIGH ACCURACY IN PROTEIN QUANTITATION

3.1 *PUBLICATION:* ABSOLUTE SILAC FOR ACCURATE QUANTITATION OF PROTEINS IN COMPLEX MIXTURES DOWN TO THE ATTOMOLE LEVEL

This article presents the results of the project on absolute protein quantitation and was published in 2008, in the March issue of Journal of Proteome Research on pages 1118-1130.

The project was a joint effort together with Dr. Hüseyin Besir and Prof. Dr. Dieter Oesterhelt from the department of membrane biochemistry at the Max-Planck-Institute for biochemistry.

The following pages contain the published version of the article.

Absolute SILAC for Accurate Quantitation of Proteins in Complex Mixtures Down to the Attomole Level

Stefan Hanke,[†] Hüseyin Besir,[‡] Dieter Oesterhelt,[‡] and Matthias Mann*,[†]

Department of Proteomics and Signal Transduction and Department of Membrane Biochemistry, Max Planck Institute for Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Munich, Germany

Mass spectrometry based proteomics can routinely identify hundreds of proteins in a single LC-MS run, and methods have been developed for relative quantitation between differentially treated samples using stable isotopes. However, absolute quantitation has so far required addition of a labeled standard late in the experimental workflow, introducing variability due to sample preparation. Here we present a new variant of the stable isotope labeling by amino acids in cell culture (SILAC) technique termed "Absolute SILAC" that allows accurate quantitation of selected proteins in complex mixtures. SILAC-labeled recombinant proteins produced in vivo or in vitro are used as internal standards, which are directly mixed into lysates of cells or tissues. This minimizes differences in sample processing between the isotope-labeled standard and its endogenous counterpart. We show that it is possible to quantify over several orders of magnitude, even in the background of a whole cell lysate. We furthermore devise a strategy to quantify peptides at or below their signal-to-noise level on hybrid ion trap instruments, shown here for the LTQ-Orbitrap. The data system triggers on peptides of the SILAC-labeled protein, initiating ion collection in a narrow mass range including the endogenous and labeled peptide. This strategy extends the regular detection limit of an LTQ-Orbitrap by at least an order of magnitude and accurately quantifies down to 150 attomole of protein in a cell lysate without any fractionation prior to LC-MS. We use Absolute SILAC to determine the copy number per cell of growth factor receptor-bound protein 2 (Grb2) in HeLa, HepG2, and C2C12 cells to 5.5×10^5, 8.8×10^5, and 5.7×10^5, respectively, in the exponential growth phase.

Keywords: mass spectrometry • quantitation • quantification • isotope labeling • LC-MS • selected ion monitoring • proteomics • diagnostics • biomarkers • detection limit

Introduction

The literature of molecular biology contains an overwhelming amount of information on the function of proteins in a cell or the body, alone and in their interplay with each other. However, in the majority of cases, the published results are limited to qualitative information or relative quantitation of the proteins of interest, while quantitative data on their absolute amounts are lacking. Ultimately, it would be highly desirable to obtain exact quantitative values of each protein in a system, e.g., their copy number per cell or their concentration in nanogram per milliliter of body fluid. While this kind of basic information about a protein is already per se valuable for the biologist, systems biology even requires it as input for modeling. In a medical context, knowing the exact amounts of certain proteins in blood or other common sources of biomarkers can provide diagnostically relevant information for patient treatment. To this end, most clinically important markers are measured by quantitative immunoassays. These approaches can be very accurate and easy to perform once a suitable antibody has been generated, but assay development generally lags behind the increasing pace of discovery of new interesting biomarkers by modern proteomic and microarray methods. In addition, immunoassays can be affected by unspecific reactions and nonlinearity of quantitation, and they can fail if a relevant epitope has been removed by conformational changes, proteolysis, or posttranslational modifications. Since the quantitation of each protein normally requires a separate experiment, the procedure can also be very laborious and expensive. A generally applicable and robust method, comparably or even more sensitive and specific as an immunoassay but without its disadvantages, would therefore be highly desirable.

Mass spectrometry has become a widespread technology for relative quantitation between two or more experimental conditions.[1] While peak areas of peptide signals can be compared between consecutive LC-MS runs,[2–4] more sophisticated methods exploit stable isotope labeling. Differentially treated samples are labeled with heavy and light isotopes, creating distinct signals in a mass spectrum for every proteolytic peptide present in labeled and nonlabeled form. By comparing their peak areas, one can directly deduce the effect of the differential treatment

* Corresponding author. Tel.: +49 89-8578-2557. Fax: +49 89-8578-2219. E-mail: mmann@biochem.mpg.de.
† Department of Proteomics and Signal Transduction.
‡ Department of Membrane Biochemistry.

on protein amount. Several methods have been developed for introducing heavy isotopes into a sample. Digestion in heavy water ($H_2^{18}O$)[5,6] or chemical addition of labeled reagents in the case of ICAT,[7] iTRAQ,[8] ICPL,[9] and HysTag[10] integrate the label before, during, or after protein digestion. Metabolic labeling techniques such as ^{15}N-labeling[11] and SILAC[12–14] even allow the incorporation of the label preharvest during culture. In cases where metabolic labeling is hard to accomplish, like with tissue samples, a modified SILAC approach such as CDIT[15] can be applied.

Recently, several methods have been developed that combine isotope labeling and mass spectrometry for absolute quantitation of proteins. On the instrumentation side, improvements in sensitivity and sequencing speed have ameliorated the quantitation of low abundant proteins, and for triple quadrupole mass spectrometers in particular, the multiple reaction monitoring (MRM) method[16–19] has gained popularity for quantitation of selected proteins in complex mixtures. In each quantitation approach, a known amount of an isotope-labeled standard which resembles the protein to be quantified or a peptide thereof is spiked into the biological sample. Desiderio et al. were the first to use this principle for the absolute quantitation of peptides in 1983, by digestion in heavy water.[20] A report in 1996 describes the first application of synthetic peptide standards containing [13]C- and [2]H-labeled amino acids for the quantitation of the parent protein,[21] a concept expanded further in 2003 and termed AQUA.[22] The absolute quantitation of ^{15}N-labeled peptide hormones by top-down-MS was demonstrated in 1997.[23] In 2005, the QconCAT principle was introduced,[24–26] which is based on the expression of artificial genes comprising peptides from different proteins in medium containing [15]NH$_4$Cl or ^{13}C-arginine and ^{13}C-lysine.

Inherent problems of ^{15}N-labeling include the wide isotope distributions of labeled peptides,[1] and difficulties of digestion in heavy water include the control of incorporation of one or two [18]O atoms as well as the necessity to add the standard at a very late stage in sample preparation. AQUA has the advantage that labeled synthetic peptides are easy to obtain. However, when the absolute quantitation of a protein is based on only one or two peptides as is typically the case in AQUA and QconCAT, accuracy may suffer. Furthermore, not all peptides are amenable for use as standards because amino acids that could introduce side products need to be avoided. Because the standards do not behave the same as the endogenous protein during fractionation steps, these methods require that they are added at the level of protein digestion introducing several pitfalls in quantitation. First, each experimental step is subject to sample losses, and therefore, strictly speaking, one only determines the absolute amount of protein present at the stage of mixing in the standard. Second, all variations due to sample handling influence the result as long as sample and standard are separated. Third, incomplete digestion and, when using in-gel digestion, incomplete recovery of proteolytic peptides can further complicate the picture.[27]

We therefore set out to develop an MS-based absolute quantitation method which addresses these problems by spiking in the isotope-labeled standard at the earliest possible stage. To achieve this, the standard needs to be highly similar if not identical to the endogenous protein to be quantified. We adapt the SILAC approach that has proven to be very powerful for relative quantitation[28] and show that it can also be applied to absolute quantitation of individual proteins. We express recombinant proteins both in a cell-free system and in an

Escherichia coli strain auxotrophic for arginine and lysine. This yields stable isotope-labeled proteins when expressed in minimal medium supplemented with heavy labeled analogues of these two amino acids. Following purification and quantitation of the standard proteins, we add defined amounts of those proteins directly to lysates from cells or tissues. Since they have the same physicochemical properties as their endogenous unlabeled counterparts, any losses or modifications during sample fractionation and processing will not affect the quantitation (Figure 1). In addition, since all proteolytic peptides except for the C-terminal one are labeled, the chance of detecting several of them is high, improving statistical confidence of quantitation. Once the standard proteins have been generated, several of them can in principle be added to a biological sample. This multiplexed format can provide a quantitative readout for many different proteins in one single experiment. We show that quantitation via Absolute SILAC is highly accurate and linear over a wide range by mixing a protein in its normal and isotope-labeled versions at defined ratios from 1:300 to 300:1. In addition, we describe a novel strategy to improve sensitivity in complex mixtures by more than an order of magnitude by triggering narrow mass range scans upon detection of the labeled peptides. This is demonstrated with the accurate quantitation of a protein in a complex lysate down to 150 attomole, without any fractionation prior to LC-MS. Finally, as an example of relevance to systems biology, we determine the copy number per cell of the signaling adaptor protein Grb2 in HeLa, HepG2, and C2C12 cells.

Materials and Methods

Clones and Vectors. The cDNA for growth factor receptor-bound protein 2 (Grb2) in a Gateway compatible vector was purchased from the RZPD German resource center for genome research (Berlin, Germany) (clone RZPDo834A0934D). It was transferred into pDEST17 via Gateway cloning for T7-RNA polymerase dependent expression as the N-terminally His$_6$-tagged protein. The maltose binding protein (MBP) was expressed without a tag under control of a lac-promotor from the plasmid pMAL-*C*2 (New England Biolabs).

In Vitro Expression and Purification of Proteins. Recombinant N-terminally His$_6$-tagged Grb2 was expressed using a cell-free system prepared as described[29] with slight modifications. *E. coli* S30 lysate was prepared from BL21(DE3)RIL cells (Novagen). T7 RNA polymerase was also expressed in BL21(DE3) RIL containing the vector pAR1219 as described,[30] but no purification of the enzyme was performed. Instead, we prepared another lysate from this IPTG-induced culture and added 60 μL of this lysate with 400 μL of the standard lysate to 1 mL of reaction volume. The concentration of the reaction components was adjusted to 57 mM HEPES-KOH buffer (pH 8.2), 2 mM DTT, 1.2 mM ATP, 0.85 mM each of CTP, GTP, and UTP, 100 mM creatine phosphate, 130 μg/mL of creatine kinase, 2.0% PEG 8000, 0.64 mM 3′,5′-cyclic AMP, 34 μM L-(−)-5-formyl-5,6,7,8-tetrahydrofolic acid, 175 μg/mL of *E. coli* total tRNA, 90 mM potassium glutamate, 80 mM ammonium acetate, 12 mM magnesium acetate, 2.0 mM each of the 20 amino acids, and 6.7 μg/mL of plasmid DNA. The reaction mixture was incubated with shaking at 600 rpm at 30 °C for 2 h.

By exchanging normal arginine and lysine with their heavy-isotope counterparts [13]C$_6$[15]N$_4$-arginine and [13]C$_6$[15]N$_2$-lysine in the reaction, the SILAC-labeled protein was obtained. All amino acids were applied at a concentration of 2 mM. Labeled Arg and Lys were purchased from Sigma Isotec (at least 98% atom
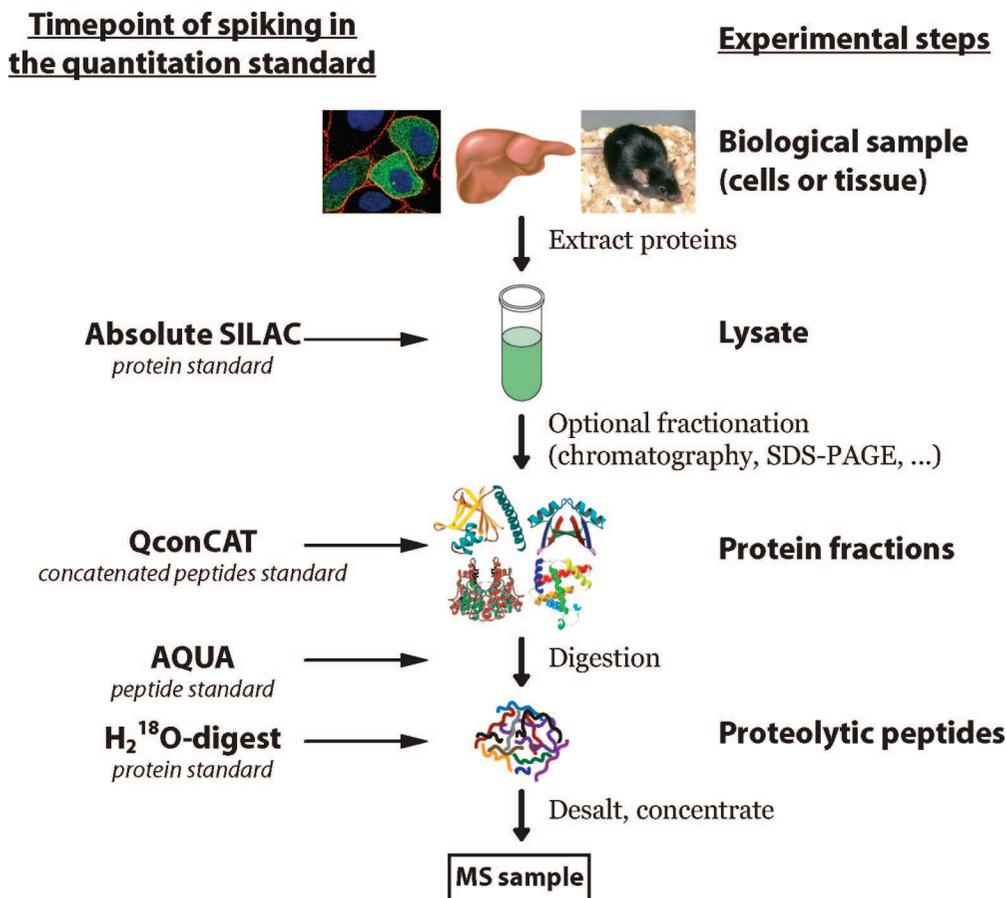
Figure 1. Experimental workflow in MS-based absolute quantitation of proteins. Different methods for introducing an isotope-labeled standard are shown. Absolute SILAC allows spiking the standard at an early level in the experimental procedure, thereby minimizing the quantitation error arising from losses and modifications during sample handling.

percent isotope enrichment). The reaction mixture contained the target protein mostly as insoluble precipitate which was solubilized with 6 M guanidinium chloride and subsequently purified employing Ni^{2+}-affinity chromatography according to the manufacturer's protocol for purification of denatured proteins using Ni-NTA sepharose (Qiagen, Germany) and spin columns (MoBiTec, Germany). The purity of the protein was checked by SDS-PAGE, and it was dialyzed against distilled water to remove imidazole and guanidinium. After refolding, protein aggregates were removed by centrifugation.

In Vivo Expression and Purification of Proteins. For expression of SILAC-labeled proteins, the *E. coli* strain AT713 was used (*E. coli* Genetic Resource Center, New Haven, strain number 4529), which is auxotrophic for arginine and lysine. Expression was conducted in M9 minimal medium supplemented with 0.2% glucose and 100 μg/mL of each amino acid. For SILAC-labeled proteins, $^{13}C_6{}^{15}N_4$-arginine and $^{13}C_6{}^{15}N_2$-lysine from Sigma Isotec were used. After harvesting, the bacteria were lysed in 20 mM TrisHCl, pH7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 1 mM AEBSF-HCl, 2 μg/mL of Aprotinin, 1 μg/mL of Leupeptin, and 1 μg/mL of Pepstatin. Debris was removed by centrifugation, and MBP was enriched from the supernatant by affinity chromatography on an amylose resin, followed by size exclusion chromatography using a Superdex75 column. The purity of the protein was checked by SDS-PAGE, and the main fractions containing MBP were pooled and stored in PBS containing 10% glycerol.

Quantitation of Purified Proteins and Storage. The concentration of purified Grb2-protein was measured by amino

acid analysis (Biotronik LC3000). The concentration of purified MBP protein was measured by UV absorption at 280 nm, using a double-beam spectrophotometer (Lambda 19, Perkin-Elmer).

We averaged five independent measurements (all within 2% deviation) to calculate the concentration based on the formula $A = c \times d \times \epsilon$, where the molar absorption coefficient ϵ was calculated from the protein sequence applying the formula $\epsilon = \#Trp \times 5500 + \#Tyr \times 1490 + \#Cystine \times 125$, as suggested by Pace et al.[31]

The stock solution was aliquoted, and different dilutions down to 500 fmol/μL were prepared in storage buffer or in 6 M urea, 2 M thiourea, 20 mM Tris, pH 8.0. If required, further dilutions were carried out directly before use to minimize losses. For MBP, solutions with fixed ratios from 1:300 to 300:1 between labeled and normal proteins were generated, along with dilution series down to the attomole range in steps of a factor of 10.

All protein solutions were prepared in protein LoBind-tubes from Eppendorf which were presaturated with a solution of 1 mg/mL of BSA in the respective buffer overnight to eliminate any unwanted adsorption sites. Before use, the solution was removed, and after a short centrifugation, the remainder was removed again. The samples were snap-frozen in liquid nitrogen and stored at -80 °C.

Cell Culture, Counting, and Lysis for Determination of Grb2 Copy Number. HeLa, HepG2, and C2C12 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and antibiotics. Cells were harvested at 70% confluency by trypsination and resus-

pended in an ice-cold growth medium. The concentration of cells was determined in a Neubauer counting chamber by 24 rounds of counting, and the average value was used for subsequent calculations. The cell suspension was aliquoted to microcentrifuge tubes and centrifuged at 400g. The cell pellet was washed twice with ice-cold PBS and either stored at −80 °C or lysed directly in 1% Igepal (NP-40), 150 mM NaCl, 50 mM Tris, pH 7.5, and protease inhibitor cocktail Roche complete. After repeated vortexing and incubation on ice for 15 min, several freeze–thaw cycles were conducted until no intact cells could be observed. Following centrifugation at 10 000g for 15 min at 4 °C, the supernatant was removed and the pellet washed 3 times with cold PBS. The desired amounts of SILAC-labeled Grb2 were then spiked into supernatant and pellet. The supernatant was precipitated with acetone before the protein pellet was resuspended in SDS-sample buffer for SDS-PAGE. The centrifugation pellet was partially resuspended in 2x SDS-sample buffer and treated with Benzonase (Roche) for degradation of DNA.

1D-SDS-PAGE and In-Gel Digestion of Proteins. The protein samples were separated by one-dimensional SDS-PAGE using NuPage Novex 4–12% Bis-Tris gels and NuPage MOPS running buffer (Invitrogen) according to the manufacturer's instructions. The gels were stained with colloidal Coomassie (Initrogen). Gel slices were cut into 1 mm^3 cubes, washed twice with 50 mM ammonium bicarbonate and 50% ethanol, dehydrated with 100% ethanol, and incubated with 10 mM DTT in 50 mM ammonium bicarbonate for 1 h at 56 °C for protein reduction. The resulting free thiol (−SH) groups were subsequently alkylated by incubating the samples with 55 mM iodoacetamide in 50 mM ammonium bicarbonate for 1 h at 25 °C in the dark. The gel pieces were washed twice with 50 mM ammonium bicarbonate, dehydrated with 100% ethanol, and dried in a vacuum concentrator. The gel pieces were rehydrated with 12.5 ng/μL of trypsin (Promega) in 50 mM ammonium bicarbonate and incubated for 16 h at 37 °C for protein digestion. Supernatants were transferred to fresh tubes, and the remaining peptides were extracted by incubating the gel pieces in 30% acetonitrile (MeCN) with 3% trifluoroacetic acid (TFA), followed by dehydration with 100% MeCN. The latter two steps were repeated. The extracts were combined, and organic solvent was removed in a vacuum concentrator. Desalting and concentration were carried out on RP-C$_{18}$ (Empore disks, 3M) StageTip columns,[32,33] and the eluted peptides were subjected to LC-MS.

In-Solution Digestion of Proteins. Proteins in 20 μL of 6 M urea, 2 M thiourea, 20 mM TrisHCl, pH 8.0, were reduced by adding 1 μg of DTT for 30 min, followed by alkylation of cysteines by incubating with 5 μg of iodoacetamide for 20 min. Digestion was started by adding endoproteinase LysC (Wako). After 3 h, the sample was diluted with 4 volumes of 50 mM NH$_4$HCO$_3$, and trypsin (Promega) was added for overnight incubation. The proteases were applied in a ratio of 1:50 to protein material, and all steps were carried out at room temperature. The digestion was stopped by acidifying with TFA, and the samples were loaded onto StageTips[32,33] for desalting and concentration prior to LC-MS analysis.

NanoLC-MS/MS. The digested peptide mixtures were separated by online reversed-phase (RP) nanoscale capillary liquid chromatography (nanoLC) and analyzed by electrospray tandem mass spectrometry (ES-MS/MS). The experiments were performed with an Agilent 1100 nanoflow system connected to an LTQ-Orbitrap mass spectrometer (Thermo Electron,

Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). Binding and chromatographic separation of the peptides took place in a 15 cm fused silica emitter (75 μm inner diameter) in-house packed[34] with reversed-phase ReproSil-Pur C$_{18}$-AQ 3 μm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany).

Peptide mixtures were injected onto the column with a flow of 500 nL/min and subsequently eluted with a flow of 250 nL/min from 2% to 40% MeCN in 0.5% acetic acid, in a 100 min gradient, or 20 min in the case of single protein digests. The mass spectrometer was operated in data-dependent mode to automatically switch between MS and MS/MS (MS^2) acquisition. Survey full scan MS spectra (from m/z 300 to 1600) were acquired in the Orbitrap with a resolution of 60 000 at m/z 400 (after accumulation to a target value of 1 000 000 charges in the linear ion trap) using the lock mass option for internal calibration of each spectrum.[35] The most intense ions (up to five, depending on signal intensity) were sequentially isolated for fragmentation in the linear ion trap using collisionally induced dissociation with a normalized collision energy of 30% at a target value of 5000. Target ions already selected for MS/MS were dynamically excluded for 60 s. The resulting fragment ions were recorded in the linear ion trap with unit resolution.

For experiments involving SIM scans, we constructed an inclusion list with the "heavy" masses of previously observed peptides from Grb2 or MBP. Detection of an ion from the inclusion list within 5 ppm triggered four narrow mass range (selected-ion-monitoring, SIM) scans and four MS/MS events at different collision energies (35, 40, 45, and 50%) at target values of 100 000 and 10 000, respectively. SIM scans were granted a maximum fill time of 0.5 s in the ion trap and were recorded in the Orbitrap at a resolution of 30 000 at m/z 400 and at a width from −6 to +3 m/z units around the selected mass.

Peptide Identification and Quantitation. Peak lists for database searching were generated from the raw data using in-house developed software called Raw2msm. From each fragment spectrum, the seven most intense peaks per 100 Thomson were extracted. Proteins were identified by automated database searching (Mascot, Matrix Science) against an in-house curated version of the human IPI database, version 3.19 (60 428 entries). This database was complemented with MBP and frequently observed contaminants like porcine trypsin and human keratins. Carbamidomethyl-cysteine was used as a fixed modification, and variable modifications were oxidation of methionine, protein N-acetylation, deamidation of Asn and Gln, arginine-^{13}C$_6$^{15}N$_4$, lysine-^{13}C$_6$^{15}N$_2$, and N-pyroglutamate. We required full tryptic specificity (cleavage at Arg-Pro and Lys-Pro as well as Asp-Pro was included), a maximum of two miscleavages, and mass accuracies of 5 ppm for the parent ion and 0.5 Da for fragment ions.

For the relative quantitation of the peptides against their SILAC-labeled counterparts, our in-house developed software MSQuant[36] was used (http://msquant.sourceforge.net).

Results and Discussion

Generation of Protein Standard in Vitro and in Vivo. The aim of this work was to develop and characterize Absolute SILAC as an MS-based method for the absolute quantitation of individual proteins in complex mixtures. Using metabolically labeled full length proteins as a quantitation reference has the major advantage of exposing standard and endogenous proteins to the same conditions beginning at the level of cell or
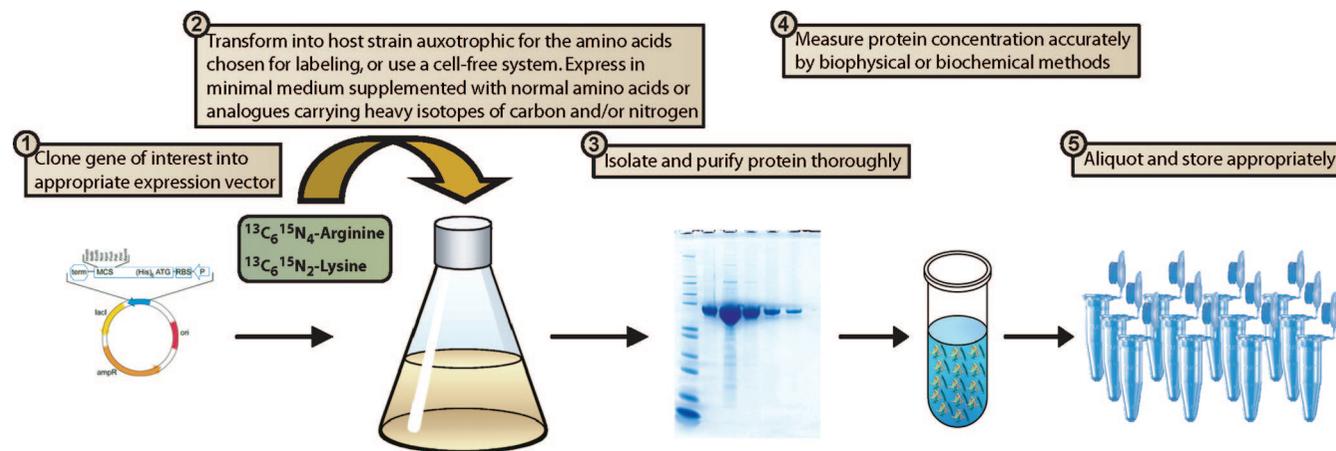
Figure 2. Production of SILAC-labeled proteins for absolute quantitation. The protein of interest is recombinantly produced in an expression host strain which is auxotrophic for the amino acids chosen for labeling. Alternatively, a cell-free approach is employed. The stable isotope-labeled analogues of those amino acids are added to a minimal medium along with the remaining other amino acids. High purity of the isolated protein and accurate measurement of its concentration are requirements for Absolute SILAC.

tissue lysis (Figure 1). This enables the quantitation of proteins present in the lysate, as opposed to quantitation of proteins present after fractionation and digestion as is the case in other MS-based methods. We decided in favor of arginine (Arg) and lysine (Lys) as the amino acids to carry the isotope label, since trypsin cleaves C-terminal to both of them, rendering all proteolytic peptides except the C-terminal one accessible for quantitation. In a first step, we established protein expression methods which allowed us to produce SILAC-labeled proteins. We worked with a cell-free in vitro expression system as well as with a bacterial in vivo expression system. In both cases, we substituted normal arginine and lysine with their heavy isotope-labeled counterparts. While this is trivial to achieve in the cell-free approach by adding heavy instead of light amino acids to a minimal medium, in vivo it required working with an *E. coli* mutant that is auxotrophic for arginine and lysine to avoid contamination by endogenously produced unlabeled Arg and Lys. Janecki et al.[37] tried to produce isotope-labeled proteins in a regular *E. coli* BL21-strain, but they reported an incorporation efficiency of only 70%. This underscores that, unless the chosen amino acid(s) are essential to the host, auxotrophic mutants should be employed. We therefore used the *E. coli* strain AT713, which is deficient for the ArgA and the LysA genes. Since LysA catalyzes the last step in the biosynthesis pathway of lysine, this mutation should prevent the synthesis of lysine very efficiently. ArgA, however, is located further upstream in the biosynthesis pathway of arginine. We therefore challenged bacterial growth in the arginine-deficient medium to check auxotrophy. Since no growth occurred, we accepted AT713 as a strain suitable for the recombinant expression of SILAC-labeled proteins. Metabolic conversion from unlabeled precursors to Arg or Lys proved to be vanishingly low, with far less than 1% of the proteolytic peptides from the expressed proteins containing unlabeled Arg or Lys as determined by quantitative mass spectrometry (see below). Conversion of labeled arginine to proline can occur in certain strains or cell types[13] but was not observed here. The auxotrophic *E. coli* strain used in these experiments is not a dedicated expression strain with features like protease deficiency or optimized tRNA-composition. Nevertheless, we obtained good yields with all proteins tested. For the future, the construction of a SILAC version of an expression strain such as BL21 is still desirable. This strain would be auxotrophic for

arginine and lysine and would not be limited to lac-promotor-dependent expression like AT713 but could use T7-RNA-polymerase-dependent promotors. Absolute SILAC is in principle amenable to any expression system, including eukaryotic hosts, provided that they can be SILAC labeled and that the amino acids chosen for labeling are not affected by metabolic conversion.

For this study, we expressed the maltose-binding protein (MBP) in vivo in *E. coli* AT713, and growth factor receptor-bound protein 2 (Grb2) employing the in vitro approach. Figure 2 shows the basic workflow for the production of stable isotope-labeled protein standards. Several aspects in the procedure require special consideration. While cell-free production is rather economic due to the small volume required (mL range), the in vivo approach will consume several hundreds of dollars due to the larger culture volumes (L range) and attendant higher amounts of labeled amino acids. Since only small amounts of protein are consumed in MS experiments, the cell-free approach usually provides sufficient material for well-expressed proteins and is anyway the method of choice for toxic proteins. For multiple and long-term use as well as for poorly expressed proteins, the in vivo approach may be preferable. In any case, one should avoid an expression procedure that leads to degradation products which still carry the tag for purification. This would not only cause problems in quantitation of the standard prior to use but also result in different ratios for peptides derived from the N-terminal versus the C-terminal part. Purity of the standard protein is an important requirement because any contaminant proteins would lead to an overestimation of the concentration of the standard. Depending on the experiment, a purification tag on the recombinant SILAC protein may induce separation of standard and endogenous proteins during fractionation. Therefore, it should be cleavable to avoid differences in the physicochemical properties between the two proteins. The protease used for cleavage can either be chosen to likewise contain a tag for removal after cleavage or be added at a substoichiometric amount which does not impair the quantitation of the stock solution afterward.

Quantitation and Handling of the Standard Protein. No matter how sophisticated the mass spectrometric analysis and quantitation, total quantitation accuracy (rather than precision) can never be better than the biophysical or biochemical

quantitation of the standard protein. For MBP, we chose to measure the absorption at 280 nm, which is known to provide the most accurate values with an error margin of approximately 3% for proteins containing tryptophane.[31] Grb2 contained some contamination by nucleic acids as judged from the UV spectrum; therefore, we used amino acid analysis for this protein instead (error margin of around 10%).

Incorporation levels of labeled amino acids into the proteins were very close to 100%. Specifically, the in vitro expressed Grb2 incorporated 99.8% heavy isotope label, and the in vivo expressed MBP incorporated 100% arginine and 99.2% lysine in labeled form. This renders the error stemming from the labeled standard protein insignificant.

We found it to be highly important to properly aliquot and store the standard protein, especially at low concentrations. Any losses due to precipitation or adsorption to surfaces which happen after the standard quantitation will distort the quantitation values of the sample later on. We addressed this issue by subjecting labeled MBP to different storage conditions and by comparing it to unlabeled MBP freshly diluted from the stock, using quantitative MS as readout. Remarkably, for a solution of 15 fmol/μL ($\approx$1 μg/mL), the protein loss was more than 90% when stored at 4 °C for 24 h in a regular plastic tube. Storage at −20 and −80 °C increased the recovery substantially, but storage in special low-adsorption plastic tubes was even better (see Materials and Methods). The highest recovery, however, was achieved by presaturating the tubes with a solution of 1 mg/mL of BSA. In this case, it did not matter which kind of tubes we used. Presaturating with 10 mg/mL of BSA or storing the protein in 0.1 mg/mL of BSA did not further improve the yield. Under these conditions, even repeated freeze–thaw cycles did not show any negative effect. Note that addition of BSA to our standard would not influence quantitation results, as the amount of labeled protein has already been determined at this point. If no presaturation was carried out, it proved to be beneficial to store the protein under denaturing conditions (6 M urea, 2 M thiourea). Clearly this environment is not compatible with most subsequent fractionation steps and is only recommended if a direct in-solution digestion of the sample is envisioned. This observation also demonstrates that Absolute SILAC is a powerful tool for choosing between or optimizing different protein protocols.

Characteristics of Quantitation by Absolute SILAC. Figure 3 shows the quantitation of a protein in a whole cell lysate by Absolute SILAC. The total ions current trace and all mass spectra are the same as in a nonlabeling experiment. The only difference is that peptides of the protein to be quantified appear in SILAC doublets. This also provides an easy way to identify the protein, in principle even without sequencing or with sequencing only the peaks that appear as doublets.

One of the main advantages of Absolute SILAC is the fact that all tryptic peptides in a protein, except for the C-terminal one, appear in quantifiable peptide pairs when labeling with Arg and Lys. Compared to methods based on the MS quantitation of only one or two peptides per protein, this permits a more accurate quantitation and more robust statistical assessment. Since most peptides will usually not be identifiable due to sample complexity, dynamic range, or ionization issues, using a fully labeled protein increases the chance of obtaining a sufficient number of peptides for quantitation.

Post-translational modifications or alternative splicing are possible confounding factors in protein quantitation. In the case of Absolute SILAC, stoichiometric post-translational modi-
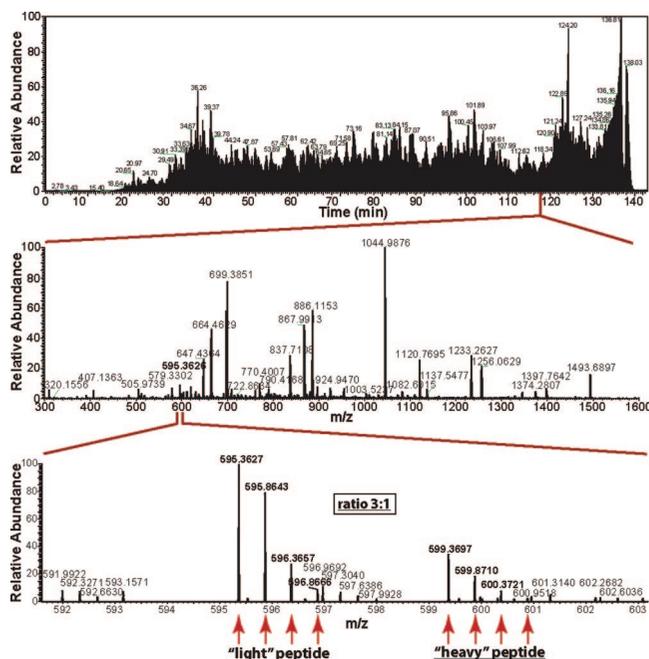
Figure 3. LC-MS-based quantitation of a peptide in a complex sample. Whole cell HeLa-lysate (1 μg) was spiked with 50 fmol of labeled and 150 fmol of unlabeled MBP. The upper panel illustrates the total ion signal across the elution time of the LC-MS experiment. The middle panel shows a survey scan acquired at the indicated time point. Zooming into this spectrum, the lower panel highlights a doubly charged SILAC−peptide pair with a ratio of 3:1 between light and heavy versions. The mass offset corresponds to the expected difference of 8.0142 Da for a peptide containing $^{13}C_6\,^{15}N_2$-lysine.

fications on the endogenous proteins will not lead to a peptide pair as the recombinant protein will not be modified, and hence they will not lead to a quantitation error. Substoichiometric modifications or splice variants may be recognizable because of the altered ratio of the nonmodified peptide pair compared to the ratios of peptides that carry no modification.[37] While this is a potential pitfall for methods which rely on only one or two peptides for quantitation, it can be exploited as an advantage in the case of Absolute SILAC. For example, if 50% of the endogenous protein is phosphorylated at a particular location, the ratio for the nonphosphorylated peptide will be twice that of the other peptides, indicating that the peptide in question may be modified.

To characterize the behavior of Absolute SILAC across a large range of possible ratios between the "heavy" standard and the "light" endogenous protein, we analyzed defined ratios of heavy and light labeled MBP, keeping the amount of light MBP constant at 150 fmol. We created a titration curve ranging from 1:300 to 300:1 (heavy to light protein) in steps of a factor of 3 (Figure 4). Defined ratios of MBP were digested in solution, either in pure form or spiked into 1 μg of whole cell HeLa-lysate. Automated quantitation of the mass spectrometric data was carried out by MSQuant (see Materials and Methods). The linearity between expected and experimental ratios proved to be excellent over a broad range from 1:100 to 100:1 with correlation coefficients of 0.9994 and 0.9964, respectively. Thus, the dynamic range of Absolute SILAC can span 4 orders of magnitude, providing a large window for the researcher who sometimes may only have a vague idea of the amount of standard protein required to mix into the sample. Table 1
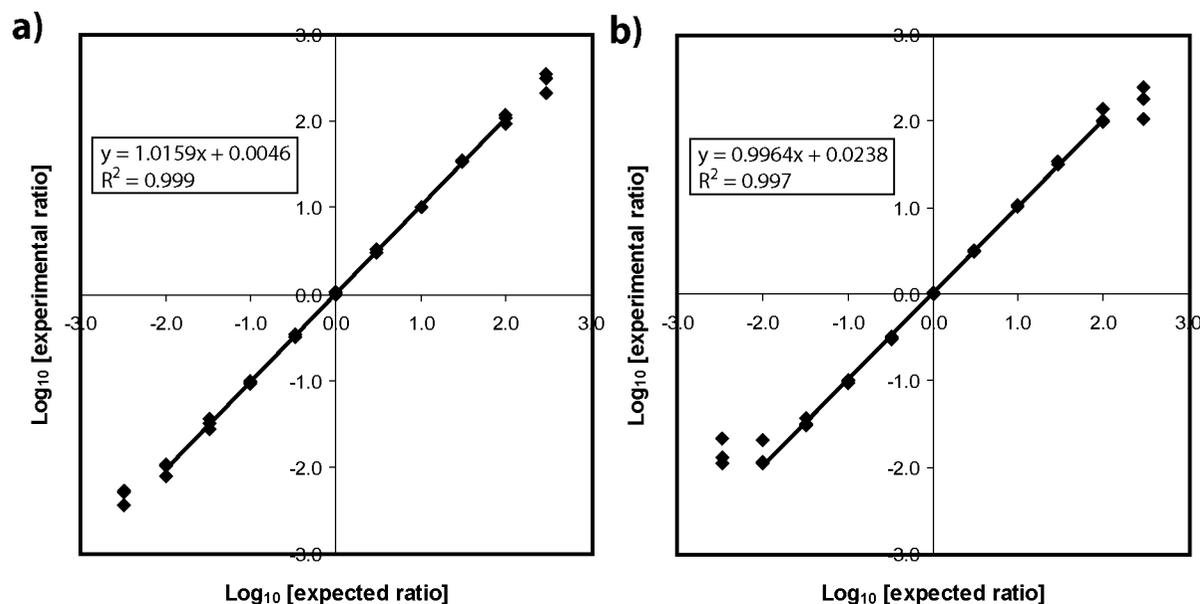
Figure 4. Linearity of quantitation of a protein by Absolute SILAC. The figure shows titration curves of a protein (maltose-binding protein) in pure form (a) and in the background of 1 μg of total cell lysate (b). The heavy, SILAC-labeled version of the protein was kept constant at 150 fmol, whereas the amount of unlabeled version varied from 500 amol to 45 000 fmol, resulting in ratios between 1:300 and 300:1. The regression line was calculated based on the values from 1:100 to 100:1. For each ratio, three independent experiments were performed, and all data points are indicated in the figure and in Table 1.

contains the data displayed in Figure 4. Within a 10-fold difference between labeled and unlabeled proteins, the deviation from the expected ratio is typically below 5%. In technical replicates, we generally observe fluctuations of 2–3% (data not shown), which indicates that around half of the error stems from the quantitative MS readout, and the other half is due to upstream sample handling. The standard deviations arising from fluctuations of the quantitation of individual peptides amount to approximately 10% of the average value. We were surprised that even ratios of 300 could still be quantified relatively well. In the background of a complex mixture, however, the quantitative accuracy at this ratio is poor (up to a factor of 2–3 uncertainty). Under these circumstances, the low abundant peptide partner is barely detectable, and background ions can sometimes appear even in the narrow quantitation window (here set to 0.02 Da) and contribute to the ratio. In most cases, manual inspection can prevent this, and software may be able to automatically subtract such noise as well. In any case, one can employ an iterative strategy by adjusting the amount of standard protein to be spiked into the sample after a first rough measurement. We generally recommend to aim for an excess of 3 to 10 times of the labeled versus unlabeled protein. While this is still within the range of a highly accurate quantitative readout, it increases the likelihood that more peptide pairs can be identified during the LC-MS experiment. Another important point is the quality of instrumentation. High-resolution mass spectrometers are capable of separating coeluting peptides with similar masses, and high sensitivity aids in increasing the signal of weak peaks. Fast sequencing cycles allow the identification and subsequent quantitation of a protein on the basis of several peptides. Since accuracy of quantitation is mainly influenced by signal-to-noise, larger peaks can be quantified more accurately than smaller peaks.[1]

Targeted Narrow Mass Range Scans for Enhanced Sensitivity, Signal-to-Noise Ratio, and Quantitative Accuracy. A central challenge for mass spectrometry especially in the biomarker field is the quantitation of peptides in the presence

of a very complex background. In triple quadrupole-type instruments, this problem can be alleviated by monitoring a fragment of the ion of interest rather than the ion itself. This technique can be multiplexed (Multiple Reaction Monitoring, MRM) and allows the detection of substances much below their signal-to-noise ratio in a full scan.[16–19] Unfortunately, it cannot be used on a trapping instrument such as the LTQ-Orbitrap employed here. We therefore developed a data acquisition scheme that addresses this challenge and is able to substantially improve the usual limit of detection and quantitation. In experiments on the LTQ-Orbitrap, where the full mass range is trapped, total dynamic range is limited by the fact that the C-trap can only accommodate a finite number of ions (one million). However, selected ion monitoring (SIM) scans accumulate only a very narrow mass range. Consequently, ions outside of this range do not limit the measurement; the desired number of ions is devoted to this mass window; and the dynamic range is solely determined by ions within it. We reasoned that the mass spectrometer could be programmed to recognize peptides belonging to the labeled protein standard based on their precise mass. Upon this trigger, it could start acquiring a series of narrow mass range windows including the labeled peptides and the calculated location of the endogenous peptide, even if it was not visible in the spectrum. In this way, the effective accumulation time in complex mixtures, which is usually only a few milliseconds, should increase at least 10-fold with concomitant increase in sensitivity for the protein to be quantified. Therefore, using data-dependent acquisition with an inclusion list containing the masses of the expected peptides of the SILAC-labeled standard protein, we instructed the mass spectrometer to record several SIM scans as soon as an ion from the inclusion list was detected. Resolution, amount of ions, fill time, and dimensions of the isolation window can be defined individually according to the type of sample and the goal of the experiment. We chose a window width for the SIM scans from −6 to +3 m/z units around the detected mass from the inclusion list, covering the complete isotope pattern of the

Table 1. Titration Experiments of a Protein (Maltose-Binding Protein) in Pure Form and in the Background of 1 μg of Total Cell Lysate (see Figure 4)[a]

expected ratio	1 μg of cell	peptides	exptl ratio
heavy:light	lysate added	quantified	heavy:light (avg ± SD)
1:300	-	4	0.0036 ± 0.0013
1:300	-	5	0.0052 ± 0.0021
1:300	-	4	0.0051 ± 0.0021
1:300	×	5	0.0213 ± 0.0182
1:300	×	3	0.0110 ± 0.0068
1:300	×	4	0.0129 ± 0.0113
1:100	-	6	0.0105 ± 0.0045
1:100	-	4	0.0109 ± 0.0057
1:100	-	8	0.0080 ± 0.0021
1:100	×	5	0.0115 ± 0.0045
1:100	×	7	0.0111 ± 0.0030
1:100	×	8	0.0203 ± 0.0138
1:30	-	7	0.0317 ± 0.011
1:30	-	4	0.0355 ± 0.021
1:30	-	6	0.0274 ± 0.014
1:30	×	3	0.0301 ± 0.012
1:30	×	9	0.0372 ± 0.023
1:30	×	11	0.0313 ± 0.011
1:10	-	6	0.0940 ± 0.017
1:10	-	10	0.0998 ± 0.021
1:10	-	10	0.0907 ± 0.011
1:10	×	13	0.0982 ± 0.017
1:10	×	11	0.0946 ± 0.012
1:10	×	10	0.1012 ± 0.013
1:3	-	12	0.314 ± 0.035
1:3	-	15	0.312 ± 0.046
1:3	-	16	0.338 ± 0.031
1:3	×	8	0.324 ± 0.052
1:3	×	13	0.304 ± 0.029
1:3	×	9	0.316 ± 0.051
1:1	-	18	1.024 ± 0.081
1:1	-	15	0.991 ± 0.100
1:1	-	14	1.040 ± 0.043
1:1	×	22	1.026 ± 0.067
1:1	×	19	1.005 ± 0.110
1:1	×	5	1.010 ± 0.121
3:1	-	25	3.078 ± 0.19
3:1	-	26	3.085 ± 0.22
3:1	-	18	3.221 ± 0.38
3:1	×	29	3.137 ± 0.25
3:1	×	29	3.161 ± 0.27
3:1	×	32	3.195 ± 0.30
10:1	-	10	9.965 ± 0.99
10:1	-	12	10.19 ± 2.22
10:1	-	35	10.16 ± 1.33
10:1	×	36	10.67 ± 1.83
10:1	×	28	10.41 ± 1.27
10:1	×	11	10.07 ± 0.79
30:1	-	34	34.83 ± 5.83
30:1	-	38	33.76 ± 6.03
30:1	-	34	34.14 ± 7.91
30:1	×	35	34.56 ± 6.07
30:1	×	34	34.18 ± 5.61
30:1	×	22	31.47 ± 5.22
100:1	-	30	108.1 ± 44.7
100:1	-	34	115.5 ± 54.9
100:1	-	23	94.6 ± 34.7
100:1	×	31	99.6 ± 40.9
100:1	×	11	137.6 ± 53.3
100:1	×	24	98.5 ± 33.9
300:1	-	23	353.6 ± 252.5
300:1	-	29	316.4 ± 204.3
300:1	-	27	213.9 ± 133.9
300:1	×	24	105.9 ± 41.6
300:1	×	23	183.3 ± 107.1
300:1	×	24	247.7 ± 169.2

[a] The heavy, SILAC-labeled version of the protein was kept constant at 150 fmol, whereas the amount of unlabeled version varied from 500 amol to 45 000 fmol, resulting in ratios between 1:300 and 300:1. For each ratio, three independent experiments were performed.

SILAC pair for both Arg and Lys and doubly and higher-charged peptides. This SIM width is sufficiently narrow to exclude the

vast majority of coeluting ions, typically eliminating more than 99% of the total ion current when measuring low concentrations of the protein of interest. We checked if the SIM isolation had any effect on the measured peptide ratios but did not find this to be the case.

Figure 5 illustrates the power of SILAC-directed SIM scans for absolute quantitation close to the detection limit. While the signal of the endogenous peptide was not detected in full scans, it was readily recorded in SIM scans, showing the boost in sensitivity gained by this approach. In addition, the signal-to-noise ratio is markedly increased, which is of great importance for accurate quantitation. The fluctuations in peak intensity and shape of the isotope pattern that are observed for low signals in consecutive full scans were also largely reduced in SIM scans. Injection times of electrosprayed ions into the ion trap to reach the nominal target value of 10^6 ions for the full scan were 5 ms on average. To reach the 10^5 ions in the SIM scan required an average injection time of 100 ms, in agreement with the 10- to 20-fold higher signal-to-noise ratio observed (Figure 5). This example also highlights the benefits of spiking the SILAC-labeled protein at a higher amount than the level of endogenous protein. In fact, in the example given in Figure 5, the peptide would not have been detected if the standard had been present at the same amount as the unlabeled protein because the data systems would not have triggered the SIM window acquisition. Once the more abundant standard peptide is identified, the quantitation software can easily pinpoint where in the spectrum the lower abundant endogenous peptide peak is located. Thus, regardless of the endogenous level of the protein of interest, the labeled standard should be added at a level enabling robust detection by the data system. Our targeted SIM window approach bypasses the normal detection limit in complex mixtures, which is given by sequencing speed, ion count, and fragment spectrum quality.[38] It therefore enables the identification and quantitation of proteins even below the identification threshold in our standard proteomic experiments.

Taking advantage of this strategy, we identified and quantified as little as 150 attomole of a protein (MBP) in a very complex matrix (whole cell HeLa lysate) without any fractionation prior to the LC-MS experiment (Figure 5). This corresponds to a detection and quantitation limit of an endogenous protein of roughly 10 000 molecules per cell. Fractionation such as one-dimensional gel electrophoresis or peptide fractionation would be expected to allow the quantitation of even lower amounts. Even at this very low level in a very complex mixture, the quantitation error was only 7% (ratio of 9.3 instead of 10.0 as expected). Quantitation precision seems to be limited by the short elution times, which permit only a few scans, and by the low number of peptides of the protein (two in this case) rather than reduced quantitative accuracy in single scans.

So far, only multiple reaction monitoring (MRM) measurements on triple quadrupole mass spectrometers have been reported to reach such low levels for quantitation.[16,17] With the approach presented here, the identification and accurate quantitation of proteins in the attomole range in highly complex samples are now accessible to other types of mass spectrometers. The SIM scan strategy is a valuable tool for absolute quantitation, which shifts the detection limit in complex mixtures by at least a factor of 10 and allows for unprecedented accuracy of quantitation at these low levels compared to full scan quantitation on the same instruments.

Determination of Copy Number per Cell of Grb2. It should be possible to use Absolute SILAC to reveal the copy number

78
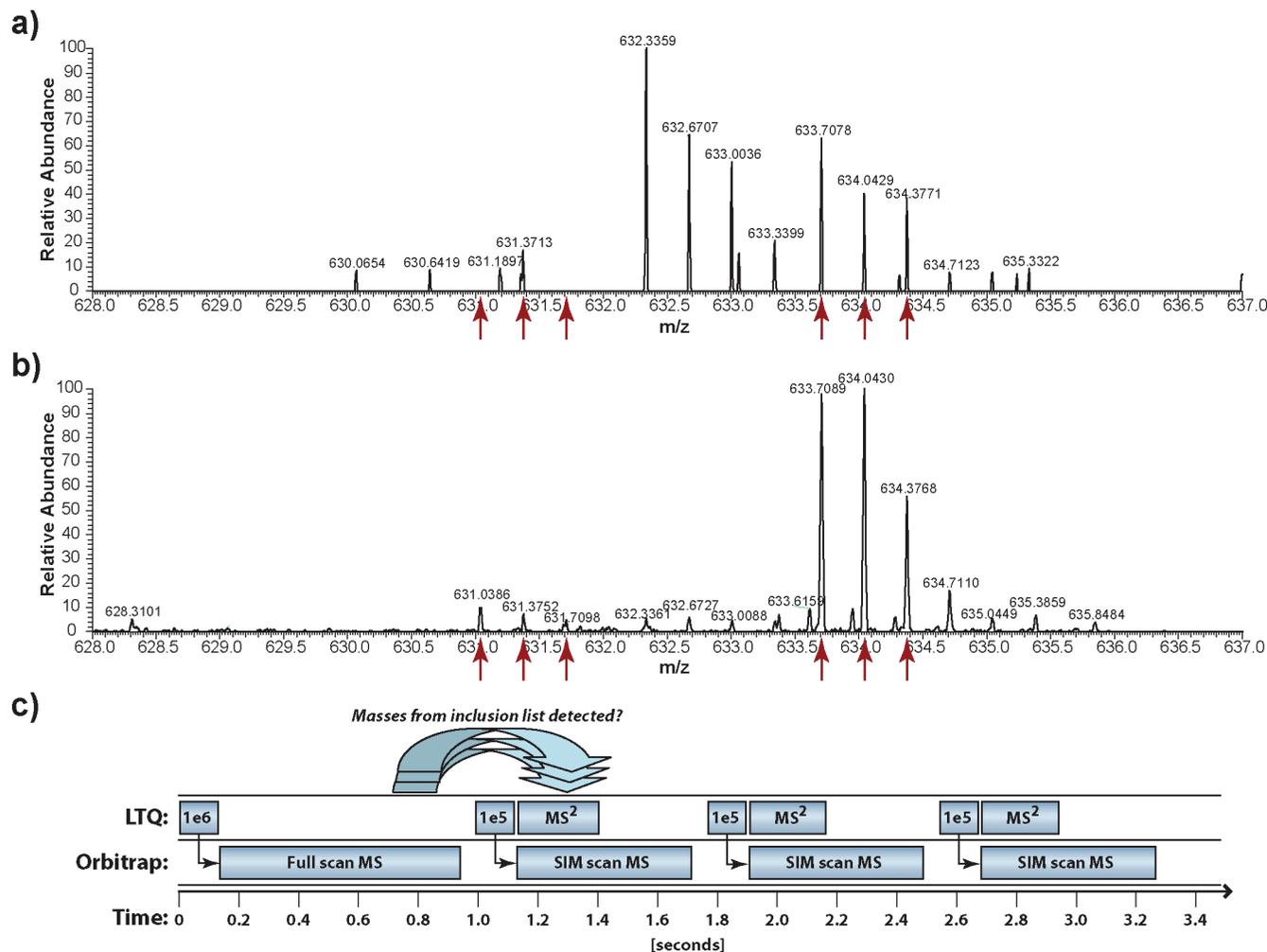
Figure 5. Targeted narrow mass range scans enhance the sensitivity of detection and quantitation. The figure shows analysis of 1.5 fmol of heavy MBP together with 150 amol of light MBP in 1 μg of HeLa-cell lysate, digested in solution and analyzed by reversed phase LC–MS using a 2 h gradient. In the upper panel, the acquisition software detected an ion from the SILAC-labeled standard (LIAYPIAVEALSLIYNK, m/z = 633.7078 in the inclusion list) and automatically triggered a SIM scan. The red arrows indicate the isotope pattern of the labeled peptide (right) and the calculated position at which the unlabeled peptide to be quantified would be expected. In the SIM scan (middle panel), the unlabeled peptide is clearly visible (left group of arrows) due to increased accumulation time. (The coeluting peptide with m/z 632.3359 disappeared during the second between the full scan and SIM scan.) The bottom panel explains the acquisition strategy. Rectangles in the first row represent target values for ion accumulation and triggering of MS/MS events. The second row depicts transient accumulation in the orbitrap for full scans and SIM scans.

per cell of any protein that can be expressed recombinantly. As an example, we chose the signaling adaptor protein growth factor receptor bound protein 2 (Grb2). This 25 kDa protein contains phosphotyrosine and polyproline binding domains and is a key player in the transduction of growth signals from receptors to intracellular effectors. We harvested HeLa, HepG2, and C2C12 cells during the exponential growth phase at 70% confluency and subjected defined numbers of cells to lysis, removal of insoluble material by centrifugation, and subsequent SDS-PAGE. We added defined amounts of labeled standard Grb2 to supernatant and pellet after the centrifugation step. Following SDS-PAGE, we excised the region corresponding to Grb2, performed in-gel digestion with trypsin, and analyzed the samples by LC-MS (Figure 6). The quantitative results from five separate aliquots of cells of each cell line are listed in Table 2. We measured the number of molecules of Grb2 per HeLa cell to be $5.55 \pm 0.26 \times 10^5$. Of these, 4.4% were retained in the pellet. The liver cell line HepG2 had $8.80 \pm 0.12 \times 10^5$ Grb2 molecules (1.2% in pellet), and the murine muscle cell line C2C12 contained $5.67 \pm 0.20 \times 10^5$ copies (3.2% in pellet) per

cell. Due to the high identity between the human and murine sequence of Grb2 (only one amino acid is different), the human Grb2 standard protein could also be used for C2C12 cells. The literature reports 1.5×10^4 copies per cell for the yeast homologue.[39] Tissue-derived murine thymocytes, which are also much smaller than HeLa cells, were shown to contain 2.3×10^4 copies,[40] and hepatocytes isolated from rats had 5.1×10^4 Grb2 molecules.[41] The finding that exponentially growing transformed cell lines contain an order of magnitude more copies of the growth-related protein Grb2 as compared to tissue-derived cells or yeast cells is not surprising. Moreover, copy numbers of growth factor receptors like epidermal growth factor receptor (EGFR), insulin-like growth factor 1 receptor (IGF1R), or insulin receptor (INSR) are in the range of 10^4 to 10^5 depending on cell type. Thus, it appears that the number of Grb2 molecules, whose function is to bind to these receptors and other signaling proteins in the cell, is not a bottleneck in propagating the signal in the cell lines investigated here.

In a recent large-scale measurement of the HeLa cell proteome, Grb2 was ranked as protein number 996 out of more
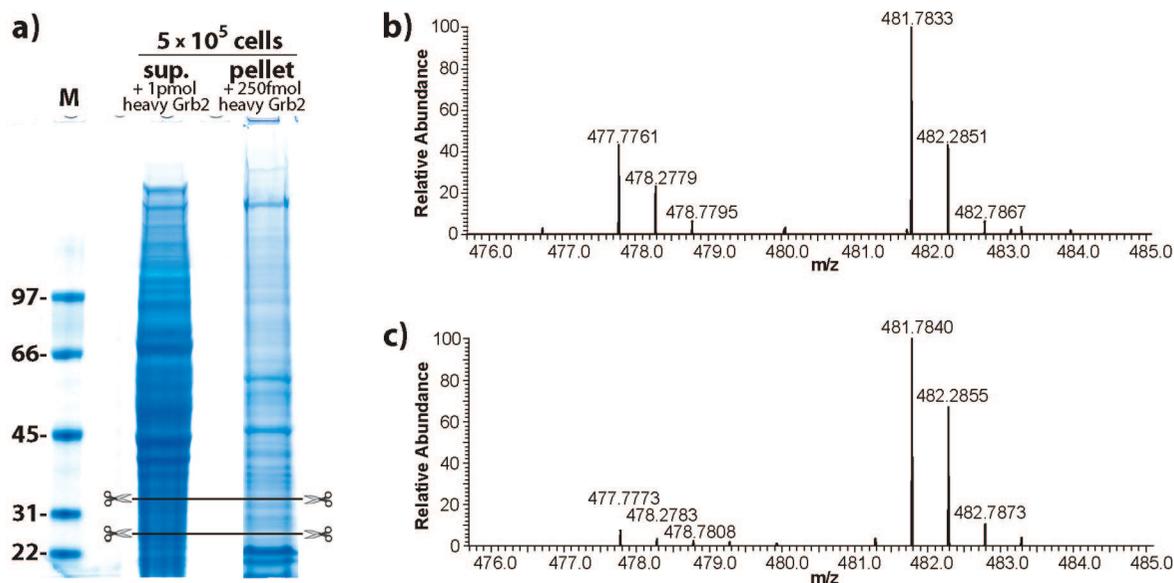
Figure 6. Copy number of Grb2 in HeLa cells. The pellet and supernatant of a cell lysate were spiked with defined amounts of SILAC-labeled Grb2 and separated by SDS-PAGE (a). The gel region around Grb2 was in-gel digested and analyzed. Representative spectra with SILAC pairs of the doubly charged peptide YFLWVVK from the supernatant (b) and pellet (c) are depicted. The entire experiment was repeated five times.

than 4000 proteins based on summed peptide intensities.[42] This rough measure of protein abundance therefore suggests that more than a thousand proteins have copy numbers of several hundred thousand.

The small fluctuations of the results in samples 1–5 lie within the expected error range (Table 2). The pipetting error at the spiking step presumably dominates this error due to the small volume of 2 μL which was added. In general, the error arising from mass spectrometric quantitation is low compared to the crucial steps during sample processing. In particular, the care which is taken during the following steps will determine the accuracy of quantitation that one can expect: cell counting, complete cell lysis, quantitation of the recombinant SILAC-labeled protein, proper storage and handling of the protein, and precision in pipetting.

It is important to avoid errors due to differences between recombinant and endogenous proteins, for example, due to aggregation of the standard. In our study of Grb2, this turned out to be important because when we spiked the protein before cell lysis, 35% of the labeled protein ended up in the pellet, as opposed to only 4% of the endogenous protein. This is presumably because Grb2 was mainly present in inclusion bodies and might not have refolded completely afterward. We therefore recommend addressing this issue in case spiking before removal of the pellet is desired or biochemical fractionation of the supernatant is planned. Because we used a non-native method for fractionation, SDS-PAGE, our analysis was not affected by this observation. We also did not remove the His-tag used for purification but instead excised a slightly broader slice from the gel.

Isotope Distributions in Absolute Quantitation by MS-Based Proteomics. No matter which approach involving stable isotope labeling is employed, one should keep in mind that the labeled amino acids are not provided at complete isotope content by the manufacturer. Typically, an isotope enrichment of 98 atom percent or higher is specified. For $^{13}C_6{}^{15}N_2$-lysine, this implies that 0.98^8 or 85.1% of the molecules are fully labeled. 12.9% contain one ^{12}C or ^{14}N atom; 1.8% contain two;

and 0.2% contain three or more of them. Those 14.9% which are not fully labeled will accordingly shift one unit to lower mass in the isotope pattern of the peptide. This does not mean that 14.9% of the signal is lost, but instead it implies that this percentage of each peak moves to the neighboring peak on the left. Depending on peptide mass, this effect can either increase or decrease the signal of a peak and create an additional peak to the low mass side of the monoisotopic peak. For maximum accuracy, the quantitation algorithm should take this fact into account. However, there is also the antagonistic effect of the labeled atoms not participating in the isotopic distribution (they cannot be ^{12}C or ^{14}N atoms). This renders the monoisotopic peak slightly higher than it would otherwise be. In our experience, those two effects largely compensate for each other at the level of the monoisotopic peak, which is why we based the quantitation on the peak areas of the monoisotopic peaks. We are currently developing a program which takes into account both of the above-mentioned phenomena, eliminating any bias caused by the quality of the labeled amino acids.

Practical Aspects of Using Absolute SILAC Compared to Other Methods. Several different methods exist for absolute quantitation of proteins using stable isotope-labeled amino acids. Any of them may be the most appropriate depending on the quantitation task, but there are some general properties that define their potential for absolute quantitation. Differentiating aspects are the time point in the workflow when the standard can be mixed in, the amount of peptides per protein amenable to quantitation, completeness of labeling, as well as practical aspects such as costs and labor associated with the method. In any of the methods, the heavy isotopes should preferably introduce a mass difference of at least 4 Da to avoid overlapping isotope clusters, and it should ensure chromatographic coelution. Furthermore, each method needs suitable bioinformatic software that can deal with the data and provide a quantitative output.

The AQUA method, among the methods using stable isotope-labeled amino acids, is the most straightforward to apply but also the least accurate for absolute quantitation of protein in

Table 2. Copy Number of Grb2 per Cell for HeLa, HepG2, and C2C12 Cells[a]

HeLa cells	analyzed fraction	spiked amount of heavy Grb2	peptides quantified	ratio heavy:light (avg ± SD)	fmol light Grb2	% Grb2 in pellet	copy number per cell
1a	supernat.	1 pmol	12	2.47 ± 0.16	405	4.3	5.10×10^5
1b	pellet	250 fmol	7	13.72 ± 2.31	18.2		
2a	supernat.	1 pmol	10	2.52 ± 0.20	397	3.3	4.95×10^5
2b	pellet	250 fmol	6	18.37 ± 1.91	13.6		
3a	supernat.	1 pmol	12	2.34 ± 0.16	427	4.9	5.41×10^5
3b	pellet	250 fmol	9	11.46 ± 2.04	21.8		
4a	supernat.	1 pmol	11	1.99 ± 0.24	503	3.5	6.28×10^5
4b	pellet	250 fmol	8	13.71 ± 2.21	18.2		
5a	supernat.	1 pmol	13	2.12 ± 0.17	472	5.9	6.04×10^5
5b	pellet	250 fmol	8	8.44 ± 1.06	29.6		
					average (±SEM):	4.4 ± 0.49	$5.55 \pm 0.26 \times 10^5$

HepG2 cells	analyzed fraction	spiked amount of heavy Grb2	peptides quantified	ratio heavy:light (avg ± SD)	fmol light Grb2	% Grb2 in pellet	copy number per cell
1a	supernat.	1 pmol	14	1.36 ± 0.11	735	1.4	8.98×10^5
1b	pellet	100 fmol	8	9.65 ± 1.90	10.4		
2a	supernat.	1 pmol	11	1.33 ± 0.07	752	1.1	9.16×10^5
2b	pellet	100 fmol	9	11.46 ± 2.32	8.7		
3a	supernat.	1 pmol	14	1.42 ± 0.14	704	1.2	8.58×10^5
3b	pellet	100 fmol	7	12.22 ± 1.04	8.2		
4a	supernat.	1 pmol	14	1.43 ± 0.11	699	1.2	8.52×10^5
4b	pellet	100 fmol	9	11.66 ± 1.71	8.6		
5a	supernat.	1 pmol	11	1.39 ± 0.11	719	1.2	8.76×10^5
5b	pellet	100 fmol	8	11.70 ± 1.58	8.5		
					average (±SEM):	1.2 ± 0.05	$8.80 \pm 0.12 \times 10^5$

C2C12 cells	analyzed fraction	spiked amount of heavy Grb2	peptides quantified	ratio heavy:light (avg ± SD)	fmol light Grb2	% Grb2 in pellet	copy number per cell
1a	supernat.	1 pmol	11	2.44 ± 0.19	410	3.4	5.11×10^5
1b	pellet	100 fmol	7	6.92 ± 0.82	14.5		
2a	supernat.	1 pmol	11	2.06 ± 0.22	485	3.1	6.03×10^5
2b	pellet	100 fmol	8	6.51 ± 0.77	15.4		
3a	supernat.	1 pmol	11	2.36 ± 0.21	424	3.1	5.27×10^5
3b	pellet	100 fmol	10	7.38 ± 1.28	13.6		
4a	supernat.	1 pmol	8	2.10 ± 0.23	476	3.4	5.93×10^5
4b	pellet	100 fmol	10	6.06 ± 0.49	16.5		
5a	supernat.	1 pmol	10	2.06 ± 0.14	485	3.0	6.02×10^5
5b	pellet	100 fmol	9	6.72 ± 0.79	14.9		
					average (±SEM):	3.2 ± 0.09	$5.67 \pm 0.20 \times 10^5$

[a] The quantitative data from five separate experiments are shown, each divided into supernatant and pellet.

the original sample. The loss of protein to be quantified during sample processing can be substantial, and the recovery of different peptides of a protein can vary tremendously[43] leading to different quantitative results depending on the AQUA peptide chosen. Especially in the case of gel-based procedures, much material can be lost due to incomplete digestion and also due to incomplete extraction, since different peptides are extracted with different efficiencies from the gel.[27] Therefore, even though this method can be very valuable to obtain a rough estimate of protein amount, it does not directly lead to knowledge of the copy number or concentration of the protein. However, for simple samples which do not require any fractionation and if complete digestion in solution is assured, these limitations may not be serious.

Relying on only one or two peptides for quantitation of a protein is not only a limitation for statistical assessment but also, unless MRM methods are employed, risky in terms of detectability. Even if one decides to synthesize more peptides of a protein, one can encounter obstacles. Some sequences are difficult to synthesize chemically, and several amino acids must be avoided because they are susceptible to modifications occurring in vitro. Furthermore, optimal AQUA peptides for one

ionization method and mass spectrometer may not be optimal for another set-up.

QconCAT is an interesting approach combining features of peptide and protein standards by designing a genetic construct consisting of many peptide sequences from different proteins fused together into one artificial protein. When assembling the concatenated sequence, it is easily possible to optimize the codons for the desired expression host. One QconCAT protein can be used for the quantitation of several different proteins; however, the desired amount to add might be very different for the various target proteins. While this technique is potentially an advance over peptide standards, it is still affected by similar problems such as the peptide sequences to be selected, the time point of spiking, the low number of peptides per protein to be quantified, as well as the necessity of complete digestion. However, compared to AQUA, this method may more accurately recreate the peptide elution step during in-gel digestion of proteins.[26]

Absolute SILAC is unaffected by these problems, and conceptually its only limitations are the steps preceding the time point of spiking (i.e., unknown cell lysis efficiency). The method is of course still susceptible to general sources of error such as

81

pipetting, etc., but those should not add up to more than 5–10% when done with care. While this manuscript was in preparation, an elegant study of the comparison of peptides, concatenated peptides, and intact proteins as standards for absolute quantitation was published.[44] The authors found that the QconCAT approach was superior to AQUA but that both of them severely underestimated the actual protein amount. In contrast, the in vitro expressed intact protein standard provided a correct quantitation. We therefore recommend the use of Absolute SILAC unless specific experimental reasons suggest a different approach. The major practical challenge in Absolute SILAC is the need to recombinantly express and purify the proteins of interest. Some proteins are very difficult to express and purify, and if one cannot settle for the expression of a domain instead, the best alternative would be to employ AQUA or QconCAT. Obtaining purified SILAC-labeled proteins can of course be labor-intensive. If properly stored and aliquoted, however, this should be a one-time effort because only minute amounts are needed per experiment. If desired, one can even get around expressing SILAC-labeled proteins and instead use unlabeled proteins by spiking it into a labeled cell line. This even works for tissue samples, if a SILAC-labeled animal is available (Krueger et al., submitted), and for researchers working with SILAC cell lines, it is in any case a very convenient approach. Furthermore, unlike AQUA and QconCAT and other peptide-based approaches, Absolute SILAC is fully compatible with top-down approaches where complete proteins instead of their proteolytic peptides are measured by mass spectrometry.[45] In principle, modified versions of recombinant proteins can also be produced, which should allow the accurate quantitation of processed proteins, protein isoforms, or proteins with post-translational modifications.

Conclusions and Perspectives

Accurate absolute quantitation of proteins in complex environments as presented here by Absolute SILAC can deliver excellent quantitative information. We have introduced a targeted approach involving SIM scans, which makes some of the advantages of MRM scans available on trapping instruments and which extends the detection and quantitation limit in complex mixtures by at least one order of magnitude. Importantly, this approach allows retrieving previously unobservable endogenous peptides not only in Absolute SILAC but also in any method employing labeled standard peptides, such as AQUA. As an example, narrow range scans could be triggered on labeled analogues of peptides from glyco-captured plasma proteins.[46]

Data obtained by Absolute SILAC can supply biology in general with a central piece of information on proteins of interest and systems biology in particular with input for modeling. Here we measured the copy number of a protein per cell, but the stoichiometry within protein complexes and the concentration of potential biomarkers in body fluids can also be determined. We believe that Absolute SILAC may have a possible role in biomarker quantitation in clinical diagnostics. Biomarkers would have to be expressed recombinantly once and distributed into many protein standards, similar to standards for ELISA tests used today. The absolute quantitation of many different proteins at once in single LC-MS experiments should be possible in a multiplexed format, after generating a "master-mix" of known amounts of several diagnostically relevant proteins in labeled form. This might eventually also

be a cheaper approach as compared to conducting multiple immunoassays for the various proteins.

In academia, most research laboratories focus on a small number of proteins for their investigations. They are often experts in expressing and purifying them from recombinant sources and will have a strong interest in their absolute amounts in the biological system studied. Given the general availability of in vitro expression systems as well as auxotrophic strains from bacteria or yeast and the growing access to MS facilities, the scientific community should be able to determine absolute quantities of many proteins in different cell lines or tissues. The data could be gathered by a central database such as Swissprot and included in the protein information sheet online.

Abbreviations: MS, mass spectrometry; LC, liquid chromatography; MS/MS, tandem mass spectrometry; SIM, selected-ion-monitoring; LTQ, linear quadrupole ion trap; Da, Dalton; RP, reversed phase; SILAC, stable isotope labeling by amino acids in cell culture; ICAT, isotope coded affinity tag; iTRAQ, isobaric tag for relative and absolute quantification; CDIT, culture derived isotope tag; AQUA, absolute quantitation; QconCAT, quantitative concatamer; Arg, arginine; Lys, lysine; Grb2, growth factor receptor-bound protein 2; MBP, maltose-binding protein; BSA, bovine serum albumin; fmol, femtomole; amol, attomole; UV, ultraviolet.

References

(1) Ong, S. E.; Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005**, *1* (5), 252–62.

(2) Chelius, D.; Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **2002**, *1* (4), 317–23.

(3) Wang, G.; Wu, W. W.; Zeng, W.; Chou, C. L.; Shen, R. F. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility linearity, and application with complex proteomes. *J. Proteome Res.* **2006**, *5* (5), 1214–23.

(4) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **2003**, *75* (18), 4818–26.

(5) Staes, A.; Demol, H.; Van Damme, J.; Martens, L.; Vandekerckhove, J.; Gevaert, K. Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-18. *J. Proteome Res.* **2004**, *3* (4), 786–91.

(6) Yao, X.; Afonso, C.; Fenselau, C. Dissection of proteolytic 18O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates. *J. Proteome Res.* **2003**, *2* (2), 147–52.

(7) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17* (10), 994–9.

(8) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–69.

(9) Schmidt, A.; Kellermann, J.; Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, *5* (1), 4–15.

(10) Olsen, J. V.; Andersen, J. R.; Nielsen, P. A.; Nielsen, M. L.; Figeys, D.; Mann, M.; Wisniewski, J. R. HysTag--a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol. Cell. Proteomics* **2004**, *3* (1), 82–92.

(11) 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* (12), 6591–6.

(12) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture. SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376–86.

(13) Ong, S. E.; Kratchmarova, I.; Mann, M. Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* **2003**, *2* (2), 173–81.

(14) Ong, S. E.; Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **2006**, *1* (6), 2650–60.

(15) Ishihama, Y.; Sato, T.; Tabata, T.; Miyamoto, N.; Sagane, K.; Nagasu, T.; Oda, Y. Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nat. Biotechnol.* **2005**, *23* (5), 617–21.

(16) Anderson, L.; Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **2006**, *5* (4), 573–88.

(17) Onisko, B.; Dynin, I.; Requena, J. R.; Silva, C. J.; Erickson, M.; Carter, J. M. Mass spectrometric detection of attomole amounts of the prion protein by nanoLC/MS/MS. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (6), 1070–9.

(18) Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* **2007**.

(19) Stahl-Zeng, J.; Lange, V.; Ossola, R.; Eckhardt, K.; Krek, W.; Aebersold, R.; Domon, B. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell. Proteomics* **2007**, *6* (10), 1809–17.

(20) Desiderio, D. M.; Kai, M. Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed. Mass Spectrom.* **1983**, *10* (8), 471–9.

(21) Barr, J. R.; Maggio, V. L.; Patterson, D. G., Jr.; Cooper, G. R.; Henderson, L. O.; Turner, W. E.; Smith, S. J.; Hannon, W. H.; Needham, L. L.; Sampson, E. J. Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin. Chem.* **1996**, *42* (10), 1676–82.

(22) 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* (12), 6940–5.

(23) Kippen, A. D.; Cerini, F.; Vadas, L.; Stocklin, R.; Vu, L.; Offord, R. E.; Rose, K. Development of an isotope dilution assay for precise determination of insulin. C-peptide, and proinsulin levels in non-diabetic and type II diabetic individuals with comparison to immunoassay. *J. Biol. Chem.* **1997**, *272* (19), 12513–22.

(24) Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods* **2005**, *2* (8), 587–9.

(25) Pratt, J. M.; Simpson, D. M.; Doherty, M. K.; Rivers, J.; Gaskell, S. J.; Beynon, R. J. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* **2006**, *1* (2), 1029–43.

(26) Rivers, J.; Simpson, D. M.; Robertson, D. H.; Gaskell, S. J.; Beynon, R. J. Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol. Cell. Proteomics* **2007**, *6* (8), 1416–27.

(27) Havlis, J.; Shevchenko, A. Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry. *Anal. Chem.* **2004**, *76* (11), 3029–36.

(28) Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (12), 952–8.

(29) Kigawa, T.; Yabuki, T.; Matsuda, N.; Matsuda, T.; Nakajima, R.; Tanaka, A.; Yokoyama, S. Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics* **2004**, *5* (1–2), 63–8.

(30) Davanloo, P.; Rosenberg, A. H.; Dunn, J. J.; Studier, F. W. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81* (7), 2035–9.

(31) Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **1995**, *4* (11), 2411–23.

(32) 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* (3), 663–70.

(33) Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2007**, *2* (8), 1896–906.

(34) Ishihama, Y.; Rappsilber, J.; Andersen, J. S.; Mann, M. Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **2002**, *979* (1–2), 233–9.

(35) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010–21.

(36) Schulze, W. X.; Mann, M. A novel proteomic screen for peptide-protein interactions. *J. Biol. Chem.* **2004**, *279* (11), 10756–64.

(37) Janecki, D. J.; Bemis, K. G.; Tegeler, T. J.; Sanghani, P. C.; Zhai, L.; Hurley, T. D.; Bosron, W. F.; Wang, M. A multiple reaction monitoring method for absolute quantification of the human liver alcohol dehydrogenase ADH1C1 isoenzyme. *Anal. Biochem.* **2007**, *369* (1), 18–26.

(38) de Godoy, L. M.; Olsen, J. V.; de Souza, G. A.; Li, G.; Mortensen, P.; Mann, M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **2006**, *7* (6), R50.

(39) Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. Global analysis of protein expression in yeast. *Nature* **2003**, *425* (6959), 737–41.

(40) Takaki, S.; Watts, J. D.; Forbush, K. A.; Nguyen, N. T.; Hayashi, J.; Alberola-Ila, J.; Aebersold, R.; Perlmutter, R. M. Characterization of Lnk. An adaptor protein expressed in lymphocytes. *J. Biol. Chem.* **1997**, *272* (23), 14562–70.

(41) Kholodenko, B. N.; Demin, O. V.; Moehren, G.; Hoek, J. B. Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.* **1999**, *274* (42), 30169–81.

(42) Cox, J.; Mann, M. Is proteomics the new genomics. *Cell* **2007**, *130* (3), 395–8.

(43) Kuhn, E.; Wu, J.; Karl, J.; Liao, H.; Zolg, W.; Guild, B. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards. *Proteomics* **2004**, *4* (4), 1175–86.

(44) Brun, V.; Dupuis, A.; Adrait, A.; Marcellin, M.; Thomas, D.; Court, M.; Vandenesch, F.; Garin, J. Isotope-labeled protein standards: Towards absolute quantitative proteomics. *Mol Cell Proteomics* **2007**.

(45) Waanders, L. F.; Hanke, S.; Mann, M. Top-Down Quantitation and Characterization of SILAC-Labeled Proteins. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (11), 2058–2064.

(46) Zhou, Y.; Aebersold, R.; Zhang, H. Isolation of N-linked glycopeptides from plasma. *Anal. Chem.* **2007**, *79* (15), 5826–37.

PR7007175

4 MOTIF DECOMPOSITION OF THE PHOSPHOTYROSINE PROTEOME

4.1 *PUBLICATION:* MOTIF DECOMPOSITION OF THE PHOSPHOTYROSINE PROTEOME REVEALS A NEW N-TERMINAL BINDING MOTIF FOR SHIP2

This article presents the results of the project on global analysis of pTyr-motifs and was published in 2008, in the January issue of Molecular and Cellular Proteomics on pages 181-192.

The project was carried out as a collaboration with Dr. Martin Miller and Prof. Dr. Søren Brunak from the center for biological sequence analysis at the Technical University of Denmark.

The following pages contain the published version of the article.

Motif Decomposition of the Phosphotyrosine Proteome Reveals a New N-terminal Binding Motif for SHIP2*□

Martin Lee Miller‡, Stefan Hanke§, Anders Mørkeberg Hinsby‡, Carsten Friis‡, Søren Brunak‡, Matthias Mann§, and Nikolaj Blom‡¶

Advances in mass spectrometry-based proteomics have yielded a substantial mapping of the tyrosine phospho-proteome and thus provided an important step toward a systematic analysis of intracellular signaling networks in higher eukaryotes. In this study we decomposed an uncharacterized proteomics data set of 481 unique phosphotyrosine (Tyr(P)) peptides by sequence similarity to known ligands of the Src homology 2 (SH2) and the phosphotyrosine binding (PTB) domains. From 20 clusters we extracted 16 known and four new interaction motifs. Using quantitative mass spectrometry we pulled down Tyr(P)-specific binding partners for peptides corresponding to the extracted motifs. We confirmed numerous previously known interaction motifs and found 15 new interactions mediated by phosphosites not previously known to bind SH2 or PTB. Remarkably, a novel hydrophobic N-terminal motif ((L/V/I)(L/V/I)pY) was identified and validated as a binding motif for the SH2 domain-containing inositol phosphatase SHIP2. Our decomposition of the *in vivo* Tyr(P) proteome furthermore suggests that two-thirds of the Tyr(P) sites mediate interaction, whereas the remaining third govern processes such as enzyme activation and nucleic acid binding. *Molecular & Cellular Proteomics 7:181–192, 2008.*

Phosphorylation-dependent protein-protein interaction is one of the key organizing principles in intracellular signaling events. The phosphotyrosine binding (PTB)[1] domain and the Src homology 2 (SH2) domain are modular domains that typically bind phosphotyrosine (Tyr(P))-containing peptides (1, 2). "Linear motifs" (unstructured sequence recognition patches with conserved residues at specific positions (3)) that direct Tyr(P)-dependent interaction have traditionally been studied using degenerate oriented peptide libraries. Such studies revealed that PTB and SH2 domains have preference for specific amino acids N- and C-terminal to the Tyr(P) residue, respectively (4, 5).

Recent methodological developments in MS-based proteomics have made it possible to identify hundreds to thousands of protein phosphorylation sites in a single project (6–14). Extensive mapping of the phosphoproteome is an important step toward analyzing the regulatory components of the cell. Because the majority of newly identified phosphopeptides are uncharacterized with respect to signaling context, there is now a unique opportunity to mine the phosphoproteome for novel phosphorylation motifs. Methods have been developed that successfully mine for overrepresented motifs from large protein data sets in general (15–17) and more recently also from phosphoproteomics data sets (18). However, these methods do not partition the data set into smaller subsets with high sequence similarity prior to motif extraction. Sequence patches flanking the motif also govern phosphorylation-dependent recognition (19); consequently there is a risk of extracting false positive motifs from functionally unrelated peptides. Furthermore the above mentioned methods are *in silico* approaches and do not combine prediction with experimental validation.

To overcome such limitations in the area of Tyr(P) motif discovery and classification, one may partition the data set into smaller subsets *e.g.* by sequence similarity with known kinase or binding substrates prior to motif extraction. Thus, the risk of retrieving false positive motifs is minimized because overrepresented motifs are extracted from peptides closely related in sequence and function. Besides Tyr(P) recognition motifs for kinases and interaction domains, there may also potentially exist Tyr(P) motifs that mediate other processes than binding such as *e.g.* enzyme activation and nucleic acid binding. Thus, it is essential to validate the extracted motifs both by experimental and bioinformatical means to obtain a functional classification.

With this in mind we developed a motif extraction and validation methodology and classified Tyr(P) motifs on a pro-

[1] The abbreviations used are: PTB, phosphotyrosine binding; PAM, partitioning around medoids; ITIM, immunoreceptor tyrosine-based inhibition motif; SH2, Src homology 2; GO, Gene Ontology; IPI, International Protein Index; zf-C_2H_2, Cys_2-His_2 zinc finger protein; N-WASP, neural Wiskott-Aldrich syndrome protein; PI3K, phosphatidylinositol 3-kinase; GAP, GTPase-activating protein; STAT, signal transducers and activators of transcription; PIP_3, phosphatidylinositol 3,4,5-trisphosphate.

teome level. Operating in sequence space, we stretched the MS-mapped Tyr(P) peptides over a backbone of ligands already known to be involved in Tyr(P)-dependent interaction. Using experimentally verified Tyr(P) ligands of the PTB and SH2 domains as both a clustering backbone and as a control for the partitioning, we split a literature-extracted data set of mammalian Tyr(P) peptides into 20 different clusters. We obtained a meaningful clustering because the controls partitioned correctly into separate clusters.

From the 20 clusters we extracted both known and unknown phosphorylation motifs, and peptides matching these motifs were synthesized and assayed for phosphorylation-specific interaction partners using a peptide pulldown assay based on quantitative proteomics (20). In contrast to the oriented peptide library approach that uses artificial degenerate peptides, we used naturally occurring peptides as baits to pull down binding partners from the cell lysate. Moreover because the interaction partners are in competition for binding, mimicking the *in vivo* binding situation, the risk of finding kinetically unfavorable interaction motifs is minimized. Finally this technique can potentially identify new types of domains with modification-specific binding capability.

Using the pulldown assay we identified the expected binding partners for numerous known C-terminally directed SH2 domain motifs. We also found 15 new phosphorylation-dependent interactions mediated by phosphosites not previously shown to direct interaction. Surprisingly we identified a new N-terminal hydrophobic motif ((L/V/I)(L/V/I)pY where pY is phosphotyrosine) for the SH2 domain-containing inositol phosphatase SHIP2. The specificity of the motif was confirmed by mutational analysis. Surprisingly this motif is N-terminally directed, which is in contrast to previous observations showing that binding of prototypical SH2 domains are directed by C-terminal recognition (21). Until now the only other known SH2 domain binding motif that is partly directed by N-terminal recognition is the immunoreceptor tyrosine-based inhibition motif (ITIM) (I/L/V)XpYXX(I/L/V) (22, 23).

On a proteome level we analyzed which Gene Ontology (GO) categories were overrepresented in proteins matching the extracted motifs. We found that motifs that mediate interaction in the pulldown assay are typically found in proteins involved in signal transduction, whereas non-binding motifs are found in enzymes and ion- and nucleic acid-binding proteins. Thus, we estimate that one-third of the *in vivo* Tyr(P) sites are not directly involved in interaction via domains such as SH2 and PTB but rather are sites that could alter the catalytic activity of enzymes or modulate the DNA binding affinity of *e.g.* transcription factors.

EXPERIMENTAL PROCEDURES

Data Set Preparation—Large scale data sets of tyrosine phosphorylation sites mapped in MS/MS experiments with mammalian cell lines were collected from the literature (8, 10, 12–14, 24, 25) yielding a total of 847 tyrosine phosphorylation sites. To filter out phosphopeptides from closely related homologs and orthologs only

unique 13-mer peptide sequences with the Tyr(P) centrally positioned were considered. This reduced the MS-based data set to 481 phosphopeptides distributed in 380 proteins. Furthermore 162 experimentally verified Tyr(P) peptide ligands of one PTB domain and 10 different SH2 domains were extracted from the Phospho.ELM database (26). The 162 peptides were included in the data set as positive controls, resulting in a data set of 643 Tyr(P) peptides (see supplemental Table 3). The criteria for selecting the positive controls were the existence of a consensus binding motif and that a suitable amount of examples could be obtained.

Generation of Weight Matrices—13-mers of the 162 phosphopeptide ligands of the 11 respective PTB and SH2 domains (see Table I) were used to create 11 weight matrices using the weight matrix mode of EasyGibbs 1.0 (27). Default settings were used except motif length was set to 13 fixed around the central Tyr(P) residue. Subsequently all phosphopeptides in the MS-based data set (481) and the positive control data set (162) were scored by each of the 11 weight matrices, and thus each phosphopeptide could be represented as a vector of the 11 weight matrix scores.

Clustering Using Partitioning around Medoids (PAM)—A matrix consisting of the 11 weight matrix scores and the 643 phosphopeptides was generated and subsequently clustered by the PAM method (28) using the cluster package in R. The PAM algorithm is a robust version of k-means, and it searches for a specified number of medoids (representatives), k, around which clusters are constructed. The clusters are generated by minimizing the sum of the dissimilarities of all observations and assigning them to their closest medoid. Using a hypergeometrical test (see "Statistics") the optimal number of clusters ($k = 20$) was inferred because this resulted in the best partitioning of the positive controls. We use z-scores, *i.e.* multiples of standard deviations from the mean, to account for the different numeric ranges of the measured parameters.

The choice of an appropriate clustering algorithm is a complex one because no given algorithm is universally superior (29, 30). Rather the best choice will depend on the data set and in particular on what constitutes a good distance measure for it. Another relevant concern is the desired outcome and whether a hierarchical or partitional result is preferable. Many sophisticated methods exist that are capable of automatically determining the number of "natural clusters" in the data like the popular density-based clustering algorithms that can describe very complex non-circular relations in the data (31). It is, however, not clear whether the ability to recognize non-circular structures in the data is beneficial in this case. Proteins that share the same features are likely to be related and will form a circular relation in feature space. On the other hand, an elongated cluster in feature space will contain proteins that share only some features but not others, and the biological implications thereof can be quite diverse. Other than being computationally effective and easy to implement, the PAM algorithm was selected because it satisfies the need for a robust clustering algorithm and because its reliance on an Euclidean distance measure ensures that the result can be easily interpreted. The primary weakness of PAM is the need to arbitrarily select a number of clusters for the data, which in this case is overcome by the mentioned application of the hypergeometric test.

Dendrogram and Sequence Logo Plots—Weight matrices of the peptides in the 20 clusters were made using positional weighting of the three residues flanking the central Tyr(P) residue (27) and used to calculate distance matrices as described previously (32). The distance matrices were used as input to the program neighbor from version 3.5 of PHYLIP (Phylogeny Inference Package). To estimate the significance of the neighbor-joining clustering we used the bootstrap method and estimated the consensus tree by bootstrapping for 1000 repetitions as described earlier (32).

The frequencies of amino acids at particular positions in each

cluster were calculated, and subsequently sequence logo plots were used for graphic visualization (33). Each position in the aligned sequences corresponds to a column in the logo plot. The height of the column represents the degree of conservation at that position, whereas the height of the individual letters is proportional to the relative frequency of this amino acid residue. The maximal height of the column for the 20-amino acid alphabet is $\log_2 20 = 4.32$ bits.

Extraction of Motifs and Selection of Peptides to Synthesize—The identified phosphomotifs in each of the 20 clusters were found using the publicly available TEIRESIAS pattern discovery tool from IBM Bioinformatics (17). Parameters were set so the extracted motifs were within a window of 13 residues centered on the phosphoresidue. The minimal number of literals in the motif was set to 4, and the amino acids were grouped according to their chemical nature (Ala/Gly, Asp/Glu, Phe/Tyr, Lys/Arg, Ile/Leu/Met/Val, Gln/Asn, Ser/Thr, Pro, Trp, His, and Cys (17)). For each of the 20 clusters the most abundant motif was selected, and subsequently one peptide matching the motif was chosen from the respective cluster. Because multiple peptides in each cluster matched the extracted motif, peptides from mouse and peptides not previously known to be involved in phosphorylation-dependent interaction were preferred. In the few cases (three) where mouse sequences could not be obtained, peptides from humans with high homology in mouse were chosen.

Gene Ontology Analysis—Gene Ontology categories were obtained from Gene Ontology Annotation mouse database version 29.0. The extracted motifs were matched to proteins in the International Protein Index (IPI) mouse proteome version 3.20. Using a hypergeometrical test (see "Statistics") with the total proteome as background we found the 10 most overrepresented GO terms in retrieved proteins. The hits were inspected manually, and the consensus GO term was assessed for each motif. For the purpose of the hypergeometrical test, each annotated GO category was taken to include all of its ancestral terms to avoid problems with diverging levels of annotation.

Statistics—To determine whether the positive controls were significantly overrepresented in specific clusters compared with the whole data set, hypergeometric sampling without replacement (34) was performed. The hypergeometric test is a statistical test used to describe the arbitrariness of a sampling without replacement from a background of true or false examples. The probability (p) to observe a given or more extreme situation by a pure coincidence is given by the hypergeometric distribution,

$$P(X = x | N, M, K) \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \quad \text{(Eq. 1)}$$

where N is the total number of peptides, M is the number of peptides in the given set, K is the number of peptides in a particular cluster, and x is the number of K that belongs to M. A Bonferroni correction was performed to correct for multiple comparisons. In the case of GO analysis, we performed the test once for each GO category present in the data and evaluated the probability of sampling the set of retrieved proteins from the background of the total proteome by mere chance, considering a protein 'true' or 'false' depending on whether it had been assigned the category in question. The end result of this test was one p value for each GO category, describing the degree of overrepresentation of that particular assignment in the retrieved set against the background of the entire proteome.

Cell Culture—Mouse C2C12 muscle cells were grown in arginine- and lysine-deficient Dulbecco's modified Eagle's medium with 10% dialyzed fetal bovine serum for at least five passages and then switched to 2% dialyzed fetal bovine serum to differentiate the cells for 8 days. In accordance with the stable isotope labeling by amino acids in cell culture (SILAC) procedure, one cell population was

supplemented with normal isotopic abundance L-arginine (Sigma) and L-lysine, and the other was supplemented with >99% isotopic abundance [$^{13}C_6$, $^{15}N_4$]arginine and [$^{13}C_8$]lysine (Aldrich) as described previously (35). Thereby full labeling of all proteins was achieved.

Peptide Synthesis and Pulldowns—Desthiobiotinylated peptides were synthesized on a solid-phase peptide synthesizer using amide resin (Intavis, Koeln, Germany). All peptides were designed as 15-mers with the Tyr(P) residue placed centrally at position 7 or 8 expect for one peptide from cluster 1 (see Table I) that was 20 amino acids long. The peptides were synthesized with an N-terminal biotin and an SG dipeptide linker. Peptides were synthesized as pairs in phosphorylated and a non-phosphorylated "control" form. The identity and purity of the synthesized peptides was confirmed by mass spectrometric analysis. For pulldowns, 1.5 nmol of immobilized peptide was added to an average of 1.5 mg of cell lysate. Dynabeads MyOne Streptavidin were saturated with biotinylated peptide prior to incubation with cell lysates. Cells were lysed as described previously (36), and equal amounts of protein were incubated overnight with the respective immobilized peptides at 4 °C. After three rounds of washing with lysis buffer, beads of pulldown pairs with the phosphorylated form and control were combined (20), and bound proteins were eluted using 16 mM biotin. Eluted proteins were precipitated and subsequently digested with trypsin for LC-MS/MS analysis.

LC/MS/MS, Database Searching, and Quantitation—After reduction in 1 μg of DTT and alkylation with 5 μg of iodoacetamide the eluted proteins were in-solution digested with 1 μg of endoproteinase Lys-C (Wako) for 3 h at room temperature. Subsequently samples were diluted with 4 volumes of 50 mM NH_4HCO_3 and further digested with 1 μg of trypsin (Promega) overnight at room temperature. Peptide mixtures were desalted on stop and go extraction tips (37) and loaded onto reversed phase analytical columns for liquid chromatography (38). Peptides were eluted from the analytical column by a multistep linear gradient running from 2 to 40% acetonitrile in 100 min and sprayed directly into the orifice of an LTQ-FT or an LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany). Proteins were identified by MS/MS by information-dependent acquisition of fragmentation spectra of multiply charged peptides. The peak list was generated using in-house software, raw2msm version 1.2, with default settings. The identified proteins were then searched against the mouse IPI database using the Mascot (version 2.1.0) algorithm (39). The MS/MS ion search parameters were set as follows: enzyme specificity for trypsin, trypsin/Pro + AspPro; maximum number of missed cleavages, 2; fixed modification, carbamidomethylcysteine; variable modifications, oxidation (Met), N-acetyl (protein), deamidation (NQ), [$^{13}C_6$, $^{15}N_4$]arginine, [$^{13}C_8$]lysine, and pyro (N-terminal QC); mass tolerance for precursor ions, 5 ppm; fragment mass tolerance, 0.5 Da; database version, IPI_mouse mouse_v314 with 68,655 entries. Common contaminants like human keratins were manually added. No species-specific restrictions were used. MSQuant (SourceForge) was used for quantitation and spectra validation. MSQuant uses peak area and extracted ion chromatogram for quantitation.

Determination of Significant Binding Partners—Intensity ratios of labeled to unlabeled forms of each validated tryptic peptide and the associated average ratio for the whole protein were obtained by MSQuant. We used 'crossover' experiments in which the specific interaction partners were required to have inverse ratios compared with the 'normal' experiment (20). A significant binder was defined as a protein with a log ratio at least three standard deviations over the log average ratio ($>3 \log(\sigma) + \log(\mu)$) of all the proteins identified in a pulldown experiment. Furthermore the binder had to be confirmed in at least two pulldown experiments (normal and crossover experiment), and we only report specific binders for the phosphopeptide
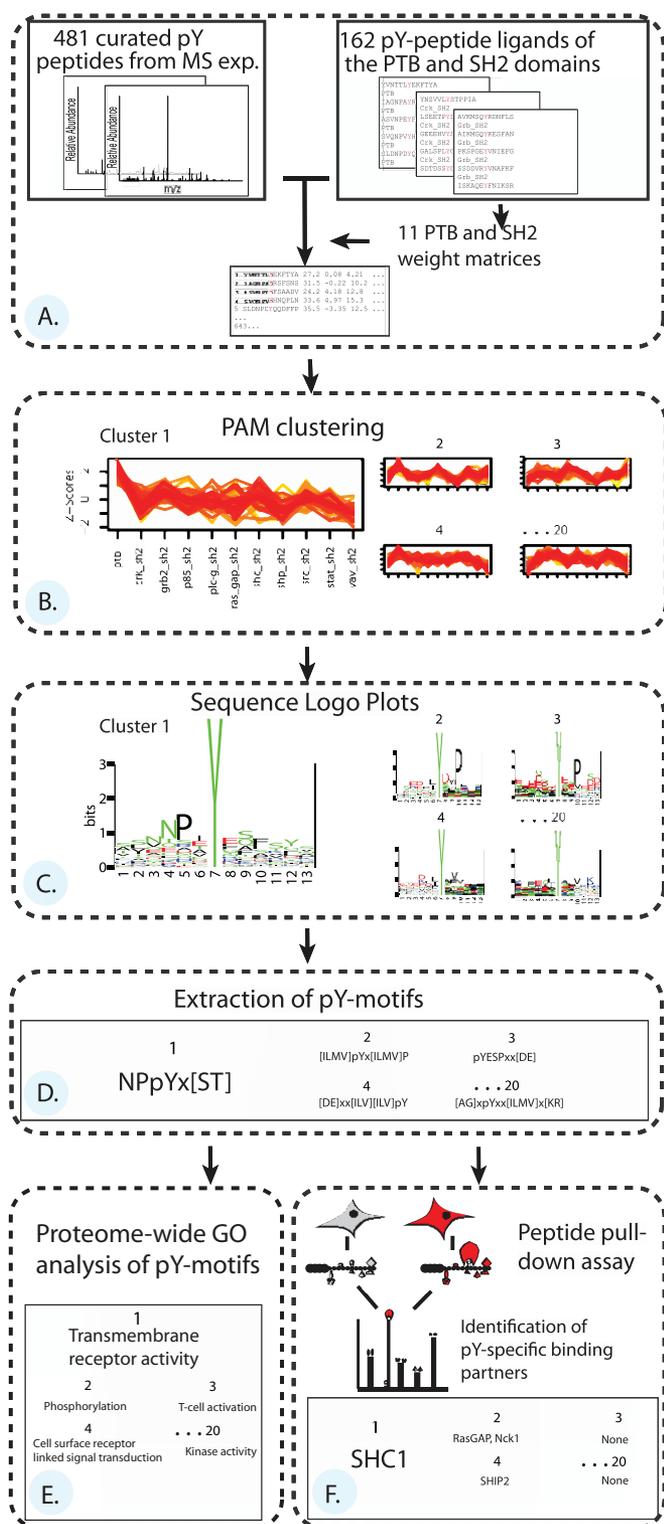
FIG. 1. **Overview of the motif extraction and validation strategy.** *A*, verified Tyr(P) peptide ligands (162) of different PTB and SH2 domains are used to create 11 specific weight matrices that subsequently score both a large scale data set of 481 Tyr(P) peptides identified in MS experiments and the 162 ligands themselves. All peptides are 13-mers with the Tyr(P) (*red*) centrally positioned. Consequently each phosphopeptide can be represented as a vector of

and not the non-phosphorylated peptides. Finally at least one peptide had to have a score above 30, corresponding to $p < 0.05$. In the 64 pulldowns performed we identified a few sequence-unspecific binders with high affinity to either the phosphorylated or non-phosphorylated peptides (staphylococcal nuclease domain-containing protein, eukaryotic translation initiation factor, peptidylprolyl isomerase B, RNA-binding protein SiahBP, and RIKEN cDNA 2410104I19). These proteins were excluded because we consider them as false positive binders, *i.e.* they bind in a sequence-unspecific manner and occasionally bind most strongly to the non-phosphorylated peptide. In all the pulldown experiments an average of 140 ± 41 proteins was quantified with an average ratio of 1.217 ± 0.529.

RESULTS

A New Clustering Approach for Motif Extraction and Classification—To discover new interaction motifs in large aligned peptide data sets, we developed a method that partitions the sequences based on similarity with proteins previously known to be involved in interaction. The method is generally applicable for peptide interaction data sets. In this work we clustered and classified motifs in the Tyr(P) proteome. A flow chart of the method is presented in Fig. 1. In short, we used verified Tyr(P) ligands of the PTB and SH2 domains as a backbone to cluster a data set of uncharacterized phosphopeptides mapped in MS-based proteomics experiments with mammalian cell lines. From each clusters the most conserved motif was extracted. To classify the function of each motifs in an unbiased and systematic manner we conducted both experimental and bioinformatical investigations. Peptides matching the motifs were assayed for binding partners using a peptide-protein interaction screen based on quantitative proteomics (20, 36). On a proteome level we analyzed which GO categories were overrepresented in proteins matching the extracted Tyr(P) motifs. Thus, we established a workflow that extracts, verifies, and classifies motifs in phosphoproteome.

Clustering—In detail, we extracted 162 known 13-mer phosphopeptide ligands of 10 different SH2 domains and a general PTB domain from the Phospho.ELM database (26). In this database, which is the key repository for high quality

the 11 weight matrix scores. *B*, this is used as input for the *k*-means-derived PAM algorithm (28), and the data set is split up into clusters (20, see text for details) depending on sequence similarity of the ligands used in the weight matrix generation. We use z-scores, *i.e.* multiples of standard deviations from the mean, to account for the different numeric ranges of the measured parameters. *C*, sequence logo plots (33) of the peptides in the different clusters are used to visually inspect the result of the partitioning (see Fig. 2). *D*, after partitioning the most conserved motif from each cluster is extracted (17). *E*, a GO analysis of the motifs on a proteome level is conducted to determine in which functional class of proteins the motifs generally are present compared with the whole proteome as background. *F*, a peptide pulldown assay based on quantitative proteomics (20) is used to evaluate the Tyr(P)-specific binding capability of the motifs. Peptides matching the motifs are synthesized in phosphorylated and non-phosphorylated pairs and used to pull down Tyr(P)-specific interaction partners from the lysate of C2C12 muscle cells.

TABLE I

Clustering and motif extraction of the Tyr(P) proteome

The PAM algorithm was used to partition the data set by sequence similarity with the known ligands of the Tyr(P) binding domains (positive controls). The first and second columns show the cluster number and the size of the clusters, respectively. The ability of the algorithm to significantly partition ($p < 0.05$) the positive controls into different clusters is shown in the third column. For example, eight out of a total of 10 ligands of the Crk SH2 domain are grouped in cluster 2, corresponding to a significant overrepresentation in a Bonferonni corrected hypergeometric sampling test ($p < 3.67e-08$). The most conserved motif in each cluster was extracted and is stated in the fourth column. The Tyr(P) residue is indicated in bold, and an "X" represents any amino acid. Also indicated is the number of occurrences of the motif in the respective cluster and in the total data set. The identified motifs were matched to a library of known Tyr(P) binding motifs, and the expected binding partner is indicated (fifth column). 15–20-mer peptides matching the motifs were synthesized in pairs, one with a Tyr(P) as indicated in bold. Furthermore the parent protein of the peptide is given by Swiss-Prot entry name, and the position of the Tyr(P) is stated. The Tyr(P)-dependent interaction partner(s) identified in a quantitative proteomics peptide-protein screen is shown in the last column (see also Supplemental Table 1). PLC, phospholipase C.

Cluster	Size	Positive controls	Extracted motifs	Matched motifs, expected partner	PubMed ID	Peptides synthesized	Identified partners
1	41	PTB 10 of 10	NP*X***pY***X*(S/T) 7 of 7	SHC PTB (NP*X***pY**)	7542744 7541030	KEVCDGWSLPNPE**pY**YTLRYA ELMO2_MOUSE, 48	SHC
2	36	Crk SH2 8 of 10	(I/L/M/V)**pY***X*(I/L/M/V)P 8 of 14	Crk/RasGAP SH2 (**pY***XX*P)	9233798 11607838	KPSTDPL**pY**DTPDTRG RIN1_MOUSE, 35	RasGAP Nck1
3	29	Vav SH2 4 of 4	**pY**ESP*XX*(D/E) 5 of 5	Vav SH2 (**pY**ESP)	9151714	TETKTIT**pY**ESPQIDG E41L2_MOUSE, 889	None
4	28		(D/E)*XXX*(I/L/V)(I/L/V)**pY** 4 of 6	New motif		RETSKVl**pY**DFIEKTG WASL_MOUSE, 253	SHIP2
5	20		(D/E)(D/E)*XXX***pY***X*N 4 of 6	Grb2 SH2 (**pY***X*N)	11994738	VYDEDSP**pY**QNIKILH SPSY_MOUSE, 147	Grb2 RasGAP
6	41	Grb2 SH2 14 of 31	**pY***X*N(I/L/M/V)*XX*L 5 of 7	Grb2 SH2 (**pY***X*N)	11994738	ELFDDPS**pY**VNIQNLD SHC1_MOUSE, 423	Grb2
7	21	Grb2 SH2 6 of 31	(D/E)**pY***X*N(I/L/M/V) 4 of 11	Grb2 SH2 (**pY***X*N)	11994738	QPASVTD**pY**QNVSFSN ITSN2_HUMAN, 858	Grb2
8	45		**pY**(I/L/M/V)*X*M*X*P 4 of 10	p85-PI3K SH2 (**pY***XX*M)	7511210 11994738	PQRVDPNG**pY**MMMSPS IRS1_MOUSE, 658	p85-PI3Kα p85-PI3Kβ
9	40		**pY**(D/E)*X*(I/L/M/V)*X*(I/L/M/V) 5 of 22	Fps/Fes SH2 (**pY**(E/D)*X*(I/V))	7511210	AGKQKLQ**pY**EGIFIKD SF3A1_MOUSE, 757	None
10	32		(D/E)*XX***pY**(D/E)*X*(I/L/M/V) 7 of 27	Fps/Fes SH2 (**pY**(E/D)*X*(I/V))	7511210	DGGSDQN**pY**DIVTIGA INP4A_HUMAN, 355	None
11	35	PI3K SH2 16 of 24	**pY**(I/L/M/V)PM*X*P 6 of 7	p85 PI3K SH2 (**pY***XX*M)	7511210 11994738	NLHTDDG**pY**MPMSPGV IRS1_MOUSE, 608	p85-PI3Kα
12	23		D**pY**(I/L/M/V)*X*(I/L/M/V) 7 of 18	SHP2 SH2 (**pY**(I/V/L)*X*(I/V/L))	7680959	DLINRMD**pY**VEINIDH VIGLN_MOUSE, 437	SHP2
13	28	SHP2 SH2 7 of 12	(I/L/M/V)*X***pY**(I/L/M/V)*X*(I/L/M/V)D 6 of 7	SHP2 SH2 (**pY**(I/V/L)*X*(I/V/L))	7680959	DIKEKLC**pY**VALDFEQ ACTB_MOUSE, 218	SHP2
14	32	PLCγ SH2 5 of 16	(I/L/M/V)**pY***XX*(I/L/M/V)(I/L/M/V) 5 of 11	General/SHP2 SH2 (**pY**(I/V/L)*X*(I/V/L))	7511210 7680959	GKSKQPL**pY**SSIVTVE O88185_MOUSE, 948	SHP2 SHIP2
15	29	RasGAP SH2 7 of 8	(A/G)(I/L/M/V)**pY***XX*P 6 of 10	Crk/RasGAP SH2 (**pY***XX*P)	9233798 11607838	GVVDSGV**pY**AVPPPAE BCAR1_HUMAN, 410	RasGAP
16	37	SHC SH2 9 of 13	P*X*E**pY***XXXXX*(I/L/M/V) 3 of 3	New motif		TTEAPGEYFFSDGVR IMDH1_MOUSE, 400	None
17	25	Src SH2 7 of 14	**pY**(D/E)*X*(I/L/M/V)H 4 of 6	Fps/Fes SH2 (**pY**(E/D)*X*(I/V))	7511210	ELTAEFL**pY**DEVHPKQ TWF2_MOUSE, 309	RasGAP
18	32	STAT SH2 15 of 19	**pY**(I/L/M/V)PQ 4 of 4	STAT SH2 (**pY***XX*Q)	14966128	SGENFVP**pY**MPQFQTC LEPR_MOUSE, 1138	None
19	33		H(S/T)G*X*KP**pY***X*C*XX*CG 10 of 10	New motif		RIHTGEKP**pY**ECVQCGK ZNF24_MOUSE, 335	None
20	36		(A/G)*X***pY***XX*(I/L/M/V)*X*(K/R) 8 of 15	New motif		KKNRIAl**pY**ELLFKEG RS10_MOUSE, 12	None

annotated phosphorylation sites, these 11 domains had the highest number of annotated ligands (see Table I for details). Thus, the 162 substrates of the 11 domains represent the broadest available *in vivo* data set of Tyr(P) ligands. These ligands were aligned with the Tyr(P) centrally positioned and used to generate a position-specific weight matrix for each of the 11 domains. The 481 phosphopeptides in the MS-based data set and the 162 ligands of the PTB and SH2 domains ("positive controls") were then scored by each of the 11 weight matrices. Consequently each phosphopeptide was

represented as a vector of the 11 weight matrix scores (Fig. 1*A*). Based on this profile of vectors we clustered the total of 643 peptides using the *k*-means-derived PAM algorithm (40). The PAM algorithm searches for a predefined number of medoids (representatives) around which clusters are constructed. We tried various cluster sizes, and by using a hypergeometrical distribution test the number of clusters was set to 20 because this gave the statistically best partition of the positive controls into different clusters (Fig. 1, *B* and *C*).

We were able to obtain a convenient fit of the model because the positive controls were distributed such that all were statistically overrepresented in separate clusters ($p < 0.05$) (see Table I). For example, 10 of a total of 10 ligands of the PTB domains were grouped in cluster 1 ($p < 7.34e-12$). Only the 31 ligands for the Grb2 SH2 domain were split up in two significant groups (clusters 6 and 7). Furthermore eight of the 20 clusters did not contain a significant overrepresentation of the known ligands used.

Motif Extraction—From each of the 20 clusters the most conserved motif was extracted with the TEIRESIAS pattern discovery tool from IBM Bioinformatics (Fig. 1*D*) (17) as described under 'Experimental Procedures.' The 20 identified motifs are presented in Table I followed by the number of matches to peptides in the particular cluster and in the total data set. For example, a unique motif, NP*X*pY*X*(S/T), was extracted from cluster 1 because all peptides (seven) that matched this motif were in this cluster. To compare the 20 identified motifs with already characterized interaction motifs, we matched each motif to a comprehensive library of Tyr(P) interaction motifs in the Human Protein Reference Database (HPRD) (41). Using the same example again, the motif extracted from cluster 1 could be matched to the NP*X*pY motif described for the SHC and IRS-1 PTB domains (1), although our extended motif contains a Ser or Thr residue at position +2 from the Tyr(P) residue. Considering that the ligands (the positive controls) of the PTB domain were grouped in cluster 1, it is not surprising that we extracted the NP*X*pY*X*(S/T) motif in this cluster; however, only three of the seven matching peptides in cluster 1 were from these PTB ligands (data not shown). Because this was a general trend it shows the ability of the clustering method to gather previously uncharacterized peptides with high sequence similarity to known ligands of the different Tyr(P) binding domains. In total, 16 of the 20 identified motifs could be matched to the motif library, showing an overall consistency between the positive controls and the extracted and matched motifs. In four clusters new Tyr(P) motifs were identified.

Gene Ontology Analysis—We used a GO analysis of the 20 extracted motifs on a proteome level to determine in which type of proteins the extracted motifs generally are present (Fig. 1*E*). We retrieved all proteins in the mouse proteome that matched the motifs and used a hypergeometrical test with the total proteome as background and thereby found the 10 most overrepresented GO terms in retrieved proteins (see "Experimental Procedures" for details). Using the same example again, the NP*XYX*(S/T) motif from cluster 1 was overrepresented in proteins involved in processes like 'receptor activity' and 'intrinsic to membrane' with the consensus parent GO term assessed to be 'transmembrane receptor activity' (see supplemental Table 2). Thus, analyzing the motif on a proteome level indicates that proteins containing the motif are involved in early signaling transduction. This makes sense in that this is a motif for the PTB domain, which is found in proteins that function as molecular scaffolds and adaptors in signaling pathways (1).

Tyr(P)-specific Interaction Partners—To experimentally verify the 20 extracted motifs we used a phosphorylation-specific peptide-protein interaction screen (Fig. 1*F*) (20, 36). This assay is based on differential labeling of proteins using stable isotope labeling by amino acids in cell culture (SILAC) making it possible to distinguish specific binders from background binders by their isotope ratios determined by quantitative mass spectrometry (35, 42). The peptides are synthesized in a phosphorylated form and a non-phosphorylated form and used as baits to pull down competing binding partners from cell lysate, thus mimicking the *in vivo* binding situation.

We synthesized peptide pairs matching the 20 extracted motifs. If there were multiple matches we chose peptides with Tyr(P) sites not previously known to mediate interaction. Using this experimental approach we could test our clustering and motif extraction method and investigate the relevance of known motifs in a near *in vivo* situation, *i.e.* endogenous proteins competing for binding, and potentially discover binding partners for novel motifs. Again using cluster 1 as an example, we synthesized a peptide pair from the engulfment and cell motility protein 2 in which the Tyr-48 residue was either phosphorylated or non-phosphorylated. This is an uncharacterized phosphosite identified in a large scale phosphoproteomics study (12), and it has not previously been shown to direct Tyr(P)-dependent interaction. Using this peptide pair as bait we retrieved one specific binder with a ratio more than three standard deviations over the log mean of a total of 162 background binders. This protein, SHC-transforming protein 1 (SHC), which contains both a PTB and an SH2 domain, had a total Mascot score of 1654 with 15 identified peptides of which nine were quantifiable (see supplemental Table 1). Theoretically it could be either the SH2 or the PTB domain that binds the bait phosphopeptide; however, because the peptide matches the consensus NP*X*pY motif known to direct PTB domain binding, it is most likely that SHC binds to the peptide through its PTB domain.

Assaying the Tyr(P) Sequence Space for Interaction Partners—The identified phosphospecific binding partners of the representative peptides from each cluster can be seen in Table I. Besides the aforementioned SHC protein, these proteins all contain SH2 domains, making it very likely that this domain governed the phosphospecific interaction. The majority (13 of 20) of the peptides retrieved one or more interaction

FIG. 2. **Dendrogram representing the *in vivo* Tyr(P) sequence space, motifs, and binding partners.** Peptides in the 20 different clusters are used to generate weight matrices that subsequently are used as input in a phylogenetic alignment. The tree is a consensus tree of 1000 bootstrap trees (32). The tree represents the distance between the clusters in sequence space, which is also visually illustrated by the sequence logo plots (33) of the peptides in each cluster. The *color* of the *branches* is based on the Tyr(P)-dependent interaction partners that were experimentally identified using a peptide-protein interaction screen. Both novel and previously known consensus motifs (see Table I, fifth column) that govern these specific interactions are indicated in the same color. Branches and extracted motifs are *gray* if no interaction partners were retrieved via the motifs. Furthermore overrepresented GO terms from proteins in the whole mouse proteome containing the motifs are stated (see text for details). Note that the motifs that do not retrieve a specific binding partner (*gray*) are typically found in proteins mediating processes such as ion binding, 'catalytic activity,' and nucleic acid binding.
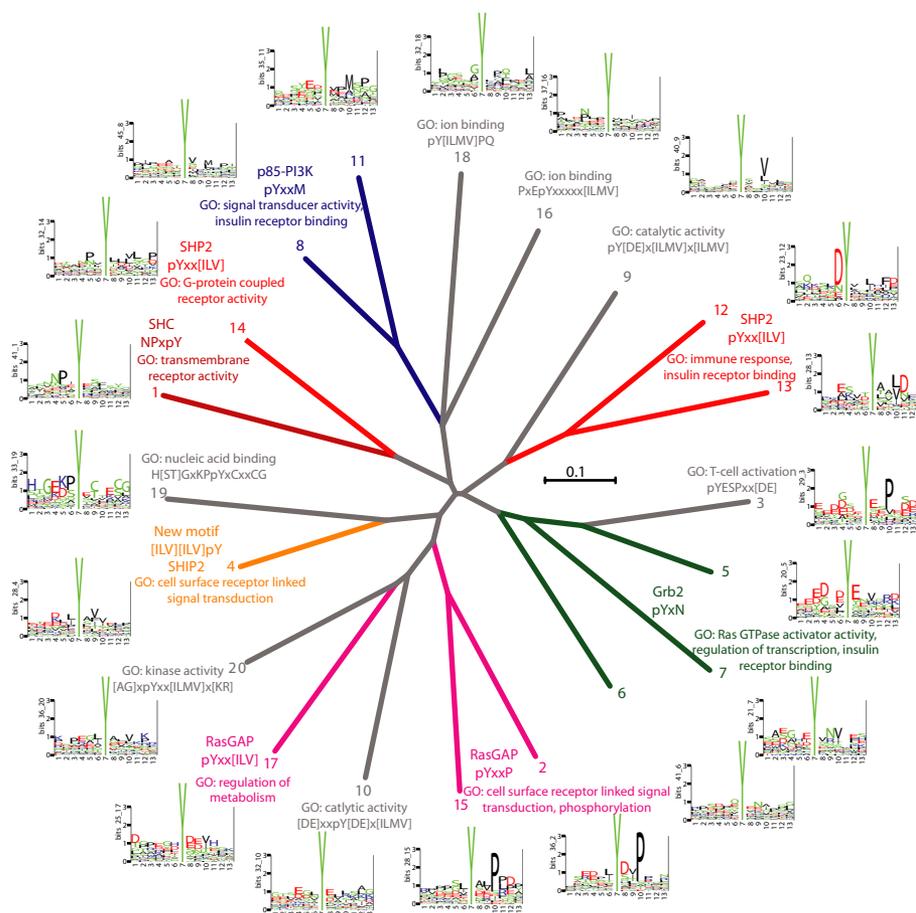
proteins. Of these proteins seven were unique because some proteins were identified several times. This is not surprising because some of the clusters were close to each other in sequence space resulting in extraction of similar motifs and ultimately retrieving the same interaction partners.

To get an overview of the results of the pulldown experiments, the GO analysis, and the sequence similarity between the different clusters, we generated weight matrices of the peptides in the individual clusters and constructed a dendrogram based on an alignment of these matrices (32). The tree can be seen in Fig. 2 together with sequence logo plots where the height of each position represents the degree of conservation (33). The logo plots visually illustrate a successful partitioning because each cluster has a distinct pattern where particular amino acids are highly abundant at specific positions flanking the central Tyr(P) residue. The tree is colored according to the identified interaction partners retrieved by peptides matching the motifs in the different clusters. For example, clusters 8 and 11 are close in sequence space with an overrepresentation of hydrophobic residues, especially methionine, at position +3 from the central tyrosine residue. Rather than being distinct clusters, these are more likely to be subsets of the same motif. Thus, from these two clusters the

motifs pY(I/L/M/V)XMXP and pY(I/L/M/V)PMXP were extracted and matched to the consensus pYXXM in the library of motifs. Accordingly the two peptide pairs synthesized from these clusters both retrieve the PI3K-p85α protein. In the same manner the majority of the expected interaction partners were identified using peptides matching to the well characterized C-terminally directed Tyr(P) motifs, such as pYXN, pYXXP, and pY(I/V/L)X(I/V/L), that retrieved Grb2, RasGAP, and SHP2, respectively. This illustrates the clear consistency between the sequence similarity of the clusters, the conserved residues in the motifs, and the interaction partners identified.

There were four clusters (clusters 3, 9, 10, and 18) where we did not identify the partners (Vav, Fps/Fes, and STAT) as expected from the signature of motifs alone (see Table I). For instance, we did not pull down the SH2 domain protein Fps/Fes when using peptides matching a pY(E/D)X(I/V/L) motif (clusters 9 and 10), which has previously been shown to direct binding (4). This motif was defined using *in vitro* oriented peptide library experiments, which have the inherent risk of defining motifs that are not relevant *in vivo*. Whether this is the case, the Fps/Fes protein was not present in the cell lysate, or because of technical limitations remains unclear. In total, here

we report 15 phosphorylation-dependent interactions mediated by phosphosites not previously known to direct protein interaction (see supplemental Table 1).

Non-binding Tyr(P) Motifs—We observed that motifs that do not mediate interaction in the pulldown assay are typically found in proteins involved in processes other than signal transduction (see Fig. 2). It is particularly interesting that we extracted a highly conserved H(S/T)GXKPp$Y$$XCXXX$CG motif from a number of closely related peptides from Cys$_2$-His$_2$ zinc finger proteins (zf-C$_2$H$_2$) concentrated in cluster 19 that did not pull down any phosphorylation-specific interaction partner. The phosphosite of the first tyrosine residue in the zf-C$_2$H$_2$ domain was also described recently in a study that mined for novel motifs in the phosphoproteome (18), although the identified motif in this work (EXXpY) was different from our top scoring motif (H(S/T)GXKPYXCXXXCG), which does not contain an acidic residue in position -3. However, our second best motif in cluster 19 (HXGEXXpY) closely resembles that reported by Schwartz and Gygi (18).

The H(S/T)GXKPYXCXXXCG motif is extremely specific for proteins containing the zf-C$_2$H$_2$ domain: of 33,758 proteins we retrieved 656 matches, all of which had the GO term 'nucleic acid binding' (GO:0003676) (see supplemental Table 2), whereas 647 of the 656 matches had the term 'zinc ion binding' (GO:0008270) ($p < 1e-100$).

Recently a role for phosphorylation of zf-C$_2$H$_2$ domains in inhibition of transcription has been suggested (43, 44), supposedly as a consequence of the negatively charged phosphomoiety that reduces DNA affinity (45). Indirectly our results support this; because we did not retrieve any interaction partner for the synthesized phosphopeptide matching the zinc finger motif, it is unlikely that this motif directs protein-protein interaction, but rather phosphorylation of this motif modulates nucleic acid binding.

Similarly the novel motifs from clusters 16 and 20 could mediate mechanisms other than protein-protein interactions, for example, a kinase motif that mediates enzyme activation, nucleic acid binding, protein folding, etc. Supporting this, we found these motifs to be present in proteins overrepresented in proteins with GO terms 'ion binding' and 'kinase activity,' respectively. Likewise the motifs from clusters where we did not find the expected partners (clusters 3, 9, 10 and 18) are all except motif 3 overrepresented in proteins involved in enzymatic processes and ion binding (see Fig. 2 or supplemental Table 2).

This indicates that the motifs not mediating protein binding could govern processes such as phosphorylation-dependent enzyme activation and nucleic acid binding. Taken together with the results from the pulldown experiment where 13 of 20 motifs mediated interaction, we estimate that one-third of the Tyr(P) motifs in the proteome mediate processes other than interaction through prototypic SH2 and PTB domains.

Identification of a New N-terminal Hydrophobic Tyr(P) Binding Motif for SHIP2—From cluster 4 we extracted a new N-terminal hydrophobic motif, (D/E)XXX(I/L/V)(I/L/V)pY. We synthesized a peptide pair from the neural Wiskott-Aldrich syndrome protein (N-WASP) matching the motif where Tyr-253 was either phosphorylated or non-phosphorylated. This phospho-Tyr-253 was identified in a large scale phosphoproteomics experiment (12). It has previously been shown that phosphorylation of Tyr-253 modulates localization of N-WASP from nucleus to cytoplasm, thereby possibly stimulating cell migration (46). Using the N-WASP peptide pair in our quantitative proteomics experiment, we found the SH2 domain-bearing inositol 5-phosphatase 2 (SHIP2) as a significant binding partner with 12 peptides (ratio, 9.5 $\pm$ 1.5) of 131 protein background partners (ratio, 1.2 $\pm$ 0.9) (see supplemental Table 1). To address the specificity of this N-terminal hydrophobic phosphomotif we repeated the experiment with the only difference being that the two hydrophobic residues were mutated to alanines (AApY) and paired this peptide with the wild type N-WASP phosphopeptide (VIpY) (Fig. 3). The SHIP2 protein bound specifically to the wild type phosphopeptide (ratio, 10.2 $\pm$ 3) but not the mutated phosphopeptide, confirming the specificity of the hydrophobic motif. In a third interaction experiment we obtained similar results by using a phosphopeptide in which all three residues in the extracted motif were mutated (see supplemental Table 1). This supports the notion that recognition is based on two hydrophobic amino acids adjacent to the Tyr(P), and we could confine the motif to (I/L/V)(I/L/V)pY.
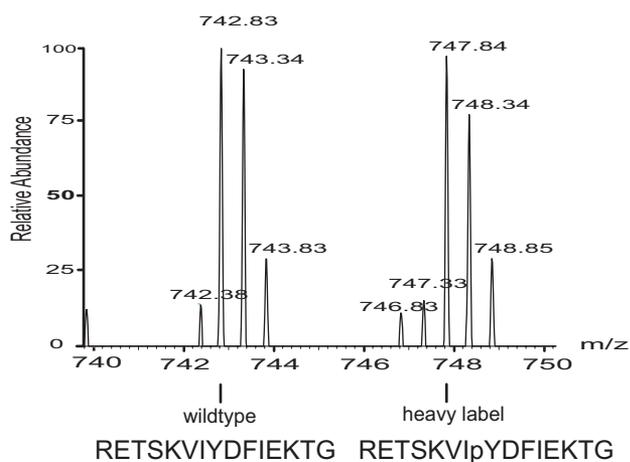
We next extracted three additional peptides from the same cluster matching the (I/L/V)(I/L/V)pY motif. Two of three of these phosphopeptide pairs also specifically pulled down SHIP2 (see supplemental Table 1). Together these experiments show that this indeed is a generic binding motif for SHIP2.

Similarly to previous studies using this phosphospecific pulldown method (20, 36) in this work we only retrieved specific interaction partners containing either an SH2 or a PTB domain. Because SHIP2 contains an SH2 domain it is highly unlikely that interaction between SHIP2 and the phosphopeptides could be mediated by domains other than the SH2 domain.
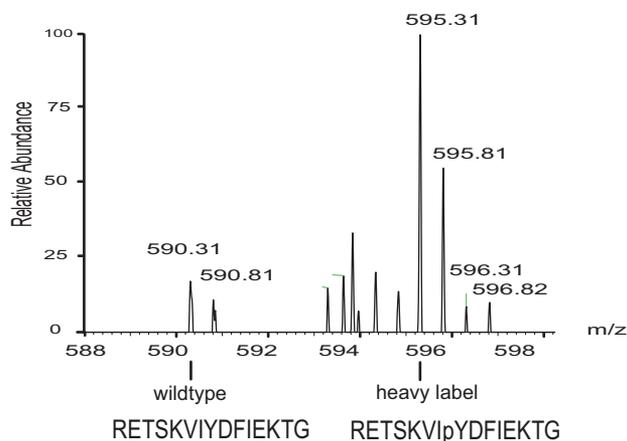
Because the observed interaction seems to be one of the first examples where N-terminal residues of the ligands guide SH2 domain binding, we wanted to exclude the possibility that the peptide was bound in a reverse fashion. We synthesized the N-WASP peptide pair with the reverse sequence (GTKEIFDpYIVKSTER) and found that SHIP2 was not retrieved using this scrambled peptide pair as bait (see supplemental Table 1). Thus, the assay has directional specificity, and only the two hydrophobic residues N-terminal and not C-terminal from the Tyr(P) direct SH2 domain-mediated SHIP2 binding.

To investigate the significance of the motif on a proteome level, we used the aforementioned GO analysis and found that proteins containing the motif indeed are involved in signal transduction. The (I/L/V)(I/L/V)Y is not particularly specific in

92

A   DQGGYEDFVEGLR from Myosin Light Chain 1

wildtype   heavy label
RETSKVIYDFIEKTG   RETSKVIpYDFIEKTG

B   GSYGLDLEAVR from SHIP2

wildtype   heavy label
RETSKVIYDFIEKTG   RETSKVIpYDFIEKTG

C   GSYGLDLEAVR from SHIP2

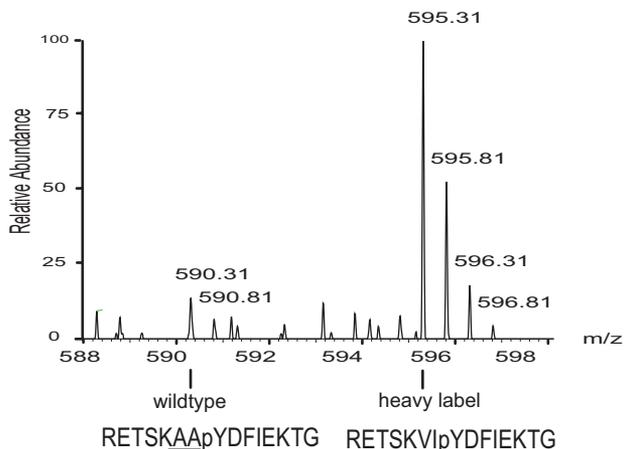

wildtype   heavy label
RETSKAApYDFIEKTG   RETSKVIpYDFIEKTG

FIG. 3. **Identification of a new hydrophobic N-terminal motif for SHIP2.** *A*, mass spectrum of a doubly charged $[^{13}C_6,^{15}N_4]$Arg-labeled tryptic peptide from myosin light chain 1 pulled down with the RETSKVIYDFIEKTG peptide from N-WASP. The myosin light chain 1

itself because it matches 3618 times, which is about 10% of all proteins in the mouse proteome. However, of the 3618 proteins, 2495 can be backtracked to the term 'cell surface receptor-linked signal transduction' (GO:0007166) ($p <$ $1e-100$). In summary, the N-terminal hydrophobic motif we characterized mediates SHIP2 interaction and is generally overrepresented in proteins involved in signal transduction.

DISCUSSION

Currently there is a unique opportunity to mine and classify Tyr(P) motifs in the proteome. 1) A significant portion of the Tyr(P) sites have been mapped by MS-based proteomics experiments. 2) Although not exhaustively, classical biochemical studies have pinpointed interaction between SH2 and PTB domain-containing proteins and specific Tyr(P) ligands. 3) Advances of high throughput validation methods, like the peptide pulldown assay used in this study, make it possible to validate dozens of interaction motifs. Together these developments constitute the basis for discovering and validating Tyr(P) motifs on a global level.

Methodological Considerations—Compared with previously published methods that mine for motifs in large scale phosphoproteomics data sets we partitioned by similarity with functionally characterized peptides prior to motif extraction. By this clustering approach we obtained high resolution and can extract meaningful patterns from functionally and physically related groups of peptides. We used the binding ligands of PTB and SH2 domains both as a clustering backbone and as positive controls in the clustering and could consequently obtain a satisfactory fit of the model. Ideally as more interaction data become available one would use independent controls; however, this was not possible due to data limitations. For all the 20 extracted motifs we could either detect a match to an existing Tyr(P) motif, retrieve a binding partner, or obtain a meaningful GO term for all proteins containing the motif, and thus, we estimate that we have an insignificantly low false positive motif extraction rate in our method.

In the few cases where we did not retrieve the expected binding partner from the Tyr(P) motif alone, *i.e.* Fps/Fes, STAT, and Vav, it can be speculated that 1) the motifs, originally defined *in vitro*, could be low affinity motifs *in vivo*, 2) the proteins were not expressed in the C2C12 muscle cell line used in this study, or 3) it could be due to a technical limitation in our assay.

protein is an unspecific binding protein because the ratio of the ion intensities of the two differentially labeled peptides is 1:1. *B*, mass spectrum of a doubly charged $[^{13}C_6,^{15}N_4]$Arg-labeled tryptic peptide from SHIP2. SHIP2 specifically binds to the phosphorylated peptide from N-WASP as indicated by the larger peak intensity of the labeled peptides. *C*, when mutating the VIpY motif to AApY and comparing these two 15-mer phosphopeptides directly in a pulldown assay, SHIP2 is still retrieved as a specific binding partner to the VIpY motif, confirming the specificity of the hydrophobic motif immediately N-terminal to the Tyr(P).

Implications for a New Motif for SHIP2—One of our findings was that the SH2 domain-containing inositol 5-phosphatase SHIP2 binds to a novel N-terminal hydrophobic motif, (I/L/V)(I/L/V)pY. The specificity of this motif was confirmed by mutational analysis. A scrambled peptide pair with the reverse sequence and thus with the motif C-terminal of the Tyr(P) did not retrieve SHIP2, confirming that the recognition lies on the N-terminal side of the Tyr(P). SHIP2 is also retrieved by two other peptide pairs matching the motifs, indicating this is a generic motif for SHIP2 binding.

In general, the C-terminal residues Tyr(P) +1 and +3 are considered as the most important for the binding specificity of prototypic SH2 domains (1, 2). Interestingly the SH2 binding motif of SHIP2 that we describe is N-terminal, indicating that the peptide binding groove of some SH2 domains also accommodates residues N-terminal of the Tyr(P). In agreement with this, the binding properties of the tandem SH2 domains of the protein-tyrosine phosphatase SHP-2 are governed by residues Tyr(P) −2 to +5 (47). SHP-1 and SHP-2 both bind the (I/L/V)XpYXX(I/L/V) ITIM motif in the cytoplasmic part of Fc receptors, and the Tyr(P) −2 hydrophobic residues have specifically been shown to mediate binding (23, 48, 49). In contrast to the prototypic SH2 domains that have an Arg in the αA2 binding pocket groove, the tandem SH2 domains of the SHP-1 and SHP-2 phosphatases instead have Gly (50). Supposedly this creates a gap that is filled by the side chain of the Tyr(P) −2 residue of the bound peptide. Supporting this hypothesis, it has been shown that a single point mutation in αA2 Gly → Arg disrupts the Tyr(P) −2-mediated binding specificity of SHP-2 (48). Furthermore it is known that signaling lymphocytic activation molecule-associated protein (SAP) and Eat2 SH2 domains in part are directed by N-terminal binding to ITIM motifs (51).

The binding motif of the SH2 domain of SHIP2 has not previously been investigated using degenerate peptide binding experiments; however, the ITIM motif has also been reported as a docking point for the SH2 domain of SHIP2 (52, 53). Combined with our observations, this indicates that the binding specificity of SHIP2 may be conferred by hydrophobic residues immediately upstream of the Tyr(P) with contributions from downstream hydrophobic patches. To our knowledge, other than that of SHP, this is the only other reported N-terminally directed Tyr(P) binding motif.

It could be speculated that the SHIP inositol phosphatases could have a binding mechanism similar to that of SHP protein-tyrosine phosphatases. However, in contrast to SHP the SH2 domains of the SHIP phosphatases resemble the prototypic SH2 domain because they have the highly conserved αA2 Arg (50), indicating that the N-terminally directed binding mechanisms differ between SHIP and SHP phosphatases. Because the crystal structure of the SHIP2 phosphatase with a bound ligand has not been resolved, the specific binding mechanisms have yet to be described.

SHIP2 is involved in membrane signaling by dephospho-rylating the 5′-phosphate group of the key secondary messenger phosphatidylinositol 3,4,5-trisphosphate (PIP$_3$). Through this action SHIP2 inhibits PI3K-mediated receptor tyrosine kinase signaling because PIP$_3$ is generated by PI3K (54). The new motif we describe for the SH2 domain of SHIP2 fits into this overall function because a GO term analysis of all proteins containing the (I/L/V)(I/L/V)pY motif revealed that these proteins are involved in cell surface receptor-linked signal transduction. Presumably SHIP2 could bind to a number of yet unidentified membrane signaling proteins through its SH2 domain and in this way be translocated to the membrane or act cooperatively with these proteins. Negative regulators of PI3K and PIP$_3$ are attractive as antiobesity and diabetes drug targets because PI3K is the main effector of insulin signaling. Recently a role for SHIP2 as a candidate for such therapeutic intervention has been proposed by studies of SHIP2 knock-out mice (55, 56). Thus, the new motif described in this report may not only be involved in regulation of SHIP2-mediated signal transduction but could also be relevant for medical use.

Conclusions—This work presents the first system-wide approach to mine the proteome for Tyr(P) interaction motifs using both bioinformatics methods and experimental validation. Strikingly 16 of the 20 motifs extracted could be matched to previously described interaction motifs. Our experimental validation shows that the majority of the Tyr(P) interaction motifs that previously have been defined *in vitro* are also able to pull down interaction partners from complex lysate mixtures. The GO analysis revealed that motifs that mediate interaction in the pulldown assay are found in proteins involved in signal transduction, whereas remaining non-binding motifs are found in enzymes and ion- and nucleic acid-binding proteins. This raises the intriguing possibility that about one-third of the *in vivo* Tyr(P) sites are not directly involved in interaction via domains such as SH2 and PTB but rather are sites that could alter the catalytic activity of enzymes or modulate the DNA binding affinity of *e.g.* transcription factors.

Perspectives—The developed clustering method is applicable to other types of complex large scale data sets of post-translational modification where a substantial amount of peptide-protein interactions have been identified. As MS-based methods map more modifications such as acetylation and methylation sites, such interactomes could be mined for conserved motifs and assayed for binding partners in a similar manner. Combining proteomics and bioinformatics enables one to do large scale screens in an unbiased way and thus allows one to reconfirm previous knowledge and discover new mechanisms at the same time. As mass spectrometric streamlining and automation advances, entire post-translational modification proteomes can be mapped, binding motifs can identified, and thus ultimately the specificity of signaling networks can be unraveled.

94

REFERENCES

1. Yaffe, M. B. (2002) Phosphotyrosine-binding domains in signal transduction. *Nat. Rev. Mol. Cell Biol.* **3,** 177–186
2. Pawson, T. (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* **116,** 191–203
3. Bork, P., and Koonin, E. V. (1996) Protein sequence motifs. *Curr. Opin. Struct. Biol.* **6,** 366–376
4. Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., and Yi, T. (1994) Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol. Cell. Biol.* **14,** 2777–2785
5. Songyang, Z., Margolis, B., Chaudhuri, M., Shoelson, S. E., and Cantley, L. C. (1995) The phosphotyrosine interaction domain of SHC recognizes tyrosine-phosphorylated NPXY motif. *J. Biol. Chem.* **270,** 14863–14866
6. Ballif, B. A., Villen, J., Beausoleil, S. A., Schwartz, D., and Gygi, S. P. (2004) Phosphoproteomic analysis of the developing mouse brain. *Mol. Cell. Proteomics* **3,** 1093–1101
7. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12130–12135
8. Brill, L. M., Salomon, A. R., Ficarro, S. B., Mukherji, M., Stettler-Gill, M., and Peters, E. C. (2004) Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal. Chem.* **76,** 2763–2772
9. Ficarro, S., Chertihin, O., Westbrook, V. A., White, F., Jayes, F., Kalab, P., Marto, J. A., Shabanowitz, J., Herr, J. C., Hunt, D. F., and Visconti, P. E. (2003) Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation. *J. Biol. Chem.* **278,** 11579–11589
10. Ficarro, S. B., Salomon, A. R., Brill, L. M., Mason, D. E., Stettler-Gill, M., Brock, A., and Peters, E. C. (2005) Automated immobilized metal affinity chromatography/nano-liquid chromatography/electrospray ionization mass spectrometry platform for profiling protein phosphorylation sites. *Rapid Commun. Mass Spectrom.* **19,** 57–71
11. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127,** 635–648
12. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* **23,** 94–101
13. Salomon, A. R., Ficarro, S. B., Brill, L. M., Brinker, A., Phung, Q. T., Ericson, C., Sauer, K., Brock, A., Horn, D. M., Schultz, P. G., and Peters, E. C. (2003) Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 443–448
14. Zheng, H., Hu, P., Quinn, D. F., and Wang, Y. K. (2005) Phosphotyrosine proteomic study of interferon α signaling pathway using a combination of immunoprecipitation and immobilized metal affinity chromatography. *Mol. Cell. Proteomics* **4,** 721–730
15. Jonassen, I., Collins, J. F., and Higgins, D. G. (1995) Finding flexible patterns in unaligned protein sequences. *Protein Sci.* **4,** 1587–1595
16. Nevill-Manning, C. G., Wu, T. D., and Brutlag, D. L. (1998) Highly specific protein sequence motifs for genome analysis. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 5865–5871
17. Rigoutsos, I., and Floratos, A. (1998) Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm. *Bioinformatics* **14,** 55–67
18. Schwartz, D., and Gygi, S. P. (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23,** 1391–1398
19. Forman-Kay, J. D., and Pawson, T. (1999) Diversity in protein recognition by PTB domains. *Curr. Opin. Struct. Biol.* **9,** 690–695
20. Schulze, W. X., and Mann, M. (2004) A novel proteomic screen for peptide-protein interactions. *J. Biol. Chem.* **279,** 10756–10764
21. Songyang, Z., and Cantley, L. C. (1995) Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem. Sci.* **20,** 470–475
22. Vely, F., and Vivier, E. (1997) Conservation of structural features reveals the existence of a large family of inhibitory cell surface receptors and non-inhibitory/activatory counterparts. *J. Immunol.* **159,** 2075–2077
23. Burshtyn, D. N., Yang, W., Yi, T., and Long, E. O. (1997) A novel phosphotyrosine motif with a critical amino acid at position −2 for the SH2 domain-mediated activation of the tyrosine phosphatase SHP-1. *J. Biol. Chem.* **272,** 13066–13072
24. Hinsby, A. M., Olsen, J. V., Bennett, K. L., and Mann, M. (2003) Signaling initiated by overexpression of the fibroblast growth factor receptor-1 investigated by mass spectrometry. *Mol. Cell. Proteomics* **2,** 29–36
25. Hinsby, A. M., Olsen, J. V., and Mann, M. (2004) Tyrosine phosphoproteomics of fibroblast growth factor signaling: a role for insulin receptor substrate-4. *J. Biol. Chem.* **279,** 46438–46447
26. Diella, F., Cameron, S., Gemund, C., Linding, R., Via, A., Kuster, B., Sicheritz-Ponten, T., Blom, N., and Gibson, T. J. (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5,** 79
27. Nielsen, M., Lundegaard, C., Worning, P., Hvid, C. S., Lamberth, K., Buus, S., Brunak, S., and Lund, O. (2004) Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics* **20,** 1388–1397
28. Kaufman, L., and Rousseeuw, P. J. (1990) *Finding Groups in Data: an Introduction to Cluster Analysis*, John Wiley, New York
29. Fraley, C., and Raftery, A. E. (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput. J.* **41,** 578–588
30. Jain, A. K., Murth, M. N., and Flynn, P. J. (1999) Data clustering: a review. *ACM Comput. Surv.* **31,** 264–323
31. Ester, M., Kriegel, H. P., Sander, S., and Xu, X. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, Portland, August 2–4, 1996* (Simoudis, E., Han, J., and Fayyad, U., eds) Vol. 1, pp. 226–231, Association for the Advancement of Artificial Intelligence (AAAI) Press, Menlo Park, CA
32. Lund, O., Nielsen, M., Kesmir, C., Petersen, A. G., Lundegaard, C., Worning, P., Sylvester-Hvid, C., Lamberth, K., Roder, G., Justesen, S., and Buus, S., and Brunak, S. (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* **55,** 797–810
33. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18,** 6097–6100
34. Johnson, L. N., Kotz, S., and Kemp, A. W. (1992) *Univariate Discrete Distributions*, 2nd Ed., Wiley-Interscience, New York
35. Ong, S. E., Kratchmarova, I., and Mann, M. (2003) Properties of ^{13}C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* **2,** 173–181
36. Schulze, W. X., Deng, L., and Mann, M. (2005) Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.* **1,** 2005 0008
37. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–670
38. Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002) Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **979,** 233–239
39. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999)

95

Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567

40. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam protein families database. *Nucleic Acids Res.* **30,** 276–280

41. Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., Jonnalagadda, C. K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T. K., Gronborg, M., Ibarrola, N., Deshpande, N., Shanker, K., Shivashankar, H. N., Rashmi, B. P., Ramya, M. A., Zhao, Z., Chandrika, K. N., Padma, N., Harsha, H. C., Yatish, A. J., Kavitha, M. P., Menezes, M., Choudhury, D. R., Suresh, S., Ghosh, N., Saravana, R., Chandran, S., Krishna, S., Joy, M., Anand, S. K., Madavan, V., Joseph, A., Wong, G. W., Schiemann, W. P., Constantinescu, S. N., Huang, L., Khosravi-Far, R., Steen, H., Tewari, M., Ghaffari, S., Blobe, G. C., Dang, C. V., Garcia, J. G., Pevsner, J., Jensen, O. N., Roepstorff, P., Deshpande, K. S., Chinnaiyan, A. M., Hamosh, A., Chakravarti, A., and Pandey, A. (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* **13,** 2363–2371

42. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386

43. Dovat, S., Ronni, T., Russell, D., Ferrini, R., Cobb, B. S., and Smale, S. T. (2002) A common mechanism for mitotic inactivation of C2H2 zinc finger DNA-binding domains. *Genes Dev.* **16,** 2985–2990

44. Saydam, N., Adams, T. K., Steiner, F., Schaffner, W., and Freedman, J. H. (2002) Regulation of metallothionein transcription by the metal-responsive transcription factor MTF-1: identification of signal transduction cascades that control metal-inducible transcription. *J. Biol. Chem.* **277,** 20438–20445

45. Jantz, D., and Berg, J. M. (2004) Reduction in DNA-binding affinity of Cys2His2 zinc finger proteins by linker phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7589–7593

46. Wu, X., Suetsugu, S., Cooper, L. A., Takenawa, T., and Guan, J. L. (2004) Focal adhesion kinase regulation of N-WASP subcellular localization and function. *J. Biol. Chem.* **279,** 9565–9576

47. Eck, M. J., Pluskey, S., Trub, T., Harrison, S. C., and Shoelson, S. E. (1996) Spatial constraints on the recognition of phosphoproteins by the tandem SH2 domains of the phosphatase SH-PTP2. *Nature* **379,** 277–280

48. Huyer, G., and Ramachandran, C. (1998) The specificity of the N-terminal SH2 domain of SHP-2 is modified by a single point mutation. *Biochemistry* **37,** 2741–2747

49. Beebe, K. D., Wang, P., Arabaci, G., and Pei, D. (2000) Determination of the binding specificity of the SH2 domains of protein tyrosine phosphatase SHP-1 through the screening of a combinatorial phosphotyrosyl peptide library. *Biochemistry* **39,** 13251–13260

50. Liu, B. A., Jablonowski, K., Raina, M., Arce, M., Pawson, T., and Nash, P. D. (2006) The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Mol. Cell* **22,** 851–868

51. Poy, F., Yaffe, M. B., Sayos, J., Saxena, K., Morra, M., Sumegi, J., Cantley, L. C., Terhorst, C., and Eck, M. J. (1999) Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. *Mol. Cell* **4,** 555–561

52. Muraille, E., Bruhns, P., Pesesse, X., Daeron, M., and Erneux, C. (2000) The SH2 domain containing inositol 5-phosphatase SHIP2 associates to the immunoreceptor tyrosine-based inhibition motif of Fc γRIIB in B cells under negative signaling. *Immunol. Lett.* **72,** 7–15

53. Bruhns, P., Vely, F., Malbec, O., Fridman, W. H., Vivier, E., and Daeron, M. (2000) Molecular basis of the recruitment of the SH2 domain-containing inositol 5-phosphatases SHIP1 and SHIP2 by fcγ RIIB. *J. Biol. Chem.* **275,** 37357–37364

54. Damen, J. E., Liu, L., Rosten, P., Humphries, R. K., Jefferson, A. B., Majerus, P. W., and Krystal, G. (1996) The 145-kDa protein induced to associate with Shc by multiple cytokines is an inositol tetraphosphate and phosphatidylinositol 3,4,5-triphosphate 5-phosphatase. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 1689–1693

55. Marion, E., Kaisaki, P. J., Pouillon, V., Gueydan, C., Levy, J. C., Bodson, A., Krzentowski, G., Daubresse, J. C., Mockel, J., Behrends, J., Servais, G., Szpirer, C., Kruys, V., Gauguier, D., and Schurmans, S. (2002) The gene INPPL1, encoding the lipid phosphatase SHIP2, is a candidate for type 2 diabetes in rat and man. *Diabetes* **51,** 2012–2017

56. Sleeman, M. W., Wortley, K. E., Lai, K. M., Gowen, L. C., Kintner, J., Kline, W. O., Garcia, K., Stitt, T. N., Yancopoulos, G. D., Wiegand, S. J., and Glass, D. J. (2005) Absence of the lipid phosphatase SHIP2 confers resistance to dietary obesity. *Nat. Med.* **11,** 199–205

5 MS-ANALYSIS OF INTACT SILAC-LABELED PROTEINS

97

5.1 *PUBLICATION:* TOP-DOWN QUANTITATION AND CHARACTERIZATION OF SILAC-LABELED PROTEINS

This article presents the investigation of top-down analysis of SILAC-labeled proteins and was published in 2007, in the November issue of Journal of the American Society for Mass Spectrometry on pages 2058-2064.

The project was a joint effort together with Leonie Waanders in our group.

The following pages contain the published version of the article.

Top-Down Quantitation and Characterization of SILAC-Labeled Proteins

Leonie F. Waanders, Stefan Hanke, and Matthias Mann
Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Martinsried, Germany

Stable isotope labeling by amino acids in cell culture (SILAC) has become a popular labeling strategy for peptide quantitation in proteomics experiments. If the SILAC technology could be extended to intact proteins, it would enable direct quantitation of their relative expression levels and of the degree of modification between different samples. Here we show through modeling and experiments that SILAC is suitable for intact protein quantitation and top-down characterization. When SILAC-labeling lysine and/or arginine, peaks of light and heavy SILAC-doublets do not interfere with peaks of different charge states at least between 10 and 200 kDa. Unlike chemical methods, SILAC ensures complete incorporation—all amino acids are labeled. The isotopic enrichment of commercially available SILAC amino acids of nominally 95% to 98% shifts the mass difference between light and heavy state but does not lead to appreciably broadened peaks. We expressed labeled and unlabeled Grb2, a 28 kDa signaling protein, and showed that the two forms can be quantified with an average standard deviation of 6%. We performed on-line top-down sequencing of both forms in a hybrid linear ion trap orbitrap instrument. The quantized mass offset between fragments provided information about the number of labeled residues in the fragments, thereby simplifying protein identification and characterization.   (J Am Soc Mass Spectrom 2007, 18, 2058–2064) © 2007 American Society for Mass Spectrometry

The use of mass spectrometry to analyze biological samples has evolved tremendously over the last decade, and mass spectrometry-based proteomics, in particular, has become an indispensable part of modern life sciences [1]. In recent years, the need for quantitative as opposed to qualitative proteomics experiments has become apparent, and a large number of different approaches have been developed, mainly based on incorporation of a "light" and "heavy" stable isotope tag [2]. In chemical approaches, such as in the original ICAT and in the iTRAQ method, the label is reacted with a functional group of an amino acid. In metabolic approaches, the label is instead incorporated by living cells through protein turnover. Our laboratory has previously described a metabolic labeling method called stable isotope labeling by amino acids in cell culture (SILAC, [3]). In SILAC, an essential amino acid has been substituted by its stable isotope counterpart in the medium in which the cells grow, and this "heavy" amino acid is hence incorporated into all expressed proteins. SILAC allows very accurate peptide quantitation in an automated, high throughput experimental setup.

In contrast to this "bottom-up" approach, "top-down" proteomics seeks to characterize intact proteins. Although not widely used in biological research yet, top-down proteomics has unique potential because it can characterize the complete primary structure of the proteins, including modifications that may be missed in the bottom-up approach. Until now, most top-down studies have focused on mass measurement of the protein, its identification in databases using MS or MS/MS data, or on measuring protein modifications. Hardware and software developments, such as improved resolution and sensitivity, better fragmentation techniques, and increasingly automated software have recently made top-down identification and characterization much faster [4–7]. These developments should make quantitative rather than qualitative top-down proteomics more feasible. So far, however, very few studies have combined the top-down approach with protein quantitation. Gordon et al. demonstrated that relative molecular ion intensities can be used for intact protein quantitation [8]. Kelleher and coworkers investigated the use of ^{15}N labeling of yeast proteins for intact protein quantitation and determined 50 protein ratios [9]. In the same paper, the authors also chemically labeled yeast proteins with acrylamide and iodoacetamide. They concluded that the use of stable isotopes is preferred, since it prevents chromatographic shifts during LC-MS separation and make quantitation easier and more accurate. Furthermore, a general problem of chemical labeling strategies is the fact that it is not 100% complete in terms of incorporation. Due to steric hindrance in the intact proteins, not all amino acids react evenly well with the isotope labeled reagent, and therefore the degree of labeling is difficult to control.

Although quantitative approaches are rare in the top-down proteomics field, stable isotopes have been used for intact protein analysis for a number of years. In 2000, Smith and coworkers used deuterated leucine to improve the identification of *E. coli* proteins as the mass offset in the full scan indicated the number of leucines present in the protein [10]. They also substituted other essential amino acids such as Ile, Phe, Arg, His, and Lys to determine their number in the protein [11]. "Heavy" and "light" labeled proteins were well separated as long as the labeled amino acid was present at least three times per protein.

We have recently investigated the LTQ-Orbitrap, a hybrid linear ion trap orbitrap mass spectrometer, for top-down proteomics [12]. We found that the good sensitivity and excellent mass accuracy and resolution are sufficient for fast and reliable protein identification and characterization. We now extend this work by investigating whether the SILAC technology is also applicable to intact protein quantitation and characterization. We show feasibility by modeling SILAC-labeling for proteins up to 220 kDa, and by quantifying a 28 kDa signaling protein that was expressed in medium with normal and with heavy arginine and lysine residues.

Experimental

Modeling

The theoretical applicability of SILAC for protein quantitation was tested by modeling the isotopic and charge state clusters of proteins of different molecular weight. Based on the amino acid frequencies as determined by McCaldon and Argos [13], we calculated the mass of a hypothetical sequence with 100 residues (~11 kDa) and extrapolated this to 55 and 220 kDa. For the modeling we used the Isotopica web application, developed by the groups of Fernandez-de-Cossio and Takao [14]. We first simulated SILAC-labeling with $^{13}C_6^{15}N_4$-arginine and $^{13}C_6^{15}N_2$-lysine, inducing a mass increment of 10.008 and 8.014 Da, respectively, assuming that 100% of the substituted carbon and nitrogen atoms were "heavy."

The influence of incomplete labeling and isotope enrichment were tested by calculating the frequency of different protein forms in an arginine-labeled 55 kDa protein with approximately average amino acid distribution. To test the effects of imperfect isotope enrichment of the SILAC amino acids, we varied the relative abundance of the substituted ^{13}C and ^{15}N in the labeled amino acid from 100% to a lower value such as 98% or 95%.

To simulate incomplete amino acid incorporation, we calculated the relative abundance of protein forms when the chance of incorporation of light labels was $P = 0.02$ or $P = 0.05$, assuming a binomial distribution. For every protein form, the isotope distribution of the 34th charge state was predicted with Isotopica (resolution 60,000) and protein forms were weighted according to their probabilistic frequencies.

In Vitro Expression and Purification of Grb2

The gene for growth factor receptor-bound protein 2 (Grb2) was purchased from RZPD German Resource Center for Genome Research, Berlin, Germany (clone RZPDo834A0934D). It was transferred into pDEST17 via Gateway-cloning for T7-RNA polymerase dependent expression as an N-terminally His$_6$-tagged protein.

All chemicals and enzymes were purchased from Sigma (Taufkirchen, Germany) or Roche Applied Science (Mannheim, Germany) at the highest purity.

Recombinant N-terminally His$_6$-tagged Grb2 was expressed using a cell-free system prepared as described in reference [15] with slight modifications. *E. coli* S30 lysate was prepared from BL21(DE3)RIL cells (Novagen, Darmstadt, Germany). T7 RNA polymerase was also expressed in BL21(DE3)RIL containing the vector pAR1219 described in ref [16], but the enzyme was not purified. Instead, we prepared another lysate from this IPTG-induced culture and added 60 μL for this lysate to 400 μL of the standard lysate to 1 mL of reaction volume. The concentration of the reaction components were adjusted to 57 mM HEPES-KOH buffer (pH 8.2), 2 mM DTT, 1.2 mM ATP, 0.85 mM each of CTP, GTP and UTP, 100 mM creatine phosphate, 130 μg/mL creatine kinase, 2.0% PEG 8000, 0.64 mM 3′,5′-cyclic AMP, 34 μM L(-)-5-formyl-5,6,7,8-tetrahydrofolic acid, 175 μg/mL *E. coli* total tRNA, 90 mM potassium glutamate, 80 mM ammonium acetate, 12 mM magnesium acetate, 2.0 mM each of the 20 amino acids, and 6.7 μg/mL of plasmid DNA. The reaction mixture was incubated for 2 h while shaking at 600 rpm at 30 °C.

For the heavy SILAC labeled Grb2, the normal arginine and lysine were replaced with their heavy-isotope counterparts $^{13}C_6^{15}N_4$-Arginine and $^{13}C_6^{15}N_2$-Lysine, also at a concentration of 2 mM. The reaction mixture contained the target protein mostly as precipitate, which was solubilized with 6 M guanidinium chloride and subsequently purified employing Ni^{2+}-affinity chromatography according to the manufacturer's protocol for purification of denatured protein using Ni-NTA sepharose (Qiagen, Hilden, Germany) and spin columns (MoBiTec, Goettingen, Germany). The purified protein was dialyzed against distilled water to remove imidazole and guanidinium, which again led to precipitation of the protein.

Sample Preparation and Mass Spectrometry

Heavy and light labeled forms of Grb2 were aliquoted separately and stored at −80°C before use. Before mass spectrometric analysis, the proteins were washed on RP-C$_{18}$ StageTips [17], eluted, and mixed in the desired ratio.

All experiments were done on a LTQ-Orbitrap (Thermo Fisher, Bremen, Germany), coupled to a nanoLC system (Agilent, Waldbronn, Germany). Online protein separation was performed by use of 75 μm × 150 mm IntegraFrit columns (New Objective, Berlin, Germany) packed in-house with 5 μm RP-C18 beads (Reprosil-Pur Aq, 200 Å pore size, Dr. Maisch, Ammerbuch-Entringen,

Germany). The column was connected to a short nano-spray needle and spraying voltage was kept low (<2 kV) to prevent oxidation [18]. During loading and washing, the flow was set to 500 nL/min; whereas during the actual gradient the flow rate was 250 nL/min. Buffers compositions were 0.5% acetic acid in mQ distilled water (Buffer A) and 0.5% acetic acid in acetonitrile (Buffer B). The proteins eluted in a 30 min gradient from 40% to 90% of Buffer B. CID fragmentation was performed in the LTQ, but MS/MS ions were detected in the orbitrap mass analyzer.

Data Analysis

The MS scans were deconvoluted with Xtract software (Thermo Fisher) and matched with the expected sequence. The light form of Grb2 was used to determine the sequence of the linker between the His-tag and the protein.

CID fragments observed in light or in both light and heavy MS/MS spectra were searched in ProSight PTM (developed by The Kelleher group, University of Illinois, Urbana-Champaign, IL), but without success. We therefore suspected they were internal fragments and searched them in Mascot (Matrix Science Ltd., London, UK) after increasing light fragment masses by one water molecule (18.0152 Da). A peptide mass fingerprint search was performed against the human IPI database (version 3.19) and a small database with the predicted His-tagged Grb2 sequence with 10 ppm maximum mass deviation and without enzyme specificity.

Results and Discussion

Simulation of SILAC Quantitation for Intact Proteins

We first wanted to investigate whether SILAC would be applicable for proteins of all sizes and whether any overlap between charge states and labeled states could occur. To this end, we used Isotopica, a software package freely available on the Internet, and developed by Fernandez-de-Cossio and Takao [14], to model the effect of SILAC labeling. Isotopica can predict the isotopic cluster of specific charge states and can also calculate the mono-isotopic and average mass of all charge states (up to 50) for both heavy and light labeled proteins. As shown in Figure 1a–c, SILAC labeling with heavy arginine and lysines results in mass offsets that are clearly distinguishable in a spectrum. The thickness of the lines is approximately equal to the full width of the isotopic cluster. Even for a 220 kDa protein (Figure 1c) the two lines are separate, indicating that at least theoretically the isotopic clusters of the light and heavy protein forms do not overlap with consecutive charge states.

Because the mass offset induced by SILAC is in the range of a few hundred Da, when both arginine and lysine are labeled, post-translational modifications with

a mass less than this would not overlap with either the heavy or the light form of the protein. For example, several phosphorylation sites on a 50 kDa protein would not overlap with the heavy version, which is ~500 Da higher in mass.

The simulation also allowed us to visualize the effect of incomplete isotope enrichment. In principle, SILAC amino acids should be labeled with ^{13}C or ^{15}N to 100% at the substituted sites but in practice commercial sources guarantee isotope enrichment between 95% and 98%. As described in the Experimental section, we modeled this by varying isotope abundances in the heavy form of an average 55 kDa arginine-labeled protein. Interestingly, incomplete isotope enrichment does not broaden the peaks but it does shift the mass of the heavy form to a lower value (Figure 1d). It follows from this that accurate mass measurement to determine the molecular weight of the intact protein should be performed on the light and not on the heavy form of the SILAC protein pair.

Next, we modeled the effect of incomplete labeling, that is, we modeled the case that not every light amino acid was replaced by its heavy counterpart. This could happen during SILAC cell adaptation, while light amino acids are still present. As Figure 1e shows, even incomplete labeling at the two percent level (no more than one amino acid in our example), causes the signal to split into at least three states and this reduces the overall intensity and signal to noise (S/N) dramatically. At the 5% level, severe splitting into at least four states occurs, and the heavy peak is broadened at least 3-fold. Therefore, complete labeling is a precondition for successful quantitative top-down analysis. Fortunately, it is not difficult in SILAC experiments to achieve complete labeling; the only requirement is to grow the cells for a sufficient number of cell doubling.

Note, however, that this broadening is very likely to occur in chemical labeling strategies because it is very difficult to achieve close to 100% labeling on the desired amino acids while preventing any labeling of untargeted sites.

Expression and Mass Spectrometry of SILAC-Labeled Grb2

To experimentally test the feasibility and accuracy of quantitation by SILAC for an intact protein, we expressed HIS-tagged Grb2, a signaling protein with a calculated monoisotopic mass of 27,789.758 Da in vitro in normal media and in media with heavy arginine and lysine. Light and heavy protein forms were measured separately and mixed in ~1:1 or 2:1 ratio. The mixture was analyzed by online HPLC MS on the LTQ-Orbitrap (see Figure 2).

The resolving power of the orbitrap was sufficient to observe the isotopic clusters. By measuring the heavy labeled protein only, we obtained a single population and conclude that we achieved 100% label incorporation (data not shown).
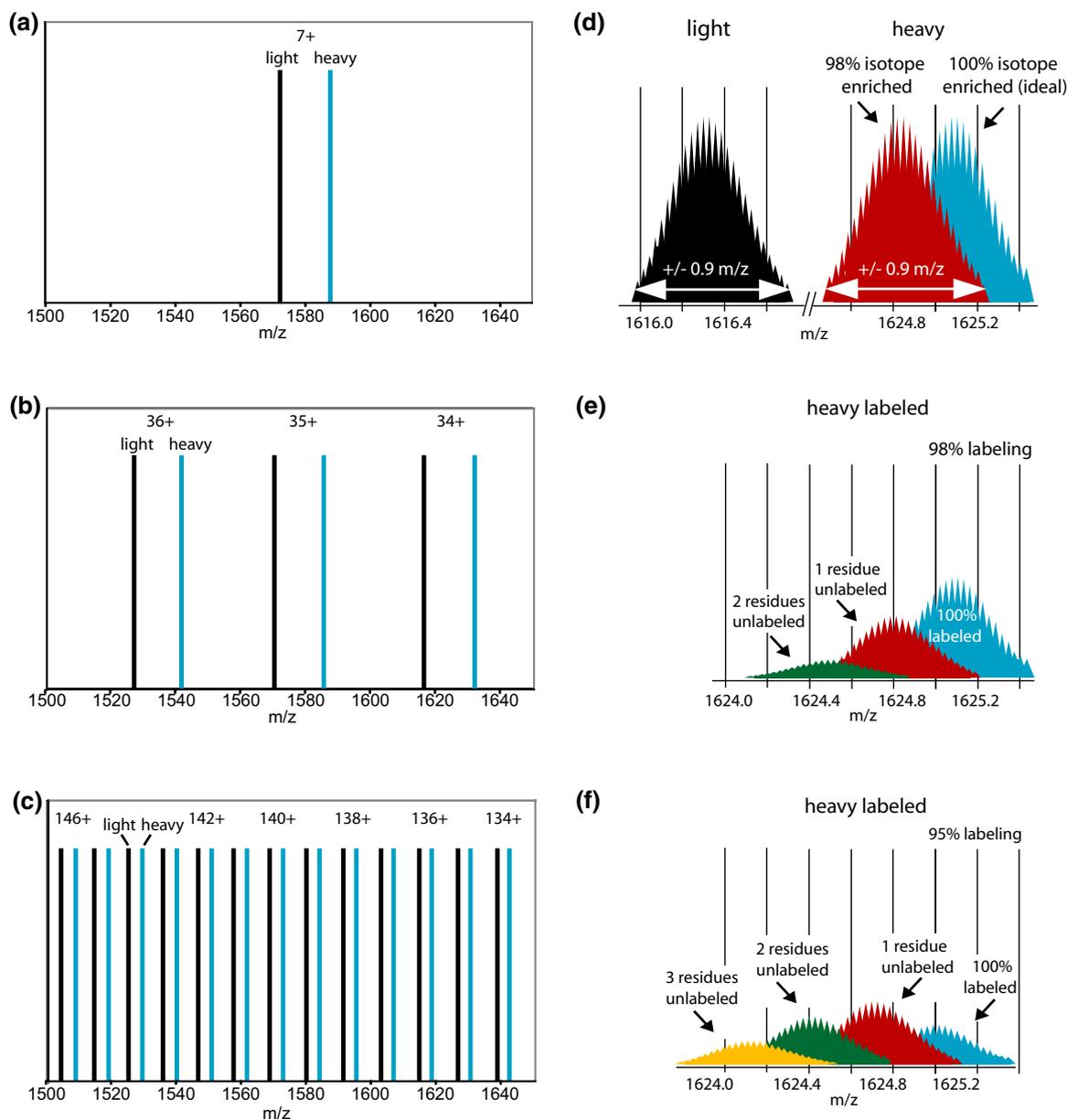
Figure 1. Simulation of heavy and light SILAC labeled proteins of various molecular weights with an average amino acid distribution. (**a**)–(**c**) SILAC-labeling with $^{13}C_6{}^{15}N_4$-Arg and $^{13}C_6{}^{15}N_2$-Lys increases the protein mass by approximately 1%, resulting in well separable isotope clusters, independent of protein size. The isotopic clusters of light and heavy labeled proteins of 11, 55, and 200 kDa in the *m/z* range 1500 to 1650 are depicted in panels (**a**)–(**c**), respectively. The thickness of the lines indicates the full width of the isotope cluster, which for all size proteins remains below 1 Thomson. (**d**)–(**f**) The effects of incomplete isotope enrichment and incomplete mass labeling are modeled, based on the 34+ peak of a theoretical protein of 55 kDa with average amino acid composition, labeled with heavy arginine and measured with 60,000 resolving power. (**d**) A 98% isotope enrichment of ^{13}C and ^{15}N in the heavy amino acids results in a shift of the total heavy isotopic cluster from 100% enrichment. The width and the intensity of the isotopic cluster remain unchanged. (**e**) In contrast, if only 98% of the arginines and lysines are labeled with heavy amino acids, the result is a significant spread of the signal. (**f**) With 95% labeling the signal is even more spread and the total intensity is reduced to 33% of the original signal.

Influence of Isotope Enrichment in Heavy Label

Above, we modeled the influence of incomplete isotope enrichment of the SILAC amino acid. The SILAC amino acids used in our experiment were specified at 98% for ^{13}C. For ^{15}N they were specified as 95% for Lys and 98% for Arg. Thus about 2% of the labeled carbon and nitrogen atoms in their stable isotope labeled amino acids should be light (^{12}C and ^{14}N). However, when we used this value to predict the mass offset between the heavy and light labeled Grb2 protein forms, we could not explain an additional mass difference of 3 Da for the heavy labeled Grb2 form in comparison with the calcu-

101
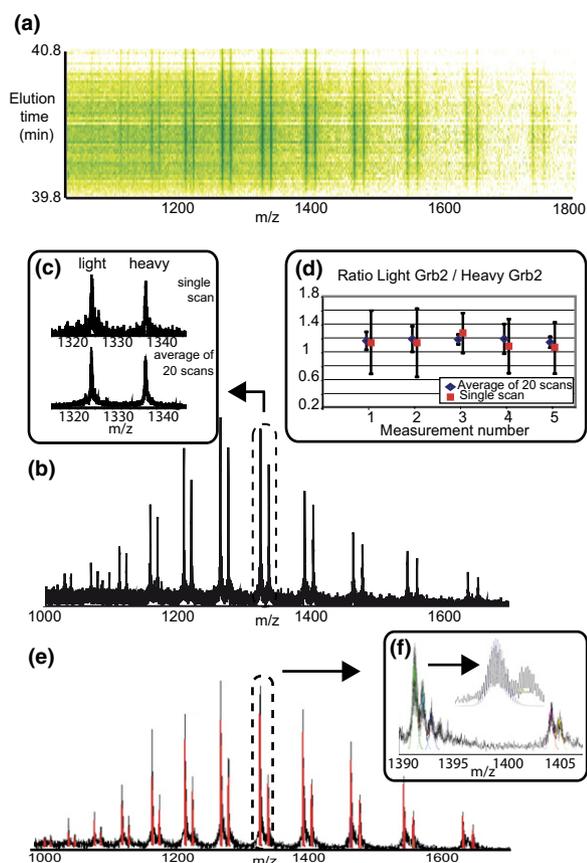
Figure 2. Experimental spectra of mixed Grb2 forms, showing the experimental feasibility of SILAC. (**a**) Contour plot of the eluting SILAC labeled protein pair. The graph is color coded with light green corresponding to low and dark green to high intensity. (**b**) Single spectrum of mixed light and heavy Grb2 forms, showing multiple charge states of both forms with isotopic clusters that do not overlap. The ratio between heavy and light Grb2 is constant over the different charge states and allows accurate quantitation based on relative intensities. (**c**) Averaging of spectra leads to significant noise reduction and improves quantitation accuracy. (**d**) Intensity ratios between light and heavy labeled Grb2, based on the eight most intense charge states of single or averaged spectra (20 scans) for 5 different LC-MS-runs were determined with their 95% confidence interval error bars. (**e**) An experimentally obtained spectrum was overlaid with the Isotopica prediction (red), calculated using 99% heavy isotope enrichment. (**f**) The inserts show that the simulation is accurate down to the isotopic resolution, indicating higher isotope enrichment of Arg and Lys than specified by manufacturers.

lated one of 264.3 Da. Since the sequence of the protein is known and confirmed by the light labeled protein, this mass difference had to be caused by the heavy amino acids.

When overlaying the experimentally obtained spectra with the prediction made by Isotopica (Figure 2b), we discovered that the matching became very accurate when isotope enrichment was 99%. Thus, in this case, the isotope enrichment of the SILAC amino acids was almost complete and significantly higher than specified. Furthermore, our measurements confirmed the simulation in Figure 1d. The rare presence of ^{12}C and ^{14}N instead of ^{13}C and ^{15}N in the heavy amino acids did not

cause problems for the analysis or quantitation, since the signal intensity and the width of the isotopic distribution remained the same. Note that the percentage of light atoms in the heavy label needs to be determined only once as it will be the same for all proteins that are labeled with that specific batch of heavy amino acids.

Quantitation at the Protein Level Using SILAC

A contour plot of the SILAC labeled Grb2 forms showed that they eluted as a 1 min peak with no discernable retention time shift between them (Figure 2a). Complete isotopic separation and coelution should allow for very accurate quantitation. The spreading of the protein signal into multiple charge states reduces the S/N but improves quantitation, since all charge states should show the same ratio between heavy and light labeled forms. To determine the quantitation precision, the ratio between Grb2-light and Grb2-heavy was calculated for a single scan and for an averaged scan, composed of 20 separate scans over the elution profile. In both cases the eight most intense charge states were used for the calculation. We then repeated this measurement in five separate HPLC runs to demonstrate the reproducibility of the measurement. In the insert of Figure 2, the average ratio is displayed with 95% confidence interval error bars. When averaging 20 scans, the mean standard deviation was 6%, whereas ratios determined from single scans had a standard deviation of typically 18%.

The ratio calculation can be further improved by using the complete elution profile of the proteins (see Figure 2a). Furthermore, as Ong et al. [19] and Du et al. [9] have noted, results become more accurate if they are corrected for noise, especially in case of low S/N peaks.

Improved Assignment of Top-Down Fragments by SILAC

Besides the accurate quantitation, SILAC can also be used for identification and characterization purposes. The information in fragmentation spectra is valuable in assigning modifications and in improving the reliability of protein identification. Modifications and truncations are more easily detected and more likely correctly assigned. Due to the speed and sensitivity of the orbitrap, such MS/MS spectra are readily obtainable on small proteins [12].

To demonstrate this in a SILAC experiment, we performed data-dependent CID fragmentation in the online format, collecting MS and MS/MS scans in the orbitrap, without microscanning to keep the duty cycle short. The MS/MS scans were of relatively low intensity (Figure 3), but the resolution and mass accuracy of the fragments was sufficiently high that even low abundant ions could be clearly distinguished from the noise.

When comparing the MS/MS scans of heavy and light labeled Grb2, we observed very similar fragmen-
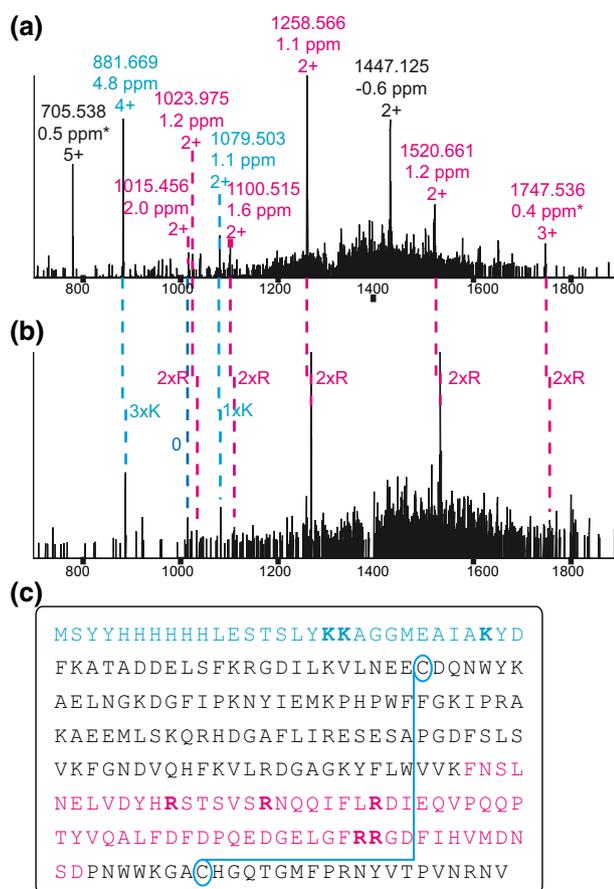
Figure 3. Identification of Grb2 is supported by direct comparison of the heavy and light CID fragmentation spectra. Mass differences between corresponding fragments observed in MS/MS spectra of the (**a**) light and (**b**) heavy Grb2 protein indicate the number of lysines and arginines per peptide and thereby simplify and confirm the peptide and protein identification. Assigned fragments have an average absolute mass accuracy of 3.1 ± 2.3 ppm. The mass deviation of the ion indicated with an asterisk has been derived from the ^{13}C isotope. N-terminal fragments are shown in blue and internal fragments in red. (**c**) In total the fragments covered 40% of the total protein sequence.

tation that appeared very useful for assigning the fragments. The mass offset between the heavy and light fragments should exactly represent the number of lysines and arginines of every fragment. Table 1 lists the fragments and their mass deviations from the calculated values. Most of the fragments were internal and likely observed because the disulphide bridge was still present in the protein.

The absolute mass deviation of the fragments was very low, on average 3.2 ppm, and the number of lysines and arginines determined by the mass increment of the heavy labeled fragments matched exactly with the proposed identification. One potential false positive hit could be eliminated both because of its mass deviation of 14 ppm and also because of the fact that its counterpart in the spectrum of the heavy form of Grb2 did not indicate the correct number of lysines. In total the single scan CID fragmentation depicted in the figure resulted in fragments covering 40% of the protein sequence. Clearly the SILAC information at the fragment level would have been very valuable to assign any protein modification, had it been present.

Conclusions and Perspectives

Here we have investigated the applicability of the SILAC technology to top-down proteomics. Through theoretical modeling, we found that heavy and light SILAC-doublets do not interfere with each other even for very large proteins. The incomplete isotope enrichment of commercial SILAC amino acids does not cause peak broadening but does shift the mass to lower values. This effect can be modeled very precisely. Incomplete labeling, on the other hand, would lead to distribution of the signal into several peaks, substantially broadening them and decreasing the signal to noise. Fortunately, incomplete labeling can easily be avoided in SILAC-metabolic labeling experiments.

Table 1. Fragments of Grb2 identified by Mascot

Light m/z	Heavy m/z	Charge	Δmass	No. of heavy residues	Mass deviation (ppm)	Sequence
954.775		3			5.49	MSYYHHHHHHLESTSLY**KK**AGGME
978.788		3			3.90	MSYYHHHHHHLESTSLY**KK**AGGMEA
1039.831	1045.504	3	16.03	2 Lys	(1.88)[a]	MSYYHHHHHHLESTSLY**KK**AGGMEAIA
881.669	887.681	4	24.04	3 Lys	4.79	MSYYHHHHHHLESTSLY**KK**AGGMEAIA**K**YD
705.538		5			(0.50)[a]	MSYYHHHHHHLESTSLY**KK**AGGMEAIA**K**YD
1175.223		3			4.81	MSYYHHHHHHLESTSLY**KK**AGGMEAIA**K**YD
1015.456	1015.455	2	0.01	0	2.00	SYYHHHHHHLESTSLY
1079.503	1083.511	2	8.02	1 Lys	1.10	SYYHHHHHHLESTSLY**K**
1577.204		2			1.67	LFDFDPQEDGELGF**RR**GDFIHVMDNSD
1520.661	1530.668	2	20.02	2 Arg	1.20	FDFDPQEDGELGF**RR**GDFIHVMDNSD
1447.125		2			−0.57	DFDPQEDGELGF**RR**GDFIHVMDNSD
1100.515	1110.520	2	20.04	2 Arg	1.59	PQEDGELGF**RR**GDFIHVMD
1258.566	1268.573	2	20.03	2 Arg	1.07	QEDGELGF**RR**GDFIHVMDNSDP
1023.975	1033.983	2	20.03	2 Arg	1.18	GELGF**RR**GDFIHVMDNS
1747.536	1757.540	3	30.01	3 Arg	(0.42)[a]	FNSLNELVDYH**R**STSVS**R**NQQIFL**R**DIE-QVPQQPTYVQALFDFD

[a]^{13}C-peaks were used for mass accuracy determination.

We have shown that a 28 kDa signaling protein, Grb2, can be readily quantified in the labeled versus the unlabeled form. The quantitation in a one to one mixture had a typical standard deviation of six percent and is mainly limited by the lower signal to noise in protein measurements compared with peptide measurements. Grb2 was fragmented and analyzed in an LTQ-Orbitrap instrument using single scan data from online experiments. The high mass accuracy combined with the quantized mass offsets significantly improves fragment identification when comparing tandem mass spectra of light and heavy SILAC labeled protein. The unambiguous information about the number of labeled residues per fragment or protein is a clear advantage of the SILAC technology over ^{15}N labeling.

In our experiment we chose to use two amino acids, arginine and lysine that were isotopically labeled, leading to a 1% mass offset between light and heavy form. Often, it may be more convenient to label with a single amino acid, for example, lysine. This would allow direct "counting" of the number of lysines in the protein and between fragments in the "heavy" and "light" tandem spectrum. Double amino acids may be valuable when analyzing proteins with potentially a high number of modifications.

Extension of the SILAC technology to intact protein analysis should allow direct quantitation of endogenous and unprocessed proteins. However, the most interesting application could be in the direct quantitation of multiple and combinatorial modifications of regulatory proteins as a function of cellular state. This goal will require SILAC quantitation of fragments isolating these modifications. It may also require nonergodic fragmentation techniques such as ECD or ETD as well as several stages of fragmentation.

Acknowledgments

References

1. Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422*, 198–207.
2. Ong, S. E.; Mann, M. Mass Spectrometry-Based Proteomics Turns Quantitative. *Nat. Chem. Biol.* **2005**, *1*, 252–262.
3. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteom.* **2002**, *1*, 376–386.
4. Kelleher, N. L. Top-Down Proteomics. *Anal. Chem.* **2004**, *76*, 197A–203A.
5. Bogdanov, B.; Smith, R. D. Proteomics by FTICR Mass Spectrometry: Top-Down and Bottom-Up. *Mass Spectrom. Rev.* **2005**, *24*, 168–200.
6. Whitelegge, J.; Halgand, F.; Souda, P.; Zabrouskov, V. Top-Down Mass Spectrometry of Integral Membrane Proteins. *Exp. Rev. Proteom.* **2006**, *3*, 585–596.
7. Zamdborg, L.; Leduc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: Improved Protein Identification and Characterization for Top-Down Mass Spectrometry. *Nucleic Acids Res.* **2007**.
8. Gordon, E. F.; Mansoori, B. A.; Carroll, C. F.; Muddiman, D. C. Hydropathic Influences on the Quantification of Equine Heart Cytochrome *c* Using Relative Ion Abundance Measurements by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Mass Spectrom.* **1999**, *34*, 1055–1062.
9. Du, Y.; Parks, B. A.; Sohn, S.; Kwast, K. E.; Kelleher, N. L. Top-Down Approaches for Measuring Expression Ratios of Intact Yeast Proteins Using Fourier Transform Mass Spectrometry. *Anal. Chem.* **2006**, *78*, 686–694.
10. Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic, L.; Smith, R. D. Proteome Analysis Using Selective Incorporation of Isotopically Labeled Amino Acids. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 78–82.
11. Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Pasa-Tolic, L.; Smith, R. D. Selective Incorporation of Isotopically Labeled Amino Acids for Identification of Intact Proteins on a Proteome-Wide Level. *J. Mass Spectrom.* **2002**, *37*, 99–107.
12. Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M. Top-Down Protein Sequencing and MS3 on a Hybrid Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer. *Mol. Cell. Proteom.* **2006**, *5*, 949–958.
13. McCaldon, P.; Argos, P. Oligopeptide Biases in Protein Sequences and Their Use in Predicting Protein Coding Regions in Nucleotide Sequences. *Proteins* **1988**, *4*, 99–122.
14. Fernandez-de-Cossio, J.; Gonzalez, L. J.; Satomi, Y.; Betancourt, L.; Ramos, Y.; Huerta, V.; Besada, V.; Padron, G.; Minamino, N.; Takao, T. Automated Interpretation of Mass Spectra of Complex Mixtures by Matching of Isotope Peak Distributions. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2465–2472.
15. Kigawa, T.; Yabuki, T.; Matsuda, N.; Matsuda, T.; Nakajima, R.; Tanaka, A.; Yokoyama, S. Preparation of *Escherichia coli* Cell Extract for Highly Productive Cell-Free Protein Expression. *J. Struct. Funct. Genomics* **2004**, *5*, 63–68.
16. Davanloo, P.; Rosenberg, A. H.; Dunn, J. J.; Studier, F. W. Cloning and Expression of the Gene for Bacteriophage T7 RNA Polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 2035–2039.
17. 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.
18. Morand, K.; Talbo, G.; Mann, M. Oxidation of Peptides During Electrospray Ionization. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 738–743.
19. Ong, S. E.; Kratchmarova, I.; Mann, M. Properties of 13C-Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *J. Proteome Res.* **2003**, *2*, 173–181.

6 FURTHER PUBLISHED AND UNPUBLISHED WORK

Apart from the printed publications mentioned above, some projects were presented at international conferences in the form of the following oral presentations and posters:

- S. Hanke, H. Besir, D. Oesterhelt, M. Mann. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. Oral presentation at HUPO-conference in Amsterdam 08-2008

- S. Hanke, H. Besir, D. Oesterhelt, M. Mann. Absolute quantitation of proteins in proteomics: measuring the copy number per cell. Poster at Interact-PhD-symposium in Munich, 12-2007

- S. Hanke, H. Besir, D. Oesterhelt, M. Mann. Absolute SILAC: Absolute quantitation of proteins in complex mixtures using recombinant stable isotope labeled proteins. Poster at ASMS-conference in Indianapolis, 06-2007

- S. Hanke, WX. Schulze, H. Gausepohl, M. Mann. Microscale peptide synthesis for protein interaction experiments. Poster at HUPO-conference in Munich, 09-2005

6.1 VESICULAR TRANSPORT IN THE GOLGI-APPARATUS

As part of a collaborative project with the research group of Prof. Dr. Felix Wieland at the biochemistry center of the University of Heidelberg, mass spectrometric protein identification was conducted to determine interaction partners of ADP-ribosylation factor-1 (Arf1). Arf1 is a small GTPase that can attach to membranes and is relevant for vesicular transport in the secretory pathway of cells. Using site-directed photo-activated crosslinking, interaction partners for different regions within Arf1 could be determined.

The project culminated in a publication in the journal Traffic in May 2007 on pages 582-593.

6.2 TRAFFICKING OF CAVEOLAE

As part of a collaborative project with the research group of Prof. Dr. Miguel del Pozo at the national center for cardiovascular research in Madrid, quantitative proteomics was

applied to identify interaction partners of Caveolin-1. Peptide pulldown experiments revealed pTyr-based specific binding events that in part depend on cell culture conditions (adherent versus suspension). Biological follow-up experiments are still being performed in Madrid, and publication of the results is scheduled for the year 2009.

6.3 CAPSID PROTEINS OF HERPES SIMPLEX VIRUS

As part of a collaborative project with the research group of Prof. Dr. Baumeister at the Max-Planck-Institute for biochemistry, absolute quantitation of Herpes simplex virus proteins was carried out. Stable isotope labeled peptide standards were used to determine the copy number per viral particle for several capsid proteins and to determine the stoichiometric relationship between them. These data form part of a structural project on viral architecture and are planned to be published in the year 2009.

6.4 PHOSPHORYLATION DYNAMICS OF IRS-1 AND IRS-2

The project on the pTyr-interactome in the insulin signaling pathway yielded many interactions that can be mediated by tyrosine phosphorylated sites in IRS proteins. This raised the question which of these sites are in fact phosphorylated *in vivo*, and which stimuli trigger them. Thus, an experimental approach was designed that enables the identification of phosphorylation sites in response to hormone stimulation.

A multiplexed quantitative phosphoproteomics strategy was chosen in order to also address the time dimension and deliver dynamic profiles of phosphorylation sites depending on the timepoint after stimulation. This strategy is described by the scheme in figure 19, and the analyzed timepoints were 0, 5, 10, 20, 30, 40 and 60 minutes after stimulation.
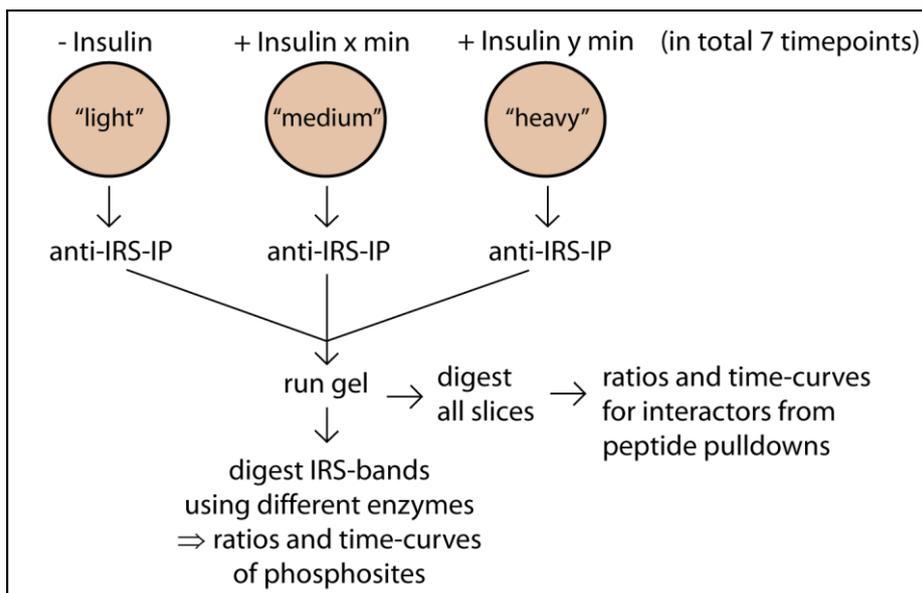
Figure 19: Simplified scheme of the experimental strategy for the investigation of time-resolved phosphorylation changes in IRS proteins after hormone stimulation.
Triple encoding SILAC allows for the quantitative comparison of 3 different stimulation times in one experiment. Immuno-precipitations against IRS-1 and IRS-2 are carried out separately, and samples are combined before SDS-PAGE.

This setup allows to simultaneously identify phosphorylation sites and IRS-interacting proteins with a readout for their quantitative increase (or decrease) in response to hormone stimulation. Different variants of individual steps in the procedure were tested and optimized. Antibodies against IRS-1 and IRS-2 were chemically crosslinked to magnetic beads, and efficient removal of IRS proteins from myotube cell lysates was achieved. Due to relatively clean IRS-bands in SDS-PAGE, further enrichment of phosphopeptides via titanium dioxide chromatography after proteolytic digestion did not increase the yield and was therefore omitted. More than 20 phosphorylation sites per IRS protein were identified with high confidence, however, only three of them were pTyr-sites despite targeted MS-acquisition via inclusion lists. Quantitative abundances of phosphorylation sites hardly changed in response to insulin stimulation. The most pronounced increase was found for the pTyr-sites, but all changes remained below a factor of 2. Accordingly, not more than two of the previously identified interaction partners of IRS proteins were found.

Our conclusion from these results is that the stimulation with insulin did not work properly. For future efforts this needs to be improved substantially. Additionally, more focused data acquisition methods in MS analysis, together with improved performance of inclusion list dependent events, will likely enhance the output of this experimental approach. Due to time constraints the experiment could not be repeated within the scale of this thesis.

7   CONCLUDING REMARKS AND PERSPECTIVES

7.1 SUMMARY AND CONCLUSIONS

The work presented in this thesis applied quantitative proteomics to different fields of biology. On the one hand, specific protein interactions based on pTyr were characterized, leading to a large array of novel possible interactions in the insulin signaling pathway and to the definition of binding motifs. On the other hand, MS-analysis of SILAC-labeled proteins was assessed. This resulted in the development of a novel technique in MS-based absolute quantitation of proteins, Absolute SILAC, which allows for unprecedented quantitative accuracy.

The insulin signal is known to be translated into a cascade of tyrosine phosphorylation at the intracellular side of target cells, which forms the molecular basis for regulated interactions between proteins. The molecular interaction platforms IRS-1 and IRS-2 as well as the receptors InsR, IGF1R and IRR were successfully screened for their pTyr-based interaction profiles in a site-specific format. The results provide novel insights into the connectivity and the concerted organization of signal transmission at this central stage of the pathway. A large body of evidence points at more extended interaction capabilities of IRS-1 and IRS-2 than currently known. Several proteins involved in signaling and metabolism interact differentially with them and thus provide an explanation for their different physiological roles. A large body of common interactors, on the other hand, contributes to their frequently observed redundant roles. The hitherto undiscovered direct binding of fatty acid degrading enzymes to IRS-proteins provides an additional and direct link to metabolic control. The data furthermore suggest the presence of novel pTyr-binding modules besides the established SH2 and PTB domains. Differential recruitment of interactors by members of the insulin receptor family yielded a molecular basis for their unequal physiological roles. Particularly intriguing was the discovery of exclusive binding events for the insulin receptor related receptor. In addition, doubly- and triply phosphorylated motifs provided insight into the combinatorial effects of phosphorylation events in close proximity to each other.

The technology applied to the insulin project was also exploited in a system-wide approach to mine the proteome for pTyr-interaction motifs using both bioinformatic methods and experimental validation. A novel algorithm clustered experimentally observed pTyr-sites into 20 different groups of motifs. Two thirds of them were found to mediate protein interactions. Interestingly, those clusters stem from proteins that are

annotated as being involved in signal transduction. The other third of clusters was derived from proteins that are annotated as enzymes or proteins involved in nucleic acid or ion binding. This suggests that approximately one third of *in vivo* pTyr-sites does not trigger interactions between proteins, but instead may serve structural functions by modulating the catalytic activity of enzymes or influencing the affinity for DNA or RNA of proteins involved in gene expression. In addition to these global insights, a new type of binding motif was discovered for the inositol phosphatase SHIP-2. Remarkably, residues N-terminal of the pTyr govern the binding event, which has been observed for very few SH2 domains so far.

Conventional methods for absolute quantitation of proteins by MS-based proteomics suffer from questionable accuracy due to the different nature of the quantitation standard as compared to the endogenous protein to be quantified. This work has endowed the field of protein quantitation with a novel technique called Absolute SILAC, which delivers unprecedented accuracy in absolute quantitation. The generation of SILAC-labeled recombinant proteins for use as quantitation standards was demonstrated *in vivo* and *in vitro*. Handling of the standard was optimized to guarantee high quality of results. The approach was characterized in depth, proving excellent accuracy and linearity. Beyond the biochemical and analytical features, a novel mass spectrometric data acquisition method was designed that improves the detection limit by more than an order of magnitude, enhances the success of identification and optimizes the quality of quantitation especially for low abundant proteins. Accurate quantitation of a protein down to 150 attomole could be demonstrated in the complex background of a whole cell lysate without any upfront fractionation. Application of Absolute SILAC to the quantitation of the signaling adaptor protein Grb2 in several different cell lines resulted in a copy number of close to 500,000 molecules per cell. Comparison against data from whole proteome analysis suggested that approximately 1000 proteins in a cell can be estimated to have a copy number higher than this.

Finally, the analysis of intact proteins as SILAC-pairs by top-down proteomics yielded valuable findings for applications that might use top-down measurements on SILAC-labeled material in the future. It was demonstrated that isotope clusters of light and heavy versions of a protein never overlap. Despite the high number of isotopic peaks in case of large proteins, isotope clusters are always confined within the width of one Thomson. Incomplete isotope enrichment in the amino acid(s) used for SILAC-labeling was shown to shift the isotope cluster of the labeled protein slightly to a lower mass without affecting the performance of the method. In contrast, incomplete labeling efficiency through incorporation of unlabeled components has a strong negative effect by spreading the total signal across several isotope clusters. This confirms the importance of using metabolic

labeling because chemical labeling can hardly reach complete labeling efficiency. Moreover the behavior in MS/MS was investigated, which demonstrated the potential of top-down analysis of SILAC-labeled proteins for improved certainty in identification and for the quantitative assessment of combinatorial posttranslational modifications.

7.2 PERSPECTIVES

With the accomplishment of a comprehensive pTyr-dependent interaction profile of IRS-1, IRS-2, InsR, IGF1R and IRR, this study extends the known pTyr-interactome of these proteins and provides solid hypotheses for a plethora of biological follow-up projects. In the future, it will be exciting to find out which of the pTyr-sites that were found to attract interaction partners indeed become phosphorylated *in vivo*. This is a multidimensional space, given the large number of different tissues, stimuli and physiological conditions that can decide on the occurrence of phosphorylation events. Thorough investigation of the precise role of novel interactions is likely to yield valuable information on the way the insulin signal is spread in a cell. Particularly interesting insights can be expected from the exploration of the pathways linking insulin to the regulation of transcription and lipid catabolism. The fact that several proteins showed pTyr-dependent interactions that are not mediated by SH2 or PTB domains provides promising ground for the discovery of novel pTyr-binding modules.

The clustering method developed within the proteome-wide pTyr-motif project is not restricted to pTyr-sites. It can be applied to other types of posttranslational modifications and supply those areas with an assessment of conserved motifs that can then be assayed for interactions using a similar experimental setup like the one reported here.

Absolute SILAC as a novel method for absolute quantitation of individual proteins in complex mixtures shows great promise for applications in research and clinical diagnostics. Accurate copy numbers of proteins per cell can provide systems biology with input for modeling and enable a more realistic view of connectivities in signaling networks and quantitative flux in metabolic pathways. Ultimately, contemporary immunoassays employed in patient diagnosis might be complemented or replaced by high-throughput multiplexed MS-based quantitation in the future. The high data quality of Absolute SILAC provides an important prerequisite for this kind of application. In addition, features of the newly developed data acquisition method are of generic nature and can be applied for other targeted approaches that aim at signal enhancement and improvements in identification and quantitation.

Analysis of proteins derived from SILAC-labeled cultures in intact form (top-down) holds great promise for the quantitative assessment of combinatorial modifications that change dynamically as a function of the cellular state.

8   BIBLIOGRAPHY

1.   Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The protein kinase complement of the human genome. *Science* **2002,** 298, (5600), 1912-34.
2.   Jackson, M. D.; Denu, J. M., Molecular reactions of protein phosphatases--insights from structure and chemistry. *Chem Rev* **2001,** 101, (8), 2313-40.
3.   Roque, A. C.; Lowe, C. R., Lessons from nature: On the molecular recognition elements of the phosphoprotein binding-domains. *Biotechnol Bioeng* **2005,** 91, (5), 546-55.
4.   Yaffe, M. B.; Elia, A. E., Phosphoserine/threonine-binding domains. *Curr Opin Cell Biol* **2001,** 13, (2), 131-8.
5.   Alonso, A.; Sasin, J.; Bottini, N.; Friedberg, I.; Friedberg, I.; Osterman, A.; Godzik, A.; Hunter, T.; Dixon, J.; Mustelin, T., Protein tyrosine phosphatases in the human genome. *Cell* **2004,** 117, (6), 699-711.
6.   Schlessinger, J.; Lemmon, M. A., SH2 and PTB domains in tyrosine kinase signaling. *Sci STKE* **2003,** 2003, (191), RE12.
7.   Pawson, T., Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* **2004,** 116, (2), 191-203.
8.   Neduva, V.; Russell, R. B., Peptides mediating interaction networks: new leads at last. *Curr Opin Biotechnol* **2006,** 17, (5), 465-71.
9.   Liu, B. A.; Jablonowski, K.; Raina, M.; Arce, M.; Pawson, T.; Nash, P. D., The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Mol Cell* **2006,** 22, (6), 851-68.
10.   Lappalainen, I.; Thusberg, J.; Shen, B.; Vihinen, M., Genome wide analysis of pathogenic SH2 domain mutations. *Proteins* **2008,** 72, (2), 779-792.
11.   Huang, H.; Li, L.; Wu, C.; Schibli, D.; Colwill, K.; Ma, S.; Li, C.; Roy, P.; Ho, K.; Songyang, Z.; Pawson, T.; Gao, Y.; Li, S. S., Defining the specificity space of the human SRC homology 2 domain. *Mol Cell Proteomics* **2008,** 7, (4), 768-84.
12.   Miller, M. L.; Hanke, S.; Hinsby, A. M.; Friis, C.; Brunak, S.; Mann, M.; Blom, N., Motif decomposition of the phosphotyrosine proteome reveals a new N-terminal binding motif for SHIP2. *Mol Cell Proteomics* **2008,** 7, (1), 181-92.
13.   Hu, J.; Hubbard, S. R., Structural characterization of a novel Cbl phosphotyrosine recognition motif in the APS family of adapter proteins. *J Biol Chem* **2005,** 280, (19), 18943-9.
14.   Pascal, S. M.; Singer, A. U.; Gish, G.; Yamazaki, T.; Shoelson, S. E.; Pawson, T.; Kay, L. E.; Forman-Kay, J. D., Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-gamma 1 complexed with a high affinity binding peptide. *Cell* **1994,** 77, (3), 461-72.
15.   Li, S. C.; Gish, G.; Yang, D.; Coffey, A. J.; Forman-Kay, J. D.; Ernberg, I.; Kay, L. E.; Pawson, T., Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A. *Curr Biol* **1999,** 9, (23), 1355-62.
16.   Bradshaw, J. M.; Waksman, G., Calorimetric examination of high-affinity Src SH2 domain-tyrosyl phosphopeptide binding: dissection of the phosphopeptide sequence specificity and coupling energetics. *Biochemistry* **1999,** 38, (16), 5147-54.
17.   Felder, S.; Zhou, M.; Hu, P.; Urena, J.; Ullrich, A.; Chaudhuri, M.; White, M.; Shoelson, S. E.; Schlessinger, J., SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. *Mol Cell Biol* **1993,** 13, (3), 1449-55.

18.   Rameh, L. E.; Chen, C. S.; Cantley, L. C., Phosphatidylinositol (3,4,5)P3 interacts with SH2 domains and modulates PI 3-kinase association with tyrosine-phosphorylated proteins. *Cell* **1995,** 83, (5), 821-30.

19.   Uhlik, M. T.; Temple, B.; Bencharit, S.; Kimple, A. J.; Siderovski, D. P.; Johnson, G. L., Structural and evolutionary division of phosphotyrosine binding (PTB) domains. *J Mol Biol* **2005,** 345, (1), 1-20.

20.   Smith, M. J.; Hardy, W. R.; Murphy, J. M.; Jones, N.; Pawson, T., Screening for PTB domain binding partners and ligand specificity using proteome-derived NPXY peptide arrays. *Mol Cell Biol* **2006,** 26, (22), 8461-74.

21.   Harrison, S. C., Peptide-surface association: the case of PDZ and PTB domains. *Cell* **1996,** 86, (3), 341-3.

22.   Sondermann, H.; Kuriyan, J., C2 can do it, too. *Cell* **2005,** 121, (2), 158-60.

23.   Christofk, H. R.; Vander Heiden, M. G.; Wu, N.; Asara, J. M.; Cantley, L. C., Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* **2008,** 452, (7184), 181-6.

24.   Sesti, G., Pathophysiology of insulin resistance. *Best Pract Res Clin Endocrinol Metab* **2006,** 20, (4), 665-79.

25.   Bryant, N. J.; Govers, R.; James, D. E., Regulated transport of the glucose transporter GLUT4. *Nat Rev Mol Cell Biol* **2002,** 3, (4), 267-77.

26.   Saltiel, A. R.; Kahn, C. R., Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **2001,** 414, (6865), 799-806.

27.   Taniguchi, C. M.; Emanuelli, B.; Kahn, C. R., Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol* **2006,** 7, (2), 85-96.

28.   White, M. F., Insulin signaling in health and disease. *Science* **2003,** 302, (5651), 1710-1.

29.   Saltiel, A. R., New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. *Cell* **2001,** 104, (4), 517-29.

30.   Sladek, R.; Rocheleau, G.; Rung, J.; Dina, C.; Shen, L.; Serre, D.; Boutin, P.; Vincent, D.; Belisle, A.; Hadjadj, S.; Balkau, B.; Heude, B.; Charpentier, G.; Hudson, T. J.; Montpetit, A.; Pshezhetsky, A. V.; Prentki, M.; Posner, B. I.; Balding, D. J.; Meyre, D.; Polychronakos, C.; Froguel, P., A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **2007,** 445, (7130), 881-5.

31.   Summers, S. A., Ceramides in insulin resistance and lipotoxicity. *Prog Lipid Res* **2006,** 45, (1), 42-72.

32.   Petersen, K. F.; Shulman, G. I., Etiology of insulin resistance. *Am J Med* **2006,** 119, (5 Suppl 1), S10-6.

33.   Houstis, N.; Rosen, E. D.; Lander, E. S., Reactive oxygen species have a causal role in multiple forms of insulin resistance. *Nature* **2006,** 440, (7086), 944-8.

34.   Zierath, J. R.; Krook, A.; Wallberg-Henriksson, H., Insulin action and insulin resistance in human skeletal muscle. *Diabetologia* **2000,** 43, (7), 821-35.

35.   Cantley, L. C., The phosphoinositide 3-kinase pathway. *Science* **2002,** 296, (5573), 1655-7.

36.   Korsheninnikova, E.; Voshol, P. J.; Baan, B.; van der Zon, G. C.; Havekes, L. M.; Romijn, J. A.; Maassen, J. A.; Ouwens, D. M., Dynamics of insulin signalling in liver during hyperinsulinemic euglycaemic clamp conditions in vivo and the effects of high-fat feeding in male mice. *Arch Physiol Biochem* **2007,** 113, (4-5), 173-85.

37.   Mendez, R.; Myers, M. G., Jr.; White, M. F.; Rhoads, R. E., Stimulation of protein synthesis, eukaryotic translation initiation factor 4E phosphorylation, and PHAS-I phosphorylation by insulin requires insulin receptor substrate 1 and phosphatidylinositol 3-kinase. *Mol Cell Biol* **1996,** 16, (6), 2857-64.

38.   Mendez, R.; Kollmorgen, G.; White, M. F.; Rhoads, R. E., Requirement of protein kinase C zeta for stimulation of protein synthesis by insulin. *Mol Cell Biol* **1997,** 17, (9), 5184-92.

39.   Harrington, L. S.; Findlay, G. M.; Lamb, R. F., Restraining PI3K: mTOR signalling goes back to the membrane. *Trends Biochem Sci* **2005,** 30, (1), 35-42.

40.   Love, D. C.; Hanover, J. A., The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci STKE* **2005,** 2005, (312), re13.

41.   Yang, X.; Ongusaha, P. P.; Miles, P. D.; Havstad, J. C.; Zhang, F.; So, W. V.; Kudlow, J. E.; Michell, R. H.; Olefsky, J. M.; Field, S. J.; Evans, R. M., Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* **2008,** 451, (7181), 964-9.

42.   Kraegen, E. W.; Cooney, G. J.; Turner, N., Muscle insulin resistance: a case of fat overconsumption, not mitochondrial dysfunction. *Proc Natl Acad Sci U S A* **2008,** 105, (22), 7627-8.

43.   Maasen, J. A., Mitochondria, body fat and type 2 diabetes: What is the connection? *Minerva Med* **2008,** 99, (3), 241-51.

44.   Nakae, J.; Kido, Y.; Accili, D., Distinct and overlapping functions of insulin and IGF-I receptors. *Endocr Rev* **2001,** 22, (6), 818-35.

45.   Kim, J. J.; Accili, D., Signalling through IGF-I and insulin receptors: where is the specificity? *Growth Horm IGF Res* **2002,** 12, (2), 84-90.

46.   Urso, B.; Cope, D. L.; Kalloo-Hosein, H. E.; Hayward, A. C.; Whitehead, J. P.; O'Rahilly, S.; Siddle, K., Differences in signaling properties of the cytoplasmic domains of the insulin receptor and insulin-like growth factor receptor in 3T3-L1 adipocytes. *J Biol Chem* **1999,** 274, (43), 30864-73.

47.   Kalloo-Hosein, H. E.; Whitehead, J. P.; Soos, M.; Tavare, J. M.; Siddle, K.; O'Rahilly, S., Differential signaling to glycogen synthesis by the intracellular domain of the insulin versus the insulin-like growth factor-1 receptor. Evidence from studies of TrkC-chimeras. *J Biol Chem* **1997,** 272, (39), 24325-32.

48.   Lammers, R.; Gray, A.; Schlessinger, J.; Ullrich, A., Differential signalling potential of insulin- and IGF-1-receptor cytoplasmic domains. *Embo J* **1989,** 8, (5), 1369-75.

49.   Hirayama, I.; Tamemoto, H.; Yokota, H.; Kubo, S. K.; Wang, J.; Kuwano, H.; Nagamachi, Y.; Takeuchi, T.; Izumi, T., Insulin receptor-related receptor is expressed in pancreatic beta-cells and stimulates tyrosine phosphorylation of insulin receptor substrate-1 and -2. *Diabetes* **1999,** 48, (6), 1237-44.

50.   Klammt, J.; Garten, A.; Barnikol-Oettler, A.; Beck-Sickinger, A. G.; Kiess, W., Comparative analysis of the signaling capabilities of the insulin receptor-related receptor. *Biochem Biophys Res Commun* **2005,** 327, (2), 557-64.

51.   White, M. F., Regulating insulin signaling and beta-cell function through IRS proteins. *Can J Physiol Pharmacol* **2006,** 84, (7), 725-37.

52.   White, M. F., IRS proteins and the common path to diabetes. *Am J Physiol Endocrinol Metab* **2002,** 283, (3), E413-22.

53.   Cai, D.; Dhe-Paganon, S.; Melendez, P. A.; Lee, J.; Shoelson, S. E., Two new substrates in insulin signaling, IRS5/DOK4 and IRS6/DOK5. *J Biol Chem* **2003,** 278, (28), 25323-30.

54.   Yamauchi, T.; Kaburagi, Y.; Ueki, K.; Tsuji, Y.; Stark, G. R.; Kerr, I. M.; Tsushima, T.; Akanuma, Y.; Komuro, I.; Tobe, K.; Yazaki, Y.; Kadowaki, T., Growth hormone and prolactin stimulate tyrosine phosphorylation of insulin receptor substrate-1, -2, and -3, their association with p85 phosphatidylinositol 3-kinase (PI3-kinase), and concomitantly PI3-kinase activation via JAK2 kinase. *J Biol Chem* **1998,** 273, (25), 15719-26.

55.   Lanzino, M.; Garofalo, C.; Morelli, C.; Le Pera, M.; Casaburi, I.; McPhaul, M. J.; Surmacz, E.; Ando, S.; Sisci, D., Insulin receptor substrate 1 modulates the transcriptional activity and the stability of androgen receptor in breast cancer cells. *Breast Cancer Res Treat* **2008**.

56.   Liang, L.; Zhou, T.; Jiang, J.; Pierce, J. H.; Gustafson, T. A.; Frank, S. J., Insulin receptor substrate-1 enhances growth hormone-induced proliferation. *Endocrinology* **1999,** 140, (5), 1972-83.

57.   Senthil, D.; Ghosh Choudhury, G.; Bhandari, B. K.; Kasinath, B. S., The type 2 vascular endothelial growth factor receptor recruits insulin receptor substrate-1 in its signalling pathway. *Biochem J* **2002,** 368, (Pt 1), 49-56.

58.   Vuori, K.; Ruoslahti, E., Association of insulin receptor substrate-1 with integrins. *Science* **1994,** 266, (5190), 1576-8.

59.   Shaw, L. M., Identification of insulin receptor substrate 1 (IRS-1) and IRS-2 as signaling intermediates in the alpha6beta4 integrin-dependent activation of phosphoinositide 3-OH kinase and promotion of invasion. *Mol Cell Biol* **2001,** 21, (15), 5082-93.

60.   Yenush, L.; White, M. F., The IRS-signalling system during insulin and cytokine action. *Bioessays* **1997,** 19, (6), 491-500.

61.   Dearth, R. K.; Cui, X.; Kim, H. J.; Hadsell, D. L.; Lee, A. V., Oncogenic transformation by the signaling adaptor proteins insulin receptor substrate (IRS)-1 and IRS-2. *Cell Cycle* **2007,** 6, (6), 705-13.

62.   Whitehead, J. P.; Clark, S. F.; Urso, B.; James, D. E., Signalling through the insulin receptor. *Curr Opin Cell Biol* **2000,** 12, (2), 222-8.

63.   Inoue, G.; Cheatham, B.; Emkey, R.; Kahn, C. R., Dynamics of insulin signaling in 3T3-L1 adipocytes. Differential compartmentalization and trafficking of insulin receptor substrate (IRS)-1 and IRS-2. *J Biol Chem* **1998,** 273, (19), 11548-55.

64.   Baserga, R., Is cell size important? *Cell Cycle* **2007,** 6, (7), 814-6.

65.   Sesti, G.; Federici, M.; Hribal, M. L.; Lauro, D.; Sbraccia, P.; Lauro, R., Defects of the insulin receptor substrate (IRS) system in human metabolic disorders. *Faseb J* **2001,** 15, (12), 2099-111.

66.   Withers, D. J.; Burks, D. J.; Towery, H. H.; Altamuro, S. L.; Flint, C. L.; White, M. F., Irs-2 coordinates Igf-1 receptor-mediated beta-cell development and peripheral insulin signalling. *Nat Genet* **1999,** 23, (1), 32-40.

67.   Kido, Y.; Burks, D. J.; Withers, D.; Bruning, J. C.; Kahn, C. R.; White, M. F.; Accili, D., Tissue-specific insulin resistance in mice with mutations in the insulin receptor, IRS-1, and IRS-2. *J Clin Invest* **2000,** 105, (2), 199-205.

68.   Huang, C.; Thirone, A. C.; Huang, X.; Klip, A., Differential contribution of insulin receptor substrates 1 versus 2 to insulin signaling and glucose uptake in l6 myotubes. *J Biol Chem* **2005,** 280, (19), 19426-35.

69.   Araki, E.; Lipes, M. A.; Patti, M. E.; Bruning, J. C.; Haag, B., 3rd; Johnson, R. S.; Kahn, C. R., Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature* **1994,** 372, (6502), 186-90.

70.   Taniguchi, C. M.; Ueki, K.; Kahn, R., Complementary roles of IRS-1 and IRS-2 in the hepatic regulation of metabolism. *J Clin Invest* **2005,** 115, (3), 718-27.

71.   Withers, D. J.; Gutierrez, J. S.; Towery, H.; Burks, D. J.; Ren, J. M.; Previs, S.; Zhang, Y.; Bernal, D.; Pons, S.; Shulman, G. I.; Bonner-Weir, S.; White, M. F., Disruption of IRS-2 causes type 2 diabetes in mice. *Nature* **1998,** 391, (6670), 900-4.

72.   Previs, S. F.; Withers, D. J.; Ren, J. M.; White, M. F.; Shulman, G. I., Contrasting effects of IRS-1 versus IRS-2 gene disruption on carbohydrate and lipid metabolism in vivo. *J Biol Chem* **2000,** 275, (50), 38990-4.

73.   Ma, Z.; Gibson, S. L.; Byrne, M. A.; Zhang, J.; White, M. F.; Shaw, L. M., Suppression of insulin receptor substrate 1 (IRS-1) promotes mammary tumor metastasis. *Mol Cell Biol* **2006,** 26, (24), 9338-51.

74.   Ogihara, T.; Shin, B. C.; Anai, M.; Katagiri, H.; Inukai, K.; Funaki, M.; Fukushima, Y.; Ishihara, H.; Takata, K.; Kikuchi, M.; Yazaki, Y.; Oka, Y.; Asano, T., Insulin receptor substrate (IRS)-2 is dephosphorylated more rapidly than IRS-1 via its association with phosphatidylinositol 3-kinase in skeletal muscle cells. *J Biol Chem* **1997,** 272, (19), 12868-73.

75.   Lee, Y. H.; Giraud, J.; Davis, R. J.; White, M. F., c-Jun N-terminal kinase (JNK) mediates feedback inhibition of the insulin signaling cascade. *J Biol Chem* **2003,** 278, (5), 2896-902.

76.   Zick, Y., Uncoupling insulin signalling by serine/threonine phosphorylation: a molecular basis for insulin resistance. *Biochem Soc Trans* **2004,** 32, (Pt 5), 812-6.

77.   Herschkovitz, A.; Liu, Y. F.; Ilan, E.; Ronen, D.; Boura-Halfon, S.; Zick, Y., Common inhibitory serine sites phosphorylated by IRS-1 kinases, triggered by insulin and inducers of insulin resistance. *J Biol Chem* **2007,** 282, (25), 18018-27.

78.   Mann, M., Can proteomics retire the Western blot? *J Proteome Res* **2008,** 7, (8), 3065.

79.   Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, (6928), 198-207.

80.   Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988,** 60, (20), 2299-301.

81.   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, (4926), 64-71.

82.   March, R. E., An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry* **1997,** 32, (4), 351-369.

83.   Steen, H.; Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **2004,** 5, (9), 699-711.

84.   Schrader, W.; Klein, H. W., Liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry (LC-FTICR MS): an early overview. *Anal Bioanal Chem* **2004,** 379, (7-8), 1013-24.

85.   Amster, I. J., Fourier transform mass spectrometry. *Journal of Mass Spectrometry* **1996,** 31, (12), 1325-1337.

86.   Scigelova, M.; Makarov, A., Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **2006,** 6 Suppl 2, 16-21.

87.   Marshall AG, H. C., High-resolution mass spectrometers. *Annual Review of Analytical Chemistry* **2008,** 1, 579-599.

88.   Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **2005,** 4, (12), 2010-21.

89.   Zubarev, R.; Mann, M., On the proper use of mass accuracy in proteomics. *Mol Cell Proteomics* **2007,** 6, (3), 377-81.

90.   Gao, J.; Friedrichs, M. S.; Dongre, A. R.; Opiteck, G. J., Guidelines for the routine application of the peptide hits technique. *J Am Soc Mass Spectrom* **2005,** 16, (8), 1231-8.

91.   Florens, L.; Washburn, M. P.; Raine, J. D.; Anthony, R. M.; Grainger, M.; Haynes, J. D.; Moch, J. K.; Muster, N.; Sacci, J. B.; Tabb, D. L.; Witney, A. A.; Wolters, D.; Wu, Y.; Gardner, M. J.; Holder, A. A.; Sinden, R. E.; Yates, J. R.; Carucci, D. J., A proteomic view of the Plasmodium falciparum life cycle. *Nature* **2002,** 419, (6906), 520-6.

92.   Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **2004,** 76, (14), 4193-201.

93.   Rappsilber, J.; Ryder, U.; Lamond, A. I.; Mann, M., Large-scale proteomic analysis of the human spliceosome. *Genome Res* **2002,** 12, (8), 1231-45.

94.   Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M., Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **2005,** 4, (9), 1265-72.

95.   Chelius, D.; Bondarenko, P. V., Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **2002,** 1, (4), 317-23.

96.   Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H., Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem* **2003,** 75, (18), 4818-26.

97.   Staes, A.; Demol, H.; Van Damme, J.; Martens, L.; Vandekerckhove, J.; Gevaert, K., Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-18. *J Proteome Res* **2004,** 3, (4), 786-91.

98.   Yao, X.; Afonso, C.; Fenselau, C., Dissection of proteolytic 18O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates. *J Proteome Res* **2003,** 2, (2), 147-52.

99.   Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999,** 17, (10), 994-9.

100.   Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004,** 3, (12), 1154-69.

101.   Schmidt, A.; Kellermann, J.; Lottspeich, F., A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005,** 5, (1), 4-15.

102.   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, (12), 6591-6.

103.   Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002,** 1, (5), 376-86.

104.   Ong, S. E.; Kratchmarova, I.; Mann, M., Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res* **2003,** 2, (2), 173-81.

105.   Ong, S. E.; Mann, M., A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* **2006,** 1, (6), 2650-60.

106.   Ishihama, Y.; Sato, T.; Tabata, T.; Miyamoto, N.; Sagane, K.; Nagasu, T.; Oda, Y., Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nat Biotechnol* **2005,** 23, (5), 617-21.

107.   Desiderio, D. M.; Kai, M., Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed Mass Spectrom* **1983,** 10, (8), 471-9.

108.   Barr, J. R.; Maggio, V. L.; Patterson, D. G., Jr.; Cooper, G. R.; Henderson, L. O.; Turner, W. E.; Smith, S. J.; Hannon, W. H.; Needham, L. L.; Sampson, E. J., Isotope

dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin Chem* **1996,** 42, (10), 1676-82.

109.   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, (12), 6940-5.

110.   Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J., Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* **2005,** 2, (8), 587-9.

111.   Pratt, J. M.; Simpson, D. M.; Doherty, M. K.; Rivers, J.; Gaskell, S. J.; Beynon, R. J., Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat Protoc* **2006,** 1, (2), 1029-43.

112.   Rivers, J.; Simpson, D. M.; Robertson, D. H.; Gaskell, S. J.; Beynon, R. J., Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol Cell Proteomics* **2007,** 6, (8), 1416-27.

113.   Kito, K.; Ota, K.; Fujita, T.; Ito, T., A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J Proteome Res* **2007,** 6, (2), 792-800.

9 ABBREVIATIONS

Abl	Abelson murine leukemia viral oncogene homolog
AGC	automatic gain control
Akt	=PKB (protein kinase B)
ATP	adenosine 5´-triphosphate
AQUA	absolute quantification
Cbl	Casitas B-lineage lymphoma proto-oncogene
CDIT	culture-derived isotope tag
CID	collisionally induced dissociation
CoA	coenzyme A
Crk	chicken tumor virus no. 10 regulator of kinase
Da	Dalton
Dab	disabled homolog
DOK	downstream of kinase
eIF4E	eukaryotic translation initiation factor 4E
ERK	extracellular signal-regulated kinase
ESI	electrospray ionization
Fes	feline sarcoma virus oncogene homolog
FFAs	free fatty acids
FHA	forkhead associated domain
FOXO1	forkhead box protein O1
FOXA2	forkhead box protein A2
FT-ICR	Fourier transform ion cyclotron resonance
Fyn	tyrosine-protein kinase Fyn
Gab1	Grb2-associated binding protein 1
GAP	GTPase activating protein
GLUT-4	glucose transporter 4
GPR40	G protein-coupled receptor 40
Grb2	growth factor receptor-bound protein 2
GSV	GLUT-4 storage vesicle
HPLC	high performance liquid chromatography
ICAT	isotope-coded affinity tag
ICPL	isotope-coded protein label
iTRAQ	isobaric tag for relative and absolute quantification
IGF1R	insulin-like growth factor 1 receptor
InsR	insulin receptor

IRR	insulin receptor related receptor
IRS	insulin receptor substrate
JAK	janus kinase
JNK	c-Jun NH2-terminal kinase
LC-MS	liquid chromatography mass spectrometry
MALDI	matrix assisted laser desorption/ionization
MAPK	mitogen activated protein kinase
MEK	MAP/ERK kinase
MODY	maturity-onset diabetes of the young
MS	mass spectrometry
MS/MS	tandem mass spectrometry
mTOR	mammalian target of rapamycin
Nck	Nck adaptor protein
O-GlcNAc	O-linked N-acetylglucosamine
PCS	peptide concatenated standard
PDB	protein database
PDK1	phosphoinositide dependent kinase
PH	pleckstrin homology domain
PHAS-I	phosphorylated heat- and acid-stable protein regulated by insulin 1 (= eIF4E-binding protein 1)
PI3K	phosphatidylinositol-3-kinase
PKC	protein kinase C
PLC	phospholipase C
PTB	phosphotyrosine binding domain
PTM	posttranslational modification
pTyr	phosphotyrosine
QconCAT	quantitative concatemer
Raf	Raf protein
Ras	GTPase Ras
ROS	reactive oxygen species
RTK	receptor tyrosine kinase
S6K	S6 kinase
SAP	signaling lymphocyte activation molecule-associated protein
SGK	serum/glucocorticoid-regulated kinase 1
Shc	SH2 domain containing transforming protein C1
SILAC	stable isotope labeling by amino acids in cell culture
Src	Rous sarcoma virus oncogene
SH2	Src homology 2 domain

SH3	Src homology 3 domain
SHP2	SH2 domain containing protein tyrosine phosphatase 2
SNAP23	synaptosomal-associated protein 23
SOCS	suppressor of cytokine signaling
Sos	son of sevenless
STAT	signal transducer and activator of transcription
TNFα	tumor necrosis factor alpha
UDP	uridine 5´-diphosphate
VAMP2	vesicle-associated membrane protein 2
VLDL	very low density lipoprotein
WW	WW domain (2 conserved tryptophanes spaced 20-22 amino acids apart)
XIC	extracted ion chromatogram

SH3	Src homology 3 domain
SHP2	SH2 domain containing protein tyrosine phosphatase 2
SNAP23	synaptosomal-associated protein 23
SOCS	suppressor of cytokine signaling

10 ACKNOWLEDGEMENTS

My achievements build on a large array of support from many sources, and with respect to this I would like to deeply thank the following people:

Prof. Dr. Matthias Mann for highly interesting projects and generous support, for the rewarding opportunity to learn the trade of MS-based proteomics from a leader in the field, for his congenial way of leadership, for his marvelous sense of humor and much more,

Prof. Dr. Elena Conti for being the co-referee for this thesis,

Prof. Dr. Dieter Oesterhelt for his excellent idea of the Absolute SILAC project,

Dr. Hüseyin Besir for sharing his very skilled knowledge on cloning and protein expression with me, as well as for his kind and relaxed way of communication,

Dr. Martin Miller for a collaboration that opened out into a groovy friendship and a terrific time together,

Leonie Waanders for the collaboration on Top-Down SILAC, for her friendship and for forming one of the joyful pillars sustaining a cool time at the MPI,

Dr. Waltraud Schulze for mentoring me during my starting phase in the department,

Dr. Michiel double-Ballermann Mali-Losinj Venice Indianapolis Gardalake Spetses Amsterdam Vermeulen for years of joy and laughter, great friendship as well as valuable professional support,

Dr. Matthias Selbach and Dr. Tizia Bonaldi for the same as Dr. M.d-B.M-L.V.I.G.S.A. Vermeulen,

Other members of the department for creating a very comfortable atmosphere and for their excellent and plentiful professional advice, especially Dr. Boris Macek, Dr. Jesper Olsen, Dr. Johannes Graumann and Peter Bandilla,

My office mates Maxi Hilger, Falk Butter, Dr. Michael Nielsen, Dr. Michiel Vermeulen and Christian Eberl for fruitful discussions and jokes,

My first lady Mag. ♥♀ Rosemarie Buchegger for her love, and for her patience and support for this thesis despite the deprivations it meant for her,

My parents Klaus and Christine Hanke for supporting the many different crossroad-decisions during my professional education, which was preceded by a fantastic childhood and adolescence,

Odense in Denmark for yet another great experience abroad,

Oktoberfest, Isar, Flaucher, Englischer Garten, Nightlife, Augustiner and other essential Munich-features for their constant support in the quest for something like a work-life-balance

11 CURRICULUM VITAE

STEFAN HANKE

DATE OF BIRTH:	MAY 10,1978
PLACE OF BIRTH:	ROSENHEIM
FAMILY STATUS:	UNMARRIED
NATIONALITY:	GERMAN

DISSERTATION

12/2004 – 09/2008	Department of Proteomics and Signal Transduction, Prof. Dr. Matthias Mann, Max-Planck-Institute for Biochemistry, Martinsried. Topic: „*Characterization of signal transduction pathways by MS-based quantitative proteomics*". The first 7 months were conducted at the Center for Experimental Bioinformatics at the University of Southern Denmark in Odense, Denmark. Departmental speaker of PhD-students from 2005-2007

UNIVERSITY EDUCATION

BIOLOGY (DIPLOM)

2001 – 2004	Advanced study period, University of Heidelberg, „Passed with distinction" Major subject: Biochemistry/Biophysics Subsidiary subjects: Molecular biology, Cell biology Diploma thesis: With Prof. Dr. Felix Wieland, Biochemistry-Center, Topic: „*Mechanisms of Hsp-mediated presentation of exogenous antigen*". Biochemical analysis of the activation of immune cells by cross-priming.
2000 - 2001	Abroad studies, Universidad de Granada/Spain
1998 - 2000	Basic study period, University of Bayreuth

INTERNSHIPS

2003	1 week at Hoffmann-La Roche in Basel, Applied Immunology, Center for Medical Genomics. *Specialized lab-course on the work with immune cells.*
2001	2 months at GPC Biotech in Martinsried, Department of Microbiology. *Generation and characterization of conditionally lethal bacterial mutants applying genetic and microbiological methods.*
2000	6 weeks at the German Cancer Research Center (DKFZ), Department Biology of mitosis. *Training in various techniques in molecular biology and biochemistry.*

SOCIAL SERVICE

1997 - 1998	Neurological clinic in Bad Aibling; Occupation in organization and care.

SCHOOL EDUCATION

1988 - 1997	High school (Gymnasium) Bad Aibling
1984 - 1988	Primary school in Bruckmühl-Heufeld

STIPENDS

09/2005 – 06/2008	PhD-scholarship holder of the „Studienstiftung des Deutschen Volkes"
09/2001 – 06/2004	Scholarship holder of the „Studienstiftung des Deutschen Volkes"