Amino Acids (2011) 41:329–341
DOI 10.1007/s00726-010-0669-1

ORIGINAL ARTICLE

Functional analysis of proteins and protein species using shotgun proteomics and linear mathematics

Wolfgang Hoehenwarter · Yanmei Chen · Luis Recuenco-Munoz · Stefanie Wienkoop · Wolfram Weckwerth

Received: 1 April 2010 / Accepted: 16 June 2010 / Published online: 3 July 2010

Abstract Covalent post-translational modification of proteins is the primary modulator of protein function in the cell. It greatly expands the functional potential of the proteome compared to the genome. In the past few years shotgun proteomics-based research, where the proteome is digested into peptides prior to mass spectrometric analysis has been prolific in this area. It has determined the kinetics of tens of thousands of sites of covalent modification on an equally large number of proteins under various biological conditions and uncovered a transiently active regulatory network that extends into diverse branches of cellular physiology. In this review, we discuss this work in light of the concept of protein speciation, which emphasizes the entire post-translationally modified molecule and its interactions and not just the modification site as the functional entity. Sometimes, particularly when considering complex multisite modification, all of the modified molecular species involved in the investigated condition, the protein species must be completely resolved for full understanding. We present a mathematical technique that delivers a good approximation for shotgun proteomics data.

Keywords Mass spectrometry · Proteomics · Protein function · Protein species · PTM

W. Hoehenwarter (✉) · Y. Chen · L. Recuenco-Munoz · S. Wienkoop · W. Weckwerth
Department of Molecular Systems Biology, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria
e-mail: wolfgang.hoehenwarter@univie.ac.at

Proteomes and proteomics: an update

The dynamic total protein complement of a biological system, the proteome, is of enormous complexity and size. A well-defined concept to deal with this situation is the speciation of the proteome (Jungblut et al. 2008). Each and every polypeptide is defined by the sum of its covalent chemical bonds meaning its primary structure and in addition any covalently bonded moieties. Thus, every protein, every expressed member of a multigene family, i.e. protein isoform, and post-translationally modified form of a protein is understood as a unique protein species. It is not clear if every protein species has its own function, the paradigm, however, captures the extent of function in the proteome. Newest insights, however, are that one primary structure can result in multiple secondary structures and conformations and thus possible multiple functions due to bias towards synonymous, slowly translated codons (Zhang et al. 2009).

The meta-analysis of large scale genomics, transcriptomics and proteomics studies and the use of central data repositories such as National Center for Biotechnology Information (NCBI) and The Arabidopsis Information Resource (TAIR) allow a relatively good survey of the proteome of higher eukaryotes (Fig. 1). The lower proteome entries (red) intersect with the upper genome entries (black) on their common axis reflecting the manifold greater complexity of the proteome. The human being, mouse and rat as well as the small flowering plant *Arabidopsis thaliana* all have around 25,000 protein encoding genes (Collins et al. 2004; Gibbs et al. 2004; Kaul et al. 2000; Waterston et al. 2002). Transcriptome analysis of 32 human tissues revealed that only about half of these are expressed at one time (Jongeneel et al. 2005). Transcriptome-based monitoring of more than 10,000 multi-exon

genes in 52 human tissues and cell lines showed that nearly every gene is alternatively spliced (Johnson et al. 2003) and other studies suggest many genes will generate multiple, sometimes up to a hundred or thousand alternative mRNAs (Schmucker et al. 2000; Tress et al. 2008). Therefore, the amount of translated expression products present in a cell at any one time was estimated to be approximately 30,000. More than 200,000 post-translational modifications (PTM) of proteins are reported in the literature (Seo and Lee 2004) and numerous proteomics studies have identified tens to hundreds of modifications of a single protein (Garcia et al. 2007; Larsen et al. 2001; Scheler et al. 1997). The average number of differently processed expressed gene products is estimated at 5–7 in humans (Humphery-Smith 2004), ultimately allowing a conservative estimate that the proteome of the higher eukaryotic cell comprises 150,000 protein species at one time. Others put the number at 1 million (Humphery-Smith 2004). As an afterthought, we note that we have not considered the dynamic range of protein abundance which has been shown to be four to five orders of magnitude in *E. coli* and yeast (around <100 to 2,000,000 copies per cell) (de Godoy et al. 2008; Lu et al. 2007; Usaite et al. 2008) and ten or more in human plasma (Anderson and Anderson 2002).

The current proteomics technologies have had some success in dealing with this complexity. Shotgun proteomics is the most popular application for unbiased, discovery type analysis of the proteome. The proteins are first digested into peptides followed by an online combination of liquid chromatography (LC) for peptide separation and electrospray ionization mass spectrometry (ESI-MS) for the measurement of peptides mass to charge ratios (*m/z*). It has been used to quantify all of the translated expression products of yeast (de Godoy et al. 2008) and major coverage of all of the open reading frames of *Drosophila* (Brunner et al. 2007), mouse (Graumann et al. 2008) and *Arabidopsis* (Baerenfaller et al. 2008) has also been achieved. The large scale identification and quantification of PTMs has had some encouraging results as well. Several groups have reported thousands of phosphorylation sites (Beausoleil et al. 2004; Olsen et al. 2006, 2010; Schmidt et al. 2008; Sugiyama et al. 2008; Villen et al. 2007), acetylation sites (Choudhary et al. 2009) and other modifications (MacCoss et al. 2002; Tsur et al. 2005) and hundreds of ubiquitination sites (Igawa et al. 2009; Maor et al. 2007; Peng et al. 2003). Nevertheless, the proteome-wide mapping of PTMs remains a difficult task because the modifications are transient, modified peptides are often not very abundant and require specific enrichment strategies and their tandem mass spectra (MS/MS spectra) are difficult to interpret.

Two-dimensional gel electrophoresis (2DE), tried and tested for more than 30 years, still provides the best resolution of the proteome on the protein level. It is capable of separating the heterogenous mixture into up to 10,000 components according to their isoelectric points (PI) and molecular weights (M_w) (Klose and Kobalz 1995). The protein species remain intact giving researchers access to the entire molecule. Analysis with liquid chromatography for further separation of the proteins at distinct positions in the gel matrix (2DE spots) and the highly sensitive new generation of mass spectrometry machines has revealed that generally a few protein species co-migrate (Jungblut et al. 2010). This precludes accurate quantification of individual protein species by measuring polypeptide staining intensity; they can, however, be quantified by labeling them with stable isotopes either directly or by covalently attaching chemical moieties and then comparing their mass spectrometric signals (Schulze and Usadel 2010). Despite these technological advances, the maximum entry on the proteomics axis (gray) in Fig. 1 is still more than one order of magnitude below the maximum entry for proteomes highlighting the need for further innovation and that proteomics science will remain challenging for some years to come.

PTMs: switches that control protein function

Post-translationally modified proteins have a vast array of functions in the cell, which are often quite different from those of their unmodified counterparts. Covalent modification of primary structure can alter protein conformation thereby producing new recognition motifs, hiding existing motifs or making previously hidden motifs accessible. PTMs can have allosteric effects, controlling protein function via modification of non-active sites. Initial modification can recruit subsequent additional modifications by either the same or different chemical moieties which may act in concert. Two or more different PTMs can compete for modification of the same residue in protein primary structure. All of these mechanisms affect protein interaction via specific PTM-binding domains which defines function (Seet et al. 2006).

The covalent attachment of phosphate moieties donated by ATP and catalyzed by kinases via esterification of side chain hydroxyl-groups of primarily serine, threonine and tyrosine residues is the most widely studied PTM. It occurs on approximately one-third of eukaryotic proteins and is mostly a monoadduct (Cohen 2000; Hubbard and Cohen 1993). The phosphoryl-group can be removed by phosphatase catalyzed hydrolysis so protein phosphorylation is reversible and in most cases transient. It is a central switch of protein function. It controls the activity of the components of signaling cascades such as the canonical mitogen-associated protein kinase (MAPK) cascade downstream of the tyrosine auto-phosphorylated epidermal growth factor

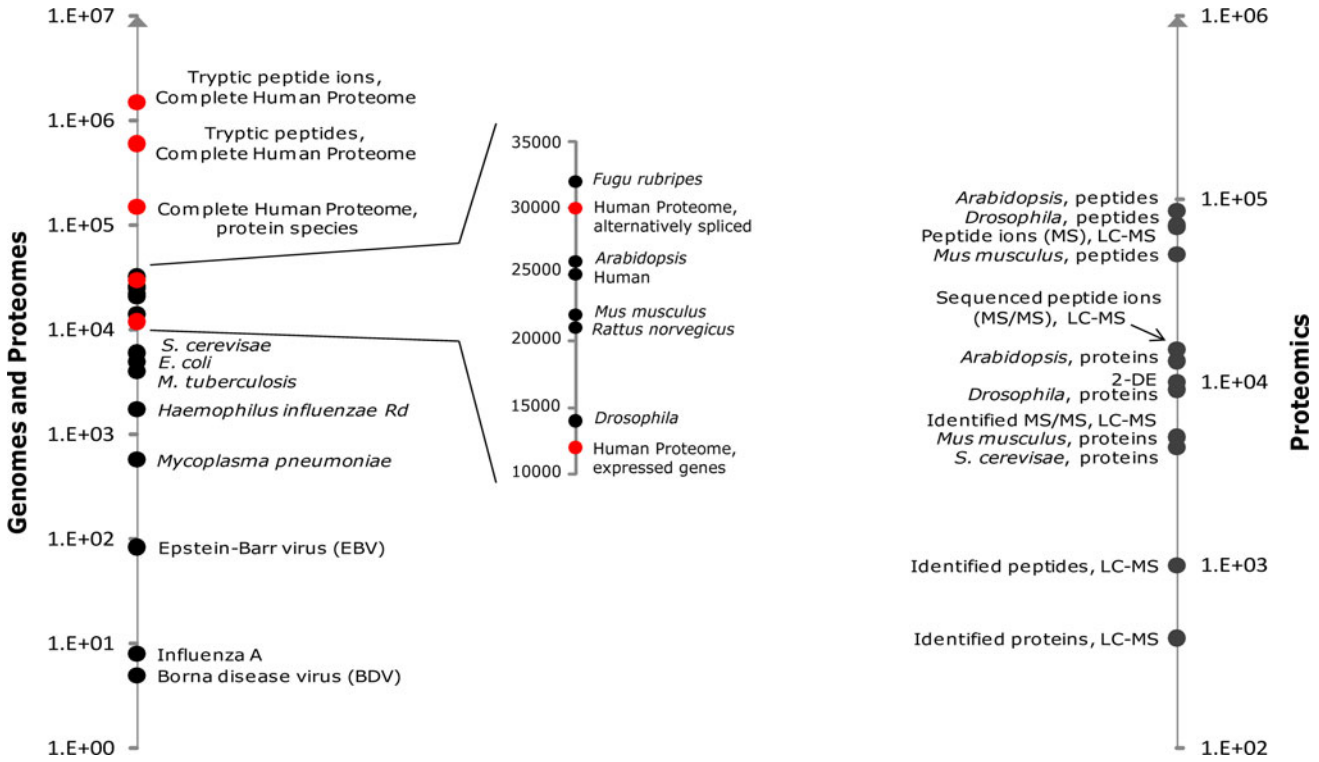

Fig. 1 Some important reference values regarding genomes in *black*, proteomes in *red* and proteomics in *gray*. The axes are in logarithmic scale. The values are from pertinent publications (Adams et al. 2000; Ahn et al. 2007; Aparicio et al. 2002; Baer et al. 1984; Baerenfaller et al. 2008; Bogdanov and Smith 2005; Brunner et al. 2007; Collins et al. 2004; Cox and Mann 2008; Cubitt et al. 1994; de Godoy et al. 2008; Fleischmann et al. 1995; Garcia et al. 2007; Ghaemmaghami et al. 2003; Gibbs et al. 2004; Goffeau et al. 1996; Graumann et al. 2008; Johnson et al. 2003; Jongeneel et al. 2005; Jungblut et al. 2008; Kaul et al. 2000; Klose and Kobalz 1995; Larsen et al. 2001; Misra et al. 2002; Myers et al. 2000; Ono et al. 2006; Scheler et al. 1997; Schmucker et al. 2000; Tress et al. 2008; Waterston et al. 2002) and from our experience. The RNA viruses have very small genomes with less than ten genes; the Epstein–Barr virus has a double-stranded DNA genome of 84 open reading frames. Unicellular organisms all have less than 10,000 genes. Peptide ions (MS), LC–MS refers to all recorded *m/z* in a 1D shotgun proteomics analysis, sequenced peptide ions denotes the total number of recorded MS/MS spectra

(EGF) receptor (Pawson and Nash 2000) that consists of the MAPK kinase kinase (MAPKKK) RAF, the MAPK kinase (MAPKK) MEK/MKK1 and the MAPK ERK (Marshall 1994). It also controls the MAPK cascade MEKK1-MKK4/MKK5-MPK3/MPK6 that administers response to bacterial pathogens in plants (Asai et al. 2002) following perception of the flg22 epitope of bacterial flagellin by the leucine rich repeat receptor kinase (LRR-RK) FLS2 (Chinchilla et al. 2006). MAP kinases are modular as well as promiscuous. This means that cascades can be activated by diverse stimuli, that components of different cascades are interchangeable and act on several substrates to communicate a host of different signals and that they produce a wide range of functional responses (Nakagami et al. 2005). The MAPKs are activated and deactivated by phosphorylation and dephosphorylation at multiple sites which gives rise to a large number of functional protein species whose abundance and very existence is in a high state of flux.

Phosphorylation can determine the translocation of proteins to organelles and subcellular compartments for instance in the case of protein phosphatase-1 (PP1) which is released from glycogen particles to the cytosol upon phosphorylation of one of its subunits, G_M (Hubbard and Cohen 1993). Apoptosis is inhibited by binding and sequestering of the phosphorylated proapototic BAD by 14-3-3 protein (Lizcano et al. 2000). Gene transcription is initiated by phosphorylation of the active subunit of the transcription factor NFκB, p65–p50 (Wang and Baldwin 1998), phosphorylation of members of the Jun, Fos and ATF transcription factor families (Karin et al. 1997; Murphy et al. 2002; Ventura et al. 2003) as well as other DNA-associated proteins. Transcription is inhibited by phosphorylation of HY5, a b-ZIP transcription factor that controls light induced gene expression in plants (Hardtke et al. 2000).

Reversible protein phosphorylation ties in with ubiquitination, and indeed, there is extensive regulatory crosstalk between these two PTMs (Hunter 2007). Protein species such as the inhibitory subunit of NFκB, IκBα, phosphorylated at serine residues 32 and 36 (Karin 1999; Yaron et al. 1998), phosphorylated HIV-1-Vpu (Margottin et al.

1998), cyclins and cyclin-dependent kinase (CDK) inhibitors (Koepp et al. 1999) as well as dephosphorylated β-catenin (Kitagawa et al. 1999) and the anti-apototic Bcl-2 protein (Dimmeler et al. 1999) are targeted for ubiquitination and ultimately degradation by the 26S proteasome. The phosphate moiety and surrounding amino acids, which has generally become known as a phosphodegron, is the recognition signal for the ubiquitin protein ligase E3. The enzyme binds the phosphodegron on the substrate protein via an F-box subunit that contains the WD40 or LRR-binding domains and catalyzes the covalent attachment of ubiquitin to an ε-amino group of an internal lysine residue or to the α-amino group (Breitschopf et al. 1998). Alternatively, the two main types of E3 ubiquitin protein ligases, the single or multisubunit RING finger ligases that lack catalytic activity and function by recruiting the E2-ubiquitin conjugated intermediate and the target protein substrate and the catalytic HECT domain E3 ligases can themselves be controlled by phosphorylation. The anaphase promoting complex type E3 enzyme is phosphorylated conferring substrate specificity and recognition properties (Shteinberg et al. 1999) and the Itch E3 ligase is allosterically activated by phosphorylation (Gallagher et al. 2006).

Protein ubiquitination is reversible through nucleophile attack of the peptide bond between ubiquitin and the substrate catalyzed by deubiquitinating enzymes (Dubs) (Sowa et al. 2009; Ventii and Wilkinson 2008). Reversible ubiquitination controls cellular processes and in particular kinase activity and protein phosphorylation primarily by controlling the stoichiometry of their components via degradation. The regulatory phosphate moiety on the active kinase becomes a phosphodegron that is recognized by the enzymes of the ubiquitination machinery leading to covalent attachment of polyubiquitin moieties branched at lysine residue 48 of ubiquitin followed by degradation by the proteasome. In some cases activated kinases are targeted to the lysosome/vacuole or subcellular compartments, which is mediated by lysine 63 branched polyubiquitination and multiubiquitination (Huang et al. 2006). Polyubiquitin of kinases branched at lysine residues 29 or 33 can directly inhibit kinase activity and kinases can be activated by polyubiquitin promoted transphosphorylation (Chen 2005). In plants, the ubiquitin proteasome system may play a role in sucrose related processes possibly via substrate proteins phosphorylated by Snf-1-related protein kinases (SnRKs) (Ellis et al. 2002). It has also been implicated in activation of the auxin response by degradation of Aux/IAA proteins and in suppression of transcription of light responsive genes by degradation of unphosphorylated HY5 (Ellis et al. 2002).

Acetylation of lysine as well as serine and threonine residues is another PTM that has far reaching impact on cellular functions. The formation of the amide bond between the lysine ε-amino group and the hydroxyl group of the acetate moiety that is donated by Acetyl-CoA is catalyzed by histone acetyltransferases (HAT, also known as lysine acetyltransferases [KAT]). Its hydrolysis is mediated by histone deacetylases (HDAC or in analogy KDAC for lysine acetyltransferase). As for phosphorylation and sumoylation but not for ubiquitination, primary structure consensus sequences have been identified and it has been ascertained that lysine acetylation is favored in ordered secondary structure as well as in macromolecular complexes (Choudhary et al. 2009; Kim et al. 2006) such as the nuclear HAT complexes themselves (Thompson et al. 2004) or the major actin nucleation complex ARP2/3.

There is ample evidence that protein function is regulated by acetylation in conjunction with phosphorylation, ubiquitination and other PTMs. The abundance of phosphorylated protein species is influenced by acetylation exercising control over kinases, for example the phosphoinositide-3-kinase related protein kinases (PIKK) that are integral to DNA damage repair (Jiang et al. 2006; Sun et al. 2005). The activity of CDC2, a kinase involved in cell cycle progression and mitosis, and CDK9 may be abolished by acetylation in their kinase domains (Sabo et al. 2008). Competitive acetylation of serine and threonine residues in the activation loop of MAPKKs is used by *Yersinia* species to shut down signal transduction via the MAPK cascade and overcome the immune response (Mukherjee et al. 2006). Binding of phosphorylated primary structure motifs by the 14-3-3 domain is inhibited by its acetylation (Choudhary et al. 2009) and acetylation and methylation in cis as well as in trans affects the interaction of binding domains and recognition motifs on histones (Fischle et al. 2003; Latham and Dent 2007). Ubiquitination, methylation, SUMOylation and acetylation all compete for modification of lysine residues and thereby modulate the activity of such prominent examples as the tumor suppressor p53, the transcriptional cofactor p300 and the nuclear transport protein RANGAP1 (Bouras et al. 2005; Mahajan et al. 1997; Yang and Seto 2008). Ubiquitin ligases and DUBs are themselves extensively acetylated which may affect their function and the ubiquitination of other proteins (Choudhary et al. 2009).

In summary, phosphorylated, ubiquitinated and acetylated protein species regulate nearly all aspects of cell life (Choudhary et al. 2009; Ciechanover et al. 2000; Cohen 2000). This includes metabolism (Cohen 1999; Kempa et al. 2007; Kim et al. 2006; Polge and Thomas 2007), the cell cycle (Brooks and Gu 2003; Choudhary et al. 2009; Glotzer et al. 1991; Inze and De Veylder 2006; O'Connell et al. 2000; Zhang et al. 2008) and cytokinesis (Takahashi et al. 2004) and growth and death (Haas et al. 1995; Kim et al. 2006). Signal transduction and concurrently the response to infection (Kotlyarov et al. 1999; Rock and

Goldberg 1999), pathogens (Asai et al. 2002; Chisholm et al. 2006; Devoto et al. 2003; Kunkel and Brooks 2002) and abiotic stress (Ichimura et al. 2000; Kempa et al. 2007; Nakagami et al. 2005) is also highly regulated by PTMs.

Reversible, PTM at multiple sites by different moieties produces a tremendous amount of transiently abundant protein species that are the primary conveyors of function in the cell. These protein species greatly amplify the functional repertoire inherent in the proteome beyond the unmodified expressed genome. Indeed, the unmodified proteins and protein complexes may be seen as the molecular machinery, the PTM as the switches used to operate it.

Shotgun proteomics unravels the cellular signaling network

The function of phosphorylated protein species was traditionally investigated with reverse genetic approaches combined with enzyme activity and protein interaction assays, particularly in the case of signaling cascades and the phosphorylation status of their components (Ahlfors et al. 2004; Asai et al. 2002; Chinchilla et al. 2007; Meskiene et al., 1998, 2003; Nuhse et al. 2000; Schweighofer et al. 2007; Takahashi et al. 2007). Recent advances in shotgun proteomics have made the detection and quantification of site-specific phosphorylation on a proteome-wide scale feasible (Amanchy et al. 2005; Beausoleil et al. 2004; Benschop et al. 2007; Chen et al. 2010; Chi et al. 2007; Nuhse et al. 2004, 2007; Olsen et al. 2006, 2010; Schmidt et al. 2008; Schulze 2010; Sugiyama et al. 2008; Thelemann et al. 2005; van Bentem and Hirt 2007; Villen et al. 2007).

A landmark is the study of the phosphorylation events following EGF stimulation of HeLa cells that employed subcellular fractionation to analyze the nuclear and cytosolic protein complement, strong cation exchange and titanium dioxide affinity chromatography to enrich phosphorylated peptides and stable isotope labeling (SILAC) and shotgun proteomics to identify and quantify the phosphorylation of amino acid residues over time (Olsen et al. 2006). It describes the kinetics of 6,600 phosphorylation sites mapped to 2,244 proteins, around 20% of the expressed open reading frames, and of EGF signaling in unprecedented detail, including the entire MAPK cascade, a large number of transcription factors and associated proteins that were not known to be involved in growth factor signaling, cytoskeletal proteins such as actin and GTP-associated proteins, the ubiquitination machinery and RNA-binding proteins. Phosphorylation at multiple sites in many cases with different kinetics was shown to predominate and to regulate protein function as exemplified by the early, activating tyrosine autophosphorylation of the EGF receptor that declined concomitantly with later phosphorylation of serine and threonine residues, which is known to attenuate the signal via negative feedback (Schlessinger 2000).

Two studies of the plasma membrane-associated proteome that applied similar methodologies as the work on the EGF receptor describe the early response to perception of flg22 in *Arabidopsis* cell culture (Benschop et al. 2007; Nuhse et al. 2007). Like EGF receptor signaling it is hallmarked by transient protein phosphorylation. Benschop et al. (2007) demonstrated that the measured changes in the phosphorylation levels of amino acid residues were due to reaction kinetics and not changes in protein abundance by normalizing the abundance ratios of phosphorylated peptides to peptides that did not contain a phosphorylation site on the same protein. Both studies identified phosphorylation sites on numerous receptor-like kinases (RLK), implicating phosphorylation in sensitizing the plant to further pathogen challenge and priming of defense response (Conrath et al. 2002; Zipfel et al. 2004). Phosphorylation of the components of the MAPK cascade, regulatory protein phosphatases such as PP2C (Schweighofer et al. 2007) and auxin signaling proteins was also found. The latter connects MAPKs with the down regulation of auxin signaling which is known to play a role in the plant immune response (Kovtun et al. 1998; Navarro et al. 2006).

The induced phosphorylation of many sites on a single protein was also reported by both authors. Differential kinetics of several residues of the H^+-ATPases AHA1 and AHA2 and reduction of phosphorylation of the penultimate threonine residue 948 which directly controls protein activity (Palmgren 2001) was detected following treatment with flg22. More pronounced, the respiratory oxidative burst protein RBOHD, an NADPH oxidase involved in the production of reactive oxygen intermediates in the apoplast following pathogen perception and integral to defense response was found to be differentially phosphorylated at up to seven distinct residues. This suggests the protein is controlled by various interacting partners and that it administers distinct functions in manifold physiological pathways in a complex regulatory network.

The widespread incidence and far reaching affects of acetylation in all cellular compartments was discovered for the first time by two shotgun proteomics studies (Choudhary et al. 2009; Kim et al. 2006). The more recent work describes the kinetics of 3,600 sites on 1,750 proteins in three human cell lines including one affected by acute myeloid leukemia in response to two KDAC inhibitors, one of which (SAHA) is in clinical use. The studies show that multisite acetylation of proteins is a central regulator of all nuclear and DNA-associated processes. This includes chromatin remodeling and DNA replication, transcription,

splicing, DNA damage repair and nuclear transport. They detected the hitherto unknown acetylation of numerous cytoplasmic metabolic enzymes including aldolase which also binds actin and provide evidence for a major role of acetylated proteins in cytoskeleton architecture as actin-, micro- and intermediate filament proteins were all functionally acetylated (Anastasiadis et al. 2000; Posern et al. 2004; Zhang et al. 2007). They also showed that the modification is prominent in the mitochondrion and uncovered a strong link to its regulation of energy metabolism (Kim et al. 2006). Many of the proteins of the TCA cycle, oxidative phosphorylation, lipid, carbohydrate, amino acid and nucleotide metabolism as well as the mitochondrial dehydrogenase protein complexes were found to be acetylated. The activity of the dehydrogenase complexes is controlled by NAD^+ to NADH ratios and Acetyl-CoA levels which also serve as substrates or cofactors for HATs and HDACs (Blander and Guarente 2004) suggesting a regulatory feedback loop.

Acetylation of multiple lysine residues of the same protein was prevalent like in the studies of phosphorylation. The herpesvirus-associated ubiquitin-specific protease (HAUSP), a DUB that influences the nuclear cytoplasmic partitioning of the tumor suppressor PTEN via deubiquitination was acetylated at five different residues. The cytoskeletal protein cortactin was modified at 7 sites and 14 acetyl-lysine residues were identified for the HSP90-α subunit. This further indicates that intricate combinatorial switches that consist of reversible modification at distinct residues by different interchangeable moieties are common controls of protein function.

The largest work on PTM sites to date was published recently. It defines the kinetics of over 20,000 phosphorylation sites on 6,000 proteins in the cell cycle and highlights what is possible with mass spectrometry-based proteomics today (Olsen et al. 2010). While not yet fully comprehensive, shotgun proteomics has provided substantial information on the in vivo post-translationally modified proteome, recording the state of thousands of residues on thousands of proteins under various conditions. Shotgun proteomics studies have elucidated an extensive regulatory network based on the interactions of transiently abundant highly modified protein species that is universally active in the cell and given a global perspective on the function and profound importance of PTMs.

Shotgun proteomics and protein species function

The measurements of the state of modification of individual amino acid residues reflect the effects of the experimental condition under investigation such as exposure to a growth factor or temperature stress. In most cases they will also reflect the effects on individual proteins. An example is the differential phosphorylation of the EGF receptor described above. The observed kinetics are in agreement with the known changes in the phosphorylation state of several residues in response to EGF perception. In cases of more complex multisite modification, such as the differential phosphorylation of the RBOHD protein at seven or the acetylation of HSP90-α at 14 different residues, the effects of the experimental condition on the protein may not be so clear. Although the induced changes in the modification state at multiple sites can undoubtedly be brought into causal connotation with the condition, it is not straightforward to assay if all of the sites are truly localized on one protein, or more precisely on one protein species, or are shared by several protein species that administer function in concert.

This point must not be ignored. As we know, the polypeptide molecule which has its conformation defined by primary structure and which in many cases is a subunit of higher order macromolecular assembly is the basic functional entity and not the peptide or amino acid residue (we will not consider the point of multiple secondary structures with the same primary structure mentioned in the introduction).To truly pinpoint function in some cases it may be necessary to have the knowledge of the full primary structure of all polypeptides involved in a physiological process in addition to modification kinetics. This can be achieved experimentally by separating the proteome into its protein species using top-down technologies such as 2-DE, or, which may be the next true breakthrough in proteomics research, top-down LC-ESI-MS proteomics (Chait 2006; Siuti and Kelleher 2007).

Mathematical procedures are very powerful and can give good approximations of protein species function from shotgun proteomics data. The proteome is digested into peptides so the protein species are not separated prior to analysis. Nevertheless, we routinely use the variance and covariance of the abundance of the identified proteins, which is inferred by the measured abundance of the peptides assigned to them by comparative search of translated nucleic acid sequences or de novo spectral interpretation to explore a multitude of biological questions. We do this primarily with a combination of the canonical principal component analysis (PCA) (Pearson 1901) and independent component analysis (ICA) (Comon 1994). These techniques make it possible to determine correlations between the shotgun proteomics analyses and the phenotypes under different experimental conditions (Scholz et al. 2004).

Data consisting of multiple observations of a large number of variables can be seen as a number of vectors or points in space. PCA and ICA compress and project this data into a lower number of dimensions, in a direction so

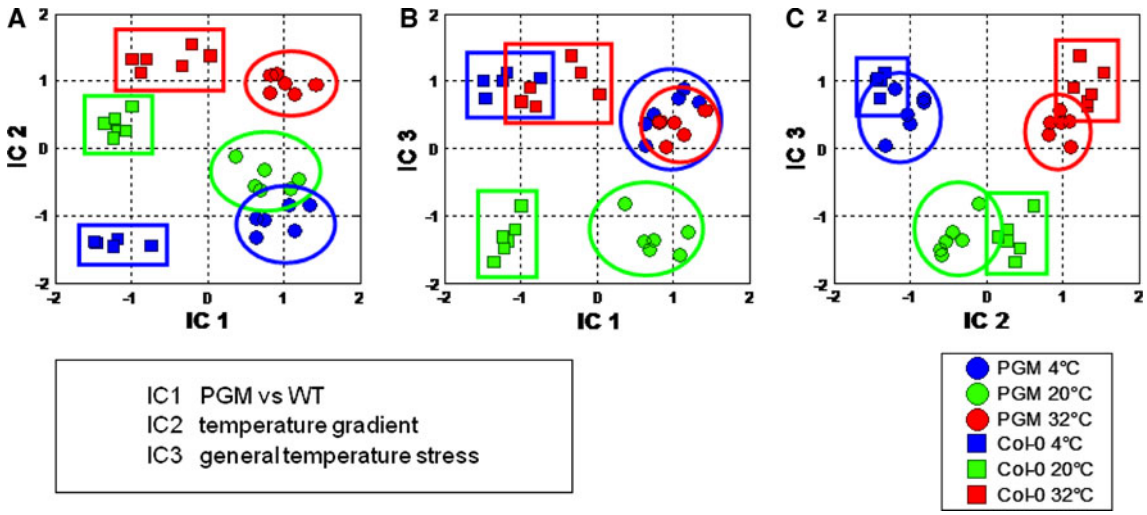

Fig. 2 Visualization of shotgun proteomics data with ICA. Thirty-six combined shotgun proteomics and metabolomics analyses of two genotypes of *Arabidopsis thaliana*, the wild type Col-0 and a starch-deficient mutant PGM, adapted to three temperatures, 4, 20 and 32°C, for 3 days, are plotted in three 2-dimensional plots (**a**–**c**) of an optimal three dimensional space using combined principal and independent component analysis (PCA/ICA). Four principal components from the PCA which was applied first were used as the input for subsequent ICA. The first dimension, IC1 contains information on the proteins that are discriminatory for the genotypes, the second, IC2, on proteins indicative of temperature adaption, the third, IC3, on proteins expressed under temperature stress. The sample pattern in all three dimensions, i.e. the complete data structure nearly perfectly reflects the applied experimental conditions. The figure was taken from Wienkoop et al. (2008) and is reprinted here with permission

that the maximum amount of information is retained, the principal or independent components (PC or IC). The data is visualized in up to three dimensions and the user can judge structure and inherent patterns that can shed light on relationships that may be interesting. The PC or ICs themselves can be expressed in terms of the original variables making it clear which parts of the original data are prominent and determine the observed structure and relationships. These can then be interpreted in a biological context.

In more mathematical terms, PCA can be explained as an eigenvalue decomposition of the covariance matrix. The principal components are the orthonormal eigenvectors that indicate the directions of maximal variance in decreasing order of their eigenvalues and span the lower dimensional space of linear combinations of the original variables for mapping with the minimum reconstruction error. Their coordinates in the original higher dimensional space are the loadings that place a scalar value on the contribution of the variables, in our case the identified proteins, to the data structure. ICA is an extension of PCA where the components are independent, a stronger condition than uncorrelation in PCA, and no longer restricted to orthonormality. The independent components (IC) are sorted according to their kurtosis, the fourth auto-cumulant of a distribution of parameter values following the mean, variance/standard deviation and skewness and which is an indicator of its Gaussian fit. A sub-Gaussian or flat distribution can be the result of a single or few significant differences in parameter values and is therefore often of biological interest, i.e. a significant difference in protein abundance under two or more conditions. As in PCA, the loadings are the coordinates of the independent components in the higher dimensional space.

Figure 2 taken from a recent study of ours (Wienkoop et al. 2008), shows a plot of 36 combined shotgun proteomics and metabolomics analyses of two *A. thaliana* genotypes, the Col-0 wild type and a plastidic PGM mutant, under cold (4°C), normal ambient (20°C) and heat (32°C) conditions reduced to three dimensions using PCA/ICA. Figure 2a, b both contain the first dimension, IC1, and looking at the data from this angle, it is clear that IC1 separates the two investigated genotypes. Figure 2a additionally shows the second dimension, IC2, of the compressed data which separates the applied temperatures in a gradient that goes from low to high. Figure 2b contains the third dimension, IC3, in addition to the first, which discriminates the temperature stressed conditions of 4 and 32°C from the normal environmental condition of 20°C. Figure 2c combines the second and third dimensions; the former again shows the gradient from low to high temperature, the latter again distinguishes the stressed conditions from the normal ambient temperature. These two independent components do not contain any information regarding the genotypes so they are congruent.

The plotted reduced data contains 95% of the extent (to be mathematically precise, the variance) and also 95% of the information of the original data. Therefore, the

correlation between the shotgun proteomics analyses which are fully quantitative and the investigated phenotypes is as equally nearly perfect as the correlation between the lower dimensional data and the phenotypes as seen in the plot. This means that the analyses can be used to extrapolate function. It can also be shown analytically, that there is a fundamental relationship between the covariance of the measured proteomics data and the fluctuating concentrations of protein species in physiological pathways which can be expressed as entries in the Jacobian matrix (Steuer et al. 2003). In the given example low-temperature induced protein 78 and cold-regulated protein COR6.6 (KIN2) as well as several RNA-binding proteins were shown to regulate the plants' adaption to temperature (Kim et al. 2005; Smallwood and Bowles 2002).

To rapidly determine the sites of functional PTMs, we have developed an application for comparison of hundreds of shotgun proteomics analyses called mass accuracy precursor alignment (MAPA) (Hoehenwarter et al. 2008). It works by measuring the *m/z* of peptides with very high accuracy (average error ~1.5 ppm, SD ~1 ppm) which is a unique identifier and allows the abundance of every measured peptide in any number of shotgun proteomics analyses to be aligned in a quantification matrix. It is done with a program we developed called ProtMAX. It places the *m/z* ratios of all peptides ions for which at least one MS/MS spectrum is available in all of the compared analyses in the rows, the analysis identifiers in the columns and the number of MS/MS spectra for each *m/z* ratio in each analysis in the cells. The number of MS/MS spectra is the spectral count of each peptide ion which correlates linearly with peptide abundance over two orders of magnitude (Abdi et al. 2006; Old et al. 2005; Wienkoop et al. 2006) so the matrix contains a quantitative value for each peptide in each analysis. The strength of MAPA is that it produces an accurate quantitative comparison of any number of shotgun proteomics measurements in minimal time with minimal computing power.

PCA/ICA as well as clustering or supervised classification schemes can be employed to find the peptides that are especially characteristic of the experimental conditions. The quantification matrix is unbiased. It is inclusive of all of the measured peptides and does not omit post-translationally modified peptides that are difficult to identify by searching databases of translated DNA sequences. Therefore, post-translationally modified peptides that are correlated with the different experimental conditions can be readily detected and then identified by de novo spectral interpretation.

We employed this strategy to explore the processes involved in hormone signaling in *A. thaliana* cell culture, 1, 3 and 6 h after exposure to the phytohormones abscisic acid (ABA), gibberellic acid (GA), auxin (IAA), jasmonate (JA) and kinetin (Chen et al. 2010). In total, 152 phosphorylated peptides were identified and quantified using mass spectrometry. These peptides contained 170 phosphorylation sites. They were all differentially responsive to at least one of the hormones and could be mapped to 130 proteins. Many of the sites were induced by several hormones which is indicative of activated protein species transmitting signals via several pathways in the regulatory network. The abundance of the phosphorylated peptides at all of the time points after exposure to the phytohormones was used to produce the MAPA quantification matrix. This quantitative data was then used to model the phosphorylation kinetics with the combined PCA/ICA.

The activation of MAPK and calcium-dependent kinase (CDPK) cascades in response to phytohormone perception is well known (Alonso and Stepanova 2004; Kovtun et al. 1998; Ludwig et al. 2005; Navarro et al. 2006; Schweighofer et al. 2007; Takahashi et al. 2007). As a small example from the study, Fig. 3 shows the phosphorylation state of four sites on the central modules of cell signaling from the membrane to the nucleus 3 h after treatment with each of the hormones. The RLK is phosphorylated in response to the continued perception of IAA, JA and kinetin. The cytosolic CDPK is induced under all of the conditions which can be expected in light of its promiscuous nature. The exclusive phosphorylation of the ABA responsive element-binding factors (AREB) in response to ABA indicates the signal has been transmitted through the network and reached its highly specific destination. The sites on the AREBs had not been described previously showing the power of MAPA and PCA/ICA to uncover novel regulatory modifications because of their causal and potentially functional correlation to the phenotype.

Concluding remarks

The proteome is vastly more complex and dynamic than the genome. PTM in particular expand its functional potential immensely. Shotgun proteomics has measured the modification kinetics of a very large number of amino acid residues on an equally substantial number of proteins and uncovered a regulatory network that extends into all aspects of cellular physiology.

Transient modifications of multiple sites by different interchangeable moieties are recognized by PTM-binding domains that are organized as modules in proteins (Seet et al. 2006; Yang 2005). The dynamic interaction of recognition motifs and binding domains executes a powerful regulatory program that is flexible, sensitive and far reaching. These insights, where the modular-binding domains "read" the state of modification of proteins that

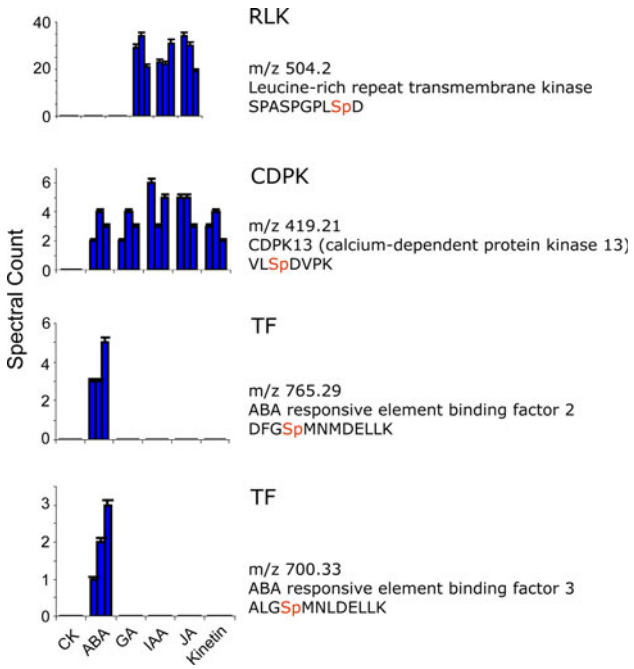

Fig. 3 Phosphorylation state of key components of the signal transduction cascade 3 h after induction with five phytohormones in *Arabidopsis thaliana* cell culture. The phosphorylation kinetics of four serine residues on a receptor-like kinase (RLK), a CDPK and transcription factors (TF) were determined with MAPA. The *m/z* and peptide primary structure are shown, the phosphorylation site is colored *red*

controls function are expressed in the emerging concept of the protein code (Gimona 2006; Sims and Reinberg 2008). Considering the large number of known PTMs, the common phenomenon of dynamic multisite modification and the equally large number of interaction domains, this "protein code", if one wishes to use the term, may be much more complex than the genetic code and constitute a primary functional dimension in the molecular biology of the cell.

To truly understand the mechanisms of molecular interaction and function in some cases it may be necessary to resolve the proteome on a molecular level, the protein species level. Shotgun proteomics studies in conjunction with higher mathematical procedures can give good approximations of protein species function. Ideally they should be employed to gain a more or less comprehensive overview of the question at hand. They should be complemented with top-down technologies such as 2-DE to focus on the protein species of interest and genetic and biochemical techniques to validate the functional hypothesis.

Acknowledgments We thank Peter R Jungblut for his kind solicitation of the manuscript and for the many years of stimulating discussion. WH is generously supported by EraNet Plant Genomics.

References

Abdi F, Quinn JF, Jankovic J, McIntosh M, Leverenz JB, Peskind E, Nixon R, Nutt J, Chung K, Zabetian C et al (2006) Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders. J Alzheimers Dis 9:293–348

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF et al (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195

Ahlfors R, Macioszek V, Rudd J, Brosche M, Schlichting R, Scheel D, Kangasjarvi J (2004) Stress hormone-independent activation and nuclear translocation of mitogen-activated protein kinases in Arabidopsis thaliana during ozone exposure. Plant J 40:512–522. doi:10.1111/j.1365-313X.2004.02229.x

Ahn NG, Shabb JB, Old WM, Resing KA (2007) Achieving in-depth proteomics profiling by mass spectrometry. Acs Chem Biol 2:39–52

Alonso JM, Stepanova AN (2004) The ethylene signaling pathway. Science 306:1513–1515. doi:10.1126/science.1104812

Amanchy R, Kalume DE, Iwahori A, Zhong J, Pandey A (2005) Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). J Proteome Res 4:1661–1671. doi:10.1021/Pr050134h

Anastasiadis PZ, Moon SY, Thoreson MA, Mariner DJ, Crawford HC, Zheng Y, Reynolds AB (2000) Inhibition of RhoA by p120 catenin. Nat Cell Biol 2:637–644. doi:10.1038/35023588

Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1:845–867

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A et al (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297:1301–1310. doi:10.1126/science.1072104

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J (2002) MAP kinase signalling cascade in Arabidopsis innate immunity. Nature 415:977–983. doi:10.1038/415977a

Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, Hatfull G, Hudson GS, Satchwell SC, Seguin C et al (1984) DNA-sequence and expression of the B95–8 Epstein–Barr virus genome. Nature 310:207–211

Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S (2008) Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. Science 320:938–941

Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li JX, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci USA 101:12130–12135. doi:10.1073/pnas.0404720101

Benschop JJ, Mohammed S, O'Flaherty M, Heck AJR, Slijper M, Menke FLH (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol Cell Proteomics 6:1198–1214. doi:10.1074/mcp.M600429-MCP200

Blander G, Guarente L (2004) The Sir2 family of protein deacetylases. Annu Rev Biochem 73:417–435. doi:10.1146/annurev.biochem.73.011303.073651

Bogdanov B, Smith RD (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. Mass Spectrom Rev 24:168–200. doi:10.1002/mas.20015

Bouras T, Fu M, Sauve AA, Wang F, Quong AA, Perkins ND, Hay RT, Gu W, Pestell RG (2005) SIRT1 deacetylation and

repression of p300 involves lysine residues 1020/1024 within the cell cycle regulatory domain 1. J Biol Chem 280:10264–10276. doi:10.1074/jbc.M408748200

Breitschopf K, Bengal E, Ziv T, Admon A, Ciechanover A (1998) A novel site for ubiquitination: the N-terminal residue, and not internal lysines of MyoD, is essential for conjugation and degradation of the protein. EMBO J 17:5964–5973. doi:10.1093/emboj/17.20.5964

Brooks CL, Gu W (2003) Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. Curr Opin Cell Biol 15:164–171

Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, Deutsch EW, Panse C, de Lichtenberg U, Rinner O et al (2007) A high-quality catalog of the Drosophila melanogaster proteome. Nature Biotechnology 25:576–583

Chait BT, Chait BT (2006) Chemistry. Mass spectrometry: bottom-up or top-down? Science 314:65–66. doi:10.1126/science.1133987

Chen ZJ (2005) Ubiquitin signalling in the NF-kappaB pathway. Nat Cell Biol 7:758–765. doi:10.1038/ncb0805-758

Chen Y, Hoehenwarter W, Weckwerth W (2010) Comparative analysis of phytohormone—responsive phosphoproteins in *Arabidopsis thaliana* using TiO_2 phosphopeptide enrichment and MAPA. Plant J

Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JEP, Bai DL, Shabanowitz J, Burke DJ, Troyanskaya OG, Hunt DF (2007) Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci USA 104:2193–2198. doi:10.1073/pnas.0607084104

Chinchilla D, Bauer Z, Regenass M, Boller T, Felix G (2006) The Arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. Plant Cell 18:465–476. doi:10.1105/tpc.105.036574

Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nurnberger T, Jones JDG, Felix G, Boller T (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature 448:497. doi:10.1038/Nature05999

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124:803–814. doi:10.1016/j.cell.2006.02.008

Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science 325:834–840. doi:10.1126/science.1175371

Ciechanover A, Orian A, Schwartz AL (2000) Ubiquitin-mediated proteolysis: biological regulation via destruction. Bioessays 22:442–451

Cohen P (1999) The Croonian Lecture (1999). Identification of a protein kinase cascade of major importance in insulin signal transduction. Philos T Roy Soc B 354:485–495

Cohen P (2000) The regulation of protein function by multisite phosphorylation—a 25 year update. Trends Biochem Sci 25:596–601

Collins FS, Lander ES, Rogers J, Waterston RH, Conso IHGS (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945. doi:10.1038/Nature03001

Comon P (1994) Independent component analysis, a new concept. Signal Process 36:287–314

Conrath U, Pieterse CMJ, Mauch-Mani B (2002) Priming in plant-pathogen interactions. Trends Plant Sci 7:210–216

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372

Cubitt B, Oldstone C, Delatorre JC (1994) Sequence and genome organization of borna-disease virus. J Virol 68:1382–1396

de Godoy LMF, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, Walther TC, Mann M (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455:1251–1260

Devoto A, Muskett PR, Shirasu K (2003) Role of ubiquitination in the regulation of plant defence against pathogens. Curr Opin Plant Biol 6:307–311. doi:10.1016/S1369-5266(03)00060-8

Dimmeler S, Breitschopf K, Haendeler J, Zeiher AM (1999) Dephosphorylation targets Bcl-2 for ubiquitin-dependent degradation: a link between the apoptosome and the proteasome pathway. J Exp Med 189:1815–1822

Ellis C, Turner JG, Devoto A (2002) Protein complexes mediate signalling in plant responses to hormones, light, sucrose and pathogens. Plant Mol Biol 50:971–980

Fischle W, Wang Y, Allis CD (2003) Binary switches and modification cassettes in histone biology and beyond. Nature 425:475–479. doi:10.1038/nature02017

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512

Gallagher E, Gao M, Liu YC, Karin M (2006) Activation of the E3 ubiquitin ligase itch through a phosphorylation-induced conformational change. Proc Natl Acad Sci USA 103:1717–1722. doi:10.1073/pnas.0510664103

Garcia BA, Pesavento JJ, Mizzen CA, Kelleher NL (2007) Pervasive combinatorial modification of histone H3 in human cells. Nat Methods 4:487–489. doi:10.1038/Nmeth1052

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. Nature 425:737–741. doi:10.1038/nature02046

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE et al (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521. doi:10.1038/Nature02426

Gimona M (2006) Protein linguistics—a grammar for modular protein assembly? Nat Rev Mol Cell Biol 7:68–73. doi:10.1038/nrm1785

Glotzer M, Murray AW, Kirschner MW (1991) Cyclin is degraded by the ubiquitin pathway. Nature 349:132–138

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M et al (1996) Life with 6000 genes. Science 274:546, 563–567

Graumann J, Hubner NC, Kim JB, Ko K, Moser M, Kumar C, Cox J, Scholer H, Mann M (2008) Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. Mol Cell Proteomics 7:672–683

Haas AL, Baboshina O, Williams B, Schwartz LM (1995) Coordinated induction of the ubiquitin conjugation pathway accompanies the developmentally programmed death of insect skeletal-muscle. J Biol Chem 270:9407–9412

Hardtke CS, Gohda K, Osterlund MT, Oyama T, Okada K, Deng XW (2000) HY5 stability and activity in Arabidopsis is regulated by phosphorylation in its COP1 binding domain. EMBO J 19:4997–5006

Hoehenwarter W, van Dongen JT, Wienkoop S, Steinfath M, Humme J, Erban A, Sulpice R, Regierer B, Kopka J, Geigenberger P et al (2008) A rapid approach for phenotype-screening and database independent detection of cSNP/protein polymorphism using mass accuracy precursor alignment. Proteomics 8:4214–4225

Huang F, Kirkpatrick D, Jiang X, Gygi S, Sorkin A (2006) Differential regulation of EGF receptor internalization and

degradation by multiubiquitination within the kinase domain. Mol Cell 21:737–748. doi:10.1016/j.molcel.2006.02.018
Hubbard MJ, Cohen P (1993) On target with a new mechanism for the regulation of protein phosphorylation. Trends Biochem Sci 18:172–177
Humphery-Smith I (2004) A human proteome project with a beginning and an end. Proteomics 4:2519–2521. doi:10.1002/pmic.200400866
Hunter T (2007) The age of crosstalk: phosphorylation, ubiquitination, and beyond. Mol Cell 28:730–738. doi:10.1016/j.molcel.2007.11.019
Ichimura K, Mizoguchi T, Yoshida R, Yuasa T, Shinozaki K (2000) Various abiotic stresses vapidly activate Arabidopsis MAP kinases ATMPK4 and ATMPK6. Plant J 24:655–665
Igawa T, Fujiwara M, Takahashi H, Sawasaki T, Endo Y, Seki M, Shinozaki K, Fukao Y, Yanagawa Y (2009) Isolation and identification of ubiquitin-related proteins from Arabidopsis seedlings. J Exp Bot 60:3067–3073. doi:10.1093/jxb/erp134
Inze D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet 40:77–105. doi:10.1146/annurev.genet.40.110405.090431
Jiang X, Sun Y, Chen S, Roy K, Price BD (2006) The FATC domains of PIKK proteins are functionally equivalent and participate in the Tip60-dependent activation of DNA-PKcs and ATM. J Biol Chem 281:15741–15746. doi:10.1074/jbc.M513172200
Johnson JM, Castle J, Garrett-Engele P, Kan ZY, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302:2141–2144
Jongeneel CV, Delorenzi M, Iseli C, Zhou DX, Haudenschild CD, Khrebtukova I, Kuznetsov D, Stevenson BJ, Strausberg RL, Simpson AJG et al (2005) An atlas of human gene expression from massively parallel signature sequencing (MPSS). Genome Res 15:1007–1014
Jungblut PR, Holzhutter HG, Apweiler R, Schluter H (2008) The speciation of the proteome. Chem Cent J 2:16
Jungblut PR, Schiele F, Zimny-Arndt U, Ackermann R, Schmid M, Lange S, Stein R, Pleissner KP (2010) Helicobacter pylori proteomics by 2-DE/MS, 1-DE-LC/MS and functional data mining. Proteomics 10:182–193. doi:10.1002/pmic.200900361
Karin M (1999) How NF-kappaB is activated: the role of the IkappaB kinase (IKK) complex. Oncogene 18:6867–6874. doi:10.1038/sj.onc.1203219
Karin M, Liu Z, Zandi E (1997) AP-1 function and regulation. Curr Opin Cell Biol 9:240–246
Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin XY et al (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815
Kempa S, Rozhon W, Samaj J, Erban A, Baluska F, Becker T, Haselmayer J, Schleiff E, Kopka J, Hirt H et al (2007) A plastid-localized glycogen synthase kinase 3 modulates stress tolerance and carbohydrate metabolism. Plant J 49:1076–1090. doi:10.1111/j.1365-313X.2006.03025.x
Kim YO, Kim JS, Kang H (2005) Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana*. Plant J 42:890–900. doi:10.1111/j.1365-313X.2005.02420.x
Kim SC, Sprung R, Chen Y, Xu Y, Ball H, Pei J, Cheng T, Kho Y, Xiao H, Xiao L et al (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Mol Cell 23:607–618. doi:10.1016/j.molcel.2006.06.026
Kitagawa M, Hatakeyama S, Shirane M, Matsumoto M, Ishida N, Hattori K, Nakamichi I, Kikuchi A, Nakayama K (1999) An F-box protein, FWD1, mediates ubiquitin-dependent proteolysis of beta-catenin. EMBO J 18:2401–2410. doi:10.1093/emboj/18.9.2401
Klose J, Kobalz U (1995) 2-Dimensional electrophoresis of proteins—an updated protocol and implications for a functional-analysis of the genome. Electrophoresis 16:1034–1059
Koepp DM, Harper JW, Elledge SJ (1999) How the cyclin became a cyclin: regulated proteolysis in the cell cycle. Cell 97:431–434
Kotlyarov A, Neininger A, Schubert C, Eckert R, Birchmeier C, Volk HD, Gaestel M (1999) MAPKAP kinase 2 is essential for LPS-induced TNF-alpha biosynthesis. Nat Cell Biol 1:94–97
Kovtun Y, Chiu WL, Zeng W, Sheen J (1998) Suppression of auxin signal transduction by a MAPK cascade in higher plants. Nature 395:716–720. doi:10.1038/27240
Kunkel BN, Brooks DM (2002) Cross talk between signaling pathways in pathogen defense. Curr Opin Plant Biol 5:325–331
Larsen MR, Larsen PM, Fey SJ, Roepstorff P (2001) Characterization of differently processed forms of enolase 2 from *Saccharomyces cerevisiae* by two-dimensional gel electrophoresis and mass spectrometry. Electrophoresis 22:566–575
Latham JA, Dent SY (2007) Cross-regulation of histone modifications. Nat Struct Mol Biol 14:1017–1024. doi:10.1038/nsmb1307
Lizcano JM, Morrice N, Cohen P (2000) Regulation of BAD by cAMP-dependent protein kinase is mediated via phosphorylation of a novel site, Ser155. Biochem J 349:547–557
Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 25:117–124. doi:10.1038/Nbt1270
Ludwig AA, Saitoh H, Felix G, Freymark G, Miersch O, Wasternack C, Boller T, Jones JD, Romeis T (2005) Ethylene-mediated cross-talk between calcium-dependent protein kinase and MAPK signaling controls stress responses in plants. Proc Natl Acad Sci USA 102:10736–10741. doi:10.1073/pnas.0502954102
MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A et al (2002) Shotgun identification of protein modifications from protein complexes and lens tissue. Proc Natl Acad Sci USA 99:7900–7905
Mahajan R, Delphin C, Guan T, Gerace L, Melchior F (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. Cell 88:97–107
Maor R, Jones A, Nuhse TS, Studholme DJ, Peck SC, Shirasu K (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol Cell Proteomics 6:601–610. doi:10.1074/mcp.M600408-MCP200
Margottin F, Bour SP, Durand H, Selig L, Benichou S, Richard V, Thomas D, Strebel K, Benarous R (1998) A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. Mol Cell 1:565–574
Marshall CJ (1994) MAP kinase kinase kinase, MAP kinase kinase and MAP kinase. Curr Opin Genet Dev 4:82–89
Meskiene I, Bogre L, Glaser W, Balog J, Brandstotter M, Zwerger K, Ammerer G, Hirt H (1998) MP2C, a plant protein phosphatase 2C, functions as a negative regulator of mitogen-activated protein kinase pathways in yeast and plants. Proc Natl Acad Sci USA 95:1938–1943
Meskiene I, Baudouin E, Schweighofer A, Liwosz A, Jonak C, Rodriguez PL, Jelinek H, Hirt H (2003) Stress-induced protein phosphatase 2C is a negative regulator of a mitogen-activated protein kinase. J Biol Chem 278:18945–18952. doi:10.1074/jbc.M300878200
Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnik SE et al (2002) Annotation of the *Drosophila melanogaster*

euchromatic genome: a systematic review. Genome Biol 3:RESEARCH0083
Mukherjee S, Keitany G, Li Y, Wang Y, Ball HL, Goldsmith EJ, Orth K (2006) Yersinia YopJ acetylates and inhibits kinase activation by blocking phosphorylation. Science 312:1211–1214. doi:10.1126/science.1126867
Murphy LO, Smith S, Chen RH, Fingar DC, Blenis J (2002) Molecular interpretation of ERK signal duration by immediate early gene products. Nat Cell Biol 4:556–564. doi:10.1038/ncb822ncb822
Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA et al (2000) A whole-genome assembly of Drosophila. Science 287:2196–2204. doi:8395
Nakagami H, Pitzschke A, Hirt H (2005) Emerging MAP kinase pathways in plant stress signalling. Trends Plant Sci 10:339–346. doi:10.1016/j.tplants.2005.05.009
Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JD (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312:436–439. doi:10.1126/science.1126088
Nuhse TS, Peck SC, Hirt H, Boller T (2000) Microbial elicitors induce activation and dual phosphorylation of the *Arabidopsis thaliana* MAPK 6. J Biol Chem 275:7521–7526
Nuhse TS, Stensballe A, Jensen ON, Peck SC (2004) Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. Plant Cell 16:2394–2405. doi:10.1105/tpc.104.023150
Nuhse TS, Bottrill AR, Jones AM, Peck SC (2007) Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. Plant J 51:931–940. doi:10.1111/j.1365-313X.2007.03192.x
O'Connell MJ, Walworth NC, Carr AM (2000) The G2-phase DNA-damage checkpoint. Trends Cell Biol 10:296–303
Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4:1487–1502
Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127:635–648
Olsen JV, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, Gnad F, Cox J, Jensen TS, Nigg EA et al (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3:ra3. doi:10.1126/scisignal.2000475
Ono M, Shitashige M, Honda K, Isobe T, Kuwabara H, Matsuzuki H, Hirohashi S, Yamada T (2006) Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. Mol Cell Proteomics 5:1338–1347
Palmgren MG (2001) PLANT PLASMA MEMBRANE H+-ATPases: Powerhouses for Nutrient Uptake. Annu Rev Plant Physiol Plant Mol Biol 52:817–845. doi:10.1146/annurev.arplant.52.1.817
Pawson T, Nash P (2000) Protein-protein interactions define specificity in signal transduction. Genes Dev 14:1027–1047
Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2:559–572
Peng JM, Schwartz D, Elias JE, Thoreen CC, Cheng DM, Marsischky G, Roelofs J, Finley D, Gygi SP (2003) A proteomics approach to understanding protein ubiquitination. Nat Biotechnol 21:921–926. doi:10.1038/Nbt849
Polge C, Thomas M (2007) SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control? Trends Plant Sci 12:20–28. doi:10.1016/j.tplants.2006.11.005
Posern G, Miralles F, Guettler S, Treisman R (2004) Mutant actins that stabilise F-actin use distinct mechanisms to activate the SRF coactivator MAL. EMBO J 23:3973–3983. doi:10.1038/sj.emboj.7600404
Rock KL, Goldberg AL (1999) Degradation of cell proteins and the generation of MHC class I-presented peptides. Annu Rev Immunol 17:739–779
Sabo A, Lusic M, Cereseto A, Giacca M (2008) Acetylation of conserved lysines in the catalytic core of cyclin-dependent kinase 9 inhibits kinase activity and regulates transcription. Mol Cell Biol 28:2201–2212. doi:10.1128/MCB.01557-07
Scheler C, Muller EC, Stahl J, Muller-Werdan U, Salnikow J, Jungblut P (1997) Identification and characterization of heat shock protein 27 protein species in human myocardial two-dimensional electrophoresis patterns. Electrophoresis 18:2823–2831
Schlessinger J (2000) Cell signaling by receptor tyrosine kinases. Cell 103:211–225
Schmidt A, Gehlenborg N, Bodenmiller B, Mueller LN, Campbell D, Mueller M, Aebersold R, Domon B (2008) An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures. Mol Cell Proteomics 7:2138–2150
Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101:671–684
Scholz M, Gatzek S, Sterling A, Fiehn O, Selbig J (2004) Metabolite fingerprinting: detecting biological features by independent component analysis. Bioinformatics 20:2447–2454
Schulze WX (2010) Proteomics approaches to understand protein phosphorylation in pathway modulation. Curr Opin Plant Biol. doi:10.1016/j.pbi.2009.12.008
Schulze WX, Usadel B (2010) Quantitation in mass-spectrometry-based proteomics. Annu Rev Plant Biol 61:491–516. doi:10.1146/annurev-arplant-042809-112132
Schweighofer A, Kazanaviciute V, Scheikl E, Teige M, Doczi R, Hirt H, Schwanninger M, Kant M, Schuurink R, Mauch F et al (2007) The PP2C-type phosphatase AP2C1, which negatively regulates MPK4 and MPK6, modulates innate immunity, jasmonic acid, and ethylene levels in Arabidopsis. Plant Cell 19:2213–2224. doi:10.1105/tpc.106.049585
Seet BT, Dikic I, Zhou MM, Pawson T (2006) Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol 7:473–483. doi:10.1038/nrm1960
Seo J, Lee KJ (2004) Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. J Biochem Mol Biol 37:35–44
Shteinberg M, Protopopov Y, Listovsky T, Brandeis M, Hershko A (1999) Phosphorylation of the cyclosome is required for its stimulation by Fizzy/cdc20. Biochem Biophys Res Commun 260:193–198. doi:10.1006/bbrc.1999.0884
Sims RJ, Reinberg D (2008) Is there a code embedded in proteins that is based on post-translational modifications? Nat Rev Mol Cell Bio 9:815–820. doi:10.1038/Nrm2502
Siuti N, Kelleher NL (2007) Decoding protein modifications using top-down mass spectrometry. Nat Methods 4:817–821. doi:10.1038/nmeth1097
Smallwood M, Bowles DJ (2002) Plants in a cold climate. Philos Trans R Soc Lond B Biol Sci 357:831–847. doi:10.1098/rstb.2002.1073
Sowa ME, Bennett EJ, Gygi SP, Harper JW (2009) Defining the human deubiquitinating enzyme interaction landscape. Cell 138:389–403. doi:10.1016/j.cell.2009.04.042
Steuer R, Kurths J, Fiehn O, Weckwerth W (2003) Observing and interpreting correlations in metabolomic networks. Bioinformatics 19:1019–1026

Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. Mol Syst Biol 4:193. doi:10.1038/Msb.2008.32

Sun Y, Jiang X, Chen S, Fernandes N, Price BD (2005) A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. Proc Natl Acad Sci USA 102:13182–13187. doi:10.1073/pnas.0504211102

Takahashi Y, Soyano T, Sasabe M, Machida Y (2004) A MAP kinase cascade that controls plant cytokinesis. J Biochem 136:127–132. doi:10.1093/Jb/Mvh118

Takahashi F, Yoshida R, Ichimura K, Mizoguchi T, Seo S, Yonezawa M, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K (2007) The mitogen-activated protein kinase cascade MKK3-MPK6 is an important part of the jasmonate signal transduction pathway in Arabidopsis. Plant Cell 19:805–818. doi:10.1105/tpc.106.046581

Thelemann A, Petti F, Griffin G, Iwata K, Hunt T, Settinari T, Fenyo D, Gibson N, Haley JD (2005) Phosphotyrosine signaling networks in epidermal growth factor receptor overexpressing squamous carcinoma cells. Mol Cell Proteomics 4:356–376. doi:10.1074/mcp.M400118-MCP200

Thompson PR, Wang D, Wang L, Fulco M, Pediconi N, Zhang D, An W, Ge Q, Roeder RG, Wong J et al (2004) Regulation of the p300 HAT domain via a novel activation loop. Nat Struct Mol Biol 11:308–315. doi:10.1038/nsmb740

Tress ML, Bodenmiller B, Aebersold R, Valencia A (2008) Proteomics studies confirm the presence of alternative protein isoforms on a large scale. Genome Biol 9:R162

Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA (2005) Identification of post-translational modifications by blind search of mass spectra. Nat Biotechnol 23:1562–1567

Usaite R, Wohlschlegel J, Venable JD, Park SK, Nielsen J, Olsson L, Yates JR (2008) Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression *Saccharomyces cerevisiae* strains: the comparison of two quantitative methods. J Proteome Res 7:266–275. doi:10.1021/Pr700580m

van Bentem SD, Hirt H (2007) Using phosphoproteomics to reveal signalling dynamics in plants. Trends Plant Sci 12:404–411. doi:10.1016/j.tplants.2007.08.007

Ventii KH, Wilkinson KD (2008) Protein partners of deubiquitinating enzymes. Biochem J 414:161–175. doi:10.1042/Bj20080798

Ventura JJ, Kennedy NJ, Lamb JA, Flavell RA, Davis RJ (2003) c-Jun NH(2)-terminal kinase is essential for the regulation of AP-1 by tumor necrosis factor. Mol Cell Biol 23:2871–2882

Villen J, Beausoleil SA, Gerber SA, Gygi SP (2007) Large-scale phosphorylation analysis of mouse liver. Proc Natl Acad Sci USA 104:1488–1493. doi:10.1073/pnas.0609836104

Wang D, Baldwin AS Jr (1998) Activation of nuclear factor-kappaB-dependent transcription by tumor necrosis factor-alpha is mediated through phosphorylation of RelA/p65 on serine 529. J Biol Chem 273:29411–29416

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562. doi:10.1038/Nature01262

Wienkoop S, Larrainzar E, Niemann M, Gonzalez EM, Lehmann U, Weckwerth W (2006) Stable isotope-free quantitative shotgun proteomics combined with sample pattern recognition for rapid diagnostics. J Sep Sci 29:2793–2801

Wienkoop S, Morgenthal K, Wolschin F, Scholz M, Selbig J, Weckwerth W (2008) Integration of metabolomic and proteomic phenotypes: analysis of data covariance dissects starch and RFO metabolism from low and high temperature compensation response in *Arabidopsis thaliana*. Mol Cell Proteomics 7:1725–1736. doi:10.1074/mcp.M700273-MCP200

Yang XJ (2005) Multisite protein modification and intramolecular signaling. Oncogene 24:1653–1662. doi:10.1038/sj.onc.1208173

Yang XJ, Seto E (2008) Lysine acetylation: codified crosstalk with other posttranslational modifications. Mol Cell 31:449–461. doi:10.1016/j.molcel.2008.07.002

Yaron A, Hatzubai A, Davis M, Lavon I, Amit S, Manning AM, Andersen JS, Mann M, Mercurio F, Ben-Neriah Y (1998) Identification of the receptor component of the IkappaBalpha-ubiquitin ligase. Nature 396:590–594. doi:10.1038/25159

Zhang X, Yuan Z, Zhang Y, Yong S, Salas-Burgos A, Koomen J, Olashaw N, Parsons JT, Yang XJ, Dent SR et al (2007) HDAC6 modulates cell motility by altering the acetylation level of cortactin. Mol Cell 27:197–213. doi:10.1016/j.molcel.2007.05.033

Zhang J, Shi X, Li Y, Kim BJ, Jia J, Huang Z, Yang T, Fu X, Jung SY, Wang Y et al (2008) Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Mol Cell 31:143–151. doi:10.1016/j.molcel.2008.06.006

Zhang G, Hubalewska M, Ignatova Z (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol 16:274–280. doi:10.1038/Nsmb.1554

Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JDG, Felix G, Boller T (2004) Bacterial disease resistance in Arabidopsis through flagellin perception. Nature 428:764–767. doi:10.1038/Nature02485