REVIEW

doi:10.1038/nature19949

Mass-spectrometric exploration of proteome structure and function

Ruedi Aebersold[1,2] & Matthias Mann[3,4]

Numerous biological processes are concurrently and coordinately active in every living cell. Each of them encompasses synthetic, catalytic and regulatory functions that are, almost always, carried out by proteins organized further into higher-order structures and networks. For decades, the structures and functions of selected proteins have been studied using biochemical and biophysical methods. However, the properties and behaviour of the proteome as an integrated system have largely remained elusive. Powerful mass-spectrometry-based technologies now provide unprecedented insights into the composition, structure, function and control of the proteome, shedding light on complex biological processes and phenotypes.

Collectively, proteins catalyse and control essentially all cellular processes. They form a highly structured entity known as the proteome, the constituent proteins of which carry out their functions at specific times and locations in the cell, in physical or functional association with other proteins or biomolecules. A proliferating *Schizosaccharomyces pombe* cell contains about 60 million protein molecules, which have abundances that range from a few copies to 1.1 million copies per expressed gene[1]. Across the species, proteins constitute about 50% of the dry mass of a cell and reach a remarkable total concentration of 2–4 million proteins per cubic micrometre or 100–300 mg per ml (ref. 2). The extensive proteome network of the cell adapts dynamically to external or internal (that is, genetic) perturbations and thereby defines the cell's functional state and determines its phenotypes. Describing and understanding the complete and quantitative proteome as well as its structure, function and dynamics is a central and fundamental challenge of biology.

Two strategies that differ in principle have been used to study the proteome and the molecular mechanisms that it mediates. Conventionally, specific proteins are isolated and then analysed with respect to their structure and function through the established methods of biochemistry and biophysics. But it has also become possible to perform large-scale, systematic measurements of proteomes to generate biological insights from the computational analysis of proteomic datasets, either on their own or in combination with other 'omics' types of data. Both approaches have been transformed fundamentally by the development of powerful mass-spectrometry-based methods. Such techniques have the capability to identify conclusively and quantify accurately almost any protein that has been expressed. They can also systematically identify and localize modified amino acids in the polypeptide chain as well as determine the composition, stoichiometry and topology of the subunits of multiprotein complexes and even contribute to determining their structure.

The annotated genome identifies the entire proteome of an organism. However, the literature has focused on the small fraction of the proteome for which measurement assays are readily available[3]. This set of intensely studied proteins has remained surprisingly constant over the past few decades. Robust mass-spectrometry-based methods now enable most proteins to be measured reliably, which vastly extends the range of the classic, mechanism-focused analyses of specific components of the proteome. They also make possible the systematic analysis of the proteome to an extent that had been predicted previously[4,5].

Underlying reasons for the success of mass spectrometry in proteomics include its inherent specificity of identification, the generic nature of the proteomics workflow and its potential for extreme sensitivity that, in principle, extends to the single ion. In practice, it has been challenging to realize the full potential of the technique, and ingenious ways of implementing mass spectrometry as a universal detector of protein identity, abundance, precise chemical state and cellular context and localization are still being devised. At present, no single mass-spectrometry-based system or method can determine by itself these diverse dimensions for proteome data.

This Review highlights the achievements of mass-spectrometry-based proteomics and the challenges that remain. Efforts to catalogue systematically the proteomes of an array of species and to transform these catalogues into highly specific assays that can quantify any component are described. The analysis of post-translational modifications is discussed, especially with regard to completeness of measurement and how the research community might assign functions to the tens of thousands of modified sites that have been discovered in the past decade. The state of mass spectrometry is reviewed in the context of the study of functional modules, in which components of the proteome come together stably or temporarily in complexes to carry out a biochemical function. Last, mass-spectrometry-based techniques that are capable of quantifying thousands of proteins across collections of large numbers of samples with a high degree of reproducibility are described; these generate large datasets that can be mined by statistical machine-learning tools to determine the state of the proteome and its response to perturbations. Such datasets start to uncover systemic malfunctions at the cellular and organismal levels in diseases that have been difficult to reach through classic protein-based or nucleic-acid-based research.

The identification and quantification of the proteome

The ability to identify reliably any component of the proteome is a requirement both for mechanistic, hypothesis-driven investigations and for large-scale, omics-type studies. A comprehensive and reliable mass-spectrometry-based proteome map is also a prerequisite for the development of targeted mass spectrometry techniques, as well as for data-independent acquisition (DIA) strategies (Fig. 1 and Box 1); these rely on information from pre-existing high-quality spectral libraries. The importance of accurate quantification in proteomics is hard to overstate,

[1]Institute of Molecular Systems Biology, Department of Biology, ETH Zürich, 8093 Zürich, Switzerland. [2]Faculty of Science, University of Zürich, 8093 Zürich, Switzerland. [3]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. [4]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark.

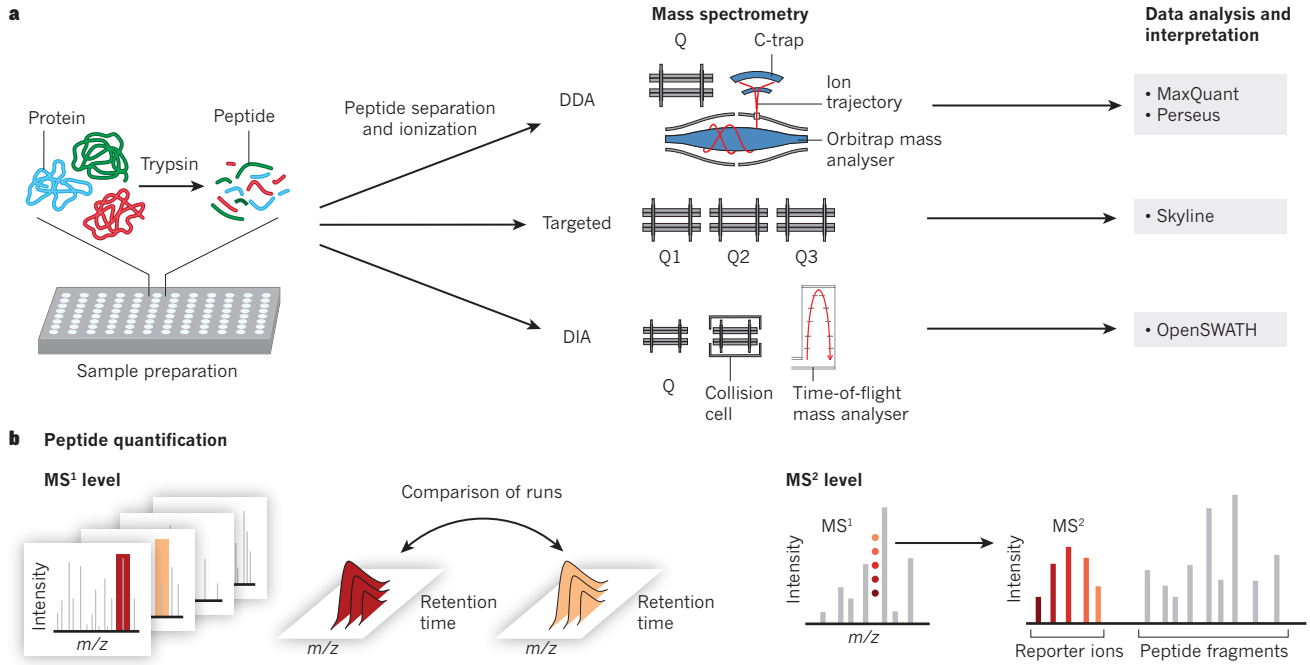

Figure 1 | Bottom-up proteomics workflows. **a**, All bottom-up proteomics workflows begin with a sample-preparation stage in which proteins are extracted and digested by a sequence-specific enzyme such as trypsin. Present methods of protein preparation are highly efficient and can be performed in 96-well plates with robotic assistance. Peptides are then separated by means of chromatography and electrosprayed, after which they are introduced into the vacuum of a mass spectrometer. Three classes of methods are shown. In DDA methods, a full spectrum of the peptides (at the MS^1 level) is acquired, followed by the collection of as many fragmentation spectra (at the MS^2 level) as possible, within a cycle time of about 1 second. A quadrupole–orbitrap mass analyser is depicted, although other types of analyser are also used in DDA. Results are interpreted using software packages such as MaxQuant[100] and the downstream Perseus environment[101]. In targeted analysis, a peptide of known mass-to-charge ratio (m/z) is selected in the first quadrupole, then the peptide is fragmented and several fragments are monitored over time. These transitions are multiplexed and their specificity is checked using software packages such as SkyLine[102]. In DIA methods, which are exemplified by sequential window acquisition of all theoretical fragment-ion spectra (SWATH)–MS[103], ranges of m/z values (that typically span 25 m/z units) are selected and peptides are fragmented, followed by the acquisition of the fragments in a time-of-flight mass spectrometer. The instrument rapidly and seamlessly cycles through the entire mass range within a few seconds. The multiplexed fragment spectra are interpreted — often with the help of known fragment spectra from large spectral libraries — by software such as OpenSWATH[104]. **b**, Peptide quantities can be determined at the MS^1 level by integrating the signal from peaks of the precursor ions that elute from the high-performance liquid chromatography column. An arbitrary number of runs (stacked mass spectra, left) can be compared using sophisticated alignment and normalization procedures. Quantitative comparison of the isotopic cluster of the same peptide over two runs can be performed. Peptide identities can also be transferred when the peptide is fragmented in only one of the runs but matches precisely the mass and elution time of an aligned peak (known as the 'match between runs' feature in MaxQuant[100]). Absolute quantities can be estimated by adding up the peak volumes of all peptides that identify a particular protein then determining the proportion of the (known) total proteome mass that has been analysed. Peptides can also be subjected to label-free quantification at the MS^2 level (right). In this case, the fragment-ion intensities that are unique to a specific peptide are used for quantification, in a way that is analogous to the use of precursor-ion signal intensities for quantification using MS^1-level data. In multiplexed shotgun proteomics, up to ten samples are labelled differentially so that they release reporter ions that can be distinguished in the MS^2 spectra. In DIA-based methods, the intensities of fragments that belong to the same precursor ion are extracted to yield a measure of peptide abundance[104,105]. Q, quadrupole.

and this has become a crucial requirement for almost all functional studies in the past 10 years.

The preferred method for proteome discovery is data-dependent acquisition (DDA) (Fig. 1) and the past decade has seen striking advances in this area. Whereas the first description of a complete model proteome[6] and the identification of more than 10,000 different proteins in human cell lines[7,8] were technological tours de force, a similar depth of coverage can now be achieved within hours and with minimal sample-preparation steps[9,10]. These developments, although still confined to a few specialized laboratories, will make proteomics increasingly applicable to everyday cell biology and biochemical research, which overwhelmingly uses classic antibody-based techniques such as western blotting. In addition to its exquisite specificity, other advantages of DDA-based proteomics include that it is unbiased and free from hypotheses; that is, the researcher does not need to know the identity of the expected proteins in advance. Furthermore, in a DDA-based proteomics experiment all proteins can be interrogated at once. As well as helping to answer a specific question, proteomics can therefore turn every experiment into a global discovery study, which enables the detection of new and unexpected molecules and connections, providing fresh biological insights. These developments are supported by publicly accessible bioinformatics tools for processing and interpreting the large amounts of data that are generated in complex projects (Fig. 1). The continued development of highly streamlined and robust proteomics workflows, including robust and economical mass spectrometers, is advocated to usher in an age of complete, accurate and ubiquitous proteomes[11], in analogy to what the introduction of next-generation sequencing has provided for genomics-related fields.

Present technology already enables analysis of the complete protein inventory of biological systems, including cell-type-specific proteomes of mammalian organs[12–14]. One outcome of in-depth proteomics studies has been a demonstration of the extent to which diverse cellular systems have similar proteomes, with few proteins being uniquely detectable in specific situations[15]. This surprising finding is supported by the Human Protein Atlas, a large-scale antibody-based study that also reports ubiquitous expression[16]. The identity of cells and tissues therefore seems to be determined primarily by the abundance at which they express their constituent proteins, and perhaps by the manner in which the proteins are organized in the proteome, rather than the presence or absence of certain proteins.

The application of DDA-based proteomics to a collection of human

BOX 1

Bottom-up proteomics

Proteins can be studied as intact entities by mass spectrometry, an approach called top-down proteomics[21]. This has the advantage that all modifications that occur on the same molecule can, in principle, be measured together, enabling identification of the precise proteoform[107]. However, bottom-up proteomics, in which peptides are generated by the enzymatic digestion of proteins, has been experimentally and computationally more tractable and is the most widespread proteomic workflow. A number of bottom-up techniques exist; each has a specific purpose, a performance profile and a range of utility. In all of the techniques, proteins are extracted from the source material then digested into peptides by a sequence-specific enzyme such as trypsin. The resulting mixture of peptides is separated by reverse-phase chromatography, which is coupled online to electrospray ionization (Fig. 1). The peptide ions are then transferred to the vacuum of a mass spectrometer, where they are fragmented in the gas phase to generate MS/MS (MS^2) spectra that contain the information to identify and quantify specific peptides. Almost always, collision-induced dissociation or higher-energy collisional dissociation[108] are used for fragmentation, but alternative methods are becoming more widely available. One such method, electron transfer dissociation[109], is particularly beneficial for the fragmentation of large and modified peptides. The resulting data are analysed by mass-spectrometry-specific computational pipelines as well as general downstream systems-biology solutions that are tailored to proteomics[101].

Three main approaches are used in bottom-up proteomics: discovery (or shotgun) proteomics by means of DDA, aimed at achieving unbiased and complete coverage of the proteome; targeted proteomics using selected reaction monitoring, aimed at the reproducible, sensitive and streamlined acquisition of a subset of known peptides of interest; and multiplexed fragmentation of all peptides that elute from the high-performance liquid chromatography column by DIA, aimed at generating comprehensive fragment-ion maps for a sample (Fig. 1a–c).

In DDA-based methods, mass spectra of all the ion species that co-elute at a specific point in the gradient elution (that is, precursor-ion spectra) are recorded at the MS^1 (or full-scan) level. The instrument alternates between the acquisition of full-scan data and the acquisition of fragment-ion spectra, in which as many precursors as possible are sequentially isolated and fragmented (at the MS^2 level). Of many possible instrument configurations, quadrupole–orbitrap analysers[110] dominate DDA proteomics but time-of-flight instruments also have unique promise. In typical 'top *N*' cycles (in which '*N*' denotes the number of MS^2 spectra that follow), an MS^1 scan is followed by about ten fragment-ion scans. Contemporary instruments transfer ions into the vacuum with greatly improved efficiency, which results in very bright beams (of more than 10^9 ions per second). The resolution of orbitraps has improved several fold, enabling very fast top *N* cycles at high resolution. However, the capacity of orbitraps is still limited to about 1 million ions, which restricts the dynamic range that can be achieved in MS^1 spectra.

In targeted proteomics, the proteins of interest are predetermined and known. Using pre-existing information, characteristic (proteotypic) peptides are selectively and recursively isolated and then fragmented over their chromatographic elution time. This is done by setting the first quadrupole of a triple quadrupole instrument to the expected precursor ion *m/z* ratio and the third quadrupole to the *m/z* ratio of an abundant fragment ion that is specific for the targeted peptide. (The second quadrupole houses the collision chamber.) To achieve selectivity, the process is multiplexed to several fragments per peptide (known as multiple reaction monitoring, MRM), and throughput is increased by multiplexing it to many peptides[111]. Alongside the robust and economical triple quadrupole instruments, high-resolution instruments such as quadrupole orbitraps are used increasingly for targeted analysis, a variant known as parallel reaction monitoring because it utilizes the entire MS^2 spectrum[112].

In DIA-based methods[113] such as SWATH[103], entire ranges of precursors are fragmented at the same time. The peptide fragmentation information is retrieved from the multiplexed MS^2 spectra either by targeted signal extraction on the basis of previously acquired single-peptide fragmentation spectra[112] or by the generation of 'pseudo' fragment-ion spectra constructed directly from the DIA data that are then subjected to classic database searching[105]. The advantage of this approach is that the entire range of possible precursor-ion masses can be analysed seamlessly and in rapid succession, which eliminates the missing value problem of DDA (in which peptides are only measured in some of a set of liquid chromatography–mass spectrometry (LC–MS^2) runs), at least within the dynamic range that is achieved in the experiment. At present, DIA is limited to a dynamic range of 4–5 orders of magnitude and it requires the *a priori* construction of fragment-ion spectra for the query peptides to deconvolve these peptides from the DIA data[104,105,114].

Each of these approaches has advantages and limitations; hybrid methods that combine the best aspects will therefore probably emerge in the near future. Entirely new methods will also be created. For instance, in the past year it has become possible to store several precursor ions in parallel in a trapped-ion mobility device, which can then be followed by serial fragmentation. Known as parallel accumulation–serial fragmentation (PASEF), this method promises to increase the speed and sensitivity of fragmentation several fold[115].

Metabolic and chemical labelling strategies have matured and can now be used for precise quantification, but they can still suffer from limitations to their accuracy and dynamic range[116–118]. Improvements in the resolution that can be achieved, combined with advances in algorithms, are making label-free quantification increasingly useful for DDA[119], selected reaction monitoring[120] and DIA[104,105] methods.

tissues, combined with the integration of data from the community, has resulted in two draft human proteomes[17,18]. Mass-spectrometric evidence for 84% (ref. 17) or 92% (ref. 18) of protein-coding sequences was reported. However, re-analysis of the data using standard and community-approved false-discovery rates for peptides and proteins leads to much lower coverage and the removal of proteins not thought to be expressed in the sampled tissues[19,20]. Extensive peptide pre-fractionation has been combined with digestion by various enzymes and peptide fragmentation methods to reach a depth of proteome coverage that should soon be on par with the comprehensiveness to which the transcriptome can be probed by next-generation sequencing[13]. Comprehensive characterization of the proteome is therefore feasible and we predict that it will soon become routine[11]. The coverage of identified proteins with sequenced peptides has also been improving, which makes it increasingly realistic to distinguish between and quantify proteoforms, the different molecular forms of a protein that originate from the same gene. A complete inventory of proteoforms cannot yet be achieved and will be a challenge to attain because of the combinatorial explosion of proteoforms that are created by even a moderate number of modifications. Top-down proteomics characterizes the actual combination of modification events for each proteoform[21]. Although attractive in principle, top-down mass spectrometry is experimentally and computationally challenging because

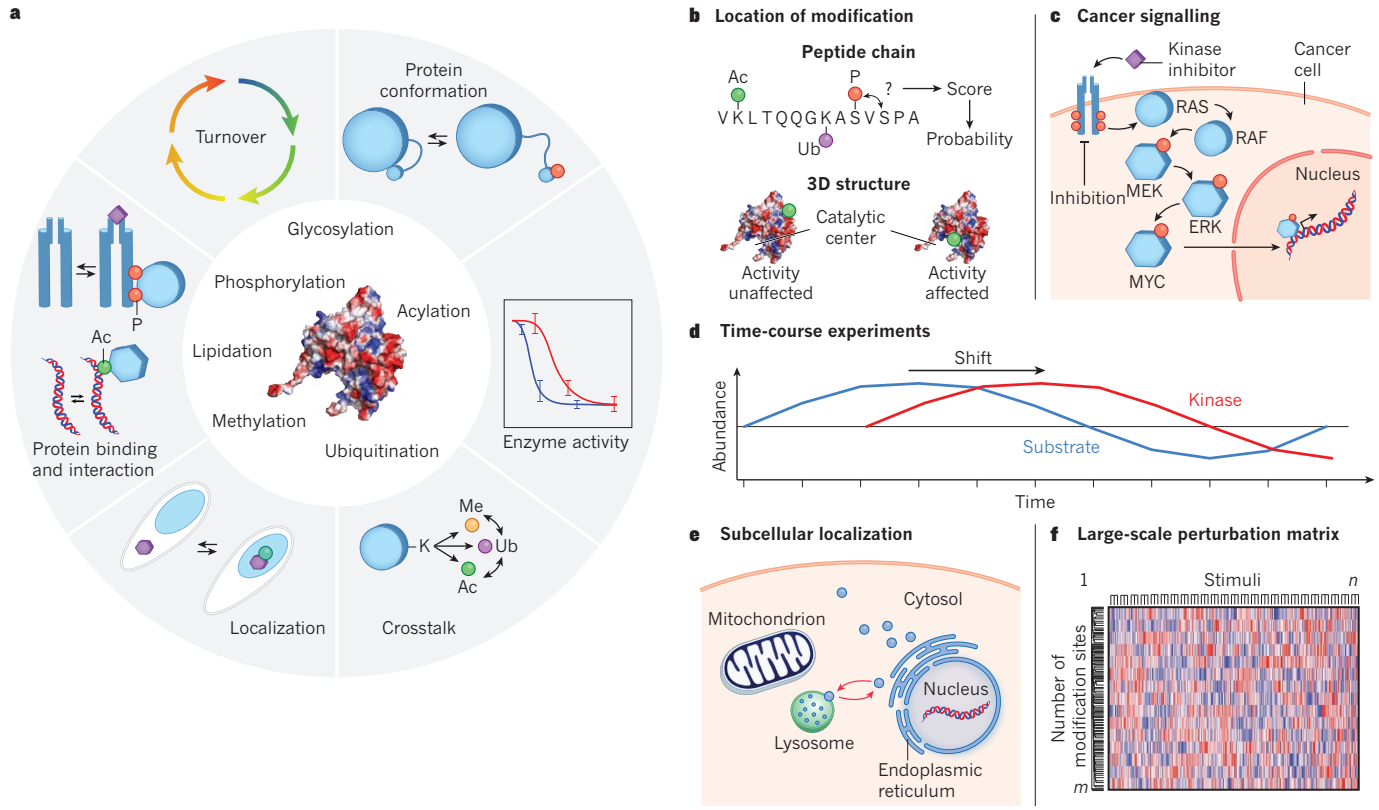

Figure 2 | Analysis of post-translational modifications. **a**, In post-translational modification, proteins are modified through the attachment of a chemical moiety such as a phosphate group, usually by a dedicated and highly specific system of enzymes. The most commonly studied post-translational modifications are listed (centre) and these are accompanied by hundreds of other less-well-studied or unknown types of modifications. Such modifications can lead to: alterations in protein conformation (through phosphorylation) and subsequent allosteric regulation; changes in enzyme activity; crosstalk that results from the same amino-acid residue being targeted by more than one type of modification; alterations in the subcellular localization of proteins; changes in protein binding; and alterations in protein lifetimes (for example, through the attachment of the small protein ubiquitin). Ac, acetyl; ERK, extracellular signal-related kinase; Me, methyl; MEK, mitogen-activated protein kinase kinase; MYC, transcription factor cMYC; P, phosphate; RAF, RAF kinase; RAS, RAS GTPase; Ub, ubiquitin. **b**, After a modified peptide has been identified from the fragment spectra, the amino acid in the peptide chain to which the post-translational modification is attached must be determined. The location of the modification within the three-dimensional structure of the protein can often also be determined, which provides clues about function. **c**, Global interrogation of the changes in a signalling pathway can be achieved readily by quantitative phosphoproteomics. For example, the suppression of aberrant signalling in cancer cells by drugs known as kinase inhibitors can be followed. **d**, Detailed time-course experiments yield information on the temporal ordering of events such as the activation of a kinase upstream of one of its substrates. The proportion of proteins that are modified by a particular post-translational modification (also termed the occupancy or stoichiometry) can change drastically depending on the biological conditions (not shown). It can be derived from the changes in protein level and the levels of the modified and unmodified peptide in two cellular states[106]. **e**, The modification of a protein often determines its subcellular localization — that is, whether it is found in the nucleus or the cytosol, for instance. Many types of stimuli can be applied to biological systems, after which the level of a particular post-translational modification can be determined. **f**, The structure of the perturbation matrix that results reveals the regulated sites and how they correlate between stimuli, as indicated by hot spots in the heat map. *m*, number of modification sites quantified; *n*, number of stimuli applied.

of the greater difficulty in analysing proteins in comparison with peptides and because each protein is distributed as multiple proteoforms that might or might not differ functionally. The array of modern mass spectrometry techniques has also been deployed to analyse unique types of sample with biological and clinical importance, including secreted proteins in the context of immunology[22], the peptidome of body fluids such as cerebrospinal fluid[23], the immunopeptidome[24] and the extracellular matrix[25].

Proteomics is sufficiently advanced to warrant the in-depth characterization of a great variety of biological systems. Along with other important information, this enables protein copy numbers or concentrations to be determined on a proteome-wide scale[26–28], which helps to improve understanding of the underlying biology.

Characterizing protein modifications and cell signalling

Mass-spectrometry-based proteomics is well suited to the study of post-translational modifications because such changes lead to characteristic shifts in mass and can be located with the resolution of a single amino acid through peptide-fragment ion spectra (Fig. 2). The only deviation from the DDA-based proteomic workflow that is used to identify unmodified peptides is the addition of an enrichment step for peptides that carry the modification of interest. Post-translational modifications that are particularly labile, such as *O*-linked β-*N*-acetylglucosamine (*O*-GlcNAc), benefit from the use of electron transfer dissociation as the fragmentation method, and certain classes of modifications, including glycosylations with large glycans and nucleotide modifications, can also be challenging to detect using mass spectrometry. The most frequently studied types of post-translational modifications are phosphorylation, ubiquitylation, the addition of ubiquitin-like proteins, glycosylation, methylation, acetylation and other types of acylation. For these, present technology enables the identification of thousands of sites of modification and their accurate quantification between proteomic states[29]. The main surprise has been the number and diversity of these post-translational modifications as well as how many of them seem to be involved in cellular regulation. For example, more than 50,000 phosphorylation events on at least 75% of the proteome have been documented in a single cell line[30]. Phosphoproteomics is used routinely to quantify the response of cells to

stimuli and such studies have reached a remarkable level of detail and sophistication. As well as providing large catalogues of sites, they have led to the discovery of sites of regulation with pivotal roles in determining the state of biological processes[31–35]. A streamlined protocol has made it possible to analyse *in vivo* signalling events with high temporal resolution[36]. This revealed that insulin signalling in the liver is unexpectedly fast: maximal phosphorylation was reached within a few seconds at many sites and transcription factors were phosphorylated fully within 30 seconds. Another message emerging from phosphoproteomics is that the proportion of sites that are functional seems to be high. This is suggested by high stoichiometry (that is, the fraction of proteins that are phosphorylated at a specific site), a large number of highly regulated sites in diverse processes, and by the tight temporal correlation of many uncharacterized sites with sites that are known to be functional. Conversely, lysine acetylation behaves very differently: the stoichiometry is extremely low for most sites and often these modifications seem to be of a non-enzymatic origin, which is also true for acylations such as succinylation[37,38]. Lysine is the most frequently modified amino-acid residue and the specific target of ubiquitylation, a modification that can be enriched efficiently and studied in a linkage-specific manner by mass spectrometry. Effective strategies also exist for characterizing SUMOylation and modification with other ubiquitin-like proteins, and these have revealed unique insights into their large-scale behaviour[39]. Histone modifications and their regulators (proteins known as 'writers', 'readers' and 'erasers' that make, recognize and edit epigenetic marks) are of great interest and specific methods have been devised for their detection[40,41].

Mass spectrometry also enables the characterization of hundreds of exotic or unknown modifications[42–44]. This emerging area builds on new instrumentation, innovative methods of fragmentation and fresh protocols for enrichment but faces the challenge of devising enrichment methods that are specific for each post-translational modification of interest. As the proteome is probed to ever increasing depths, the analysis of modifications without their enrichment is becoming more feasible, and this is already possible for methylation and phosphorylation.

Post-translational modifications and proteolytic processing events, in particular, can also be analysed using chemical proteomics approaches. These use compounds that bind to engineered small-molecule binding pockets[45] or probes that label the freshly created N termini of proteins after cleavage[46,47]. The deep, quantitative and time-resolved analysis of specific types of modifications in many systems and species has already provided a wealth of biological insights. These data also indicate that specific modification systems intersect and cooperate to generate a specific cellular state. The comprehensive analysis of proteoforms that differ in their state of modification, the determination of the functional significance of such proteoforms and the elucidation of the processes that catalyse and control their homeostasis remain challenges for the future.

Protein modules, networks and cellular functions

Proteins rarely function alone; instead, they depend on the association of various components into macromolecular complexes. The concept of modular biology, proposed by Leland Hartwell and his colleagues, states that the biological functions of the cell are carried out by multicomponent modules[48], and the modularity of the proteome has been impressively demonstrated by several classic studies[49]. An array of mass-spectrometry-based strategies, the best established of which is interaction proteomics, has made considerable contributions to integrative or hybrid approaches to yield the composition, topology and structure of specific complex macromolecular assemblies[50].

Interaction proteomics involves a pull-down assay of a bait protein with its binding partners followed by mass-spectrometric analysis, known as affinity-purification mass spectrometry (AP–MS)[51] (Fig. 3a). Thousands of proteins can be detected in such experiments owing to the high sensitivity of mass spectrometry and the propensity of the samples to contain unspecific contaminants. Proteins that bind with specificity to the bait can be distinguished effectively from the contaminants through the quantitative comparison of samples with control assays, preferably using rigorous statistical controls[52,53]. Without the ability to distinguish background binding, the reported interactomes of specific proteins often contain hundreds of purported binders with little biological importance. Versions of this basic AP–MS workflow have been implemented robustly to support large-scale mapping of the wiring diagrams of the human cellular proteome[54]. Taking advantage of the relative abundance levels of prey proteins and the endogenously expressed bait, and adding copy numbers of the entire cellular proteome, provides a human interactome in three quantitative dimensions and enables the estimation of binding stoichiometries. This helps to classify interactions into stable, regulatory

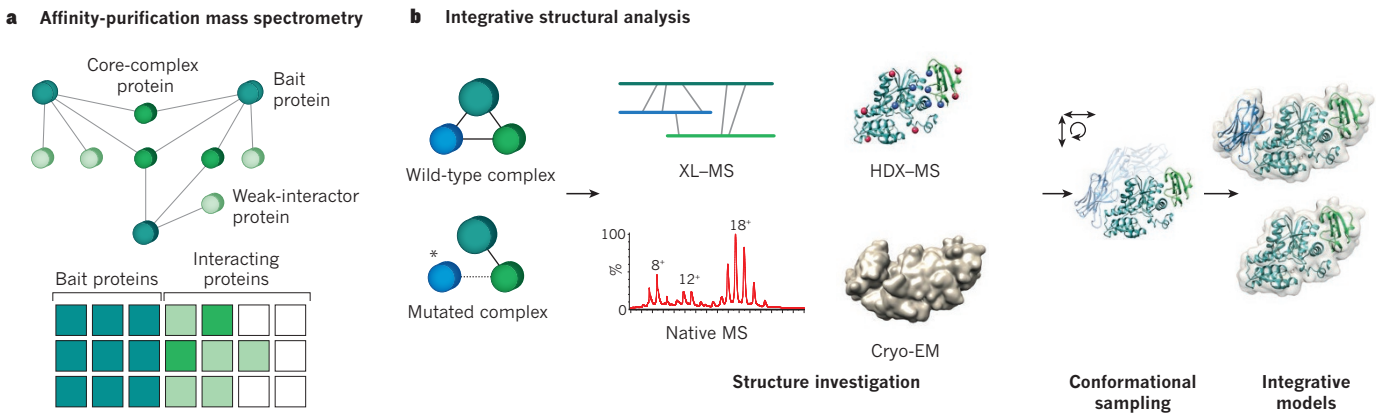

Figure 3 | Interaction proteomics and structural proteomics. a, Schematic representations of a protein interaction network with bait proteins (teal), core complex members (dark green) and weak interactors (light green). A bait protein is precipitated with its interaction partners and is measured in replicates by one of the workflows described in Fig. 1. By considering the interaction stoichiometry (the molar ratio of prey proteins and the bait protein expressed under endogenous control) and the relative cellular abundances of the proteins, stable core complexes can be distinguished from weak interactions and unspecific interactions, as well as from asymmetric interactions between proteins of different abundances[55]. **b**, A wild-type protein complex and the same complex with mutations (*) are investigated using complementary structural techniques, collectively termed integrative or hybrid structural analysis. For example, XL–MS can reveal information about subunit topology and direct domain–domain interactions. Hydrogen–deuterium exchange mass spectrometry (HDX–MS) is able to determine the interaction surfaces and solvent-exposed regions. Native mass spectrometry (native MS), in which entire protein complexes are electrosprayed into the mass spectrometer, can infer the stoichiometry and the assembly pathway of such complexes, and cryo-EM can obtain their overall shape and their density maps. The heterogeneous structural restraints are integrated in a common computational framework that evaluates subunit configurations (known as conformational sampling). Consensus models that represent the structures of the wild-type and mutated complexes can then be derived.

or transient ones and even captures client interactions such as proteins being folded by chaperone complexes[55]. This work established that networks of cells are surprisingly dominated by a large number of weak interactions and that the number of stable core complexes is limited. The emerging picture of a modular proteome in which modules have variable stoichiometric robustness is also supported by a study in which the relative changes of bona fide protein components of 182 complexes were determined in 11 cell types and 5 temporal states[56]. The covariance of the co-expression profiles for complex subunits varied considerably, which suggests that dynamic subunit associations fine-tune the composition and function of specific cellular modules[56].

Modified peptides, oligonucleotides and small molecules have also been used with success as bait proteins for AP–MS experiments[51]. For instance, transcription-factor complexes that are crosslinked to DNA can be analysed readily, as can protein complexes that are recruited to specific DNA lesions[57]. Other approaches to capture protein interactions include enzyme-meditated proximity labelling in cells followed by pull-down assays of the labelled proteins[58,59] and the accurate measurement of co-fractionation patterns[60–62]. Such measurements are also the basis of organellar proteomics, which aims to determine the subcellular location and dynamics of the proteome[63–66], a valuable complement to imaging-based technologies.

Although AP–MS and related methods indicate the composite population of proteins that is associated with a particular bait protein, other mass-spectrometry-based methods can also identify the subunit interfaces, topology, conformation and structure of protein complexes (Fig. 3b), as shown by the analysis of the nuclear pore complex[67].

Native mass spectrometry, which is the direct analysis of macromolecular assemblies by mass spectrometry, has been used both by itself[68] and as part of an integrative approach[69] to gain insights into the subunit stoichiometry, topology and structure of macromolecular assemblies. When applied to membrane protein complexes, the technique revealed an unappreciated structural role for lipids in respiratory protein complexes[70].

Integrative or hybrid approaches complement X-ray crystallography and nuclear magnetic resonance, methods that are central to structural biology, and mass spectrometry has become an essential component of the hybrid structural-biology toolbox[71]. Distance restraints that are generated by chemical crosslinking and the mass-spectrometry-based identification of crosslinked residues (an approach termed XL–MS) have proven helpful for determining the structure of large complexes[72], particularly in combination with single-particle cryo-electron microscopy (cryo-EM) data. XL–MS and cryo-EM have been used to solve longstanding problems in structural biology[71], to identify the substrate binding sites in molecular chaperones[73] and to detect steric alterations in complexes in different functional states[74]. XL–MS has also been used to analyse protein–RNA interfaces[75], to identify receptor–ligand pairs directly[76], to map physical interactions between different types of biomolecules and to identify the ligands of orphan receptors.

Integrative structural-biology methods are being adapted for use with the microgram amounts of protein complexes that are isolated by affinity purification, and this advance has been applied to mapping the organization of the protein phosphatase 2A (PP2A) enzyme system in HEK293 cells[77]. Using the two catalytic subunits, the scaffold subunit and most of the 15 regulatory subunits from which trimeric PP2A structures are combinatorially assembled as bait proteins, XL–MS identified the protein–protein interfaces, the actual subunit composition of the PP2A complexes that are concurrently expressed in the cell and their associated proteins to establish a high-granularity protein interaction network consisting of more than 150 proteins[77].

Notably, XL–MS is beginning to be used on a proteomics scale[78,79]. Although the crosslinks that are identified in such studies come primarily from highly expressed complexes, they highlight a path towards the direct measurement of protein–protein interfaces in the cell. The combination of AP–MS and XL–MS was recently refined so that chemical crosslinks could be identified from samples containing only a few million cells[80,81]. Complexes that are isolated by AP–MS can also be used to generate cryo-EM single-particle data, which opens up the possibility of linking the atomic structure and function of macromolecular assemblies that have been isolated from cells in a particular functional state. Results from cryo-electron tomography studies further extend this perspective towards the possibility of observing specific macromolecular modules by template matching *in situ*[82,83].

In a similar way to their composition, the conformation of the subunits of protein complexes can adapt to the state of the cell. Mass spectrometry techniques can detect changes in protein conformation and protein interfaces and then relate these observations to functional alterations in particular proteins. Hydrogen–deuterium exchange mass spectrometry is a classic method for determining alterations in the conformation, structure and interfaces of specific complexes[84]. By contrast, the hydroxyl radical footprinting method predominantly labels solvent-exposed side chains and is not affected by back exchange of the labelled residues[85]. The different conformations of a protein can vary in thermal stability, an observation that has been used to probe conformational changes at a proteomic scale[86]. Cells treated with a cancer drug were subjected to different temperatures, after which heat-denatured proteins were removed and the remaining soluble proteins were analysed by mass spectrometry. This pinpointed both expected and unexpected binding partners of the drug. A conceptually similar technique used the fact that conformational changes in proteins can be detected using protein digestion patterns generated under conditions of limited proteolysis[87]. Structural features of more than 1,000 yeast proteins were concurrently monitored by targeted mass spectrometry and altered conformations for about 300 proteins on a change in nutrients were detected[87]. Such examples demonstrate how structural proteomics techniques are helping to tackle the challenge of detecting often weak interactions between proteins, small-molecule ligands and co-factors on a global scale, as well as the structural effects of ligand binding.

Proteotype states and cellular phenotypes

In the 1940s, Linus Pauling established that a structural alteration in haemoglobin was related causally to a disease phenotype[88]. In that particular case, the structural variation was caused by a single amino acid change in one of the haemoglobin chains, the result of a mutation in the gene that encodes the chain. The extension of this fundamental principle of biology to the level of proteome networks suggests that genetic or external perturbations change the state of the proteome network and that such changes cause or correlate with altered phenotypes (Fig. 4). The state of a proteome that is associated with a specific phenotype can be described as a proteotype. The association between a proteotype and its corresponding phenotype can be investigated by means of two mass-spectrometry-based approaches that differ in principle. The first approach attempts to describe a phenotype mechanistically using the aggregated structure and function of the proteins or modules that constitute the underlying processes. The second approach associates a phenotype with its proteotype through advanced statistical machine-learning tools (known collectively as 'big data' analytics) but does not necessarily reach a causal or mechanistic understanding of the underlying processes. Both approaches have been greatly advanced by mass-spectrometry-based technology. In particular, the big data approach based on statistical associations has become possible only through the development of mass spectrometry techniques that are capable of quantifying sets of proteins with a high degree of reproducibility across large collections of samples, generating large data matrices of proteins measured across various samples with minimal missing values. Mass-spectrometry techniques that are used to generate such matrices include the matching of MS^1 intensity maps, using their retention time versus mass-to-charge ratio, from collections of samples and DIA-based methods, and the targeted mass spectrometry of smaller numbers of proteins (Box 1).

In a demonstration of these concepts, a yeast genetic reference panel was used to quantify the effect of genetic perturbations on a metabolic network[89]. Selected-reaction-monitoring targeted mass spectrometry measured 50 metabolic proteins in 96 genetically well-defined strains of yeast. Parental strains acquired independent genetic variations that

consistently affected levels of proteins from the same module or pathway that selective pressures favoured for the acquisition of sets of polymorphisms that maintain the stoichiometry of complexes and pathways. Similarly, 192 proteins that constituted a metabolic network were quantified by selected reaction monitoring of liver samples in two metabolic states from 40 strains of mice from a genetic reference strain compendium[90], enabling genetic and environmental perturbations to be probed effectively[91]. This established a direct mechanistic link between alleles of the gene *Dhtkd1* (a protein quantitative trait locus (pQTL)), the quantity of 2-aminoadipate (a metabolite that is controlled by Dhtkd1) and a disease risk for type 2 diabetes. Mechanistic and data-driven approaches can therefore converge to enhance understanding of complex phenotypes if multilevel omics data are integrated at the level of modular networks. Repeating the proteomics measurements of the liver samples using DIA-based mass spectrometry techniques quantified more than 2,600 proteins across the collection of samples, which led to the detection of hundreds of pQTLs as well as mechanistic insights into inborn errors of metabolism and the determination of a molecular basis for respiratory super-complex formation[92].

These examples and analogous ones from the proteogenomics of cancer[93] establish a link through association studies between genetic loci and the network state, as well as between the network state and disease phenotypes. The mass spectrometry methods of bottom-up proteomics (Box 1) represent a general experimental framework for systematically probing the proteotype at ever increasing levels of completeness and precision to support the association of proteotypes and phenotypes.

In the context of translational medicine, proteins that consistently alter their abundance in correlation with a disease phenotype are considered to be biomarker candidates for the phenotype of interest. Typically, a small number of study participants are investigated in depth to extract potential biomarkers that can be validated in larger cohorts[94,95]. Although attractive in principle, biomarker discovery using mass-spectrometry-based methods is extremely challenging in practice. However, data-driven approaches are opening fresh avenues to associating protein-expression patterns with disease states.

In particular, the detection of protein biomarkers in blood plasma as a window to the physiological state of a person has been an important goal of protein science since before the advent of mass spectrometry. Experience gained over the past decade in plasma proteome analysis by mass spectrometry has demonstrated the enormous challenges of this approach, which are rooted in the complexity of the plasma proteome, its inherent variability across a population and the prevalence of factors that affect its composition, including age, gender and lifestyle. However, several studies[94,96,97] have shown that the highly reproducible mass spectrometry techniques used for proteotype measurements in tissues can be applied to plasma proteins. Fast and reliable measurements of plasma samples will therefore be possible in collections that consist of hundreds of samples. The systematic measurement of plasma proteins in twin populations has already been used to associate observed changes in abundance in the plasma proteome with genotype[98]. Furthermore, the plasma proteome can now be probed in a broad and high-throughput manner with the aim of extracting as much information about the health or disease state of an individual as possible, effectively enabling high-throughput phenotyping of people[96]. Continuing advances in mass spectrometry technology might therefore enable the future discovery of clinically actionable protein biomarker patterns.

Outlook

Over the past decade, mass-spectrometry-based proteomics has matured from a largely technology-driven field of research into a mainstream analytical tool for the life sciences. It is a versatile approach that supports the analysis of many aspects of proteins, including sequence, quantity, state of modification, structure and macromolecular context. It also accommodates a variety of research approaches, such as mechanism-oriented exploration for determining causal relationships and big-data strategies that rely on statistical associations to discover biological relationships.

Further, dramatic improvements in the core technology of mass spectrometry are probable and will open up the field of proteomics to even more applications. Aside from a focus on signalling and structural applications, important goals for proteomics will be to build comprehensive and quantitative catalogues of proteins under many conditions and perturbations and to organize these proteoforms into a modular proteome of the cell. This will improve understanding of processes across many areas of biology and diseases and will constitute an excellent starting point for modelling the cell. For this to occur, proteomics must be tightly integrated with other technologies and it should address challenges such as single-cell analysis, an approach that was pioneered by mass cytometry[99]. The integration of different types of data is already far advanced in the case of next-generation sequencing technologies (for example, RNA sequencing, chromatin immunoprecipitation followed by sequencing (ChIP-seq) and ribosome profiling) and metabolomics, and the integration of data from structural biology and imaging-based technologies is advancing at a rapid pace. There are also considerable opportunities for bringing proteomics together with increasingly efficient tools for editing the genome — in particular, CRISPR–Cas9. We envision this to work in an iterative manner in which proteomics findings are interrogated by deleting, tagging and point-mutating one or more genes of importance, followed by further rounds of proteomics measurements to determine the effects of the genetic alterations on the proteome. This will address the fundamental question of how genotypic variability is mechanistically translated into phenotypic variability. The integration of various omics approaches and many perturbations will generate exponential flows of disparate data types. This will necessitate commensurate advances in bioinformatics and computational proteomics, which will be powered increasingly by

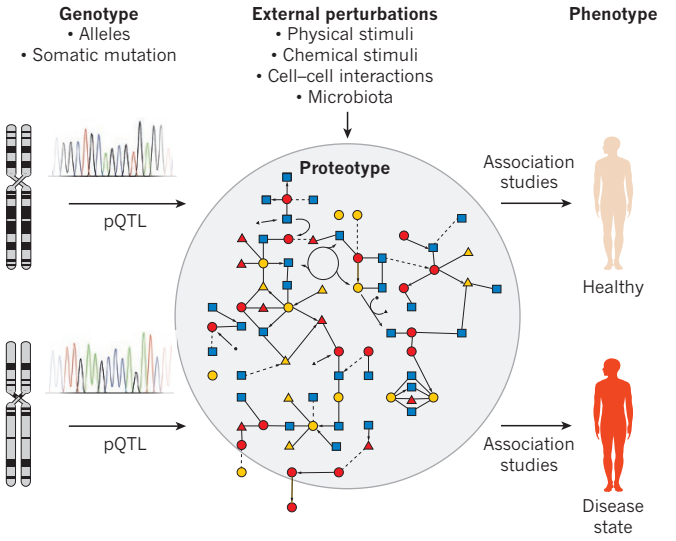

Figure 4 | Proteotype states and phenotypes. The proteotype, which is the acute state of the proteome, is shown as a modular network of interacting protein entities (coloured shapes). The composition of the proteotype and the organization of individual proteins into functional modules and interaction networks are determined by the combined effects of genotype and external perturbations, which include physical or chemical stimuli, cell–cell interactions or the microbiota. Genotypic differences such as allele differences or somatic mutations might perturb the proteotype. The relationship between genetic loci and the abundance of a protein can be described by a pQTL. These are identified by associating the abundance of a specific protein with particular alleles in genetically characterized sample populations such as genetic reference panels. In turn, the proteotype determines phenotypes, including clinical phenotypes. Association studies can identify relationships between proteotypes and phenotypes. Establishing such associations requires the generation of quantitatively accurate and highly reproducible datasets in which the same proteins are quantified across a large number of samples (for example, genetic reference panels or cohorts of patients). Datasets that support such association studies can now be generated using various mass spectrometry techniques.

machine-learning technologies while retaining their ability to generate biological insights. In this regard, the journey from single-protein analysis to a true understanding of the proteome and the importance of proteotypes will be long, challenging and exciting. ■

Received 11 January; accepted 15 July 2016.

1. Marguerat, S. *et al.* Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151,** 671–683 (2012).
2. Milo, R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays* **35,** 1050–1055 (2013).
3. Edwards, A. M. *et al.* Too many roads not taken. *Nature* **470,** 163–165 (2011).
4. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).
5. Cravatt, B. F., Simon, G. M. & Yates, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **450,** 991–1000 (2007).
6. de Godoy, L. M. F. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455,** 1251–1254 (2008).
This paper demonstrates that complete proteomes of a model organism can be obtained and quantified in different biological states.
7. Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7,** 549 (2011).
8. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7,** 548 (2011).
9. Hebert, A. S. *et al.* The one hour yeast proteome. *Mol. Cell. Proteomics* **13,** 339–347 (2014).
10. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods* **11,** 319–324 (2014).
11. Mann, M., Kulak, N. A., Nagaraj, N. & Cox, J. The coming age of complete, accurate, and ubiquitous proteomes. *Mol. Cell* **49,** 583–590 (2013).
12. Azimifar, S. B., Nagaraj, N., Cox, J. & Mann, M. Cell-type-resolved quantitative proteomics of murine liver. *Cell Metab.* **20,** 1076–1087 (2014).
13. Richards, A. L., Merrill, A. E. & Coon, J. J. Proteome sequencing goes deep. *Curr. Opin. Chem. Biol.* **24**, 11–17 (2015).
14. Sharma, K. *et al.* Cell type- and brain region-resolved mouse brain proteome. *Nature Neurosci.* **18,** 1819–1831 (2015).
15. Lundberg, E. *et al.* Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* **6,** 450 (2010).
16. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).
This paper provides an integrative analysis of the human proteome through large-scale antibody localization and transcriptomics; the findings are organized in an accompanying database.
17. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509,** 575–581 (2014).
18. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509,** 582–587 (2014).
This study aggregates data on diverse human proteomes from the authors and the research community and, like ref. 17, argues that a large part of the genome is accessible to mass-spectrometric detection.
19. Ezkurdia, I., Vázquez, J., Valencia, A. & Tress, M. Analyzing the first drafts of the human proteome. *J. Proteome Res.* **13,** 3854–3855 (2014).
20. Omenn, G. S. *et al.* Metrics for the Human Proteome Project 2015: progress on the human proteome and guidelines for high-confidence protein identification. *J. Proteome Res.* **14,** 3452–3460 (2015).
21. Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–258 (2011).
22. Meissner, F., Scheltema, R. A., Mollenkopf, H.-J. & Mann, M. Direct proteomic quantification of the secretome of activated immune cells. *Science* **340,** 475–478 (2013).
23. Secher, A. *et al.* Analytic framework for peptidomics applied to large-scale neuropeptide identification. *Nature Commun.* **7,** 11436 (2016).
24. Caron, E. *et al.* Analysis of major histocompatibility complex (MHC) immunopeptidomes using mass spectrometry. *Mol. Cell. Proteomics* **14,** 3105–3117 (2015).
25. Schiller, H. B. *et al.* Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Mol. Syst. Biol.* **11,** 819 (2015).
26. Malmström, J. *et al.* Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* **460,** 762–765 (2009).
27. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
A pioneering investigation of the degree of correlation between the transcriptome and the proteome — a question that is still unresolved.
28. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **13,** 3497–3506 (2014).
29. Doll, S. & Burlingame, A. L. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem. Biol.* **10,** 63–71 (2015).
30. Sharma, K. *et al.* Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* **8,** 1583–1594 (2014).
31. Hsu, P. P. *et al.* The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* **332,** 1317–1322 (2011).
32. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143,** 1174–1189 (2010).
33. Olsen, J. V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127,** 635–648 (2006).
34. Francavilla, C. *et al.* Functional proteomics defines the molecular switch underlying FGF receptor trafficking and cellular outputs. *Mol. Cell* **51,** 707–722 (2013).
35. Steger, M. *et al.* Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife* **5,** e12813 (2016).
This study used a combination of genetics, chemical proteomics and cutting-edge phosphoproteomics to reveal genuine, *in vivo* substrates of the Parkinson's disease kinase LRRK2, opening the way to clinical trials.
36. Humphrey, S. J., Azimifar, S. B. & Mann, M. High-throughput phosphoproteomics reveals *in vivo* insulin signaling dynamics. *Nature Biotechnol.* **33,** 990–995 (2015).
37. Weinert, B. T. *et al.* Acetyl-phosphate is a critical determinant of lysine acetylation in *E. coli*. *Mol. Cell* **51,** 265–272 (2013).
38. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature Rev. Mol. Cell Biol.* **15,** 536–550 (2014).
39. Hendriks, I. A. *et al.* Uncovering global SUMOylation signaling networks in a site-specific manner. *Nature Struct. Mol. Biol.* **21,** 927–936 (2014).
40. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115,** 2376–2418 (2015).
41. Zheng, Y., Huang, X. & Kelleher, N. L. Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. *Curr. Opin. Chem. Biol.* **33,** 142–150 (2016).
42. Savitski, M. M., Nielsen, M. L. & Zubarev, R. A. ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol. Cell. Proteomics* **5,** 935–948 (2006).
43. Jungmichel, S. *et al.* Proteome-wide identification of poly(ADP-ribosyl)ation targets in different genotoxic stress responses. *Mol. Cell* **52,** 272–285 (2013).
44. Chick, J. M. *et al.* A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nature Biotechnol.* **33,** 743–749 (2015).
45. Rix, U. & Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nature Chem. Biol.* **5,** 616–624 (2009).
46. Gawron, D., Ndah, E., Gevaert, K. & Van Damme, P. Positional proteomics reveals differences in N-terminal proteoform stability. *Mol. Syst. Biol.* **12,** 858 (2016).
47. Kleifeld, O. *et al.* Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nature Protocols* **6,** 1578–1611 (2011).
48. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402** (suppl.), C47–C52 (1999).
49. Pawson, T. Protein modules and signalling networks. *Nature* **373,** 573–580 (1995).
50. Ward, A. B., Sali, A. & Wilson, I. A. Integrative structural biology. *Science* **339,** 913–915 (2013).
51. Dunham, W. H., Mullin, M. & Gingras, A.-C. Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* **12,** 1576–1590 (2012).
52. Choi, H. *et al.* SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature Methods* **8,** 70–73 (2011).
53. Keilhauer, E. C., Hein, M. Y. & Mann, M. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol. Cell. Proteomics* **14,** 120–135 (2015).
54. Huttlin, E. L. *et al.* The BioPlex network: a systematic exploration of the human interactome. *Cell* **162,** 425–440 (2015).
A large-scale investigation of proteins binding to tagged constructs to establish a human interactome.
55. Hein, M. Y. *et al.* A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163,** 712–723 (2015).
This paper describes the characterization of a human interactome using bait proteins that are expressed under endogenous control; its analysis in several quantitative dimensions revealed a preponderance of weak interactions.
56. Ori, A. *et al.* Spatiotemporal variation of mammalian protein complex stoichiometries. *Genome Biol.* **17,** 47 (2016).
57. Räschle, M. *et al.* Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. *Science* **348,** 1253671 (2015).
58. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196,** 801–810 (2012).
59. Rhee, H.-W. *et al.* Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339,** 1328–1331 (2013).
60. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150,** 1068–1081 (2012).
61. Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nature Methods* **9,** 907–909 (2012).
62. Wan, C. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525,** 339–344 (2015).
63. Christoforou, A. *et al.* A draft map of the mouse pluripotent stem cell spatial proteome. *Nature Commun.* **7,** 8992 (2016).
64. Larance, M. & Lamond, A. I. Multidimensional proteomics for cell biology. *Nature Rev. Mol. Cell Biol.* **16,** 269–280 (2015).

65. Yates, J. R., Gilchrist, A., Howell, K. E. & Bergeron, J. J. M. Proteomics of organelles and large cellular structures. *Nature Rev. Mol. Cell Biol.* **6,** 702–714 (2005).
66. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* **5,** e16950 (2016).
67. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450,** 695–701 (2007).
68. Marcoux, J. & Robinson, C. V. Twenty years of gas phase structural biology. *Structure* **21,** 1541–1550 (2013).
69. Politis, A. *et al.* A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nature Methods* **11,** 403–406 (2014).
70. Zhou, M. *et al.* Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science* **334,** 380–385 (2011).
An elegant demonstration of native mass spectrometry in structural studies of intact membrane complexes.
71. Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. *Trends Biochem. Sci.* **41,** 20–32 (2016).
72. Liu, F. & Heck, A. J. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **35,** 100–108 (2015).
73. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159,** 1042–1055 (2014).
74. Walzthoeni, T. *et al.* xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nature Methods* **12,** 1185–1190 (2015).
75. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods* **11,** 1064–1070 (2014).
76. Frei, A. P. *et al.* Direct identification of ligand-receptor interactions on living cells and tissues. *Nature Biotechnol.* **30,** 997–1001 (2012).
77. Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337,** 1348–1352 (2012).
This study pioneered the use of chemical crosslinking to reveal the topology of an important phosphatase complex.
78. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nature Methods* **12,** 1179–1184 (2015).
79. Navare, A. T. *et al.* Probing the protein interaction network of *Pseudomonas aeruginosa* cells by chemical cross-linking mass spectrometry. *Structure* **23,** 762–773 (2015).
80. Makowski, M. M., Willems, E., Jansen, P. W. T. C. & Vermeulen, M. Cross-linking immunoprecipitation-MS (xIP-MS): topological analysis of chromatin-associated protein complexes using single affinity purification. *Mol. Cell. Proteomics* **15,** 854–865 (2016).
81. Shi, Y. *et al.* A strategy for dissecting the architectures of native macromolecular assemblies. *Nature Methods* **12,** 1135–1138 (2015).
82. Aufderheide, A. *et al.* Structural characterization of the interaction of Ubp6 with the 26S proteasome. *Proc. Natl Acad. Sci. USA* **112,** 8626–8631 (2015).
83. Mahamid, J. *et al.* Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351,** 969–972 (2016).
84. Engen, J. R. Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal. Chem.* **81,** 7870–7875 (2009).
85. Wang, L. & Chance, M. R. Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. *Anal. Chem.* **83,** 7234–7241 (2011).
86. Savitski, M. M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346,** 1255784 (2014).
In this paper, isobaric chemical labelling was used to measure the proportion of proteins that bound to a drug as a function of temperature, on a proteome-wide scale.
87. Feng, Y. *et al.* Global analysis of protein structural changes in complex proteomes. *Nature Biotechnol.* **32,** 1036–1044 (2014).
88. Pauling, L., Itano, H. A., Singer, S. J. & Wells, I. C. Sickle cell anemia, a molecular disease. *Science* **110,** 543–548 (1949).
89. Picotti, P. *et al.* A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494,** 266–270 (2013).
90. Andreux, P. A. *et al.* Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. *Cell* **150,** 1287–1299 (2012).
91. Wu, Y. *et al.* Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. *Cell* **158,** 1415–1430 (2014).
92. Williams, E. G. *et al.* Systems proteomics of liver mitochondria function. *Science* **352,** aad0189 (2016).
A demonstration of the combined use of proteomics and genetics to interrogate mitochondrial function.
93. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534,** 55–62 (2016).
This analysis of breast cancer tissues revealed that proteomics is almost on a par with transcriptomics in terms of achievable depth of coverage of gene expression.
94. Carr, S. A. *et al.* Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* **13,** 907–917 (2014).
95. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnol.* **24,** 971–983 (2006).
96. Geyer, P. E. *et al.* Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2,** 185–195 (2016).
97. Surinova, S. *et al.* Prediction of colorectal cancer diagnosis based on circulating plasma proteins. *EMBO Mol. Med.* **7,** 1166–1178 (2015).
98. Liu, Y. *et al.* Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **11,** 786 (2015).
99. Bandura, D. R. *et al.* Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* **81,** 6813–6822 (2009).
100. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26,** 1367–1372 (2008).
101. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13,** 731–740 (2016).
102. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968 (2010).
103. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11,** O111.016717 (2012).
104. Röst, H. L. *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnol.* **32,** 219–223 (2014).
105. Tsou, C.-C. *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods* **12,** 258–264 (2015).
106. Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3,** ra3 (2010).
107. Smith, L. M., Kelleher, N. L. & The Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. *Nature Methods* **10,** 186–187 (2013).
108. Olsen, J. V. *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **4,** 709–712 (2007).
109. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. USA* **101,** 9528–9533 (2004).
110. Zubarev, R. A. & Makarov, A. Orbitrap mass spectrometry. *Anal. Chem.* **85,** 5288–5296 (2013).
111. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **9,** 555–566 (2012).
112. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* **11,** 1475–1488 (2012).
113. Chapman, J. D., Goodlett, D. R. & Masselon, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **33,** 452–470 (2014).
114. Rosenberger, G. *et al.* A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci. Data* **1,** 140031 (2014).
115. Meier, F. *et al.* Parallel accumulation–serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **14,** 5378–5387 (2015).
116. Ow, S. Y. *et al.* iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J. Proteome Res.* **8,** 5347–5355 (2009).
117. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **8,** 937–940 (2011).
118. Wühr, M. *et al.* The nuclear proteome of a vertebrate. *Curr. Biol.* **25,** 2663–2671 (2015).
119. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13,** 2513–2526 (2014).
120. Ludwig, C., Claassen, M., Schmidt, A. & Aebersold, R. Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol. Cell. Proteomics* **11,** M111.013987 (2012).

Acknowledgements We thank M. Faini and R. Ciuffa for help in preparing the figures and Y. Liu for help in compiling the literature citations. M. Hein provided inspiration for this Review, read the manuscript critically and helped with preparing the figures, as did F. Hosp, P. Geyer and S. Beck. We thank members of our groups for critical discussions.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/2bqo2n8. Correspondence should be addressed to R.A. (aebersold@imsb.biol.ethz.ch) or M.M. (mmann@biochem.mpg.de).